



MATEMATICA E SCIENZE

CLASSE DI CONCORSO
A28 (ex A059)

Manuale disciplinare

completo per le prove scritte e orali
dei concorsi a cattedra e dei FIT

**I LIBRI
PER CONCORSI
DA SEMPRE
PIÙ VENDUTI
IN ITALIA**

- ▶ Sintesi delle discipline di insegnamento
- ▶ Didattica disciplinare
- ▶ Indicazioni nazionali e Obiettivi di apprendimento

EDIZIONI
SIMONE[®]

Gruppo Editoriale **Simone**
50° ANNIVERSARIO



CON **ESPANSIONI
ONLINE**



in cattedra

Concorsi Scuola & FIT

MATEMATICA E SCIENZE

Manuale disciplinare

completo per le prove scritte e orali
dei concorsi a cattedra e dei FIT

EDIZIONI
SIMONE[®]

Gruppo Editoriale **Simone**
50° ANNIVERSARIO

50 anni di editoria al servizio di
studenti, professionisti e concorsisti

Copyright © 2018 Simone s.r.l.
Via F. Russo, 33/D
80123 Napoli
www.simone.it

Tutti i diritti riservati
È vietata la riproduzione anche parziale e con
qualsiasi mezzo senza l'autorizzazione
scritta dell'editore.

Aprile 2018
526/12 • Matematica e Scienze per il Concorso a Cattedra 2018

Questo volume è stato stampato presso:
Arti Grafiche Italo Cernia
Via Capri, n. 67 - Casoria (NA)

Realizzazione del volume a cura di: *Giovanni Ciotola*

Hanno collaborato alla precedente edizione: *Andrea Ciotola e Giuseppe Milano*

Seguici su 

Collegati alla nostra pagina Facebook dedicata ai concorsi nella scuola www.facebook.com/ConcorsiACattedra/ per tenerti informato su tutto quanto verte intorno al mondo dei concorsi per docenti e sulla nostra produzione editoriale.

La pubblicazione di questo volume, pur curato con scrupolosa attenzione dagli Autori e dalla redazione, non comporta alcuna assunzione di responsabilità da parte degli stessi e della Casa editrice per eventuali errori, incongruenze o difformità dai contenuti delle prove effettivamente somministrate in sede di concorso. Tuttavia, per continuare a migliorare la qualità delle sue pubblicazioni e renderle sempre più mirate alle esigenze dei vari lettori, la Edizione Simone sarà lieta di ricevere le segnalazioni e le osservazioni dei lettori all'indirizzo info@simone.it

PREMESSA

Questo manuale è indirizzato a quanti vogliono affrontare le prove scritte e orali dei nuovi **concorsi a cattedra** e dei **FIT**.

In particolare, l'opera ripercorre per punti e snodi essenziali tutto il *programma d'esame delle discipline d'insegnamento* relative alla **classe di concorso A28 - Matematica e Scienze** (ex A059), al fine di consentire un *rapido ripasso delle materie*, nonché *approfondimenti sulla didattica disciplinare* (per gli aspetti didattici più generali si rinvia al volume 526/B).

Completano il testo le **Indicazioni nazionali** e le **Linee guida** (disponibili come **Espansioni Web** e accessibili tramite il *QR Code* posto all'interno del volume) riguardanti le discipline oggetto d'insegnamento della classe suddetta, indispensabili per una **corretta progettazione di lezioni e unità d'apprendimento**.

Sempre in **Espansioni Web** offriamo ulteriori approfondimenti su alcuni degli argomenti trattati.

Ricordiamo ai candidati che, oltre alle competenze disciplinari proprie di ciascuna classe di concorso, alle prove verrà testata anche la conoscenza delle cosiddette «**Avvertenze generali**». A tale delicata sezione del programma d'esame (che comprende argomenti di didattica, psicologia dell'età evolutiva, normativa scolastica etc.) questa casa editrice dedica due *volumi specifici*:

- il **526/B - Manuale delle metodologie e tecnologie didattiche**, nel quale sono approfonditi gli aspetti metodologici, divenuti ormai una parte fondamentale del processo di valutazione dell'aspirante docente;
- il **526/C - Legislazione e normativa scolastica**, dove vengono trattate le componenti giuridiche della professione, aggiornate alle numerose novità intervenute negli ultimi anni per effetto della *L. 107/2015* (cd. «Buona Scuola»), tra cui i *D.Lgs. nn. 59, 60, 61, 62, 63, 64, 65 e 66 del 2017*.

PER COMPLETARE LA PREPARAZIONE

Volumi consigliati per tutti i profili

- 526/B** – Manuale delle metodologie e tecnologie didattiche
- 526/C** – Legislazione e normativa scolastica
- 526/AG** – Avvertenze generali 2018 (526/B + 526/C)
- 526/D** – La prova orale del concorso per abilitati
- 526/3** – Competenze informatiche per la prova orale
- 526/4A** – Competenze linguistiche INGLESE per la prova orale

Di notevole interesse segnaliamo:

- 526/10A** – Lezioni simulate: Matematica e Fisica, Matematica e Scienze, Scienze naturali
- 526/10** – Matematica e Fisica
- 526/16** – Scienze naturali, chimiche e biologiche

La produzione editoriale per i concorsi a cattedra e i FIT 2018 si arricchirà, nei prossimi mesi, di altri volumi utili per la preparazione.

Visita il nostro sito **www.simone.it** per conoscere le ulteriori novità.

I Scienze matematiche

Sommario Parte I

1 Apprendimento matematico: l'importanza della didattica	10 Le funzioni reali di una variabile reale
1 I momenti principali dello sviluppo del pensiero matematico	11 Il calcolo differenziale per funzioni di una variabile reale
2 Il linguaggio della teoria degli insiemi ed elementi di combinatoria	12 Calcolo integrale per funzioni di una variabile reale ed elementi di teoria della misura
3 Elementi di logica matematica	13 Successioni, serie numeriche ed equazioni differenziali
4 La geometria euclidea del piano e dello spazio	14 I processi di approssimazione e di stima degli errori
5 I sistemi numerici N, Z, Q, R, C e le strutture algebriche fondamentali	15 Elementi di statistica descrittiva
6 Il linguaggio dell'algebra lineare e il calcolo vettoriale	16 Elementi di statistica inferenziale e di calcolo delle probabilità
7 Il metodo delle coordinate per la descrizione di luoghi geometrici	17 Esami, problemi e concetti di interesse interdisciplinare
8 Gli algoritmi	18 I principali software per imparare e sperimentare la matematica
9 Elementi di trigonometria e funzioni trigonometriche	
-

Introduzione

Apprendimento matematico: l'importanza della didattica

La Matematica offre strumenti per la descrizione scientifica del mondo e per affrontare situazioni e questioni della vita quotidiana; inoltre contribuisce a sviluppare le capacità di comunicare e discutere, di argomentare in modo corretto, di comprendere i punti di vista e le motivazioni degli altri. La costruzione del pensiero matematico è un processo lungo e progressivo, nel quale concetti, abilità, competenze vengono intrecciati ed elaborati a più riprese.

1 Insegnare la Matematica

La formazione degli insegnanti di Matematica è basata non solo su una solida conoscenza della disciplina stessa, nonché della storia della materia, ma anche su una fortissima formazione culturale, critica, professionale della **didattica della Matematica**.

È importante, pertanto, capire che cosa si intende per «didattica della matematica». Una delle concezioni più accreditate è quella della didattica della matematica come *divulgazione delle idee* (B. D'Amore). Sulla base di tale idea, infatti, molti ricercatori hanno indicato agli insegnanti concrete possibilità per migliorare significativamente l'apprendimento, attraverso studi innovativi, accorgimenti, attività.

Un buon insegnante cioè, oltre a conoscere profondamente la disciplina che insegna, deve anche saper trasmettere ciò che sa. Tale capacità è diversa per ogni insegnante ed è legata alle proprie capacità creative, nonché alle singolarità degli allievi. Tuttavia un buon docente deve porsi tante domande: che cosa vuol dire apprendere la matematica? Come possiamo verificare ciò che l'allievo ha appreso? Come si può favorire l'apprendimento in aula? A tali domande ricercatori e scienziati cercano di dare risposte mediante lo studio della didattica.

2 La didattica della Matematica

Ci sono tre modi di intendere la didattica della Matematica:

- la *didattica di tipo A*, basata sulla divulgazione delle idee e concentrata prevalentemente sulla fase dell'insegnamento;
- la *didattica di tipo B*, concentrata sulla fase dell'apprendimento;
- la *didattica di tipo C*, che studia le convinzioni degli insegnanti di matematica e la loro influenza sugli studenti e sul loro apprendimento.

La **didattica A** influisce principalmente sul *sapere* e pone al centro della sua attenzione l'*Allievo*, preoccupandosi soprattutto di pre confezionare proposte attraenti e appassionanti, al fine di rendere più affascinante la matematica. Nell'ambito della didattica A, hanno acquisito particolare rilevanza i *laboratori di matematica*, nei quali, partendo da un dato progetto, si producono manufatti, a cura degli allievi, di cui poi il docente verifica l'efficacia.

Il limite di questa didattica è che gli alunni non riescono a trasferire la conoscenza acquisita in altri ambiti.

Ecco che, pertanto, è arrivata in aiuto la **didattica B**, mirata all'apprendimento specifico della matematica da parte dell'allunno. Una delle linee più seguite nell'ambito di questa teoria è quella del **costruttivismo**, in base al quale l'allievo costruisce in modo attivo una sua propria

conoscenza, interagendo con l'ambiente esterno ed organizzando le sue costruzioni mentali: cioè, in buona sostanza, ciascun allievo rielabora in maniera autonoma ogni proposta. Tali modalità di apprendimento sono condizionate anche dall'ambiente in cui si lavora (contratto didattico, ostacoli, situazioni didattiche).

Fino a pochi anni fa, tuttavia, era ancora sottovalutata l'influenza dell'insegnante in questo processo cognitivo e di apprendimento dell'alunno. Oggi, grazie alla **didattica C**, sappiamo che le convinzioni dell'insegnante determinano e condizionano l'insegnamento e l'apprendimento (gli ostacoli didattici, le clausole del contratto didattico, le misconcezioni).

3 Il contratto didattico

«In una situazione di insegnamento, preparata e realizzata da un docente, l'allievo ha generalmente il compito di risolvere un problema matematico che gli è stato presentato, ma l'accesso a questo compito si fa attraverso un'interpretazione delle domande poste, delle informazioni fornite, degli obblighi imposti, che sono costanti nel modo di insegnare del maestro» (Brousseau, 1986).

Spesso queste attese non sono dovute ad accordi espliciti, imposti dalla scuola o dagli insegnanti, ma alla concezione della «scuola della matematica» e alla ripetizione di determinate modalità.

In un celeberrimo problema, Bruno D'Amore, nel 1993, chiese agli alunni: «se un pastore ha 12 pecore e 6 capre, quanti anni ha il pastore?». «18» fu la risposta.

L'insegnante non aveva mai dato precedentemente esercizi impossibili, per cui gli alunni hanno pensato che il problema avesse sicuramente una soluzione. Inoltre, essendoci dei dati numerici, gli alunni hanno autonomamente pensato che andassero presi nell'ordine e sommati. L'unica perplessità poteva sorgere per il calcolo da effettuare: avrebbero potuto fare una sottrazione o una divisione ma, come spiegato da uno di loro, il numero, e quindi l'età del pastore, sarebbe stato troppo piccolo, ridando così senso ad una soluzione che invece non aveva senso. In questo esempio si crea un contratto didattico con delle clausole ben precise: è necessaria una soluzione e devono essere utilizzati tutti i dati numerici.

L'impegno di un insegnante dovrebbe essere quello di rompere il contratto didattico, liberando il ragionamento degli alunni da questo vincolo. Come? Ricorrendo a *situazioni a-didattiche*.

4 Le situazioni «didattiche» e quelle «a-didattiche»

Brousseau ha suddiviso le situazioni in *didattiche*, *a-didattiche*, *non didattiche*.

Nelle **situazioni didattiche** l'insegnante struttura l'ambiente circostante in modo da giungere ad una determinata conoscenza. Egli dichiara l'obiettivo che vuole raggiungere e l'allievo si impegna a comprendere ciò che il docente vuole (non tanto ad apprendere la matematica in quanto tale). Le attività svolte in aula sono quasi tutte «situazioni didattiche».

Nelle **situazioni a-didattiche** l'allievo è il protagonista e cerca da solo, o in gruppo, di risolvere un problema e di fornire delle risposte. Egli verifica la validità delle stesse e discute sulle modalità più opportune per risolvere il compito. L'insegnante svolge un ruolo di regista.

Le fasi sono:

- *devoluzione*, in cui l'insegnante affida un compito all'alunno, chiedendo la sua responsabilizzazione;
- *implicazione*, in cui l'alunno accetta l'offerta e si occupa personalmente del problema senza la guida dell'insegnante;
- *costruzione*, ovvero della conoscenza privata;

- *validazione*, in cui l'alunno spiega ai compagni la sua idea;
- *istituzionalizzazione*, in cui l'insegnante dichiara che ciò che la classe ha prodotto fa parte delle conoscenze disponibili.

Secondo Brousseau, la «situazione a-didattica» è la più adatta alla creazione della conoscenza perché l'allievo si attiva ed ha la responsabilità personale. Nella situazione didattica, invece, l'allievo riproduce solo ciò che ha detto l'insegnante: non impara la matematica ma impara ad intuire ciò che vuole sentirsi dire l'insegnante. La situazione a-didattica, rompendo il contratto didattico, permette un apprendimento più consapevole e profondo, anche se ciò avviene più lentamente.

5 L'apprendimento della Matematica

Uno dei principali problemi dell'insegnante è quello di analizzare gli apprendimenti degli allievi, al fine di capire il perché di eventuali errori degli alunni stessi e le motivazioni per le quali uno studente abbia fallito, ossia non abbia appreso la matematica e, di conseguenza, quali rimedi porre in essere per recuperare.

L'apprendimento della matematica può, a tal fine, essere distinto in 5 tipologie di apprendimento distinte:

- **apprendimento concettuale**, basato sulla costruzione di concetti che gli allievi siano in grado di identificare, rappresentare e usare in maniera opportuna;
- **apprendimento algoritmico**, legato all'abilità di dare risposte alle operazioni, al calcolo, all'applicazione delle formule;
- **apprendimento strategico**, dove si cerca di convincere gli studenti del fatto che ciò che conta sono i processi e quindi le strategie utilizzate per risolvere i problemi;
- **apprendimento comunicativo**, nel quale viene data importanza alla capacità di esprimere e comunicare le idee matematiche, argomentando anche in modo visivo, ossia con figure;
- **apprendimento delle trasformazioni semiotiche**, che è trasversale a tutti i precedenti apprendimenti e mira ad individuare i segni più adatti alla rappresentazione che si vuole fare.

Questa ripartizione agevola l'analisi dell'insegnante nell'individuazione delle cause degli errori, allo scopo di porre in atto i mezzi di recupero più adatti. Bisogna sempre tener presente che tali tipologie di apprendimento sono tra loro interconnesse e che, nella prospettiva di una valutazione, è necessario tenere conto almeno delle prime quattro componenti, arrivando a concepire una tabella come quella che segue:

Nome	Test 1	Test 2	Test ...	Valutazione	Totale
Apprendimento concettuale					
Apprendimento algoritmico					
Apprendimento strategico					
Apprendimento comunicativo					

6 Le indicazioni nazionali

Sulla base delle *Indicazioni nazionali* per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione, operanti dall'anno scolastico 2013-2014, riportiamo di seguito i traguardi e gli obiettivi previsti per lo sviluppo delle competenze nell'ambito della matematica.

Le conoscenze matematiche contribuiscono alla formazione culturale delle persone e delle comunità, sviluppando le capacità di mettere in stretto rapporto il «pensare» e il «fare», e offrendo strumenti adatti a percepire, interpretare e collegare tra loro fenomeni naturali, concetti

e artefatti costruiti dall'uomo, nonché eventi quotidiani. In particolare, la matematica fornisce strumenti per descrivere scientificamente il mondo e per affrontare e risolvere problemi della vita quotidiana; contribuisce a sviluppare la capacità di comunicare e discutere, di argomentare in modo corretto, di comprendere i punti di vista e le motivazioni degli altri.

In matematica, come nelle altre discipline scientifiche, elemento fondamentale è il **laboratorio**, inteso sia come luogo fisico sia come momento in cui l'alunno attivo formula le proprie ipotesi e ne controlla le conseguenze; progetta e sperimenta; discute e argomenta le proprie scelte; impara a raccogliere dati; negozia e costruisce significati; porta a conclusioni temporanee e a nuove aperture la costruzione delle conoscenze personali e collettive. Nella scuola primaria si potrà utilizzare il gioco, che ha un ruolo cruciale nella comunicazione, nell'educazione al rispetto di regole condivise, nell'elaborazione di strategie adatte a contesti diversi.

La costruzione del pensiero matematico è un processo lungo e progressivo, nel quale concetti, abilità, competenze e atteggiamenti vengono ritrovati, intrecciati, consolidati e sviluppati a più riprese; un processo che comporta anche difficoltà linguistiche e che richiede un'acquisizione graduale del linguaggio matematico. Caratteristica della pratica matematica è la risoluzione di problemi, che devono essere intesi come questioni autentiche e significative, legate alla vita quotidiana, e non solo esercizi a carattere ripetitivo o quesiti ai quali si risponde semplicemente ricordando una definizione o una regola. Gradualmente, stimolato dalla guida dell'insegnante e dalla discussione con i pari, l'alunno imparerà ad affrontare con fiducia e determinazione situazioni problematiche, rappresentandole in diversi modi, conducendo le esplorazioni opportune, dedicando il tempo necessario alla precisa individuazione di ciò che è noto e di ciò che s'intende trovare, congetturando soluzioni e risultati, individuando possibili strategie risolutive. Nella scuola secondaria di primo grado si svilupperà un'attività più propriamente di «matematizzazione», formalizzazione e generalizzazione. L'alunno analizza le situazioni per tradurle in termini matematici, riconosce schemi ricorrenti, stabilisce analogie con modelli noti, sceglie le azioni da compiere (operazioni, costruzioni geometriche, grafici, formalizzazioni, scrittura e risoluzione di equazioni...) e le concatena in modo efficace, al fine di produrre una risoluzione del problema. Un'attenzione particolare andrà dedicata allo sviluppo della capacità di esporre e di discutere con i compagni le soluzioni e i procedimenti seguiti.

L'uso consapevole e motivato di calcolatrici e del computer deve essere incoraggiato opportunamente fin dai primi anni della scuola primaria, ad esempio per verificare la correttezza di calcoli mentali e scritti, oltre che per esplorare il mondo dei numeri e delle forme.

Estrema importanza ha lo sviluppo di un'adeguata visione della matematica, non ridotta a un insieme di regole da memorizzare e applicare, ma riconosciuta e apprezzata come contesto per affrontare e porsi problemi significativi, per esplorare e percepire relazioni e strutture che si ritrovano e ricorrono in natura, nonché nelle creazioni dell'uomo.

Traguardi per lo sviluppo delle competenze al termine della scuola secondaria di primo grado

L'alunno si muove con sicurezza nel calcolo, anche con i numeri razionali, ne padroneggia le diverse rappresentazioni, stima la grandezza di un numero e il risultato di operazioni.

Riconosce e denomina le forme del piano e dello spazio, nonché le loro rappresentazioni, e ne coglie le relazioni tra gli elementi.

Analizza e interpreta rappresentazioni di dati per ricavarne misure di variabilità e prendere decisioni.

Riconosce e risolve problemi in contesti diversi, valutando le informazioni e la loro coerenza.

Spiega il procedimento seguito, anche in forma scritta, mantenendo il controllo sia sul processo risolutivo, sia sui risultati.

Confronta procedimenti diversi e produce formalizzazioni che gli consentono di passare da un problema specifico a una classe di problemi.

Produce argomentazioni in base alle conoscenze teoriche acquisite (ad esempio, sa utilizzare i concetti di *proprietà caratterizzante* e di *definizione*).

Sostiene le proprie convinzioni, portando esempi adeguati e utilizzando concatenazioni di affermazioni.

Accetta di cambiare opinione, riconoscendo le conseguenze logiche di un'argomentazione corretta.

Utilizza e interpreta il linguaggio matematico (piano cartesiano, formule, equazioni...) e ne coglie il rapporto col linguaggio naturale.

Nelle situazioni di incertezza (vita quotidiana, giochi...) si orienta con valutazioni di probabilità.

Ha rafforzato un atteggiamento positivo, rispetto alla matematica, attraverso esperienze significative e ha capito come gli strumenti matematici appresi siano utili in molte situazioni per operare nella realtà.

Obiettivi di apprendimento al termine della classe terza della scuola secondaria di primo grado

• *Numeri*

Eseguire addizioni, sottrazioni, moltiplicazioni, divisioni, ordinamenti e confronti tra i numeri conosciuti (numeri naturali, numeri interi, frazioni e numeri decimali), quando possibile a mente, oppure utilizzando gli usuali algoritmi scritti, le calcolatrici, i fogli di calcolo e valutando quale strumento può essere più opportuno.

Dare stime approssimate per il risultato di un'operazione e controllare la plausibilità di un calcolo.

Rappresentare i numeri conosciuti sulla retta.

Utilizzare scale graduate in contesti significativi per le scienze e per la tecnica.

Utilizzare il concetto di rapporto fra numeri o misure ed esprimerlo sia nella forma decimale, sia mediante frazione.

Utilizzare frazioni equivalenti e numeri decimali per denotare uno stesso numero razionale in diversi modi, essendo consapevoli di vantaggi e svantaggi delle diverse rappresentazioni.

Comprendere il significato di percentuale e saper calcolare quest'ultima utilizzando strategie diverse.

Interpretare una variazione percentuale di una quantità data come una moltiplicazione per un numero decimale.

Individuare multipli e divisori di un numero naturale, nonché multipli e divisori comuni a più numeri.

Comprendere il significato e l'utilità del multiplo comune più piccolo e del divisore comune più grande, in matematica e in situazioni concrete.

In casi semplici, scomporre numeri naturali in fattori primi e conoscere l'utilità di tale scomposizione per diversi fini.

Utilizzare la notazione usuale per le potenze con esponente intero positivo, consapevoli del significato, e le proprietà delle potenze per semplificare calcoli e notazioni.

Conoscere la radice quadrata come operatore inverso dell'elevamento al quadrato.

Dare stime della radice quadrata adoperando solo la moltiplicazione.

Sapere che non si può trovare una frazione o un numero decimale che elevato al quadrato dà 2, o altri numeri interi.

Utilizzare la proprietà associativa e distributiva per raggruppare e semplificare, anche mentalmente, le operazioni.

Descrivere con un'espressione numerica la sequenza di operazioni che fornisce la soluzione di un problema.

Eseguire semplici espressioni di calcolo con i numeri conosciuti, essendo consapevoli del significato delle parentesi e delle convenzioni sulla precedenza delle operazioni.

Esprimere misure usando anche le potenze del 10 e le cifre significative.

- *Spazio e figure*

Riprodurre figure e disegni geometrici, utilizzando in modo appropriato e con accuratezza opportuni strumenti (riga, squadra, compasso, goniometro, software di geometria).

Rappresentare punti, segmenti e figure sul piano cartesiano.

Conoscere definizioni e proprietà (angoli, assi di simmetria, diagonali...) delle principali figure piane (triangoli, quadrilateri, poligoni regolari, cerchio).

Descrivere figure complesse e costruzioni geometriche, al fine di comunicarle ad altri.

Riprodurre figure e disegni geometrici in base a una descrizione e a una codificazione fatta da altri.

Riconoscere figure piane simili in vari contesti e riprodurre in scala una figura assegnata.

Conoscere il Teorema di Pitagora e le sue applicazioni in matematica e in situazioni concrete.

Determinare l'area di semplici figure scomponendole in figure più elementari, ad esempio triangoli, o utilizzando le più comuni formule.

Stimare per difetto e per eccesso l'area di una figura delimitata anche da linee curve.

Conoscere il numero π e alcuni modi per approssimarlo.

Calcolare l'area del cerchio e la lunghezza della circonferenza, conoscendo il raggio, e viceversa.

Conoscere e utilizzare le principali trasformazioni geometriche e i loro invarianti.

Rappresentare oggetti e figure tridimensionali in vario modo, tramite disegni sul piano.

Visualizzare oggetti tridimensionali a partire da rappresentazioni bidimensionali.

Calcolare l'area e il volume delle figure solide più comuni.

Risolvere problemi utilizzando le proprietà geometriche delle figure.

- *Relazioni e funzioni*

Interpretare, costruire e trasformare formule che contengono lettere per esprimere in forma generale relazioni e proprietà.

Esprimere la relazione di proporzionalità con un'uguaglianza di frazioni e viceversa.

Usare il piano cartesiano per rappresentare relazioni e funzioni empiriche, o ricavate da tabelle, e per conoscere, in particolare, le funzioni del tipo $y = ax$, $y = a/x$ e i loro grafici, e, inoltre, collegarle al concetto di proporzionalità.

Esplorare e risolvere problemi utilizzando equazioni di primo grado.

- *Dati e previsioni*

Rappresentare insiemi di dati, anche facendo uso di un foglio elettronico.

In situazioni significative, confrontare dati al fine di prendere decisioni, utilizzando le distribuzioni delle frequenze e delle frequenze relative.

Scegliere ed utilizzare valori medi (moda, mediana, media aritmetica) adeguati alla tipologia ed alle caratteristiche dei dati a disposizione.

Saper valutare la variabilità di un insieme di dati, determinandone, ad esempio, il campo di variazione.

In semplici situazioni aleatorie, individuare gli eventi elementari, assegnare a essi una probabilità e calcolare la probabilità di qualche evento, scomponendolo in eventi elementari disgiunti.

Riconoscere coppie di eventi complementari, incompatibili e indipendenti.

I momenti principali dello sviluppo del pensiero matematico

1 Introduzione

Per la storia della Matematica un capitolo è appena sufficiente. Lo sforzo che sarà fatto in questo contesto sarà considerare, in breve, le evoluzioni dei principali concetti o definizioni matematiche nei secoli, le negazioni di verità sembrate indiscutibili e i principali risultati conseguiti.

Saranno trattate, nel primo paragrafo, in una visione d'insieme, le origini della Matematica, quindi, nei paragrafi successivi, le idee sottostanti le evoluzioni di concetti e definizioni rilevanti ai fini del nostro studio.

2 Dalle origini

La definizione della *Matematica* come *scienza del calcolo e della misura* non è più appropriata, essa è atta, piuttosto, a denotare le origini delle diverse branche della Matematica, le quali, indubbiamente, risalgono ad un'epoca anteriore non solo alla scrittura stessa, ma addirittura alle civiltà più antiche. Tuttavia, l'individuazione incondizionata di un'origine nel tempo e nello spazio della Matematica è una pura supposizione che, in quanto tale, non è storia. Per questo motivo, per origini della Matematica nelle diverse branche si devono intendere solo quelle assolutamente documentabili.

Oggetto di studio della Matematica sono stati per secoli, il *calcolo aritmetico* e la *misura delle figure geometriche*.

Il primo concetto che si presenta alla mente è quello di *numero intero*; si tratta di un concetto antico, lo è di meno quello di *numero razionale* che si sviluppa solo più tardi e che, con la civiltà greca, si scopre non in grado di fornire la misura di tutte le grandezze, per cui trova e deve trovare assetto il concetto di *numero irrazionale*. Man mano che si assiste all'estensione del concetto di numero, dai naturali in poi, ci si allontana anche dal concetto originario di quantità per approssimarsi a quello di calcolo.

Con gli Egiziani e i Babilonesi, dal 3000 a.C. al 600 a.C., si ritrovano i primi insegnamenti di aritmetica e geometria.

Gli Egiziani sviluppano un primo concetto di aritmetica, avendo predisposto metodi elementari per scrivere numeri grandi, e tabelle per il computo veloce.

Letteralmente la parola *geometria* significa *misurazione della terra*; le origini della geometria stessa si fanno risalire all'età degli Egiziani, costretti alle continue misurazioni dei confini della proprietà, rese necessarie dalle ricorrenti inondazioni del Nilo. Sempre gli Egiziani enunciano, ma non dimostrano, la proprietà dei triangoli rettangoli che, diversi secoli dopo, è stata enunciata nel *teorema di Pitagora* dal filosofo e matematico greco Pitagora (VI sec. a.C.).

I Babilonesi, dal canto loro, fanno uso di un tipo di scrittura detta *cuneiforme*, in cui i numeri e le lettere sono rappresentati con disegni aventi la forma di cuneo, orizzontale o verticale. Diversi rinvenimenti fanno congetturare che essi siano stati a conoscenza di una qualche formula risolutiva di equazioni di secondo grado.

Dal 600 a.C. i Greci attingono le loro conoscenze dagli Egiziani e dai Babilonesi, anche se il centro della Matematica greca è la geometria: i Greci traducono in linguaggio geometrico ogni

sorta di problemi. Fondamentale è stato il contributo di Pitagora che, con gli altri matematici greci, ha fatto della Matematica una scienza deduttiva, fondata sul ragionamento.

Si deve a Pitagora il famoso teorema che porta il suo nome, la scoperta dei *numeri irrazionali* e l'enunciazione di diverse definizioni e assiomi che sono il fondamento di ogni geometria. La scuola pitagorica distingue i numeri in pari e dispari, in primi e composti.

Appartengono alla civiltà greca il filosofo Zenone d'Elea (nato nel 490 a.C. circa), il filosofo matematico e astronomo Talete da Mileto (ca. 624 - ca. 546 a.C.), il filosofo Aristotele (384-322 a.C.) e il siracusano Archimede (287-212 a.C.).

Un cenno a parte merita il matematico greco dell'epoca alessandrina Euclide (300 a.C.), che, alla base della *geometria euclidea*, definisce il *metodo assiomatico*. Negli *Elementi*, un'opera divisa in 13 libri, Euclide raccoglie più di tre secoli di sapere geometrico. Nei primi quattro libri espone i teoremi fondamentali della geometria piana. Nel quinto e nel sesto libro sviluppa la teoria delle proporzioni. Nel settimo, nell'ottavo e nel nono libro tratta di aritmetica. Nel decimo libro classifica, dal punto di vista geometrico, i numeri irrazionali risultanti da due radicali quadratici sovrapposti. Nell'undicesimo e nel dodicesimo libro espone i teoremi fondamentali della geometria solida. Nel tredicesimo libro costruisce i cinque poliedri fondamentali.

Euclide introduce gli enti geometrici fondamentali: *punto*, *retta* e *piano*; esegue tutte le costruzioni geometriche rigorosamente con riga e compasso; inoltre, si serve di ragionamenti per dedurre un insieme di verità geometriche.

L'assenza di un sistema di numerazione opportuno e il rapporto stretto con la riga e con il compasso non hanno consentito, tuttavia, alla Matematica greca in generale una successiva evoluzione.

Nei secoli a venire la Matematica sembra aver subito un periodo di stasi, nel senso che mancano scoperte rilevanti. Si deve attendere il periodo dal 600 d.C. per l'uso dello zero, del *sistema di numerazione decimale* degli indiani, per la creazione dell'*algebra* e degli *algoritmi*. I secoli successivi sono caratterizzati dal perfezionarsi delle nozioni matematiche e dal loro graduale estendersi a civiltà diverse da quelle in cui si sono formate.

Nel XV sec. il matematico italiano Luca Pacioli (1445-1509) pubblica un'opera di grande valore per i suoi tempi, la *Summa de arithmetica, geometria, proportioni et proportionalità*, che è un compendio di aritmetica, di algebra e di geometria. Pacioli traduce in latino gli *Elementi* di Euclide.

Nel 1464 l'astronomo e matematico tedesco Regiomontano, pseudonimo di Johannes Müller (1436-1476), scrive il primo trattato di trigonometria piana e sferica, pubblicato postumo nel 1533.

Il XVI sec. è caratterizzato da un rilancio della geometria con l'uso crescente del linguaggio algebrico. Questo è il secolo del matematico Niccolò Tartaglia (1499-1557), di Gerolamo Cardano (1501-1576) e di Bonaventura Cavalieri (1598-1647), che espone *procedimenti infinitistici*. Ma è anche il secolo del matematico scozzese John Napier, italianizzato in Giovanni Nepero (1550-1617), e del matematico nato nel Liechtenstein Joste Bürgi (1552-1632), che, contemporaneamente, inventano i *logaritmi*. Nepero inventa un procedimento per eseguire le operazioni di moltiplicazione e divisione, basato sull'uso di dieci piccole aste numerate, denominate *bastoncini di Nepero*. Gli studi matematici di Nepero sono alla base della successiva realizzazione, da parte dell'inglese Edmund Gunter (1581-1626), del *regolo calcolatore* imperniato proprio sui presupposti dei logaritmi.

Il XVII sec. non è da meno. Si perfeziona il *calcolo letterale* e si afferma il matematico e filosofo francese René Descartes, italianizzato in Renato Cartesio (1596-1650), il cui metodo scientifico è imperniato sul rigore formale del metodo matematico. Cartesio inventa il sistema delle *coordinate cartesiane* per individuare i punti di un piano ed è per questo considerato il fondatore della *geometria analitica*, ma entra subito in polemica con Pierre de Fermat (1601-1665), che, contemporaneamente, ma senza dare atto ad alcuna pubblicazione, si occupa anch'egli dell'argomento.

È fondatore della *geometria proiettiva* lo scienziato e filosofo francese Blaise Pascal (1623-1662), assieme al matematico francese Gérard Desargues (1593-1662).

Sulla base di due problemi, l'uno riguardante la nozione di *velocità*, l'altro la determinazione di una tangente ad una curva, nasce il *calcolo differenziale*. Tra il matematico e fisico inglese Isaac Newton (1642-1727) e il matematico e uomo d'affari tedesco Gottfried Wilhelm Leibniz (1646-1716) è subito controversia sull'antiorità della scoperta. Verso il 1665, Newton, dal canto suo, ottiene la gran parte degli sviluppi in serie delle funzioni elementari, tra cui quello binomiale. Nel 1676 anche Leibniz riottiene quasi tutti gli sviluppi in serie cui Newton è già pervenuto, ma in maniera indipendente da quest'ultimo. Leibniz introduce altresì la *logica simbolica*.

In tempi diversi Pascal e Leibniz cominciano l'ascesa verso l'automatismo dell'elaborazione. Nel 1652 Pascal progetta e costruisce la prima macchina aritmetica completamente meccanica, in grado di addizionare e sottrarre. Nel 1673 Leibniz potenzia la macchina da calcolo di Pascal, rendendola in grado di eseguire anche le operazioni di moltiplicazione e divisione.

L'italiano Pietro Mengoli (1625-1686) usa le serie nel trattare problemi di calcolo dell'area di figure piane. Nello stesso periodo lo scozzese James Gregory (1638-1675) elabora lo sviluppo in serie della funzione arcotangente.

Senza dimenticare che appartengono al XVII sec. il marchese Guillaume François Antoine de L'Hôpital (1661-1704), i fratelli Jacques (1654-1705) e Jean (1667-1748) Bernoulli.

I secoli XVIII e XIX sono quelli dei *numeri immaginari* e dei *numeri complessi*, legati ai nomi del matematico svizzero Leonhard Euler, italianizzato in Eulero (1707-1783), e del matematico e fisico tedesco Karl Friedrich Gauss (1777-1855). Eulero si occupa, tra le altre cose, delle serie, ottenendo, nel 1736, la somma della serie dei reciproci dei quadrati degli interi, per casi finiti: somma che Mengoli, Leibniz e Bernoulli non sono riusciti ad ottenere. Facendo uso delle serie, Eulero ottiene numerosi risultati sui numeri naturali. Gauss, dal canto suo, compie studi che riguardano svariati campi della Matematica pura, della geometria e della fisica; si devono a lui importanti ricerche sul calcolo integrale, sul calcolo delle probabilità e sulla teoria degli errori (*curva di Gauss* o *curva degli errori accidentali*).

Il XVIII e il XIX sono anche i secoli di Laplace (1749-1827), cui si deve il termine *derivata*, di Ruffini (1765-1822), di Fourier (1768-1830) e di Galois (1811-1832), che dimostra, basandosi sulla *teoria dei gruppi*, che le equazioni di grado superiore al quarto non possono essere risolte per via algebrica.

Si assiste alla creazione delle *geometrie non euclidee* basate sulla negazione della validità dell'*assioma delle parallele* di Euclide, ad opera del matematico russo Nikolaj Ivanovich Lobatchewsky (1793-1856), che se ne occupa indipendentemente dal matematico ungherese János Bolyai (1802-1860), e che danno vita alla *geometria iperbolica*.

Il matematico francese Augustin Louis Cauchy (1789-1857) formula il concetto di *limite*, compie importanti studi sulle equazioni, elabora, altresì, il criterio di convergenza delle serie che ha il suo nome, si occupa della teoria della probabilità e delle applicazioni della Matematica alla Fisica. Ma il suo nome è strettamente legato a quello del matematico tedesco Bernhard Riemann (1826-1866) relativamente al calcolo integrale che, con i due matematici, perde il significato alquanto ristretto di processo inverso della derivazione per divenire concetto principale. Riemann è fondamentale per lo studio delle superfici che oggi portano il suo nome e per lo sviluppo di quell'indirizzo di geometria non euclidea noto come *geometria ellittica*. Nello studio delle relazioni tra teoria delle funzioni e teoria delle superfici è il fondatore della *topologia*.

Il matematico inglese Charles Babbage (1792-1871) compie studi su due macchine calcolatrici, l'una differenziale, l'altra digitale, che, per la loro complessità, non vengono realizzate.

Il matematico inglese George Boole (1815-1864) è uno degli iniziatori della moderna logica simbolica, spiegando come i simboli che esprimono date operazioni possono essere considerati disgiuntamente dal loro significato ed essere contenuto di operazioni autonome. Con l'*Algebra di Boole* si elaborano i *valori di verità* (*vero* o *falso*) relativi ad eventi, enunciati, e alle relazioni logiche tra queste entità.

Nel XIX sec. la geometria euclidea è vista come caso subordinato di *geometria proiettiva*, la quale è nata nel XVII sec. come estensione della geometria euclidea.

Il matematico tedesco Georg Cantor (1845-1918) è il fondatore della *teoria degli insiemi*, la quale, tuttavia, contiene delle situazioni contraddittorie note come *paradossi* della teoria degli insiemi, contraddizioni che lo stesso Cantor ha individuato. Assieme al tedesco Julius Wilhelm Richard Dedekind (1831-1916), allievo di Gauss, opera uno sconvolgimento della Matematica, dando al concetto di *infinito* una definizione che sovverte un pensiero durato per ben ventitré secoli. Essi individuano una *teoria dei numeri irrazionali*.

Il matematico tedesco Felix Klein (1849-1925) compie importanti studi sui principi della geometria elaborati nel cosiddetto *Programma di Erlangen* (1872).

Con l'avvento delle geometrie non euclidee la geometria si è *disancorata* dall'oggetto del suo studio, *la misura delle figure*, per divenire studio di strutture astratte che soddisfano un certo sistema di assiomi. Analoga cosa è avvenuta per l'algebra, facendo della *Matematica lo studio di strutture formali*, quali quelle algebriche e topologiche.

È opinione ormai consolidata che oggi la Matematica, essendo studio di tali strutture, e di conseguenza allontanandosi nel pensiero astratto, è divenuta uno strumento sempre più importante per lo studio della realtà dei fatti.

3 La trigonometria: da ombra retta a cateto

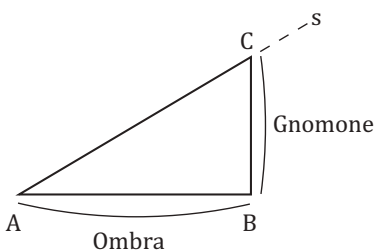
La *trigonometria* trae origine dalla risoluzione di problemi di astronomia e di calcolo di altezze attraverso la misura di ombre.

Dal 600 a.C. i Greci misurano il tempo con un semplice strumento: lo *gnomone*.

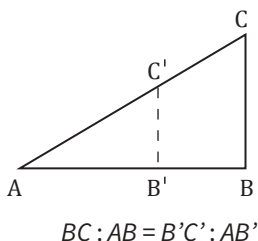
Nella sua forma più elementare, lo gnomone è composto da un'asta che proietta un'ombra su un piano orizzontale. Essendo nota la lunghezza dello gnomone, la misura dell'ombra proiettata permette di risalire all'altezza del Sole o della Luna sull'orizzonte.

Lo studio delle variazioni dell'ombra consente di riconoscere il solstizio estivo e quello invernale, e quindi consente di misurare il tempo.

Il rapporto tra BC (gnomone) e AB (ombra) è quello che, ora, si chiama *tangente dell'angolo $\hat{A}$* .



Talete ha il merito di aver misurato l'altezza di una piramide partendo dalla misura della sua ombra. Egli adotta un ragionamento molto semplice che utilizza la similitudine dei triangoli:



In questo modo si conosce la misura di una distanza che è fuori della portata degli strumenti che si hanno a disposizione in quel momento.

Nello studio dei triangoli rettangoli il cateto AB ha, per secoli, il nome di *ombra retta*, per sottolineare che la relazione tra AB e BC è entrata nella trigonometria attraverso lo studio delle ombre.

Per risolvere un triangolo i Greci lo pensano inscritto in un cerchio, i lati sono considerati come corde e queste sono calcolate in funzione del raggio. Si procede, allora, alla tabulazione delle corde di un cerchio.

Gli Arabi e i Persiani sostituiscono alla corda la semicorda ed introducono in Europa le tabulazioni delle corde.

Con il passare dei secoli, lo studio della trigonometria in Europa è distaccato dall'astronomia, assume il linguaggio simbolico dell'algebra, sottolinea le relazioni che restano fisse e giunge a definire espressamente le funzioni trigonometriche come puri numeri.

Nel 1533 è pubblicato il trattato di trigonometria piana e sferica di Johann Müller, *De triangulis omnimodis libri V* (1464), in cui è introdotto l'impiego delle tangenti, ed è usato per la prima volta il termine *seno*.

A questo punto la trigonometria diventa un capitolo dell'analisi e può essere utilizzata per i problemi che coinvolgono i triangoli rettangoli, per quelli legati a triangoli qualsiasi e per una serie di problemi diversi.

La trigonometria ha avuto origine dallo studio delle relazioni tra gli angoli del triangolo rettangolo e dei rapporti tra i cateti e l'ipotenusa del triangolo. Sotto l'influsso della nuova Matematica dell'analisi delle funzioni, si è estesa, grazie alla sua forma astratta, allo studio delle funzioni periodiche astratte semplici che esprimono tali rapporti in generale.

4 Achille e la tartaruga

«Se, in una gara di corsa, Achille concede un vantaggio $\overline{AB} = d$, ad una tartaruga, non riuscirà mai a raggiungerla», è questo il paradosso enunciato dal pensatore greco Zenone d'Elea nel V secolo a.C. Secondo la tesi, Achille impiegherebbe un tempo infinito per percorrere le infinite strisce di spazio che lo separano dalla tartaruga, man mano che quest'ultima procede dalla sua posizione di partenza.

Con il paradosso Zenone tenta di difendere la dottrina dell'unicità e dell'immobilità dell'essere.

Con le progressioni geometriche ma, soprattutto, con il *calcolo infinitesimale*, il paradosso cade. Le progressioni, infatti, insegnano che *la somma di infiniti termini può essere un numero finito*.

Il vantaggio concesso è $\overline{AB} = d$ e consideriamo che Achille proceda ad una velocità, diciamo, s volte maggiore di quella della tartaruga. Mentre Achille percorre il primo tratto, che abbiamo indicato con d , la tartaruga si troverà avvantaggiata di un tratto pari a $\frac{1}{s}d = \frac{d}{s}$. Quando Achille avrà percorso questo tratto, la tartaruga si troverà avvantaggiata, questa volta, di un tratto pari a $\frac{1}{s} \cdot \frac{d}{s} = \frac{d}{s^2}$. Pertanto, i primi n tratti percorsi da Achille saranno pari a:

$$d, \frac{d}{s}, \frac{d}{s^2}, \dots, \frac{d}{s^{n-1}}$$

Si tratta di una progressione geometrica avente per primo termine d e ragione $q = \frac{1}{s}$; essendo $s > 1$, la ragione q è minore di 1 e quindi la somma dei suoi infiniti termini è il numero finito:

$$S_{\infty} = \frac{1}{1 - \frac{1}{s}}d = \frac{s}{s-1}d$$

5 Le cifre arabe, ovvero le cifre indiane

Le cifre da uno a nove (1, 2, 3, ..., 9) appaiono nel III sec. a.C., in India.

Nel V sec. d.C. anche il sistema di numerazione posizionale e lo zero appaiono in India, e precisamente nel 458 quando è pubblicato un trattato di cosmologia, la cui traduzione in italiano è *Le parti dell'universo*. In tale opera viene citato, con un chiaro riferimento al sistema posizionale, il numero di otto cifre, 14.236.713; in realtà, le cifre sono scritte in lettere e da destra a sinistra (*tre, uno, sette, sei, tre* etc.). Nello stesso trattato si fa riferimento al *vuoto* che simboleggia lo zero.

Agli inizi del IX sec. è venuto alla luce il trattato di algebra in lingua araba *Libro dell'addizione e della sottrazione secondo il calcolo degli indiani* del matematico arabo Muhammad ibn Mûsa al-Khwarizmi, che utilizza la numerazione posizionale, diffusa, poi, in tutto il mondo occidentale cristiano. Grazie anche alle traduzioni in latino, dal nome dell'autore deriva il termine *algoritmo* e dalla sua opera, in cui in una equazione con il termine *al-giabr* intende il trasporto dall'uno all'altro membro di un addendo (cambiandolo di segno), deriva il termine *algebra*.

Le cifre indiane si diffondono nel mondo arabo, raggiungono la parte occidentale del mondo islamico per poi arrivare nella penisola iberica. Il merito dell'introduzione in Europa della numerazione indiana va al matematico e mercante Leonardo da Pisa o Fibonacci (1170-1250 ca.), con il suo *Liber abbaci* (1202). Nel passaggio dal mondo indiano alla Spagna moresca, in un periodo di ottocento anni, le cifre si modificano, assumendo la forma definita *ghobar*, ma, nel passare dalla penisola iberica ai restanti paesi europei, assumono la forma che hanno ancora oggi.

Gli arabi hanno soltanto compreso e diffuso il *calcolo degli indiani*, per cui è merito di questi ultimi l'invenzione di quei simboli che oggi, erroneamente, sono detti *cifre arabe*.

6 Lo zero e i numeri negativi

Lo zero appare nella cultura babilonese prima del III sec. a.C., vale a dire prima ancora della sua comparsa in India. I babilonesi, che rappresentano per mezzo di cunei le cifre, inventano un segno che idealmente è uno zero, utilizzato in quanto elemento di separazione nella scrittura dei numeri.

Lo zero, in ogni sua funzione, è utilizzato dai matematici indiani dal V sec. d.C.; per esigenze relative alla contabilità, essi registrano i beni e le proprietà in quantità positive (*numeri positivi*) e i debiti in quantità negative (*numeri negativi*). Il tipo di registrazione adottata, in cui si fa riferimento a numeri positivi e negativi, ovvero all'insieme degli interi relativi, non può prescindere dall'uso dello zero, indicato dagli indiani con un piccolo cerchio.

In Occidente solo durante l'Alto Medioevo ci si serve dello zero e lo si considera come cifra in un poema in latino del 1200 circa, *Carmen de Algorismo* (o *Poema sull'algoritmo*). L'abacista Raoul de Laon decide di introdurre un nuovo carattere da inserire nelle colonne vuote degli abachi, ossia di quelle tavole con scanalature a colonna nelle quali si inseriscono gettoni su cui sono scritte le cifre, e che, all'epoca, costituiscono il solo mezzo con cui si effettuano le operazioni.

L'accettazione, in qualità di numeri, degli enti numerici non positivi, in Occidente, tarda a venire. Essi sono designati come *numeri absurd* e, anche se durante il XIII sec. sono stabilite regole di calcolo e dei segni, non sono considerati possibili soluzioni di equazioni. Cartesio, ad una soluzione negativa di una equazione dà la definizione di *soluzione falsa*, pur avendo egli stesso apportato un contributo fondamentale all'algebra, con la scrittura secondo cui le prime lettere dell'alfabeto latino *a, b, c* indicano le *quantità note*, mentre le ultime *x, y, z* indicano le *incognite*, e pur avendo ottenuto la rappresentazione in *coordinate cartesiane* di ogni punto nel piano.

Solo nel XVII sec., sulla base del lavoro di Cartesio, il matematico inglese John Wallis (1616-1703) assegna coordinate negative ai punti di una curva.

Sono riconosciuti pienamente l'insieme $N \{0, 1, 2, 3, \dots\}$ dei numeri naturali e l'insieme $Z = \{\dots, -3, -2, -1, 0, +1, +2, +3, \dots\}$ dei numeri relativi.

7 I numeri razionali e il teorema di Pitagora

Il teorema che a tutti è noto come *teorema di Pitagora*, secondo cui *in un triangolo rettangolo la somma dei quadrati costruiti sui cateti è uguale al quadrato costruito sull'ipotenusa*, è stato enunciato dagli scribi babilonesi già nel 1800-1600 a.C., sebbene la sua dimostrazione puntuale è opera dei greci tra il VI e il V secolo a.C., ed esattamente del filosofo e matematico Pitagora, dal quale la sua denominazione.

E proprio il teorema di Pitagora, che costituisce la 47ª proposizione del I libro di Euclide, ha fatto crollare la relazione tra numeri interi e grandezze geometriche, di cui i numeri estrinsecano la misura. La figura geometrica che rappresenta, per antonomasia, la rottura di questo legame è una delle figure basilari del sapere antico: il quadrato.

Scomponendo un quadrato in due triangoli rettangoli isosceli uguali, e applicando il teorema di Pitagora, è possibile conoscere la misura del lato conoscendo l'altra lunghezza, la diagonale, e viceversa. Tuttavia, dato, ad esempio, un quadrato di lato 1, la sua diagonale, che costituisce l'ipotenusa dei triangoli rettangoli isosceli uguali, ha dovuto far ammettere ai greci che non esiste alcun numero razionale in grado di esprimerne la misura, in quanto essa è data da quel numero il cui quadrato è uguale a 2 (2 è la somma dei quadrati dei cateti). Pertanto, si dice, in questo caso, che il lato del quadrato e la sua diagonale sono *incommensurabili*, non ammettendo una misura comune.

Di fronte alla situazione di rottura dell'equilibrio preesistente il teorema di Pitagora, i greci hanno rifiutato di definire *numeri* le proporzioni tra grandezze geometriche; solo dopo secoli tali numeri sono stati accettati in quanto tali.

Poiché non esprime un rapporto razionale con l'unità, il numero il cui quadrato è 2 è diventato, con i secoli, il *numero irrazionale* $\sqrt{2}$.

8 I numeri reali: la continuità

Nel XII sec. il matematico e poeta persiano Omar Khayyâm (secc. XI e XII) elabora una teoria generale del numero e supera l'incapacità dei numeri razionali di rappresentare tutte le misure di grandezze, ampliando il campo dei numeri con i *numeri reali*.

Tra i numeri razionali e i numeri reali c'è una differenza: la continuità.

Sia Cantor che Dedekind enunciano il *postulato di continuità della retta*, che in ogni caso stabilisce l'esistenza di una corrispondenza biunivoca tra i punti di una retta orientata, su cui si fissa un'origine O e un'unità di misura, e i numeri reali. Ad ogni punto della retta, denominata *retta reale*, infatti, corrisponde uno e un solo numero reale, e viceversa.

La continuità con la quale si susseguono i punti della retta reale si contrappone al numerabile, che, invece, presuppone un *salto* tra un *numero numerabile* e il suo successivo; in altre parole, i numeri reali riempiono interamente la retta.

Sostituire a R la sua rappresentazione geometrica mediante i punti di una retta è, pertanto, appropriato in quanto si sostituisce un modello con un altro isomorfo, aritmeticamente ed ordinatamente, al primo.

L'esigenza dell'estensione dei numeri a quelli reali deriva dalla geometria, e propriamente dall'esigenza della misura di due segmenti incommensurabili, come il lato e la diagonale di un quadrato.

Con i numeri reali trovano soluzione problemi come quello *della misura di grandezze e dell'esistenza della radice n -esima di un numero positivo*.

Si è dovuto attendere lo studio di Cantor per un'analisi dell'insieme R dei numeri reali. Cantor dimostra che l'*infinito numerabile* non è l'unico infinito; esiste, infatti, l'infinito cosiddetto *continuo* che corrisponde a R e che ha, rispetto al numerabile, una potenza maggiore, detta *potenza del continuo*.

I punti di una retta reale sono, contrariamente a quanto suggerirebbe l'intuizione, tanti quanti i punti di un piano e addirittura dello spazio. È questo uno dei paradossi dell'infinito dei quali abbonda la teoria degli insiemi.

9 3,141592653589793..., ovvero l'irrealizzabile quadratura del cerchio

Sin dall'antichità i matematici hanno osservato una proprietà fondamentale dei cerchi: il rapporto costante tra circonferenza e diametro, che rappresenta anche l'area di un cerchio di raggio unitario. Il rapporto in esame è il numero trascendente π da *periphereia*.

Nell'Antico Testamento, ossia 2000 anni prima di Cristo, si trova l'implicita affermazione che la circonferenza è il triplo del diametro, vale a dire che π è uguale a 3.

Nel *papiro di Rhind*, un antico testo matematico di oltre 1700 anni prima di Cristo, lo scriba Ahmes sostiene che l'area di un cerchio è uguale a quella di un quadrato con lato pari a $8/9$ del diametro, vale a dire che π ha un valore pari a $(16/9)^2 = 3,16049...$

Nel III sec. a.C. Archimede si occupa del problema affermando che: «*In ogni cerchio il perimetro supera il triplo del diametro di meno di un settimo ma di più di dieci settantunesimi*»; in altre parole π è compreso tra $3 + 10/71$ e $3 + 1/7$.

Nella seconda metà del XVII sec., il matematico, fisico e filosofo tedesco Johann Heinrich Lambert (1728-1777) stabilisce, finalmente, che π è un numero irrazionale; come tale non è rappresentabile da alcuna frazione e non vi è possibilità alcuna di stabilirne uno sviluppo decimale periodico.

Nel 1882 il matematico tedesco Carl Ferdinand Lindemann (1852-1939) dimostra che π è un numero trascendente. Da allora la questione della quadratura del cerchio, ossia della possibilità di costruire un quadrato che abbia la stessa area di un cerchio dato, è chiusa, nel senso di una negativa soluzione della questione.

10 2,71828182845... oppure 10?

Nel XVI sec. Nepero inventa i logaritmi; dopo aver intuito il vantaggio che da essi si può trarre per abbreviare i calcoli, egli perviene alla compilazione di una tavola in base $e = 2,71828...$, che ha reso nota nel 1614 in un'opera intitolata «*Logarithmorum canonis descriptio, seu arithmeticarum supputationum mirabilis abbreviatio*», in cui non indica il metodo di calcolo seguito, promettendo, però, di indicarlo in seguito; la promessa risulta vana, a causa della sua morte. Nascono, così, i *logaritmi naturali* (o *neperiani*).

Nel frattempo Joste Bürgi inventa anch'egli i logaritmi, indipendentemente da Nepero, giungendo alla stessa conclusione per vie diverse. pubblica sotto forma di *antilogaritmi* e anonima la sua scoperta, solo dopo che i risultati raggiunti da Nepero sono stati resi noti.

Il Professore di geometria Enrico Briggs (1561-1631) del Gresham College di Londra è un degno successore dell'opera di Nepero. Egli espone la teoria elaborata da Nepero nelle sue lezioni al Collegio di Gresham ed intuisce che i calcoli sono molto più semplici se invece di far uso della base $e = 2,71828$, adottata da Nepero, si prende per base il numero 10. Nascono, così, i *logaritmi decimali*, i quali sono classificati in apposite tavole in uso ancora oggi. Relativamente a tale cambiamento, lo stesso Nepero approva il miglioramento.

Si deve attendere solo il XIX sec. per una dimostrazione, ad opera del matematico francese Charles Hermite (1822-1901), che il numero $e = 2,71828182845\dots$ è trascendente.

11 Il quadrato negativo

Quali soluzioni ammette l'equazione $x^2 = -1$? La soluzione dovrebbe essere quel numero che elevato al quadrato è uguale a -1 .

È certo che nell'insieme dei numeri reali, dove qualunque numero elevato al quadrato dà un numero positivo, tale soluzione non esiste; per converso, nello stesso insieme, non ha alcun significato la radice quadrata di un numero reale negativo.

Nel 1545 Gerolamo Cardano pubblica, nel celebre trattato di matematica *Ars Magna*, la formula risolutiva di una equazione di terzo grado, formula il cui merito si deve a Niccolò Tartaglia. Con essa, per la prima volta, nasce l'esigenza di dare un valore alla radice quadrata di un numero reale negativo. Trovandosi al cospetto di tale problema, Tartaglia si rende conto che, quando l'equazione ha tre soluzioni reali e distinte, conduce a radici quadrate di numeri reali negativi, pur essendo il risultato costituito da numeri reali.

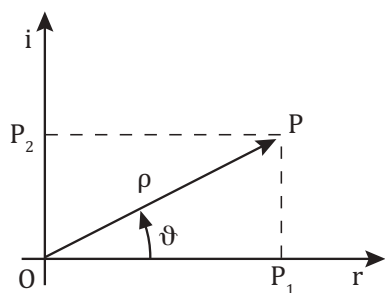
Il matematico bolognese Raffaele Bombelli (ca. 1526 - ca. 1573), nella sua opera *L'algebra*, già parla di un *più di meno*, predecessore di i . Egli apre la strada ad Eulero, che, nel 1777, esprime la funzione esponenziale e^{ix} con esponente immaginario puro mediante le funzioni seno e coseno: $e^{ix} \cos x + i \sin x$ (formula di Eulero). Eulero introduce il simbolo appropriato i al posto di $\sqrt{-1}$; i è detta *unità immaginaria* ed è la *radice immaginaria dell'unità negativa* (i moltiplicato i fa -1).

In seguito Gauss si occupa anch'egli dell'unità immaginaria. Nel *Teorema fondamentale dell'algebra* egli dimostra che una equazione algebrica di qualunque grado ammette sempre una radice del tipo $(a + ib)$, e definisce in questo modo un nuovo ente: il numero complesso $z = a + ib$, dove a è la *parte reale* del numero complesso, ib è la *parte immaginaria* e b è il *coefficiente* dell'unità immaginaria i .

L'insieme dei numeri complessi, indicato con C , è più ampio dell'insieme dei numeri reali e consente l'operazione di estrazione della radice quadrata di un numero reale negativo, in quanto, per definizione, $i^2 = -1$.

Pertanto, la risposta alla domanda posta all'inizio, appartiene all'insieme C dei numeri complessi.

La rappresentazione geometrica dei numeri complessi è un'idea risalente al 1798 ad opera del matematico norvegese Gaspar Wessel (1745-1818), che si serve dei *diagrammi di Argand*, denominati anche *diagrammi di Wessel*, una trentina di anni dopo dello stesso Gauss, il quale, con il suo *piano dei numeri*, apre la via alla rappresentazione sul *piano complesso* o *piano di Gauss*. L'opera di Wessel è rimasta pressoché sconosciuta, per cui la rappresentazione geometrica dei numeri complessi è attribuita a Gauss.



In quanto composto da due parti, il numero complesso $z = a + ib$ può essere rappresentato da un ente a due dimensioni: il *piano complesso* o *di Gauss*. Le due parti, a e b , possono essere considerate all'interno di un sistema piano di coordinate cartesiane, in cui l'asse orizzontale, detto *asse reale*, rappresenta i numeri reali e l'asse verticale, detto *asse immaginario*, rappresenta i numeri immaginari; su quest'ultimo asse si trova il numero i , a distanza di una unità dall'origine. La rappresentazione mediante i punti del piano si ottiene associando al numero complesso $z = a + ib$ il punto P di coordinate cartesiane (a, b) ; P si dice *immagine geometrica* del numero complesso, mentre l'*affissa* del

punto è il numero complesso considerato in relazione alla sua rappresentazione geometrica. Se si congiunge il punto P con l'origine O del sistema cartesiano, il numero complesso è individuato anche dal vettore $\overrightarrow{OP}$ avente per modulo $\rho = \sqrt{a^2 + b^2}$ e per argomento l'angolo $X\hat{O}P = \vartheta$.

12 Fermat: da congettura a teorema

Il matematico francese Pierre de Fermat nel 1640 scrive, in margine ad un'opera del matematico alessandrino Diofanto vissuto probabilmente nel III d.C., che non si può dividere un cubo in due cubi, né un biquadrato in due biquadrati, né, in generale, una qualsiasi potenza di grado superiore al secondo in altre due potenze dello stesso grado. In altre parole, Fermat afferma che, dati tre numeri interi x, y e z , e un numero intero n maggiore di 2, la seguente equazione:

$$x^n + y^n = z^n$$

non ammette soluzioni intere non nulle, quando n è uguale o maggiore di 3.

L'affermazione di Fermat, denominata *l'ultimo teorema di Fermat*, è stata per secoli una pura congettura e, mancando la dimostrazione, non è mai stata considerata un teorema.

Persino Eulero tenta di dare una dimostrazione dell'affermazione di Fermat, riuscendo solo per $n = 3$ e poi per $n = 4$.

Nei secoli successivi altri matematici dimostrano la verità dell'affermazione di Fermat per $n > 4$, ma non per ogni n .

Nel 1987 D. Heath Brown stabilisce la certezza per *quasi tutti gli* n . Si deve attendere il 1995, anno in cui il matematico Andrew Wiles estende il campo di verità dell'affermazione di Fermat a *tutti gli* n , e consente, finalmente, di parlare di *teorema di Fermat*.

13 Un'infinità di infiniti

Per circa duemila anni, il pensiero di Aristotele di un *infinito potenziale*, vale a dire di un infinito che non può essere raggiunto e che non ammette alcun al di là, è il pensiero prevalente.

L'asserzione principale da cui deriva il pensiero aristotelico sull'infinito è che *il tutto è più grande della parte*; in altre parole, il tutto è tale perché contiene le sue parti, le quali non possono essere messe a confronto con il tutto.

Ventitré secoli dopo, il pensiero aristotelico è sovvertito dai matematici tedeschi Dedekind e Cantor, i quali basano i loro studi su un procedimento che definisce una corrispondenza biunivoca tra due insiemi, definiti *equipotenti*.

A partire dal 1870 Cantor e Dedekind stabiliscono una serie di nuovi concetti destinati a sconvolgere l'assetto della Matematica.

La nuova concezione di *insieme infinito*, che deriva dai loro studi, è quella di *un insieme che è equipotente con una sua parte propria*.

Nel 1638 Galileo Galilei (1564-1642) osserva che è possibile creare una corrispondenza biunivoca tra l'insieme N dei numeri naturali e l'insieme P dei numeri interi pari, parte propria dell'insieme N . Infatti, a ciascun elemento dell'insieme N si fa corrispondere il suo doppio, che è un elemento dell'insieme P ; viceversa, a ciascun elemento dell'insieme P si fa corrispondere la sua metà, che è un elemento dell'insieme N . Sulla base di detta osservazione e stando alla concezione di Cantor e Dedekind sugli insiemi infiniti, l'insieme N dei numeri naturali è infinito, ed è questo l'infinito in atto, detto *numerabile* o *discreto*.

Si è rimossa una *verità* che resiste da anni, si è dimostrato che *l'infinito numerabile non è più grande di una delle sue parti*.

Cantor dimostra altresì che l'insieme Q dei numeri razionali (assoluti o relativi) è anch'esso numerabile o, ed è lo stesso, che l'insieme Q non ha più elementi dell'insieme N dei naturali.

L'infinito numerabile non è l'unico, Cantor dimostra che esiste *un altro infinito*. La dimostrazione considera l'insieme R dei numeri reali. La rappresentazione grafica dei numeri reali, la *retta reale*, contiene sicuramente più punti che numeri interi per indicarli; da ciò l'esistenza di due infiniti e quello di R è denominato *continuo*. La potenza dell'insieme R dei numeri reali, definita *potenza del continuo*, è maggiore della potenza del numerabile.

Oltre al discreto e al continuo, esistono *altri infiniti*. Cantor, sapendo che un insieme di n elementi ha 2^n parti, dimostra che un insieme ha sempre più parti che elementi; in altre parole, l'insieme di tutte le parti di un insieme ha una potenza maggiore dell'insieme stesso.

Dato, quindi, un insieme infinito è sempre possibile costruire un infinito di ordine superiore.

È questa la genesi dei *numeri transfiniti*, creati da Cantor, di cui il primo è il *numerabile* ed è indicato con $\aleph_0$ (alef con zero). Dei numeri transfiniti Cantor mette a punto un'aritmetica, realizzando la sua idea di estendere il calcolo aritmetico al di là del finito.

Nel 1878 Cantor si occupa del problema di stabilire se esiste un insieme che abbia una potenza strettamente compresa tra la potenza del numerabile e la potenza del continuo, e formula l'ipotesi, detta *ipotesi del continuo*, che non esiste alcun insieme che gode di questa proprietà.

14 Il binomio di Newton, ovvero di Pascal

Spesso accade che scoperte matematiche non siano attribuite ai giusti autori. Lo stesso vale per la formula del binomio che porta il nome del matematico e fisico inglese Isaac Newton. Essa dovrebbe essere più giustamente attribuita al matematico francese Blaise Pascal, il quale per primo nota la relazione che esiste tra i coefficienti binomiali e la formula della potenza di un binomio:

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

Tale formula sembra fosse già nota agli arabi sin dal XIII sec., ma gli occidentali ne vengono a conoscenza solo nel XVI sec.

Newton ha semplicemente esteso la formula a qualsiasi valore intero di n , da cui la denominazione *formula del binomio di Newton*.

15 Il Teorema di L'Hôpital, ovvero di Bernoulli

Il teorema che porta il nome del marchese de L'Hôpital dovrebbe essere attribuito, più esattamente, a Jean Bernoulli. Egli insegna al marchese il calcolo infinitesimale e si impegna, dietro compenso di un salario, a comunicargli tutte le scoperte matematiche, cedendogli ogni diritto sulle stesse.

Tra tali scoperte vi è anche quella che è nota come *Teorema di L'Hôpital* del 1694, la quale è semplicemente descritta dal marchese nel primo manuale di calcolo differenziale mai pubblicato.

16 Archimede e il segmento parabolico

Le origini del calcolo integrale si fanno risalire al periodo della Matematica greca, in cui, attraverso un procedimento detto *metodo di esaustione*, dovuto al geometra ed astronomo greco Eudosso di Cnido (408-355 a.C.) e sfruttato dal matematico e fisico Archimede, si ottengono soluzioni quali l'*area del cerchio* e della superficie sferica, l'area di alcune regioni piane (tra cui il *segmento parabolico*) e il volume di alcuni solidi.

Archimede giunge alla determinazione dell'area del cerchio considerando che, se si inscrivono in un cerchio poligoni regolari e si aumenta il numero dei loro lati, tali poligoni *esauriscono* (da

cui *esaustione*) la regione interna del cerchio, e le loro aree differiscono sempre meno dall'area di un triangolo rettangolo avente per base la lunghezza della circonferenza e per altezza il raggio.

Archimede giunge, altresì, all'area del cosiddetto *segmento parabolico*, ossia all'area di quella regione di piano limitata dalla parabola di equazione $y = x^2$, dall'asse delle ascisse e dalla parallela all'asse delle ordinate di equazione $y = b$, dove b è un numero reale positivo. Basandosi sul primo esempio di serie numerica, e precisamente su una *serie geometrica*, dimostra che tale area è uguale a:

$$S = \frac{b^2}{3}$$

Nel XVI sec. nasce l'Algebra e il metodo di esaustione riceve nuovi sviluppi.

Nel XVII sec., il matematico milanese Bonaventura Cavalieri, allievo di Galilei, ne la *Geometria degli invisibili* (1635), espone la teoria della *somma di infiniti indivisibili*, in cui sono riportate le misure di aree e volumi. Sulla base dei procedimenti infinitistici di Cavalieri, il matematico inglese John Wallis risolve problemi di calcoli di aree.

Successivamente, Newton, ispirandosi all'inglese Wallis, crea il *calcolo infinitesimale* e, nel 1687, pubblica i *Philosophiae naturalis principia mathematica* che, tra l'altro, contengono numerose applicazioni di tale strumento. Newton è in polemica sulla priorità della scoperta con Leibniz, in quanto nel 1684 quest'ultimo crea, indipendentemente da Newton, il *calcolo differenziale* su cui si fonda il *calcolo infinitesimale*.

Si deve, comunque, attendere il XIX sec. per una sistemazione conclusiva dell'odierno calcolo integrale da parte di Cauchy, che si basa sul concetto di *limite*, e di Riemann, cui si deve l'illustrazione del concetto di *integrale definito* o *integrale secondo Riemann*.

Il linguaggio della teoria degli insiemi ed elementi di combinatoria

1 Gli insiemi

Secondo la definizione di Cantor, un *insieme* è una raccolta di oggetti definiti e distinti del nostro pensiero. Gli oggetti sono chiamati *elementi* o *membri* dell'insieme. Gli insiemi si indicano con le lettere maiuscole dell'alfabeto e gli elementi con le lettere minuscole; per indicare che un elemento a appartiene o non appartiene ad un insieme A , si scrive rispettivamente:

$$a \in A \quad \text{e} \quad a \notin A$$

L'*insieme vuoto* è l'insieme privo di elementi e si indica con $\emptyset$. L'*insieme unitario* è l'insieme costituito da un solo elemento.

Se A e B sono due insiemi e se ogni elemento di A appartiene anche a B , si dice che A è contenuto in B , o che A è un *sottoinsieme* di B o una sua *parte*, e si scrive:

$$A \subseteq B$$

Se esiste almeno un elemento di B che non appartiene ad A , si dice che A è un *sottoinsieme proprio*, o *parte propria*, di B , e si scrive:

$$A \subset B$$

Poiché ogni elemento di A appartiene ad A , risulta:

$$A \subseteq A$$

e pertanto ogni insieme è sottoinsieme di se stesso.

Inoltre, l'*insieme vuoto* si considera come sottoinsieme di ogni insieme.

Questi due sottoinsiemi (l'insieme vuoto e l'insieme stesso) si chiamano *sottoinsiemi impropri* di A .

N indica l'insieme dei numeri naturali (P l'insieme dei numeri naturali pari, D l'insieme dei numeri naturali dispari); Z è l'insieme dei numeri interi (Z^+ l'insieme dei numeri interi positivi, Z^- l'insieme dei numeri interi negativi); Q è l'insieme dei numeri razionali (Q^+ l'insieme dei numeri razionali positivi, Q^- l'insieme dei numeri razionali negativi); R è l'insieme dei numeri reali (R^+ l'insieme dei numeri reali positivi, R^- l'insieme dei numeri reali negativi).

Si dicono *massimo* e *minimo* di un insieme di numeri reali, rispettivamente, l'*estremo superiore* e l'*estremo inferiore* dell'insieme, dove per estremo superiore e per estremo inferiore di un insieme si intendono, rispettivamente, l'elemento più grande e l'elemento più piccolo dell'insieme, e possono essere finiti o infiniti.

Due insiemi si dicono *uguali* se sono formati dagli stessi elementi, cioè se ogni elemento x dell'uno è anche elemento dell'altro.

L'*insieme delle parti* di un insieme è l'insieme formato da tutti i sottoinsiemi dell'insieme dato, propri e impropri.

Gli insiemi si possono rappresentare in vari modi:

- 1) *rappresentazione tabulare o analitica*, che consiste nell'elenco degli elementi che lo compongono, scritto fra parentesi graffe:

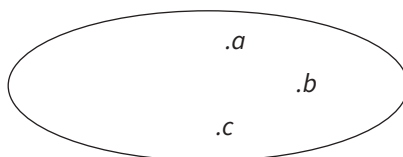
$$A = \{a, b, c\}$$

- 2) *rappresentazione caratteristica o sintetica*, mediante la scrittura simbolica della legge che individua l'insieme:

$$\{x \in R: \text{legge sulla } x\}$$

dove x è un generico elemento dell'insieme, R è l'insieme in cui x può variare, i : indicano la frase *tale che*, legge sulla x è la legge cui obbedisce la variabile;

- 3) *rappresentazione grafica* mediante i *diagrammi di Eulero-Venn*, che consistono in una linea chiusa che raccoglie i punti:



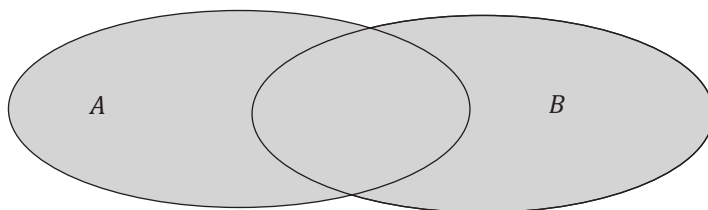
Un insieme si dice *aperto* se ogni suo punto è interno oppure se è vuoto. Un insieme aperto, quindi, non contiene nessun punto della frontiera.

Un insieme si dice *chiuso* se contiene tutta la sua frontiera, e, in particolare, se è vuoto. Un insieme è chiuso se il suo complementare è aperto.

2 Le operazioni sugli insiemi

a) Unione

Dati due insiemi A e B , si dice *unione* di A e B l'insieme i cui elementi appartengono ad uno (almeno) degli insiemi A e B , e si indica con $A \cup B$. Nel grafico $A \cup B$ è l'insieme dei punti appartenenti all'area ombreggiata.



L'unione gode delle seguenti proprietà:

- *proprietà commutativa*: $A \cup B = B \cup A$
- *proprietà associativa*: $A \cup (B \cup C) = (A \cup B) \cup C$

Inoltre:

- $A \cup A = A$
- $A \cup \emptyset = A$

ESEMPIO 1

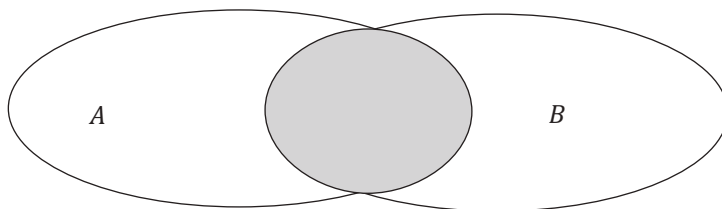
Siano $A = \{x/x \text{ numero naturale e } 12 < x < 20\}$ e $B = \{x/x \text{ numero primo e } < 23\}$ due insiemi. Determinare l'insieme *unione* $A \cup B$.

L'insieme $A = \{13, 14, 15, 16, 17, 18, 19\}$ e l'insieme $B = \{2, 3, 5, 7, 11, 13, 17, 19\}$, l'insieme costituito dagli elementi che appartengono almeno ad uno dei due insiemi considerati è:

$$A \cup B = \{2, 3, 5, 7, 11, 13, 14, 15, 16, 17, 18, 19\}$$

b) Intersezione

Si dice *intersezione* di due insiemi A e B , definiti in un certo insieme C , l'insieme degli elementi che appartengono ad A e contemporaneamente a B , e si indica con $A \cap B$. Nel grafico, $A \cap B$ è l'insieme dei punti appartenenti all'area ombreggiata.



L'intersezione gode delle seguenti proprietà:

- *proprietà commutativa:* $A \cap B = B \cap A$
- *proprietà associativa:* $A \cap (B \cap C) = (A \cap B) \cap C$

Inoltre:

- $A \cap A = A$
- $A \cap \emptyset = \emptyset$

Valgono le seguenti due proprietà:

- *Proprietà distributiva dell'unione rispetto all'intersezione:*

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
- *Proprietà distributiva dell'intersezione rispetto all'unione:*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

ESEMPIO 2

Siano $A = \{x/x \text{ divisore di } 28\}$ e $B = \{x/x \text{ multiplo di } 7 \text{ e } < 49\}$ due insiemi. Determinare l'insieme *intersezione* $A \cap B$. L'insieme costituito dagli elementi che appartengono sia ad $A = \{1, 2, 4, 7, 14, 28\}$ che a $B = \{0, 7, 14, 21, 28, 35, 42\}$ è:

$$A \cap B = \{7, 14, 28\}$$

ESEMPIO 3

Siano $A = \{x/x \text{ numero dispari e } 35 < x < 49\}$; $B = \{x/x \text{ multiplo di } 3 \text{ e } < 27\}$ e $C = \{x/x \text{ quadrato di un numero primo e } 9 < x < 169\}$ tre insiemi.

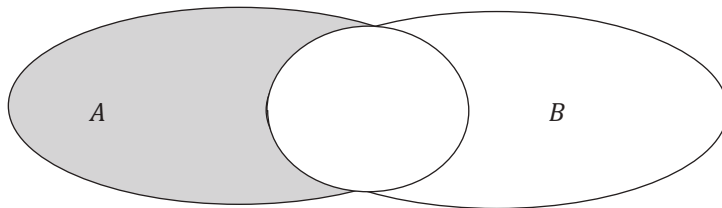
Determinare l'insieme *intersezione* $(A \cap B) \cap C$.

Si tratta dell'operazione di *intersezione* di tre insiemi (identica peraltro al caso di due insiemi); gli insiemi sono, rispettivamente, $A = \{37, 39, 41, 43, 45, 47\}$, $B = \{0, 3, 6, 9, 12, 15, 18, 21, 24\}$ e $C = \{25, 49, 121\}$. Essi non hanno elementi in comune, per cui si dice che sono *disgiunti* e la loro intersezione è l'insieme vuoto:

$$A \cap B \cap C = \emptyset$$

c) Differenza

Si dice *differenza* di due insiemi A e B l'insieme costituito dagli elementi di A che non appartengono a B , e si indica con $A \setminus B$. Nel grafico, $A \setminus B$ è l'insieme dei punti appartenenti all'area ombreggiata.



Pur essendo A e B entrambi non vuoti, si può verificare:

$$A \setminus B = A$$

oppure:

$$A \setminus B = \emptyset$$

Dato un insieme A definito in un certo insieme B , si dice *complementare* di A rispetto a B , la differenza $B \setminus A$, che è l'insieme i cui elementi appartengono a B e non ad A . Ad esempio, nell'insieme degli interi relativi, il complementare di $\mathbb{Z}^-$, insieme dei numeri interi negativi, è $\mathbb{Z}^+$ cioè l'insieme dei numeri interi positivi. Il complementare dell'insieme che contiene tutti i numeri reali è l'insieme vuoto; il complementare dell'insieme vuoto rispetto all'insieme dei numeri reali, è l'insieme dei numeri reali stesso.

Due insiemi si dicono *disgiunti* se non hanno elementi in comune, per cui, un insieme e il suo complementare sono tra loro disgiunti.

ESEMPIO 4

Siano $A = \{x/x \text{ multiplo di } 2 \text{ e } < 24\}$ e $B = \{x/x \text{ multiplo di } 6 \text{ e } < 48\}$ due insiemi. Determinare l'insieme differenza $A \setminus B$. L'insieme costituito da quegli elementi che appartengono ad $A = \{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22\}$ ma non a $B = \{0, 6, 12, 18, 24, 30, 36, 42\}$ è:

$$A \setminus B = \{2, 4, 8, 10, 14, 16, 20, 22\}$$

ESEMPIO 5

Siano $A = \{x/x \text{ divisore di } 40\}$ e $B = \{x/x \text{ multiplo di } 5 \text{ e } < 45\}$ due insiemi. Determinare l'insieme differenza $B \setminus A$. L'insieme A è uguale a $A = \{1, 2, 4, 5, 8, 10, 20, 40\}$ e l'insieme B è uguale a $B = \{0, 5, 10, 15, 20, 25, 30, 35, 40\}$; l'insieme costituito da quegli elementi che appartengono a B ma non ad A è:

$$B \setminus A = \{0, 15, 25, 30, 35\}$$

3 Il prodotto cartesiano

Siano A e B due insiemi (non vuoti), è possibile considerare l'insieme avente come elementi le *coppie ordinate* (a, b) , dove a è un elemento di A e b è un elemento di B , in simboli: $a \in A$ e $b \in B$. Ad esempio, siano $A = \{a, b, c\}$ e $B = \{d, e, f\}$ due insiemi. Tutte le possibili coppie che si possono formare prelevando da A la prima componente e da B la seconda sono:

$$\begin{array}{lll} (a, d) & (a, e) & (a, f) \\ (b, d) & (b, e) & (b, f) \\ (c, d) & (c, e) & (c, f) \end{array}$$

Nella coppia si distingue il primo dal secondo elemento, ossia si considera l'ordine con cui gli elementi sono elencati, a differenza dell'insieme, in cui tale ordine è ininfluente.

In base alle suddette considerazioni, si dice *prodotto cartesiano* l'insieme costituito da tutte le coppie ordinate (a, b) :

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

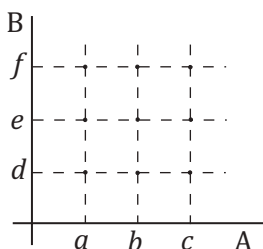
Il prodotto cartesiano non gode della proprietà commutativa, nel senso che:

$$A \times B \neq B \times A$$

Se almeno uno dei due insiemi è vuoto, allora anche il loro prodotto cartesiano è vuoto.

Il prodotto cartesiano $A \times A = A^2$ si chiama *quadrato cartesiano*.

Il prodotto cartesiano $A \times B$ si rappresenta graficamente riportando sull'asse orizzontale gli elementi del primo insieme A e sull'asse verticale gli elementi del secondo insieme B ; i punti di incontro delle parallele agli assi, condotte, rispettivamente, da a, b, c e d, e, f non sono altro che le coppie del prodotto cartesiano $A \times B$.



4 Le relazioni

Siano A e B due insiemi non vuoti e siano a e b elementi, con $a \in A$ e $b \in B$. Se esiste una legge che ad un elemento $a \in A$ associ un elemento $b \in B$, si dice che si stabilisce una *corrispondenza* o una *relazione binaria* tra gli elementi dell'insieme A e quelli dell'insieme B , in simboli:

$$a \mathfrak{R} b$$

La relazione $\mathfrak{R}$ è detta *binaria* in quanto fa riferimento ad una coppia di elementi.

Inoltre, in una relazione binaria, non necessariamente tutti gli elementi di A devono essere in relazione con B , così come non tutti gli elementi di B devono essere in qualche relazione con un elemento di A ; a tal proposito si indica con A' il sottoinsieme di A i cui elementi siano in relazione con qualche elemento di B , e con B' il sottoinsieme di B i cui elementi siano in relazione con qualche elemento di A' . La relazione binaria si dice:

- *univoca*, se ad ogni elemento di A' corrisponde un unico elemento di B' ;
- *plurivoca*, se ad ogni elemento di A' corrispondono più elementi di B' ; in altre parole, se esistono elementi di A' che hanno come *immagini* più elementi di B' .

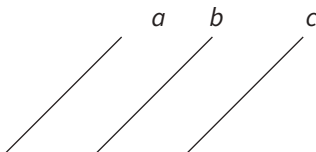
Ogni relazione può godere della:

- *proprietà riflessiva*, ossia $a \mathfrak{R} a$, per cui ogni elemento è in relazione con se stesso;
- *proprietà simmetrica*, ovvero se $a \in A$ e $b \in B$ e se $a \mathfrak{R} b$, allora $b \mathfrak{R} a$, ossia ogni volta che a è in relazione con b , anche b è in relazione con a ;
- *proprietà transitiva*, ovvero se $a \mathfrak{R} b$ e $b \mathfrak{R} c$, allora $a \mathfrak{R} c$;
- *proprietà antisimmetrica*, ovvero se $a \mathfrak{R} b$ e $b \mathfrak{R} a$, si ha necessariamente $a = b$.

a) Relazione di equivalenza

Una relazione $\mathfrak{R}$ in A si dice *relazione di equivalenza* se si verificano le proprietà riflessiva, simmetrica e transitiva.

Si consideri la relazione di equivalenza tra rette parallele.



$$\begin{array}{ll} a \mathfrak{R} a & \\ a \mathfrak{R} b & b \mathfrak{R} a \\ a \mathfrak{R} b & b \mathfrak{R} c \Rightarrow a \mathfrak{R} c \end{array}$$

L'insieme degli elementi equivalenti ad un elemento a si dice *classe di equivalenza* di tale elemento.

b) Relazione d'ordine

Una relazione $\mathfrak{R}$ in A si dice *relazione d'ordine* se si verificano le proprietà riflessiva, antisimmetrica e transitiva.

L'insieme in cui sia stata introdotta una relazione d'ordine si dice che è *ordinato* dalla relazione stessa; rispetto alla relazione d'ordine considerata, se (a, b) è una coppia che appartiene alla relazione, si dice che a *precede* b .

La relazione $<$ non è una relazione d'ordine degli insiemi numerici, in quanto non è soddisfatta la proprietà riflessiva.

La relazione $\leq$ è detta *relazione d'ordine in senso stretto* e gode delle proprietà enunciate.

Una relazione d'ordine nell'insieme numerico $\mathbb{Z}$ degli interi relativi è $a \mathfrak{R} b$ se e solo se $a \leq b$.

L'ordinamento è *totale* se ogni elemento a è confrontabile con ogni elemento b ; se ciò non accade, l'ordinamento è *parziale*.

Ogni sottoinsieme di un insieme A totalmente ordinato risulta totalmente ordinato dalla stessa relazione che ordina l'insieme A . Se, invece, l'insieme A è solo parzialmente ordinato, un suo sottoinsieme, rispetto alla stessa relazione, può risultare totalmente ordinato.

5 Le strutture d'ordine

Se in un insieme è introdotta una relazione d'ordine si dice che ad esso è stata impressa una *struttura d'ordine*. In generale, significa che in un insieme vi sono elementi confrontabili secondo date regole, come accade agli insiemi numerici (ad eccezione dell'insieme dei numeri complessi) per mezzo della relazione « $\leq$ ».

Le strutture d'ordine di insiemi finiti possono essere rappresentate con i *diagrammi d'ordine* o *diagrammi di Hasse*, che si disegnano nel modo seguente. Se $a \mathfrak{R} b$ e a è diverso da b , si pone b al di sopra di a . Inoltre si collega a a b tramite un segmento orientato, omettendo i segmenti che sono conseguenza della riflessività e anche quelli che sono conseguenza della transitività (se $a \mathfrak{R} b$ e $b \mathfrak{R} c$, allora anche $a \mathfrak{R} c$ ma non si disegna il corrispondente segmento orientato, visto che non aggiungerebbe nulla a quanto già sappiamo).

ESEMPIO 6

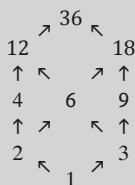
La relazione su N , definita da $a \mathrel{R} b$ se b è multiplo intero di a (cosa che si denota con $a|b$), è una relazione d'ordine; non è, però, totale (ad esempio, 2 e 3 non sono paragonabili). Spesso si considerano particolari sottoinsiemi, ordinati tramite la divisibilità, gli insiemi dei divisori di un fissato numero naturale n . In simboli:

$$D_n = \{a \in N / a|n\}$$

Ad esempio:

$$D_{36} = \{1, 2, 3, 4, 6, 9, 12, 18, 36\}$$

Il suo diagramma di Hasse è:



6 Le funzioni o applicazioni

Siano A e B due insiemi non vuoti tali che $x \in A$ e $y \in B$; si dice *funzione* o *applicazione* di A in B una legge che ad ogni elemento x di A fa corrispondere uno ed un solo elemento y di B e si indica in questo modo:

$$f: A \rightarrow B$$

La relazione f , che ad ogni $x \in A$ fa corrispondere $y \in B$, si indica in questo modo:

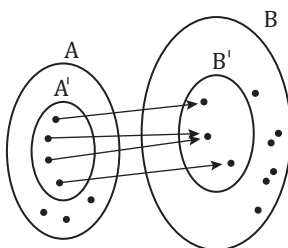
$$y = f(x)$$

in cui $y = f(x) \in B$ è l'immagine, mediante f , di x .

La x si chiama *variabile indipendente*, mentre la y si chiama *variabile dipendente*.

Il sottoinsieme A' di A , costituito dagli elementi $x \in A$ per cui esiste l'immagine $y = f(x)$ in B , si chiama *campo di esistenza* o *insieme di definizione* della funzione f .

Il sottoinsieme B' di B costituito da elementi corrispondenti di qualche elemento di A si dice *immagine* o *codominio* di f .



Una applicazione di A in B si dice *iniettiva* o una *iniezione* se per $a, b \in A$ si ha:

$$f(a) \neq f(b) \quad \text{con } f(a), f(b) \in B$$

Graficamente, da ciascun elemento di A prende il via una sola freccia e a ciascun elemento di B giunge al più una freccia.

Una applicazione di A in B si dice *suriettiva* o una *suriezione* di A su B se ogni elemento di B è l'immagine di almeno un elemento di A , ossia se il codominio $f(x)$ di f coincide con tutto l'insieme B .

Graficamente, da ciascun elemento di A prende il via una ed una sola freccia, e a ciascun elemento di B giunge almeno una freccia.

Una applicazione si dice *biunivoca* se è contemporaneamente iniettiva e suriettiva, per cui, a ciascun elemento di A corrisponde uno ed un solo elemento di B e ciascun elemento di B è il corrispondente di uno ed un solo elemento di A .

Graficamente, da ciascun elemento di A prende il via una ed una sola freccia e a ciascun elemento di B giunge una ed una sola freccia.

Poiché, in qualunque modo si fissi $y \in B$ esiste un unico elemento $x \in A$ tale che $f(x) = y$, resta individuata una funzione di B in A :

$$f^{-1} : B \rightarrow A$$

denominata *funzione inversa* di f , che, ovviamente, è biunivoca e, per ogni $x \in A$ e $y \in B$, si ha:

$$f^{-1}[f(x)] = x \quad \text{e} \quad f^{-1}[f(y)] = y$$

Siano dati tre insiemi A, B, C e due funzioni:

$$f : A \rightarrow B \quad \text{e} \quad g : B \rightarrow C$$

preso x in A , si consideri $f(x) \in B$; a quest'ultimo elemento si applichi la funzione g , ottenendo l'elemento $g[f(x)] \in C$. In tal modo si è costruita la funzione di A in C , che si indica nel modo seguente:

$$g \circ f : A \rightarrow C$$

che si chiama *funzione composta* di f e g , e si legge *g composto f* .

7 Cardinalità di un insieme, insiemi finiti e insiemi infiniti

Per stabilire se un insieme A è *finito* o *infinito* esistono due vie: l'una si basa sull'insieme dei numeri naturali e dà la nozione di insieme finito e da questa, per negazione, definisce un insieme infinito; l'altra si basa sul concetto di parte propria di un insieme. Consideriamole entrambe.

Un dato insieme A è finito se esiste un numero naturale n tale che A possa essere messo in corrispondenza biunivoca con l'insieme $\{1, 2, \dots, n\}$ dei primi n numeri naturali.

Il numero n è denominato *numero degli elementi di A* o *numero cardinale di A* e si dice che A ha *potenza n* , dove per *potenza* di un insieme si intende il numero degli elementi dell'insieme.

Per negazione, *un insieme A si dice infinito se non è finito.*

La cardinalità di un insieme finito si dice *cardinalità naturale*; quella di un insieme infinito si dice *transfinita*.

I teoremi seguenti consentono di ottenere la definizione di insieme infinito senza passare per la corrispondente definizione di insieme finito basata sui numeri naturali.

TEOREMA • *Ogni parte di un insieme finito è anch'essa finita. Di conseguenza, ogni insieme che contenga una parte infinita è necessariamente infinito.*

È ovvio che due insiemi finiti si possono mettere in corrispondenza biunivoca se e solo se sono costituiti dallo stesso numero di elementi, ossia presentano lo stesso numero cardinale; di conseguenza, un insieme finito non può essere messo in corrispondenza con una sua parte propria; da ciò discende il secondo teorema.

TEOREMA • *Ogni insieme che si possa mettere in corrispondenza biunivoca con una sua parte propria è infinito.*

Il cui inverso è il seguente:

TEOREMA • *Ogni insieme infinito può essere messo in corrispondenza biunivoca con una sua parte propria.*

Sulla base del secondo e del terzo teorema si afferma che:

Condizione necessaria e sufficiente affinché un insieme sia infinito è che esso possa essere messo in corrispondenza biunivoca con una sua parte propria.

8 Confronto tra insiemi infiniti e potenza di insiemi

Un insieme i cui elementi sono in corrispondenza biunivoca con l'insieme dei numeri naturali si dice *numerabile*.

Pur esistendo insiemi infiniti non numerabili, come l'insieme R dei numeri reali, si dimostra il seguente teorema.

TEOREMA • *Ogni insieme infinito contiene un sottoinsieme numerabile.*

In altri termini, il teorema enuncia che, effettuando un confronto tra un qualunque insieme infinito e un insieme numerabile, esiste sempre una corrispondenza biunivoca tra una parte di un insieme infinito e un insieme numerabile.

Analogamente, sulla base dei seguenti teoremi è possibile confrontare due qualunque insiemi infiniti.

TEOREMA • *Qualunque siano i due insiemi A e B , o esiste una funzione biunivoca di A in B o esiste una funzione biunivoca di B in A .*

TEOREMA • *Qualunque siano i due insiemi A e B , se esistono una funzione biunivoca di A in B e una funzione biunivoca di B in A , esiste anche una funzione biunivoca di A su tutto B .*

Sulla base dei due teoremi appena enunciati si afferma che:

- *due insiemi A e B hanno la stessa potenza, o, il che è lo stesso, sono equipotenti, se esiste una corrispondenza biunivoca tra A e B ;*
- *l'insieme A ha una potenza minore o uguale di quella di B se esiste una funzione biunivoca di A in B .*

Pertanto, dire che ogni insieme infinito contiene un sottoinsieme numerabile equivale a dire che *ogni insieme infinito ha una potenza maggiore o uguale di quella del numerabile*, dove per *potenza del numerabile* si intende la potenza dell'insieme N dei numeri naturali, o comunque, di qualunque altro insieme che si possa mettere in corrispondenza biunivoca con l'insieme N dei numeri naturali. Tale risultato è opera del matematico tedesco G. Cantor (1845-1918), i cui studi sul concetto di *infinito* e sulla *teoria degli insiemi* in generale, sono stati importantissimi.

Egli dimostra, altresì, mediante il cosiddetto *procedimento per diagonale*, che anche l'insieme dei numeri razionali (assoluti o relativi) ha la stessa potenza del numerabile.

È di Cantor la dimostrazione del seguente teorema.

TEOREMA • *L'insieme di tutte le parti di un insieme ha una potenza maggiore dell'insieme stesso.*

Il teorema, in altre parole, enuncia che, dato un qualunque insieme infinito, esiste sempre un insieme avente potenza maggiore di esso, e tale è l'insieme costituito dalle parti dell'insieme stesso.

ESEMPIO 7

Sia dato l'insieme $A = \{a, b, c, d\}$. Esso conta quattro elementi e $2^4 = 16$ parti:

$\{a, b, c, d\};$

$\{a, b, c\}, \{a, b, d\}, \{a, c, d\}, \{b, c, d\};$

$\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\};$

$\{a\}, \{b\}, \{c\}, \{d\};$

$\emptyset.$

In generale, un insieme di n elementi è costituito da 2^n parti.

Si vengono a creare dei nuovi numeri che Cantor chiama *transfiniti*, di cui il più piccolo è il numerabile, e su di essi si crea una vera e propria aritmetica.

Da tutto ciò una nuova nozione di insieme finito, basata su una negazione:

È finito un insieme che non possa essere messo in corrispondenza biunivoca con una delle sue parti proprie.

E si può affermare che *un insieme finito ha una potenza maggiore di ogni suo sottoinsieme proprio.*

Si consideri l'insieme finito $A = \{a, b, c, d, e, f\}$ e il suo sottoinsieme proprio $B = \{b, d, e\}$; tra essi non può porsi una corrispondenza biunivoca (o biiezione), in quanto A contiene un numero maggiore di elementi di B .

L'insieme i cui elementi sono in corrispondenza biunivoca con i numeri reali prende il nome di *continuo*.

È stato sempre Cantor a mostrare che l'insieme dei numeri reali (assoluti o relativi) ha una potenza maggiore del numerabile, detta *potenza del continuo*.

9 Elementi di calcolo combinatorio

a) Disposizioni

Dati n elementi distinti, si chiama *disposizione semplice degli n elementi, presi a k a k* (con $k \leq n$), o della classe k , un gruppo ordinato di k degli n elementi dati. Due di tali disposizioni si ritengono diverse quando differiscono per almeno un elemento, oppure per l'ordine con cui gli elementi si presentano.

Il numero delle *disposizioni semplici di n elementi a k a k* si indica con $D_{n,k}$ ed è pari a:

$$D_{n,k} = n(n-1)(n-2) \dots (n-k+1)$$

Dati n elementi distinti, si dice *disposizione semplice con ripetizione degli n elementi, a k a k* (con k numero intero qualunque) un gruppo ordinato formato con k degli n elementi potendo uno stesso elemento figurare nel gruppo fino a k volte. Due di tali disposizioni si ritengono diverse quando esistono due elementi distinti che occupano nelle disposizioni il medesimo posto.

Il numero delle *disposizioni con ripetizione di n elementi distinti della classe k* si indica con ${}_r D_{n,k}$ ed è pari a:

$${}_r D_{n,k} = n^k$$

b) Permutazioni

Si chiamano *permutazioni semplici di n elementi* distinti le disposizioni semplici degli n elementi, presi ad n ad n . In altre parole, si può dire anche che le *permutazioni di n elementi* distinti sono tutti i gruppi di n elementi che si possono formare con gli elementi dati, e che differiscono

tra loro soltanto per l'ordine degli elementi. Indicando con P_n il numero delle permutazioni di n elementi, si ha:

$$P_n = (n-1)(n-2) \dots 3 \cdot 2 \cdot 1$$

Le permutazioni possono essere indicate anche con la seguente scrittura:

$$P_n = n!$$

dove il simbolo $n!$ si chiama *fattoriale* del numero n ed è pari al prodotto dei primi n numeri naturali.

Per convenzione si pone $0! = 1$

Le *permutazioni cicliche di n oggetti*, indicate con ${}_c P_n$, sono le permutazioni di n elementi lungo una circonferenza (o circuito chiuso) e sono pari a:

$${}_c P_n = \frac{P_n}{n} = (n-1)!$$

c) Combinazioni semplici

Dati n elementi distinti, si chiama *combinazione semplice degli n elementi, presi a k a k , o della classe k* (con $k \leq n$), un qualunque gruppo formato da k degli n elementi dati. Due di tali combinazioni si ritengono distinte quando differiscono fra loro per almeno un elemento.

Il numero di *combinazioni semplici di n elementi a k a k* si indica con $C_{n,k}$ ed è pari a:

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n(n-1)(n-2) \dots (n-k+1)}{k!}$$

Il numero delle combinazioni semplici di n oggetti a k a k , si indica anche con:

$$\binom{n}{k}$$

e si chiama *coefficiente binomiale*.

Quando l'ordine non è importante ma è possibile avere elementi ripetuti, si parla di *combinazioni con ripetizione*. Il numero di *combinazioni con ripetizione di n oggetti di classe k* si indica con ${}_r C_{n,k}$ ed è pari a:

$${}_r C_{n,k} = C_{n+k-1,k} = \binom{n+k-1}{k}$$

d) Formula del binomio di Newton

La potenza n -esima di una espressione algebrica binomiale del tipo $a + b$ è un *polinomio omogeneo* di grado n , ordinato secondo le potenze decrescenti di a e crescenti di b , e formato da $n + 1$ termini, i cui coefficienti sono dati dal numero di combinazioni di n oggetti di classe, rispettivamente, $0, 1, 2, \dots, n$. Inoltre, in ogni termine la somma degli esponenti di a e di b è uguale a n , ossia:

$$(a + b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{n} b^n$$

Questa viene, impropriamente, detta *formula del binomio di Newton*, e più sinteticamente può essere scritta nel seguente modo:

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

Per il coefficiente binomiale valgono le seguenti proprietà notevoli:

$$- \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$- \binom{n}{k} = \binom{n}{n-k}$$

$$- \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

Dalla seconda delle proprietà notevoli appena esposte, si ha che:

$$\binom{n}{0} = \binom{n}{n} = 1$$

Inoltre:

— la somma dei coefficienti binomiali dello sviluppo della potenza $(a + b)^n$ è uguale a:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n$$

— la somma dei coefficienti dei termini di posto pari è uguale alla somma dei coefficienti dei termini di posto dispari:

$$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \dots$$

e) Regole per lo sviluppo della potenza di un binomio

I coefficienti dello sviluppo della potenza $(a + b)^n$ si possono ottenere in due modi diversi.

— Ricorrendo al *triangolo aritmetico* detto anche, impropriamente, *triangolo di Tartaglia*:

				1				
				1		1		
			1	2		1		
		1	3	3		1		
	1	4	6	4		1		
1	5	10	10	5		1		
1	6	15	20	15	6	1		

— Il coefficiente di un termine qualunque si ottiene moltiplicando il coefficiente del termine precedente per il rapporto tra l'esponente del termine precedente e il numero dei termini già scritti: ciò può essere verificato considerando che lo sviluppo della potenza $(a + b)^n$ è uguale a:

$$(a + b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{1 \cdot 2}a^{n-2}b^2 + \frac{n(n-1)(n-2) \dots (n-k+1)}{1 \cdot 2 \cdot 3 \dots k}a^{n-k}b^k + \dots + b^n$$

1 La logica

Una volta la *logica* era una parte indiscussa della filosofia, o meglio, era propedeutica non solo ad essa ma ad ogni disciplina. Oggi, la *logica* ha perso la sua posizione predominante in favore della Matematica. Infatti, tra i pensatori contemporanei, solo L. Von Mises ha riaffermato le possibilità conoscitive della logica, disprezzando il metodo matematico.

Dovendo definire in termini pratici, allora, potremmo affermare che *la logica ha a che fare con il ragionamento*, o meglio, *con tutte le attività raziozinanti in genere* che spaziano dall'Informatica alla Chimica, dalla Fisica alla Filosofia stessa.

Il ragionamento tipico di cui si occupa la logica è il *ragionamento deduttivo*, anche se raramente ci si può indirizzare verso il *ragionamento induttivo*.

Ritornando alla Matematica, può essere utile la distinzione effettuata da Peirce tra la matematica come scienza, che deriva conclusioni necessarie, e la logica, che è la scienza del modo in cui derivare conclusioni necessarie.

Quando si parla di logica non si può fare a meno di parlare dei simboli. In genere, i simboli sono un *segno* proprio della Matematica. Al di là di ogni disquisizione, che non è da farsi in questa sede, basti ricordare che, secondo gli studi di logica moderna, i linguaggi simbolici sono portatori di significati precisi e, quindi, vanno usati come sostituti di quelli naturali nel campo delle scienze esatte. A tal proposito, è doveroso ricordare la classica dicotomia *sintassi-semantica*. Per esempio, i linguaggi di programmazione sono artificiali sia nel loro lessico sia nella loro sintassi; infatti, la scrittura di programmi è guidata dalla sintassi, poiché i computer non capiscono il significato ma sono assoggettati a ben determinate e precostituite forme sintattiche.

Si ricordi che i linguaggi simbolici sono stati inventati e utilizzati da matematici e logici quali furono Peano e Frege, con lo scopo puro e semplice di avere un preciso significato; oggi sono, invece, utilizzati dai computer e dalle macchine, per cui è ovvio che il significato non esiste più.

2 La dimostrazione

Caratteristica essenziale e fondamentale della Matematica è la *dimostrazione*; non vi è Matematica senza dimostrazione. Infatti, *un teorema è un teorema solo se è riconducibile ad una dimostrazione*.

La Matematica è calcolo con i numeri nel nostro primo stadio di apprendimento, ma calcolare non è uguale a ragionare! Solo dopo anni di studio accettiamo come matematica un qualcosa di diverso rispetto a numeri, equazioni e rappresentazioni grafiche. Basta fare mente locale e ricordarsi di quante difficoltà hanno i ragazzi in prima liceo di fronte alle dimostrazioni geometriche: erroneamente molti insegnanti definiscono tali dimostrazioni come passaggio da enunciato a enunciato, da imparare a memoria.

3 Nozioni di logica matematica

Le figure e lo studio delle loro proprietà costituiscono, rispettivamente, l'oggetto e il fine della Geometria, così come i numeri con le operazioni numeriche e lo studio delle loro proprietà costituiscono, rispettivamente, l'oggetto e il fine dell'Aritmetica.

In Geometria o in Aritmetica si afferma, ad esempio, che:

- 1) due punti appartengono ad una sola retta;
- 2) in un triangolo la somma degli angoli interni è uguale ad un angolo piatto;
- 3) $3 + 2 = 2 + 3$;
- 4) i numeri primi sono infiniti.

La verità delle affermazioni 1) e 3) può essere stabilita immediatamente, ricorrendo all'intuizione, ma la verità delle affermazioni 2) e 4) non è immediata: essa è *dedotta con il ragionamento* dalla verità di altre affermazioni.

È compito della logica, anche se non l'unico, fornire quelle regole deduttive in base alle quali, muovendo da premesse vere, siamo sicuri di giungere a conclusioni vere. La logica studia i modi di ragionare correttamente.

Sia le affermazioni la cui verità è immediata, sia quelle la cui verità è dedotta sono espresse mediante il linguaggio, scritto o parlato.

Oggetto della logica è il linguaggio.

4 Le proposizioni

«Proposizione» è un insieme di termini linguistici per i quali ha senso parlare di vero o di falso.

I termini *affermazione* ed *enunciato* sono usati come sinonimi di *proposizione* e il termine *proposizione* sarà preferito agli altri due.

Secondo la precedente definizione, non si possono chiamare proposizioni tutti gli insiemi linguistici, anche se sono dotati di soggetto, verbo e predicato e hanno, quindi, sintatticamente, una tipica *forma proposizionale*. D'altra parte, si possono chiamare *proposizioni* altri insiemi linguistici che quella forma tipica non possiedono.

Si considerino, ad esempio, i seguenti insiemi linguistici:

- 1) Alla retta appartengono infiniti punti.
- 2) Sta piovendo.
- 3) La luna è vicina.
- 4) $2 + 3 = 7$.
- 5) La capra campa sopra la panca.

La 1), la 2) e la 4) sono proposizioni, perché per esse si può parlare di vero o di falso; la 5) è, chiaramente, indifferente rispetto alla verità e alla falsità; della 3) non si può dire, con certezza, se è un'affermazione vera o falsa e, dunque, non va considerata una proposizione.

Si dicono *atomiche* quelle proposizioni le cui componenti, prese singolarmente o a gruppi, non possiedono la proprietà di essere vere o false.

Ad esempio, «Il triangolo ha tre angoli» è una proposizione atomica, in quanto, è impossibile isolare, in essa, dei termini per formare altre proposizioni.

Invece, nella proposizione «Gli struzzi sono uccelli e i bisonti sono mammiferi» si riconosce una struttura *composta* o *molecolare*, perché si possono isolare le due proposizioni atomiche «Gli struzzi sono uccelli» e «I bisonti sono mammiferi».

Le proposizioni stabiliscono delle *relazioni* tra elementi. La proposizione «Il sole è più grande della terra» è un esempio di *relazione binaria*, in quanto è stabilita tra due elementi; la proposizio-

ne « $4 + 2 > 1$ », in cui intervengono tre elementi, è un esempio di *relazione ternaria*. Per estensione, le proposizioni della forma «Il sole è una stella» si dicono *relazioni unarie*.

5 I connettivi

Si analizzino le seguenti proposizioni molecolari:

- 1) Il Sole è una stella e Venere è un pianeta del Sole.
- 2) Il pianeta Marte o il pianeta Giove sono visibili di notte.
- 3) Piove, *dunque* la strada è bagnata.
- 4) Un numero « n » è pari se e solo se « n » è divisibile per due.

Ognuna di esse è formata da due proposizioni atomiche, *connesse* o *collegate* dagli elementi *e*, *o*, *dunque*, *se e solo se*.

Nell'esempio 1) la particella *e* collega le due proposizioni atomiche: *il Sole è una stella*, *Venere è un pianeta del Sole*. Nella proposizione 2), apparentemente atomica, la particella *o* collega, di fatto, due proposizioni: *il pianeta Marte è visibile di notte*, *il pianeta Giove è visibile di notte*. Nell'esempio 3) il termine *dunque* collega la proposizione *piove* con la proposizione *la strada è bagnata*. Infine, nell'esempio 4) gli elementi linguistici *se e solo se* associano la proposizione *un numero « n » è pari* alla proposizione *un numero « n » è divisibile per due*.

Dall'analisi della struttura delle proposizioni precedenti risulta che si possono eseguire certe *operazioni* mediante le quali *a due proposizioni (date in un certo ordine) si fa corrispondere una terza e nuova proposizione*.

La funzione di collegamento di due proposizioni è svolta, come si è visto, da particolari elementi linguistici, quali *e*, *o*, *dunque*, *se e solo se*: nel linguaggio della logica, questi elementi sono detti *connettivi* e ad ognuno di essi corrisponde un'*operazione elementare*.

Di seguito sarà analizzato, brevemente, il significato delle *operazioni logiche*; saranno usati nomi e simboli adoperati nella Logica Matematica. L'uso dei simboli ha il vantaggio di evitare imprecisioni ed ambiguità che si possono presentare nel linguaggio comune.

Si indicheranno le proposizioni con le lettere minuscole seguenti *p*, *q*, *r*, *s*, *t*... Volendo, poi, precisare se sono vere o false, si faranno seguire alle proposizioni, rispettivamente, i simboli *v* o *f* racchiusi tra parentesi tonde.

Scrivendo, ad esempio, *p* = Milano è una città italiana (*v*), si intenderà affermare che *p* rappresenta la proposizione «Milano è una città italiana» e che tale proposizione è vera.

a) Congiunzione

La «*coniunzione*» è l'*operazione logica mediante la quale, date due proposizioni p, q, si forma con queste una nuova proposizione r, che si considera:*

vera $\rightarrow$ se *p* e *q* sono entrambe vere,
falsa $\rightarrow$ in ogni altro caso.

N.B.: *r* è, quindi, falsa anche se una sola delle due componenti è falsa.

Si scrive:

$$r = p \wedge q$$

che si interpreta: «*r* è la proposizione ottenuta dalla congiunzione delle proposizioni *p* e *q*» e che si legge, brevemente, «*r* è uguale a *p* et *q*».

Il nome del simbolo di operazione « $\wedge$ » è, appunto, quello della congiunzione latina «et» e qualche volta, nella scrittura, si sostituirà il termine «et» al simbolo « $\wedge$ ».

Dalle proposizioni seguenti:

p = 20 è multiplo di 4 (v),

q = 20 è multiplo di 10 (v),

r = Dante è vivo (f),

s = Dante ha scritto la Divina Commedia (v),

t = Garibaldi ha dipinto la Gioconda (f),

si ottengono:

$p \wedge q$ = 20 è multiplo di 4 e di 10 (v),

$r \wedge s$ = Dante è vivo e ha scritto la Divina Commedia (f),

$r \wedge t$ = Dante è vivo e Garibaldi ha dipinto la Gioconda (f).

b) Alternazione

L'«alternazione» è l'operazione logica mediante la quale, date due proposizioni p , q , si forma con queste una nuova proposizione r , che si considera:

falsa $\rightarrow$ se p e q sono entrambe false,

vera $\rightarrow$ in ogni altro caso.

N.B.: « r » è, cioè, vera quando *almeno* una delle proposizioni è vera.

Si scrive:

$$r = p \vee q$$

che va interpretata così: « r è la proposizione ottenuta dall'alternazione delle proposizioni p e q », e che si legge, brevemente, « r è uguale a p vel q ».

Il nome, e in certo senso il significato, del simbolo di operazione « $\vee$ » è quello della particella latina «vel» (il simbolo « $\vee$ » lo sta appunto a ricordare). La particella «vel» ha, in latino, un significato più ristretto e quindi più preciso della corrispondente particella italiana «o», e come tale è più adatta a rappresentare questa particolare operazione logica.

Dalle proposizioni seguenti:

p = 5 è un numero dispari (v),

q = 6 è un numero pari (v),

t = 5 è un numero pari (f),

s = $3 < 2$ (f),

si ottengono:

$p \vee q$ = 5 è un numero dispari o 6 è un numero pari (v),

$p \vee t$ = 5 è un numero dispari o un numero pari (v),

$t \vee s$ = 5 è un numero pari o $3 < 2$ (f).

c) Implicazione

L'«implicazione» è l'operazione logica mediante la quale, date due proposizioni p , q , si forma con esse una nuova proposizione r , che si considera:

falsa $\rightarrow$ se p è vera e q è falsa,

vera $\rightarrow$ in ogni altro caso.

Si scrive:

$$r = p \Rightarrow q$$

che si interpreta così: « r è la proposizione ottenuta dall'implicazione delle proposizioni p e q », dove « $\Rightarrow$ » è il simbolo dell'operazione «implicazione».

La scrittura:

$$p \Rightarrow q$$

si legge con l'una o l'altra delle locuzioni: «se p , allora q », « p , dunque q », « p implica q », «da p segue q ».

Le perplessità che può suscitare la definizione di implicazione logica possono essere attenuate se si pensa che la logica non abbraccia solo la forma del linguaggio matematico, ma si estende a tutto il linguaggio dell'uomo. In tale ambito è facile trovare, ad esempio, affermazioni come questa: «se sono una rondine, allora volo», dove le componenti «io sono una rondine» e «io volo» sono evidentemente false, ma — in quanto legate dal connettivo «se... allora» — costituiscono un'affermazione tranquillamente accettata e considerata vera.

Interessano la Matematica i seguenti casi particolari.

a) Se, essendo p vera, q è vera, allora la proposizione:

$$p \Rightarrow q$$

è vera; in questo caso, si può anche leggere «dalla verità di p segue la verità di q ».

b) Se, essendo p vera, q è falsa, allora la proposizione:

$$p \Rightarrow q$$

è falsa. Questo caso riflette la circostanza, comunemente accettata, che dalla verità non può conseguire la falsità.

Dalle proposizioni seguenti:

p = gli angoli alla base di un triangolo isoscele sono acuti (v),

q = gli angoli alla base di un triangolo isoscele danno per somma un angolo piatto (f),

s = gli angoli alla base di un triangolo isoscele danno per somma un angolo inferiore ad un angolo piatto (v),

si ottengono:

$p \Rightarrow q$ = se gli angoli alla base di un triangolo isoscele sono acuti, allora danno per somma un angolo piatto (f),

$p \Rightarrow s$ = se gli angoli alla base di un triangolo isoscele sono acuti, allora danno per somma un angolo inferiore ad un angolo piatto (v).

d) Coimplicazione

La «coimplicazione» è l'operazione logica mediante la quale, date due proposizioni, p , q , si forma con queste una nuova proposizione r , che si considera:

vera $\rightarrow$ se p e q sono entrambe vere o entrambe false,
falsa $\rightarrow$ in ogni altro caso.

Si scrive:

$$r = p \Leftrightarrow q$$

che va interpretata così: « r è la proposizione ottenuta dalla coimplicazione delle proposizioni p e q », dove « $\Leftrightarrow$ » è il simbolo dell'operazione «coimplicazione».

La scrittura:

$$p \Leftrightarrow q$$

si legge usando l'una o l'altra delle locuzioni: « p coimplica q », « p implica reciprocamente q », « p se e solo se q », « p equivale a q ».

Si considera, ora, solo il caso — in Matematica più significativo — in cui, essendo p e q entrambe vere, la *coimplicazione* conduce ad una proposizione vera.

Dalle proposizioni:

p = un triangolo equilatero è equiangolo (v),

q = un triangolo equiangolo è equilatero (v),

si ottiene:

$p \Leftrightarrow q$ = un triangolo equilatero è equiangolo *implica reciprocamente che* un triangolo equiangolo è equilatero.

e) Negazione

Si consideri la proposizione «Roma *non* è la capitale d'Italia». Il termine «non», preposto al predicato «è la capitale d'Italia», ha la funzione di formare, data la proposizione «Roma è la capitale d'Italia», una *nuova proposizione*: «Roma *non* è la capitale d'Italia». In tal senso è giustificato affermare che anche la *negazione* è un'*operazione logica*, dando al concetto di operazione un significato più ampio di quello usato di solito in Matematica, dove il concetto di *operazione* va inteso come una corrispondenza che, a due elementi di un insieme, fa corrispondere un solo elemento del medesimo insieme.

A questo punto, è d'uopo dare la seguente definizione:

«Negazione di una proposizione» è l'*operazione logica mediante la quale, data una proposizione p , se ne ottiene un'altra, indicata con la scrittura « $\neg p$ » (si legge «non p ») che è:*

vera $\rightarrow$ se p è falsa,

falsa $\rightarrow$ se p è vera.

In luogo di « $\neg p$ » si trovano talora usate anche le scritture « $\bar{p}$ » e « $\sim p$ ».

Le due proposizioni p e $\neg p$ si dicono *proposizioni contrarie* oppure *opposte*.

Esempi:

p = l'Italia è una penisola (v),

$\neg p$ = l'Italia *non* è una penisola (f),

q = il triangolo ha due angoli (f),

$\neg q$ = il triangolo *non* ha due angoli (v).

In merito agli esempi ora citati, si nota che le proposizioni $\neg p$ e $\neg q$ si possono tradurre, nel linguaggio della Matematica, anche così: «*non è vero che* l'Italia è una penisola» e «*non è vero che* il triangolo ha due angoli».

Di seguito sono riportate delle *tabelle di operazione*, in cui sono raccolte le definizioni delle *operazioni logiche elementari* (con la notazione che p e q sono proposizioni, v = vero, f = falso).

• Congiunzione

p	q	$p \wedge q$
v	v	V
v	f	F
f	v	F
f	f	F

- Alternazione

p	q	$p \vee q$
v	v	V
v	f	V
f	v	V
f	f	F

- Implicazione

p	q	$p \Rightarrow q$
v	v	V
v	f	F
f	v	V
f	f	V

- Coimplicazione

p	q	$p \Leftrightarrow q$
v	v	V
v	f	F
f	v	F
f	f	V

- Negazione

p	$\neg p$
v	f
f	v

Con evidente abuso di linguaggio, ma con notevole semplicità ed economicità di discorso, alle *proposizioni*:

$$p \wedge q, \quad p \vee q, \quad p \Rightarrow q, \quad p \Leftrightarrow q, \quad \neg p$$

si darà, talvolta, rispettivamente il nome di *coniunzione*, *alternazione*, *implicazione*, *coimplicazione*, *negazione*, termini che spettano, invece, alle corrispondenti operazioni.

6 Dimostrazione, teorema, lemma e corollario

Se dalla verità di una proposizione p si *deduce con il ragionamento* la verità di una proposizione q , si dice, in Matematica, che l'*implicazione*:

$$p \Rightarrow q \tag{1}$$

è stata *dimostrata*.

Alla proposizione (1) allora, in Matematica, si attribuisce il nome di *teorema*; in particolare, delle due componenti la proposizione (1), p si chiama *ipotesi* o *premessa* o *precedente* o *punto di partenza*, e q si chiama *tesi* o *conseguenza* o *conseguente* o *punto di arrivo* o *conclusione*.

La *dimostrazione* è, ovviamente, l'insieme dei *passaggi logici* che dall'*ipotesi* conduce alla *tesi*.

Un teorema è, dunque, una proposizione che si può schematizzare nella forma dell'implicazione: il soggetto di tale proposizione è l'«ipotesi» e il predicato è «segue la tesi».

Spesso i teoremi non sono enunciati esplicitamente sotto forma di implicazione; è utile, tuttavia, che lo studente si abitui a rappresentarli nella forma suddetta. Così, ad esempio, la proposizione «La somma degli angoli interni di un triangolo è uguale ad un angolo piatto» può essere tradotta nella forma seguente: «Se tre angoli sono gli angoli interni di un triangolo, allora la loro somma è un angolo piatto».

Un teorema, essendo una proposizione, può essere vero o falso.

È vero quando p (ipotesi) e q (tesi) sono vere.

È falso quando p è vera e q è falsa.

Così, ad esempio, il teorema ora citato è un teorema vero, mentre la proposizione «se un triangolo è equilatero, allora la somma dei suoi angoli interni è un angolo giro» è un teorema falso.

Si chiama *lemma* un teorema il cui scopo principale sia quello di facilitare la dimostrazione di altri teoremi più importanti.

Si chiama, invece, *corollario* un teorema che consegua immediatamente dalla dimostrazione di un altro teorema.

7 La proprietà transitiva della deduzione

Si ammetta, come evidente, la verità di quest'affermazione:

«se dalla verità di p si deduce la verità di q e se dalla verità di q si deduce quella di r , allora dalla verità di p si deduce la verità di r ».

Tale affermazione esprime la *proprietà transitiva della deduzione* e si traduce, in simboli, nello schema seguente:

$$(p \Rightarrow q \text{ et } q \Rightarrow r) \Rightarrow (p \Rightarrow r)$$

L'operazione mentale che, dalla verità di p , ci conduce ad affermare la verità di r è un esempio di *ragionamento*.

8 La teoria assiomatica o ipotetico-deduttiva (i postulati)

È facile immaginare che le proposizioni che si *possono ricavare o dedurre* mediante la proprietà transitiva della deduzione sono, almeno teoricamente, innumerevoli. Per cui in un procedimento deduttivo, che si suole chiamare *in discesa*, non ci sono ultime proposizioni; invece, in un procedimento *in salita*, alla ricerca delle *proposizioni madre*, ci si deve naturalmente arrestare di fronte ad alcune *proposizioni primitive*, la cui verità deve essere ammessa perché non è possibile dedurla da alcun'altra proposizione precedente.

A tali proposizioni primitive, che costituiscono la base delle successive deduzioni, si dà — in Matematica — il nome di *postulati* o *assiomi*.

I postulati sono il fondamento dell'edificio deduttivo.

L'insieme dei postulati e delle proposizioni che da essi si deducono (i teoremi) si chiama *teoria assiomatica* o *teoria ipotetico-deduttiva*.

Si può intuire che una teoria assiomatica dipende dal sistema di postulati che si stabiliscono all'inizio e che, cambiando i postulati, cambia anche la teoria che su essi si fonda.

9 I postulati fondamentali della logica

Anche la logica ha i suoi postulati, che, di solito, sono chiamati *principi*: principio e postulato sono, in questo senso, sinonimi.

Interessa pronunciare i seguenti principi (postulati) della logica.

PRINCIPIO D'IDENTITÀ • *Ogni cosa è identica (uguale) a se stessa, e solo a se stessa.*

Tale principio si suole esprimere, simbolicamente, così:

$$A = A$$

dove il segno « $=$ » si legge «è identico» o «è uguale»; con tale segno si rappresenterà sempre e soltanto la *relazione di uguaglianza logica* o *relazione di identità*.

Allora, alla scrittura:

$$A = B$$

va data la seguente interpretazione:

«*A e B indicano la stessa cosa*»

oppure:

«*A e B sono due nomi diversi della stessa cosa*».

PRINCIPIO DI NON CONTRADDIZIONE (o **PRINCIPIO DI CONTRADDIZIONE**) • *Una cosa non può insieme «essere» e «non essere».*

Facendo riferimento alle proposizioni, tale principio si può esprimere così: «*una proposizione non può, nello stesso tempo, essere vera ed essere falsa*», in altre parole «*le due proposizioni p e $\neg p$ non possono essere contemporaneamente vere*».

Una *teoria* si dice *contraddittoria* se due proposizioni del tipo p e $\neg p$ si considerano contemporaneamente vere.

PRINCIPIO DEL TERZO ESCLUSO • *Una proposizione è vera o (vel) falsa.*

Di una proposizione, cioè, si può sostenere che è vera o che è falsa, ma non si può sostenere che è «quasi vera», «più falsa che vera», «indeterminata» etc.

Ammettere questo principio significa assumere, come si fa in Matematica, una logica a due valori (il vero e il falso).

10 Il metodo di riduzione all'assurdo

Siano t una proposizione, T una teoria e p una proposizione di T . Si consideri la teoria T' ottenuta aggiungendo a T la proposizione $\neg t$. Se da T' si deduce $\neg p$, T' è contraddittoria (contenendo p e $\neg p$) e allora da T segue t .

In altre parole:

se dall'essere vere $\neg t$ e le proposizioni di T segue che sono vere p e $\neg p$, allora dall'essere vere le proposizioni di T segue che è vera t .

Questo *principio di logica* afferma, essenzialmente, questo: se vi è contraddizione nelle conseguenze, vi è contraddizione anche nelle premesse.

Quando un teorema si dimostra ricorrendo a questo metodo, si sostiene che si è fatta una *dimostrazione per assurdo*.

11 Le implicazioni derivate

Data l'implicazione:

1) $p \Rightarrow q$,

interessano particolarmente la Matematica le seguenti *implicazioni* che da essa, denominata *implicazione diretta*, si ricavano:

2) $q \Rightarrow p$, che è detta *implicazione inversa* della 1);

3) $\neg p \Rightarrow \neg q$, che è detta *implicazione contraria* della 1);

4) $\neg q \Rightarrow \neg p$, che è detta *implicazione contrapposta* della 1).

12 Il teorema inverso o reciproco

Se « $p \Rightarrow q$ » è un teorema vero, l'implicazione inversa « $q \Rightarrow p$ » può essere vera o falsa.

Nel caso in cui dalla verità di q si può dedurre, con il ragionamento, la verità di p , si sostiene che si è dimostrato il «teorema inverso» o «reciproco» del teorema diretto « $p \Rightarrow q$ ».

L'attribuzione di *diretto* o *inverso* data ad un teorema è, di solito, soggettiva; si può, infatti, chiamare *diretto* quello che ha la forma « $q \Rightarrow p$ » ed *inverso* quello che ha la forma « $p \Rightarrow q$ ».

Esempi:

1) $p \Rightarrow q = \text{se } ABC \text{ è un triangolo equilatero, allora } ABC \text{ è un triangolo equiangolo } (v)$,
 $q \Rightarrow p = \text{se } ABC \text{ è un triangolo equiangolo, allora } ABC \text{ è un triangolo equilatero } (v)$ (teorema inverso di « $p \Rightarrow q$ »);

2) $r \Rightarrow s = \text{se } n \text{ è un numero multiplo di } 10, \text{ allora } n \text{ è un numero pari } (v)$,
 $s \Rightarrow r = \text{se } n \text{ è un numero pari, allora } n \text{ è un multiplo di } 10 \text{ (f) (essendo falsa, questa proposizione non è il teorema inverso di «} r \Rightarrow s \text{»)}.$

13 Le proposizioni equivalenti

Quando sono veri sia il teorema diretto, $p \Rightarrow q$, sia quello inverso, $q \Rightarrow p$, si usa scrivere:

$$p \Leftrightarrow q$$

che si legge adoperando l'una o l'altra delle seguenti locuzioni: «da p si deduce q e viceversa», «da p si deduce reciprocamente q », « p equivale a q », « p è vera se e solo se è vera q », « p coimplica q ».

Di p e q si dice che sono due proposizioni *logicamente equivalenti* o semplicemente *equivalenti*, e il simbolo « $\Leftrightarrow$ » è detto anche *simbolo di equivalenza logica*.

Se q indica una proprietà di un dato soggetto espresso nella p , allora, per quel soggetto, q è una *proprietà caratteristica*.

Esempi di proposizioni equivalenti:

1) $p \Leftrightarrow q = ABC \text{ è un triangolo equilatero se e solo se } ABC \text{ è un triangolo equiangolo}$;
 2) $r \Leftrightarrow s = \text{se due rette sono parallele, allora formano angoli coniugati supplementari, e viceversa}$;
 3) $t \Leftrightarrow u = \text{la proposizione «un punto } P \text{ appartiene all'asse di un segmento» implica reciprocamente la proposizione «un punto è equidistante dagli estremi di quel segmento»}.$

Sulla base del simbolo $\nLeftrightarrow$, che si legge «non è equivalente a», si consideri il seguente controesempio:

$$v \nLeftrightarrow z = \text{«un numero } n \text{ è multiplo di } 4\text{» non è equivalente a «un numero } n \text{ è pari»}.$$

Nelle dimostrazioni matematiche, ma talora anche in un discorso di carattere comune, ad una proposizione si può sostituire una proposizione equivalente.

14 La prima legge delle inverse

Se « $p \Rightarrow q$ » è un teorema diretto, allora « $\neg q \Rightarrow \neg p$ » si dice *teorema contrapposto* o *teorema contrario dell'inverso*.

Si dimostra ora che:

«se un teorema è vero, è vero anche il suo contrapposto»,

cioè, in simboli:

$$(p \Rightarrow q) \Rightarrow (\neg q \Rightarrow \neg p)$$

La dimostrazione si fa ricorrendo al *metodo di riduzione all'assurdo*.

Si suppongano vere la *coniunzione* $(p \Rightarrow q) \wedge \neg q$ e la proposizione p .

Dall'essere vere « $p \Rightarrow q$ » e p segue che è vera q : ma allora sono vere contemporaneamente, q e $\neg q$, il che è escluso dal principio di non contraddizione.

Pertanto, dall'essere vere « $p \Rightarrow q$ » e $\neg q$ segue che è vera $\neg p$.

Con lo stesso ragionamento, essendo « $p \Rightarrow q$ » contrapposto a « $\neg q \Rightarrow \neg p$ » [si ammette che $\neg(\neg p) = p$, e cioè che la negazione di una negazione è un'affermazione], segue che è vero il teorema inverso:

$$(\neg q \Rightarrow \neg p) \Rightarrow (p \Rightarrow q)$$

In definitiva, è vera l'equivalenza:

$$(p \Rightarrow q) \Leftrightarrow (\neg q \Rightarrow \neg p)$$

Tale equivalenza esprime la cosiddetta *prima legge delle inverse* che, nel linguaggio a noi ormai abituale, si esprime così:

«Se un teorema è vero, è vero anche il suo contrapposto, e viceversa».

ESEMPIO

Se un parallelogramma è un rombo, le diagonali sono perpendicolari; ciò equivale a: se le diagonali di un parallelogramma *non* sono perpendicolari, il parallelogramma *non* è un rombo.

Dalla prima legge delle inverse segue, immediatamente, il seguente corollario.

COROLLARIO • Se di un teorema è vero l'inverso, è vero anche il contrario, e viceversa.

Infatti « $q \Rightarrow p$ » (teorema inverso) e « $\neg p \Rightarrow \neg q$ » (teorema contrario) sono l'uno contrapposto all'altro.

15 La seconda legge delle inverse

Si supponga che:

- 1) $I_1, I_2, \dots, I_n$ siano tutte le possibili ipotesi che si possono formulare sopra uno stesso soggetto S ;
- 2) $T_1, T_2, \dots, T_n$ siano altrettante tesi che si deducono, rispettivamente, da $I_1, I_2, \dots, I_n$, cioè:

$$I_1 \Rightarrow T_1$$

$$I_2 \Rightarrow T_2$$

.....

$$I_n \Rightarrow T_n$$

- 3) le tesi $T_1, T_2, \dots, T_n$ si escludano a vicenda.

Allora sono veri i teoremi:

$$\begin{array}{l} T_1 \Rightarrow I_1 \\ T_2 \Rightarrow I_2 \\ \dots\dots\dots \\ T_n \Rightarrow I_n \end{array}$$

Si dimostra, ad esempio, il primo teorema $T_1 \Rightarrow I_1$ (in modo analogo si dimostrano gli altri teoremi).

La dimostrazione si fa ricorrendo al *metodo di riduzione all'assurdo*.

Se « $T_1 \Rightarrow I_1$ » non è vero (e cioè T_1 è vero e I_1 è falso), vi sarà almeno un caso in cui, assieme al verificarsi di T_1 , si verifica, per S , una delle ipotesi 1) diversa da I_1 .

Sia, ad esempio:

$$T_1 \Rightarrow I_2$$

ma, per l'ipotesi 2):

$$I_2 \Rightarrow T_2$$

quindi, per la proprietà transitiva della deduzione:

$$T_1 \Rightarrow T_2$$

cioè T_1 e T_2 sono contemporaneamente vere, contro l'ipotesi 3), secondo la quale T_1 e T_2 si escludono a vicenda.

Ad analoga considerazione si sarebbe giunti se nella implicazione si fosse associata a T_1 una qualsiasi delle ipotesi 1) diversa da I_1 .

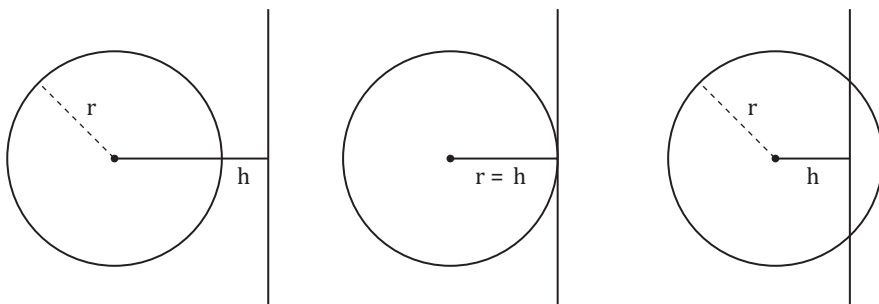
Pertanto:

$$T_1 \Rightarrow I_1$$

Il teorema ora dimostrato prende il nome di *seconda legge delle inverse* e trova larga applicazione in Geometria.

Ad esempio, osservando la figura, siano veri i teoremi:

- 1) $(I_1 \Rightarrow T_1) = (h > r \Rightarrow \text{la retta è esterna alla circonferenza})$;
- 2) $(I_2 \Rightarrow T_2) = (h = r \Rightarrow \text{la retta è tangente alla circonferenza})$;
- 3) $(I_3 \Rightarrow T_3) = (h < r \Rightarrow \text{la retta è secante alla circonferenza})$.



Siccome le ipotesi I_1, I_2, I_3 sono tutte le possibili ipotesi che si possono fare su r e h , e le tesi T_1, T_2, T_3 si escludono a vicenda, per la seconda legge delle inverse risultano veri i teoremi inversi e cioè:

$$(T_1 \Rightarrow I_1) = (\text{la retta è esterna alla circonferenza} \Rightarrow h > r) \text{ etc.}$$

16 Concetti primitivi e definizioni

Si *definisce* un *concetto*, di significato ignoto, quando lo si *mette in relazione* con altri concetti di significato noto.

Per il procedimento di *definizione* si presenta, in ogni teoria, una situazione analoga a quella del dimostrare: come è impossibile dimostrare tutte le proprietà di una teoria, così è impossibile definire tutti i concetti; e, allo stesso modo in cui si ammettono come vere alcune proposizioni (*postulati*), che servono di base per le successive deduzioni, così, all'inizio di ogni teoria, bisogna ammettere come noto il significato di alcuni concetti, che servono poi di base per le successive *definizioni*.

A tali concetti, il cui significato deve ritenersi conosciuto, si dà il nome di *concetti primitivi*.

In Geometria (*teoria delle figure*) sono primitivi, ad esempio, i concetti di *punto*, *retta* e *piano*; in Aritmetica (*teoria dei numeri*) è primitivo, ad esempio, il concetto di *insieme finito*.

Comuni a tutte le teorie sono, ad esempio, i concetti primitivi di *essere*, *concetto* o *idea*, *oggetto* o *ente* o *elemento*, *vero* e *falso*, *appartenere*; le parole *e*, *o*, *non*; *affermare*, *dedurre*; *corrispondenza* etc.

Le *definizioni* di una teoria sono i concetti non primitivi. Sono definiti, ad esempio, i concetti di *semiretta*, *segmento*, *numero irrazionale*, *numero relativo* etc.

17 Il concetto di astrazione

Non si può pensare alla natura della Matematica senza pensare al concetto di astrazione. Dire che la Matematica è astratta è diventato in verità un luogo comune, tanto che spesso si sente usare questa frase per dire che la Matematica è lontana dalla realtà, è arida, fredda, senza vita.

«*Astratto*» è diventato, nel linguaggio comune, sinonimo di *avulso dalla realtà*.

Non è questo il senso dell'astrazione matematica. Il suo significato è molto più vicino al senso originale della parola. Astratto (*abs-tractus*, «tratto fuori da») è un participio passato, rimanda ad un soggetto che compia l'azione: esiste l'astratto perché c'è qualcuno che astrae.

Allora, capire l'astrazione (azione di astrarre) vuol dire divenire consapevoli delle astrazioni che ciascuno di noi ha operato nella sua vita. Non è capire un oggetto esterno a sé, ma capire i meccanismi di una azione che si è ripetutamente compiuta.

Una indagine sull'astrazione sarà comprensibile solo se si appoggia alla riflessione su una esperienza personale realmente avvenuta; una riflessione che si basa sui meccanismi della costruzione e dell'apprendimento della Matematica, i quali spesso avvengono in noi automaticamente, senza intervento volontario e senza consapevolezza. Cogliere l'astrazione significa capire noi stessi nel nostro rapporto con la realtà.

L'astrazione è, infatti, una modalità tipica del conoscere, che l'uomo mette in atto appena gli si pone il problema della realtà; è ricerca degli elementi comuni a tanti oggetti o situazioni apparentemente diversi.

Così, quando si dice «il triangolo», non si pensa più ad un particolare triangolo reale; esso non è più né rosso né verde, né spesso né sottile, né di legno né di cartone. È un modello astratto a cui si attribuiscono solo le qualità comuni a tutti i triangoli. Si parla di triangoli perché si è riconosciuta una analogia di forma tra certe figure.

A cosa serve questo modello ideale, a cui tante qualità sono state tolte? Esso, in realtà, è più povero di attributi rispetto a qualsiasi triangolo reale, ma proprio questa semplificazione

permette di indagare più agevolmente sulle relazioni esistenti tra gli attributi rimasti. In questo modo si scoprono, più facilmente, relazioni generali (ad esempio area e perimetro) che sono valide per *tutti* i triangoli in cui ci si può imbattere nella realtà.

Quindi, in questo caso (e così succede sempre), astrazione non è astrattezza, ma è un passaggio momentaneo per tornare poi alla realtà in modo più efficace.

Facciamo, ora, un altro esempio elementare.

Cosa indica ciascuno di noi con la parola «cinque»?

Certamente non un oggetto reale, ma un concetto della nostra mente, frutto di numerose osservazioni e carico di relazioni significative.

Se dico: «ho cinque frutti nel piatto», vuol dire che quei frutti non mi interessano ciascuno per suo conto, li guardo tutti insieme, e indico che sono tanti quante le dita di una mia mano (si può porre una corrispondenza biunivoca tra i due insiemi).

Così, se dico: «prendo questi cinque libri dalla biblioteca», indico gli insiemi di oggetti che si possono mettere in corrispondenza biunivoca sia tra loro sia con l'insieme delle dita della mia mano destra (o sinistra).

Ma, se dico: «la lunghezza del viaggio dipende dal numero di chilometri», nella parola «numero» è contenuta una astrazione di ordine superiore rispetto a quella contenuta nella parola «cinque».

Premesso che la *relazione di equipotenza* (essere in corrispondenza biunivoca) è una relazione di equivalenza, dopo avere riunito nella classe di nome «uno» certi insiemi equipotenti, nella classe di nome «due» altri insiemi equipotenti e così via, si passa a considerare l'insieme delle classi così costruite.

Si individua un nuovo insieme i cui elementi sono le classi precedentemente costruite.

Se dico: «i numeri da 1 a 100», indico le classi di insiemi equipotenti che possono corrispondere, rispettivamente, a: un dito, due dita, ..., tutte le dita di 10 persone.

Se dico: «tutti i numeri naturali», allora, considero *tutte* le possibilità di corrispondenze biunivoche tra insiemi di oggetti reali. Qui c'è ancora un salto di astrazione, perché se si parla di un insieme finito di numeri naturali si può pensare di riferirsi ad una esperienza, reale o immaginata, ma se si parla di *tutti* i numeri naturali questo non è più possibile. Si può parlare di tutti i naturali quando si capisce una *legge di costruzione* («aggiungi uno») che fa passare da un numero al suo successivo. Inoltre, si scinde questa costruzione dalla sua esecuzione reale su insiemi di oggetti e se ne afferma la possibilità indefinita. È una astrazione compiuta non più attraverso l'esperienza di una azione concreta, ma attraverso la ripetizione mentale di un procedimento.

Lo schema proposto rispecchia, tuttavia, solo l'aspetto cardinale del numero.

Inoltre, è uno schema molto semplificato perché non tiene conto dei reali meccanismi psicologici che si svolgono dentro l'uomo. Questo schema fa uso del concetto di insieme che nella Matematica si è presentato abbastanza recentemente. È, quindi, un esito finale e non una riflessione su come il soggetto opera. Tuttavia, questo schema si presta ad illustrare la complessità di un procedimento che si è abituati a pensare quasi istintivo. Esso mostra bene l'esistenza di vari livelli successivi di astrazione, e rispecchia una caratteristica veramente fondamentale della Matematica, quella, cioè, di considerare non tanto i singoli oggetti, quanto le relazioni tra essi esistenti. Gli oggetti, per la Matematica, sono sempre *oggetti in relazione*. Se si riprende l'esempio precedente, ci si accorge che, dicendo «5 frutti» o «5 libri», non si pone l'attenzione sui singoli oggetti ciascuno separato dall'altro, ma si considerano insieme, in altre parole, in relazione tra loro.

Torniamo, ora, ai simboli con cui si rappresentano, usualmente, i numeri. Anche in questo caso il problema è meno scontato di quanto appaia.

Il valore posizionale delle cifre e l'invenzione dello zero (non solo con la funzione di indicare il numero cardinale di un insieme vuoto, ma anche di segnalare un posto vuoto nella rappresentazione posizionale dei numeri) sono grandissime conquiste del pensiero astratto.

In prima analisi, l'astrazione è, quindi, una manifestazione della fantasia creatrice. La Matematica è creazione di oggetti che non esistono nella realtà tangibile, ma che esistono come modelli e schemi nel nostro pensiero e nel nostro operare.

Si utilizza, ora, il linguaggio degli insiemi.

Se manipoliamo degli oggetti (quegli stessi che si classificano rispetto alla relazione di equipotenza), formando mucchi e riunendoli o dividendoli in gruppi uguali o diversi, effettuiamo delle azioni che trasformano gli insiemi, ma viene conservata complessivamente la quantità.

Se A e B sono due insiemi disgiunti ($A \cap B = \emptyset$, cioè non hanno elementi in comune) e si confrontano i numeri cardinali di A , B e $A \cup B$, si può definire:

$$C(A) + C(B) = C(A \cup B)$$

avendo chiamato $C(I)$ il numero cardinale di un insieme I . Analogamente si definisce la sottrazione, la moltiplicazione e la divisione.

Anche questo schema è una semplificazione estrema di ciò che effettivamente nella mente della persona si compie, ma come prima, anche tralasciando la complessità psicologica, questo schema fa capire come l'operazione aritmetica sia il modello astratto di certe azioni.

Anche in questo caso si possono vedere subito livelli di astrazione successivi.

Uno dei primi ampliamenti dell'insieme dei numeri avviene con l'introduzione dei numeri interi (positivi e negativi).

Nell'insieme Z dei numeri interi, esiste un sottoinsieme (Z^+) di interi positivi che si possono identificare con i numeri naturali N definiti precedentemente.

Si definiscono, ora, le quattro operazioni, nell'insieme Z dei numeri interi relativi, in modo che, quando si opera nel suo sottoinsieme Z^+ , isomorfo a N (naturali), queste operazioni si identificano con le vecchie operazioni costruite sui naturali.

Con questa estensione delle definizioni si ottiene un primo vantaggio: la sottrazione diviene sempre possibile in Z .

Qualcosa di analogo accade dopo aver definito le frazioni e le operazioni tra esse.

Gli interi relativi (Z) si possono identificare con un sottoinsieme dei numeri razionali (Q); si possono, cioè, considerare come frazioni con denominatore 1.

Tutte le operazioni sono definite in Q in modo tale che quando si opera nel suo sottoinsieme isomorfo a Z , si ritrovano le primitive operazioni definite in Z . Definita, infatti, la somma di frazioni al solito modo:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$$

se si applica questa definizione nel caso in cui $b = 1$ e $d = 1$, si ritrova la vecchia definizione di somma tra interi.

Allora, la somma definita in Q è una generalizzazione della somma definita in Z .

Questa è un'astrazione importante: sostituire ad un certo numero di definizioni un'unica definizione di validità più generale, giungendo ad una riformulazione più semplice e compatta delle operazioni.

In Matematica la semplicità è, spesso, l'esito di un lungo lavoro. Essa, a parte ogni considerazione di tipo estetico che pure è importante, rende molto più evidente la «struttura», cioè il complesso di operazioni definite e le proprietà di cui esse sono dotate, e rende più agile il lavoro.

Questo stesso spirito ha condotto ad una successiva astrazione: lo studio delle proprietà di una operazione a prescindere dalla natura particolare dell'insieme su cui essa è definita, giungendo così all'*algebra astratta*.

Questa astrazione appare senza scopo se non ci si rende conto della molteplicità di casi particolari che sono riassunti nello studio di ogni singola struttura, a meno di avere un così

grande gusto per il pensiero astratto, da godere di questa atmosfera così rarefatta, senza alcuna preoccupazione di capirne le ragioni.

Siccome l'astrazione è un'azione, non c'è da stupirsi se la si compie prima di prenderne coscienza. Questo avviene per molte azioni umane.

Un neonato succhia prima di capire cosa fa, cioè anche prima di imparare a rendere volontaria la sua azione.

Così, fortunatamente, il sangue ci scorre nelle vene anche se non sappiamo spiegare né perché né come questo avvenga e non saremo mai capaci di rendere volontaria questa azione.

Un altro problema da affrontare era costituito dai numeri negativi, che l'Europa aveva conosciuto attraverso i testi arabi; tuttavia, nel Cinquecento e nel Seicento, i matematici, per la maggior parte, non ammettevano che i negativi fossero numeri o, quando lo ammettevano, non li accettavano come radici di equazioni.

Chuquet nel Quattrocento e Stevin nel Cinquecento giudicavano assurdi i numeri negativi.

Cardano ammetteva i numeri negativi come radici di equazioni, ma li considerava soluzioni impossibili; egli li definiva numeri «ficti» e chiamava «reali» le radici positive.

Viete rifiutò totalmente i numeri negativi; Cartesio, invece, li accettò solo entro certi limiti: definiva «false» le radici negative delle equazioni in quanto pretendevano di rappresentare numeri minori del nulla. Tuttavia, egli aveva mostrato che, data una equazione, è possibile ottenerne un'altra le cui radici sono maggiori di quelle dell'equazione originaria e di ogni quantità data; perciò, una equazione con radici negative poteva essere trasformata in una con radici positive. Cartesio, dunque, affermava che si possono trasformare le radici «false» in radici «vere» e così, quantunque riluttante, accettava i numeri negativi. Nonostante ciò egli provò sempre disagio nei confronti di questi numeri.

Pascal considerava la sottrazione di 4 da 0 come una pura assurdità; nei *Pensieri* scrive: «*Ho conosciuto anche chi non riesce a capire che togliendo quattro da zero rimane sempre zero*».

Pascal, come tutti, si trovava davanti alla necessità di non affidarsi ciecamente all'intuizione, che lo portava a riempire dello stesso contenuto empirico l'idea di numero positivo e l'idea di numero negativo.

I matematici si sono sempre trovati di fronte a questo imbarazzo: creano oggetti e strutture astratte che sembrano immediatamente sfuggire loro di mano, perché fanno emergere inattese contraddizioni. Nati come conseguenza di una genialità dell'intuizione, sembrano poi sfociare in conseguenze paradossali che l'intuizione non riesce a dominare.

Così è successo quando si sono affacciate alla storia le geometrie non euclidee, con la famosa questione della natura dello spazio fisico. L'uso fatto da Einstein delle geometrie non euclidee ha man mano fatto capire che ciascuna geometria va considerata non come una rappresentazione esauriente dello spazio fisico, ma come un modello, utile o meno secondo l'ambito che si vuole studiare. Non sussiste un problema di verità assoluta, ma piuttosto di verità relativa.

Il problema, tuttavia, si era posto molto prima, già nel mondo greco, quando punti, rette e piani della geometria sembrano contraddire l'intuizione. Ad esempio, un segmento piccolo a piacere contiene sempre «infiniti» punti: si può dividere in due il segmento, e poi ciascuna metà ancora in due e così indefinitamente; ma allora che dimensione ha il punto?

Noi parliamo ora con facilità di «punto senza dimensione» senza renderci conto che in questo senso il punto non è più un ente intuitivo (granellino piccolissimo) ma un ente astratto, elaborato gradualmente proprio in seguito alle contraddizioni emerse tra le ultime conseguenze e il primo concetto, ancora molto impreciso, da cui si era partiti.

Analogamente, la scoperta di segmenti incommensurabili fece bruscamente emergere la differenza tra misura in senso «matematico» (infinita precisione) e misura in senso «fisico» (precisione significativa fino ad un certo ordine). Infatti, mentre la precisione di una misura fisica ha un limite oltre il quale non è significativo procedere, la misura in senso matematico riesce a concepire

una precisione senza limite, passando da una serie finita di operazioni concrete di misura ad una infinità virtuale di operazioni dello stesso tipo. Attraverso questo procedimento mentale, il matematico introduce un ente diverso da quelli già a sua disposizione, il numero irrazionale.

Queste ultime due questioni ripropongono una situazione già vista: noi sappiamo pensare procedimenti senza termine che in realtà non potremmo eseguire. La nostra mente contiene questa potente intuizione che sembra scontrarsi con la realtà.

Fin dall'inizio, infatti, ogni questione sull'infinito pareva condurre inevitabilmente a qualche assurdità.

Fortunatamente, l'impatto con la contraddizione è sempre stato stimolante. Ha continuamente sollecitato i matematici a ridefinire, in modo sempre più preciso, gli enti e le relazioni della Matematica, nel tentativo di poterne capire a fondo la natura o, almeno, di poterne fare uso senza eccessivo pericolo di cadere in contraddizione.

È da questo desiderio che i matematici sono stati spinti a riordinare ogni teoria in modo puramente logico, ponendo all'inizio le proprietà (*assiomi*) necessarie a sorreggere l'intera costruzione e deducendo via via tutte le proprietà successive (*teoremi*).

Quindi, i frutti delle astrazioni compiute sono stati riordinati in modo da espellere quanto più possibile l'intuizione a vantaggio della logica.

Capita, talvolta, che sia accentuata l'idea di Matematica come complesso di teorie ipotetico-deduttive rispetto alla quale la capacità di studio viene a coincidere con la capacità di dimostrare teoremi in modo rigoroso.

Uno dei modi con cui si è valorizzato l'intuizione è l'interpretazione grafica di molti problemi algebrici (equazioni e disequazioni), utilizzando l'unità profonda tra geometria analitica ed algebra.

Sia ben chiaro che la rappresentazione grafica non riporta un problema ad uno stadio empirico. Essa è un linguaggio astratto, così come lo è l'algebra, con i suoi simboli e le sue regole. Il vantaggio è che la rappresentazione grafica offre una visione sintetica della situazione in esame. Quando un problema algebrico è tradotto in un problema geometrico e poi visualizzato mediante un design, i dati del problema e le relazioni esistenti tra essi si presentano tutti contemporaneamente alla nostra attenzione, diversamente che nel calcolo. Questo facilita il confronto tra i dati, evita le contraddizioni, aiuta anche a prevedere la strada analitica più breve per la risoluzione del problema.

Non sempre, è ovvio, basta fermarsi alla rappresentazione grafica; l'intuizione è soggetta ad inganni, quindi, non sostituisce la dimostrazione. Soprattutto, si deve ricordare che il vero senso di una dimostrazione non è dire se una proposizione è vera o falsa in assoluto, ma dire se una proprietà discende da altre proprietà di una precisa teoria. Quindi, è essenziale che una proprietà algebrica sia alla fine ricondotta agli assiomi dell'algebra e dimostrata con un linguaggio algebrico e che, similmente, una proprietà geometrica sia ricondotta agli assiomi della geometria e dimostrata con linguaggio geometrico.

Questo è il senso del lavorare dentro una teoria assiomatica, in cui il problema, come si è già accennato, è riportare la «verità» di ogni proposizione alla «verità» di alcune proposizioni fissate inizialmente, i famosi assiomi (un altro esempio di verità relativa).

La struttura assiomatica della Matematica è, ormai, un dato di fatto da cui non si può prescindere, e di cui, purtroppo, gli studenti che escono dalla scuola secondaria non si rendono conto. Quanto esposto sulla formazione delle idee matematiche e sulla creazione di concetti e strutture con livelli successivi di astrazione, non è una contrapposizione alla assiomatizzazione ma un completamento.

Evidenzia l'esistenza di due momenti, dapprima la genesi, sia storica sia personale, e poi la successiva fase critica in cui tutto si vaglia, si confronta, si riordina. In questa fase l'assiomatizzazione offre un livello di rigore logico irrinunciabile.

Perché valorizziamo l'intuizione?

Perché facilita l'apprendimento, senza alcun danno per la logica. Ovviamente si intende l'apprendimento non come una memorizzazione, ma come una ricostruzione personale del sapere, che avviene attraverso un meccanismo complesso. La nostra mente lavora abbondantemente per immagini e le nostre immagini sono legate all'organizzazione spaziale della nostra percezione, quindi legate alle rappresentazioni che noi facciamo dello spazio.

Inoltre, l'atteggiamento che interessa stimolare non è la ripetitività ma la ricerca, il lavoro creativo e questo richiede capacità d'intuito oltre che di logica.

Va valorizzata, quindi, in ogni possibile maniera, anche l'intuizione, come forma particolare di fantasia creatrice.

La struttura fondamentale del pensiero scientifico consiste di due tempi: 1) formulare ipotesi; 2) verificare le ipotesi formulate. Il primo tempo consiste di un atto creativo in cui l'intuizione ha un ruolo grandissimo.

La capacità di *analisi* e la capacità di *sintesi* sono strumenti indispensabili per la costruzione delle nostre astrazioni, sia matematiche sia generali.

L'*analisi* è «capacità di vedere delle differenze fra cose vissute inizialmente come identiche, o di cogliere delle somiglianze tra cose vissute inizialmente come diverse».

Un aspetto della capacità di analisi è saper fissare l'attenzione su un particolare, isolandolo dal contesto, oppure saper fissare l'attenzione solo su un tipo di proprietà inerenti ad una situazione (ad esempio, saper separare mentalmente i rapporti topologici dai rapporti euclidei in una configurazione geometrica). Si ricordi che i rapporti topologici (dentro-fuori rispetto ad una figura, insieme chiuso-aperto etc.) non coinvolgono la misura; i rapporti euclidei coinvolgono, invece, la misura (area di una figura, angoli etc.).

È evidente che non si tratta di un'abilità che riguarda solo la Matematica, ma investe anche altre forme d'apprendimento.

La *sintesi*, invece, è «capacità di tenere mentalmente presenti più cose in una volta, e di far nascere, dalla relazione che si può allora istituire fra esse, qualcosa di nuovo, per esempio il riconoscimento di una invarianza». Ad esempio, occorre una sintesi per arrivare a definire la capacità di una bottiglia. Plaget, infatti, fece un esperimento: se si travasa un liquido da una bottiglia alta e stretta ad una bassa e larga, i bambini molto piccoli non riconoscono che la quantità di acqua non è variata, perché essi sono solo in grado di concentrarsi ogni volta su un singolo aspetto della situazione, vale a dire il livello del liquido oppure la larghezza del recipiente.

In questo esempio, capacità di sintesi è la capacità di tenere conto, contemporaneamente, di questi due aspetti (livello del liquido, larghezza del recipiente) e di giungere, così, a capire che una diminuzione del livello è stata compensata da un aumento della larghezza del recipiente.

Anche la sintesi non riguarda solo la Matematica ma molte altre forme di apprendimento.

Esaminiamo, ora, in particolare, come analisi e sintesi abbiano un ruolo fondamentale nella risoluzione di problemi.

La parola «problema» ha, oggi, nell'ambito della scuola e della cultura, più di un significato.

Si parla ancora di *problema* nel senso più tradizionale del «problema matematico», l'esercizio con una o più domande.

Si parla di *problema* anche come «situazione problematica» più generale, nel senso, cioè, di una situazione in cui ci chiediamo un perché, e questo perché ci spinge alla ricerca di nuovi dati o di nuovi rapporti tra i dati che già conosciamo.

Noi viviamo psicologicamente un problema tutte le volte che avvertiamo il *bisogno di sapere qualcosa di più* o di mettere in rapporto una cosa con le altre *in modo diverso* da come abbiamo fatto sino a quel momento.

Insistiamo su questa connotazione psicologica del problema per avvertire lo studente che risolvere un problema matematico non consiste sempre nell'applicazione meccanica di formule, come spesso avviene nelle scuole di ogni ordine.

Questo modo di lavorare non è di per sé sbagliato: applicare ripetutamente una formula serve senz'altro a capirne il funzionamento e a memorizzarla, ma questo è solo un particolare dentro un quadro più vasto.

Non si arriva ad una buona comprensione della teoria se, dopo aver studiato (analizzato e memorizzato) i singoli teoremi, non si ricostruisce un quadro sintetico in cui riemerge una sorta di questione generale, una domanda fondamentale a cui si deve dare risposta, un filone che permette di ricostruire nessi importanti tra i particolari.

È questo che rende lo studio un atto creativo, anche se non si inventa niente di nuovo.

Studiare in questo modo può non essere facile perché ciò che si trova sui manuali scolastici è la sintesi di un lavoro lungo e complesso, avvenuto in secoli di lavoro, in cui la fase critica iniziale è spesso dimenticata. L'abbondanza del materiale che si deve «tramandare» rende inevitabile agli insegnanti una rielaborazione sintetica.

Sarebbe, però, di grande aiuto agli studenti una certa abbondanza di informazioni storiche, non solo per passione di cronaca, ma proprio come valorizzazione dei problemi che hanno suscitato le scoperte.

In mancanza di questo, è molto utile intercalare la lettura dei teoremi con una serie di esercizi che facciano vedere situazioni diverse in cui i teoremi operano, per tornare, poi, ai teoremi con uno sguardo più critico, capace, cioè, di ricostruire una problematica. (Questa osservazione implica un consiglio pratico importante: nella preparazione degli esami di Matematica non separare mai la preparazione dello scritto e dell'orale. Non si capirebbero a fondo né le sottigliezze della teoria né i tranelli degli esercizi).

È questo, in fondo, ciò che si intende per metodo di studio. Sono poche le regole che si possono elencare parlando di metodo. Questo cambia, spesso, da una disciplina all'altra. L'Arte, la Biologia o la Matematica vanno studiate in maniera differente; diverso è il ruolo della memoria, dell'esperimento, della deduzione logica.

Ma anche tra diverse discipline matematiche ci può essere notevole differenza di metodo.

Ciò che guida il lavoro è capirne gli obiettivi, sia immediati sia a lunga scadenza, come ad esempio, precisione di calcolo, rigore logico o altro (parlare di obiettivi è un modo diverso di riprodurre il tema della domanda).

Naturalmente, parleremo di problema anche, e soprattutto, quando ci si trova di fronte a problemi applicativi, che non saranno mai, o quasi mai, banale applicazione di schemi, tanto più se, invece di problemi finalizzati all'acquisizione di una parte di teoria, li intendiamo come problemi reali che potrebbero presentarsi ad un ingegnere nella sua professione.

La storia della Matematica ha dimostrato che un buon esito del lavoro è anche legato ad una certa capacità di svincolarsi dagli obiettivi più immediati.

Quando si parla di domanda che caratterizza psicologicamente un problema, non ci si deve troppo legare alle domande di tipo tecnico. La curiosità non si aggancia solo a voler trovare dei modi per rispondere a domande concrete, ma anche al desiderio di far emergere relazioni o al gusto di giocare con certi giocattoli complicati come le formule matematiche.

Talvolta, le scoperte più belle sono proprio nate così, e sono state fruttifere di applicazioni impreviste e utilissime.

Questo tipo di gratuità è, al fondo, percezione della dimensione infinita del bisogno di sapere, che supera i nostri bisogni concreti.

D'altra parte, se nei matematici non ci fosse stato questo spirito, la Matematica non avrebbe assunto il suo aspetto di studio di relazioni e di strutture astratte, che fu sollecitato storicamente, soprattutto, da un bisogno psicologico e conoscitivo.

18 Il principio di induzione

Il *principio d'induzione*, in matematica, trova un ampio impiego nelle dimostrazioni, per provare che una certa proprietà è valida per tutti i numeri interi. L'idea intuitiva alla sua base è l'«effetto domino»; ossia, affinché le tessere da domino disposte lungo una fila cadano tutte, sono sufficienti due condizioni:

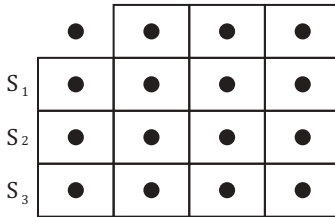
1. che cada la prima tessera;
2. che ogni tessera sia posizionata in modo tale che, cadendo, provochi la caduta della successiva.

Il principio d'induzione estende quest'idea al caso in cui la fila sia composta da infinite tessere. Vediamo ora un primo esempio, che ci permette di comprendere meglio tale principio.

Se si sommano i primi numeri dispari, si osserva una certa regolarità:

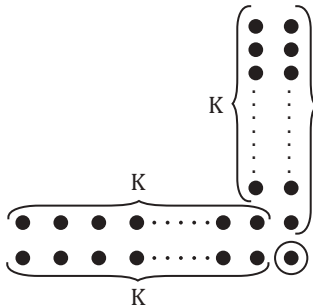
$$\begin{aligned} 1 + 3 &= 4 = 2^2 \\ 1 + 3 + 5 &= 9 = 3^2 \\ 1 + 3 + 5 + 7 &= 16 = 4^2 \\ 1 + 3 + 5 + 7 + 9 &= 25 = 5^2 \end{aligned}$$

Da tali espressioni si nota che la somma dei primi n numeri dispari vale n^2 . Se si è curiosi di sapere se questa proprietà vale *per ogni* n (intero positivo), si deve cercare di appurarlo attraverso qualche argomentazione che si sostituisca alla verifica diretta, la quale non può che avere termine. Rappresentando i numeri interi come pallini, si può visualizzare la situazione precedente:



Se si identifica il primo pallino con un quadrato di lato 1, si costruisce il quadrato di lato 2 «orlando» il quadrato precedente su due lati consecutivi. Nel primo passaggio si aggiungono 3 pallini, poi 5, poi 7 etc. Si pensi di essere giunti al k -esimo passaggio; come sarà il lato del nuovo quadrato rispetto al precedente?

Per orlare il quadrato di lato k occorrono $2k + 1$ pallini: si comincia a capire la struttura del «fenomeno».



Tutto risulta più chiaro se si scrive in modo opportuno un generico numero dispari:

$$a_h = 2h + 1 \quad (a_0 = 1; a_1 = 3; a_2 = 5; a_3 = 7 \text{ etc.})$$

dove h indica il posto del numero nella successione ordinata dei numeri dispari.

Pertanto:

$$1 + 3 + \dots + (2h + 1) = (h + 1)^2$$

Come si passa al quadrato successivo? Aggiungendo il numero dispari successivo:

$$1 + 3 + \dots + (2h + 1) + (2h + 3) = (h + 2)^2$$

Questo tipo di dimostrazione è basata sulla ricerca di una relazione fissa tra i quadrati che hanno per lato due interi successivi.

Qualcosa di analogo, ma ad un livello più astratto, è la dimostrazione per *induzione*.

Senza riferimento ai quadrati, si dimostra, ora, che se la proprietà vale per i primi $h + 1$ numeri dispari, allora vale anche nel caso successivo:

$$\begin{array}{ll} \text{ipotesi} & \rightarrow 1 + 3 + \dots + (2h + 1) = (h + 1)^2 \\ \text{tesi} & \rightarrow 1 + 3 + \dots + (2h + 1) + (2h + 3) = (h + 2)^2 \end{array}$$

La dimostrazione si ottiene riscrivendo in maniera diversa il termine $(h + 2)^2$, ossia:

$$(h + 2)^2 = [(h + 1) + 1]^2 = (h + 1)^2 + 2(h + 1) + 1 = (h + 1)^2 + 2h + 3$$

Utilizzando ora l'ipotesi, si ha immediatamente:

$$(h + 2)^2 = [1 + 3 + \dots + 2(h + 1)] + 2h + 3$$

Basta questo per essere certi che la proprietà sia sempre vera? Già si sa che essa è vera in alcuni casi iniziali. Si può utilizzare l'ultima uguaglianza per stabilire che:

$$1 + 3 + 5 + 7 + 9 = 5^2 \quad \rightarrow \quad 1 + 3 + \dots + 11 = 6^2$$

Ma allora:

$$1 + 3 + \dots + 11 = 6^2 \quad \rightarrow \quad 1 + 3 + \dots + 13 = 7^2$$

e, ancora:

$$1 + 3 + \dots + 13 = 7^2 \quad \rightarrow \quad 1 + 3 + \dots + 15 = 8^2$$

Si deduce, pertanto, che la validità generale è assicurata da due circostanze:

- in alcuni casi iniziali la proprietà è vera (basta anche un solo caso di verifica diretta);
- se la proprietà è vera nel caso n , allora è vera nel caso $n + 1$.

Si osservi che la struttura di questa dimostrazione è legata al fatto che la frase costituente la tesi dipende da n generico (n intero positivo).

Prendiamo ora in esame un altro esempio.

Sia n un intero positivo. Il numero:

$$M_n = \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$$

non può essere intero.

Poiché M dipende da n , è possibile dimostrare per induzione questa proprietà.

1^A PARTE:

ipotesi $\rightarrow M_n$ non è intero
 tesi $\rightarrow M_{n+1}$ non è intero

Si ponga $M_n = \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} = \frac{a}{b}$. L'ipotesi implica che è $a \neq kb$ (con $k \in \mathbb{N}$) per ogni intero n . In tal caso, il numero M_{n+1} è pari a:

$$M_{n+1} = M_n + \frac{1}{n+1} = \frac{a}{b} + \frac{1}{n+1} = \frac{a(n+1)+b}{b(n+1)}$$

Si supponga che M_{n+1} sia intero. Se la tesi è vera, si ottiene una contraddizione.

M_{n+1} intero significa che esiste $k \in \mathbb{N}$ tale che:

$$\frac{a(n+1)+b}{b(n+1)} = k$$

$$a(n+1)+b = kb(n+1) \quad (1)$$

$$an+a = kb(n+1) - b$$

$$an+a = b(kn+k-1)$$

$$n(a-kb) + a + (1-k)b = 0 \quad (2)$$

Se la (1) deve risultare vera indipendentemente da n , la (2) deve essere una *identità* rispetto a n ; deve perciò essere:

$$\begin{cases} a - kb = 0 \\ a + (1-k)b = 0 \end{cases}$$

La prima condizione contraddice l'ipotesi, quindi è impossibile che M_{n+1} sia intero.

2^A PARTE:

La proprietà vale per la prima somma:

$$M_3 = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

A questo punto la dimostrazione è completa.

In generale, il principio di induzione trova applicazione in Matematica per verificare la validità di enunciati che spesso si ripetono e che coinvolgono tutti i numeri interi positivi. A volte si percepisce che un enunciato valido per $n = 1, 2, \dots$ possa essere valido per tutti i numeri interi positivi.

La dimostrazione per induzione, dato un certo enunciato, si basa:

- sulla dimostrazione della proprietà dell'enunciato per $n = 1$;
- sulla congettura che esso sia valido per $n = i$, dove i è un generico numero intero positivo, e che, di conseguenza, esso sia valido per $n = i + 1$;
- sulla congettura che, essendo valido per $n = 1$, esso deve valere per $n = 1 + 1 = 2$ e così via per tutti i numeri interi positivi.

La geometria euclidea del piano e dello spazio

I

La geometria euclidea del piano

1 Gli enti fondamentali della geometria del piano

a) Rette e loro porzioni

Dato che non è possibile definire ogni oggetto, in ogni teoria matematica dobbiamo considerare alcuni elementi in modo intuitivo, associando loro solo un nome. Questi oggetti saranno chiamati *enti primitivi*.

Per la geometria del piano sono enti primitivi il *punto*, la *retta* e il *piano*.

Allo stesso modo, poiché non è possibile dimostrare ogni cosa, consideriamo delle proposizioni base che non dimostriamo e che chiamiamo *postulati* o *assiomi*.

Consideriamo i seguenti per cominciare.

POSTULATO 1 • Su ogni retta del piano possiamo stabilire un ordinamento.

POSTULATO 2 • Ogni retta è formata solo da punti e fra due punti qualsiasi di essa ci sono sempre altri punti.

POSTULATO 3 • Qualsiasi punto scegliamo sul piano, per esso passano infinite rette.

POSTULATO 4 • Comunque scegliamo due punti distinti nel piano, vi è un'unica retta che li contiene entrambi.

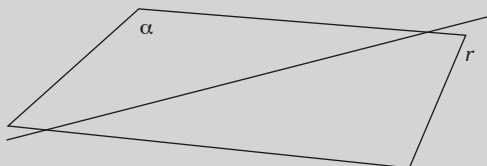
Quindi la retta è un insieme ordinato e continuo. Essa verifica la seguente proprietà, che fa sì che di rette ne esistano infinite.

TEOREMA 1 • Esistono almeno tre punti del piano non appartenenti alla stessa retta.

Una retta divide il piano in due parti.

ESEMPIO 1

Nella figura seguente, la *retta r* divide il *piano α* in due parti:



Ciascuno degli insiemi disgiunti determinati da una retta nel piano si chiama **semipiano**.

Due o più punti appartenenti a una stessa retta si dicono **allineati** o **collineari**.

Diciamo che una figura è stata sottoposta a un **movimento rigido** se, alla fine del movimento, la figura non ha variato alcuna delle sue proprietà geometriche.

Diciamo che due figure geometriche sono fra loro **isometriche** se esiste un movimento rigido che le fa sovrapporre.

Su una retta possiamo scegliere un punto determinando un nuovo oggetto.

Fissato un punto P su una retta ordinata, l'insieme formato da P e da tutti i punti che lo precedono o lo seguono si chiama **semiretta**. Il punto P si chiama **origine** della semiretta.

ESEMPIO 2

In figura, la *semiretta* di origine P :



Vale la seguente proprietà.

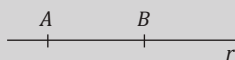
POSTULATO 5 • *Tutte le semirette sono isometriche.*

Su una retta possiamo fissare anche più di un punto, ottenendo altri oggetti.

Fissati due punti distinti A e B su una retta ordinata r , con A che precede B , diciamo **segmento** determinato da A e B l'insieme formato da A , B e da tutti i punti di r che seguono A e precedono B . I punti che determinano il segmento si chiamano **estremi** del segmento.

ESEMPIO 3

In figura, il *segmento* di estremi A e B :



Due segmenti che hanno:

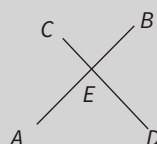
- 0 punti comuni si dicono fra loro **esterni**;
- 1 punto comune si dicono fra loro **incidenti**;
- più di un punto in comune si dicono **sovrapposti**;
- tutti i punti comuni si dicono **coincidenti**.

ESEMPIO 4

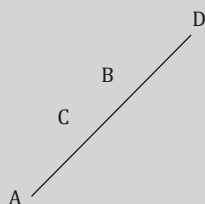
I segmenti in figura sono *esterni*:



I segmenti in figura sono *incidenti*:



I segmenti AB e CD in figura sono *sovrapposti*; il segmento BC è, infatti, loro parte comune:



Se poi A coincidesse con C e B con D , i segmenti sarebbero *coincidenti*.

Fra i segmenti incidenti possiamo distinguere una particolare classe.

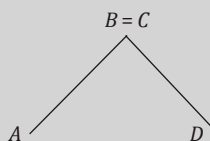
Diciamo **consecutivi** due segmenti non sovrapposti che hanno un estremo in comune.

Anche all'interno dei segmenti consecutivi possiamo distinguere un altro caso particolare.

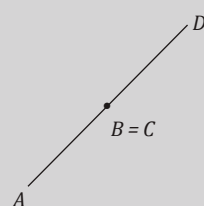
Diciamo **adiacenti** due segmenti consecutivi appartenenti a una stessa retta.

ESEMPIO 5

I segmenti in figura sono *consecutivi*:



I segmenti in figura sono *adiacenti*:



Possiamo considerare più di una retta. In questo caso abbiamo qualche semplice risultato.

TEOREMA 2 • Due rette distinte o non hanno punti in comune o hanno un solo punto in comune.

Poniamo, quindi, le seguenti definizioni.

Due rette si dicono:

- **parallele**, se non hanno punti in comune;
- **incidenti**, se hanno un solo punto in comune;
- **coincidenti**, se hanno tutti i punti in comune.

Per indicare che due rette r e s sono fra loro parallele scriveremo $r \parallel s$.

b) Angoli

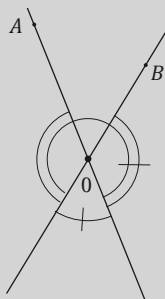
Due rette che si incontrano dividono il piano in quattro parti.

Ognuna delle quattro parti in cui un piano è diviso da due rette r e s incidenti in C si chiama **angolo**, con lati coincidenti con le semirette r e s di origine comune C .

Per indicare un angolo di vertice C , si scelgono un punto A e un punto B su ciascuno dei due lati, quindi scriviamo $\widehat{ACB}$ oppure $\widehat{BCA}$, a seconda che nel verso antiorario di lettura incontreremo prima A e poi B o viceversa.

ESEMPIO 6

In figura abbiamo tracciato i 4 *angoli* determinati dalle due rette. In particolare, l'angolo indicato con $A\hat{O}B$ è quello più in alto:



Fra i quattro angoli consideriamo due particolari coppie.

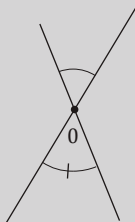
Due angoli aventi il vertice in comune e i lati disposti in modo tale che i prolungamenti dei lati dell'uno risultino lati dell'altro si dicono **opposti al vertice**.

Per questi angoli vale il seguente *teorema*.

TEOREMA 3 • Angoli opposti al vertice sono isometrici.

ESEMPIO 7

Gli angoli segnati sono *opposti al vertice* e quindi *isometrici*:



Alcuni angoli risultano più importanti di altri e quindi meritano una particolare terminologia.

Definiamo:

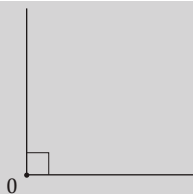
- **angolo giro** l'intero piano;
- **angolo piatto** un semipiano;
- **angolo retto** la metà di un angolo piatto;
- **angolo acuto** un angolo minore di un angolo retto;
- **angolo ottuso** un angolo maggiore di un angolo retto.

ESEMPIO 8

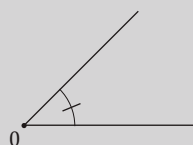
Ecco un *angolo piatto*:



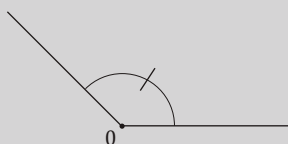
Un angolo retto:



Un angolo acuto:



Un angolo ottuso:



Vale la seguente proprietà.

TEOREMA 4 • Se due rette incidenti formano un angolo retto, tutti e quattro gli angoli formati sono retti.

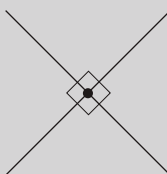
In conseguenza di ciò poniamo le seguenti definizioni:

- due **rette** si dicono fra loro **perpendicolari** o anche **ortogonali** se, incontrandosi, formano quattro angoli retti;
- diciamo **distanza di un punto da una retta** il segmento condotto perpendicolarmente dal punto alla retta;
- la retta **perpendicolare** a un dato segmento e passante per il suo punto medio si chiama **asse del segmento**.

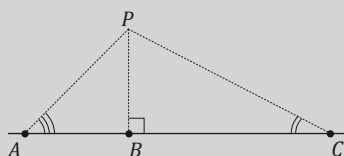
Per indicare che due rette sono fra di loro perpendicolari usiamo il simbolo $\perp$.

ESEMPIO 9

1. Due rette perpendicolari:



2. La **distanza di un punto da una retta** è il segmento minore che possa condursi dal punto alla retta. In questo caso corrisponde al segmento PB :



La regola seguente caratterizza l'asse di un segmento come luogo geometrico.

REGOLA • *I punti dell'asse di un segmento sono equidistanti dagli estremi dello stesso segmento.*

Definiamo un altro importante luogo geometrico.

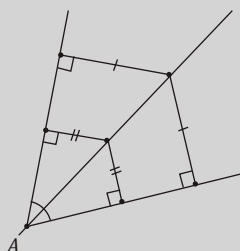
La semiretta che divide un angolo in due parti di uguale misura si chiama **bisettrice dell'angolo**.

La regola seguente caratterizza la bisettrice di un angolo.

REGOLA • *I punti della bisettrice di un angolo sono equidistanti dai lati dello stesso angolo.*

ESEMPIO 10

In figura abbiamo un *angolo* e la sua *bisettrice*:



I segmenti tracciati sono le distanze di punti della bisettrice dai lati dell'angolo e quelli ugualmente segnati sono fra loro *isometrici*.

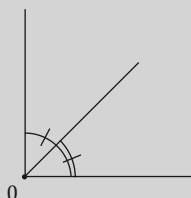
Vediamo alcune particolari coppie di angoli.

Due **angoli** dati sono tra loro:

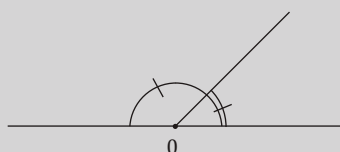
- **complementari**, se la loro somma è un angolo retto;
- **supplementari**, se la loro somma è un angolo piatto;
- **esplementari**, se la loro somma è un angolo giro.

ESEMPIO 11

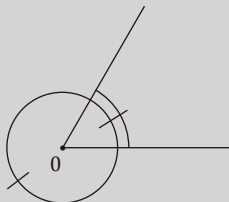
Gli angoli in figura sono *complementari*:



Gli angoli in figura sono *supplementari*:



Gli angoli in figura sono *esplementari*:



Consideriamo adesso tre rette che si incontrano.

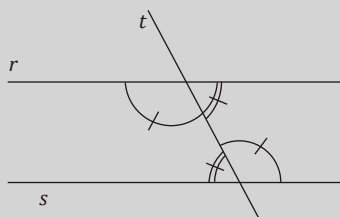
Date due **rette**, una terza retta che le incontra entrambe in punti diversi si chiama retta **trasversale** rispetto alle precedenti.

Date due rette r e s , tagliate da una trasversale t , chiamiamo:

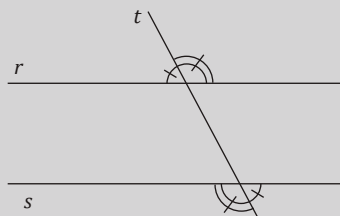
- **angoli alterni interni** quelli che sono contenuti fra r e s e si trovano da parti opposte rispetto a t ;
- **angoli alterni esterni** quelli che sono contenuti al di fuori della striscia limitata da r e s e si trovano da parti opposte rispetto a t ;
- **angoli coniugati interni** quelli che sono contenuti fra r e s e si trovano dalla stessa parte rispetto a t ;
- **angoli coniugati esterni** quelli che sono al di fuori di r e s e si trovano dalla stessa parte rispetto a t ;
- **angoli corrispondenti** quelli che sono uno interno e uno esterno rispetto alla striscia limitata da r e s e si trovano dalla stessa parte rispetto a t .

ESEMPIO 12

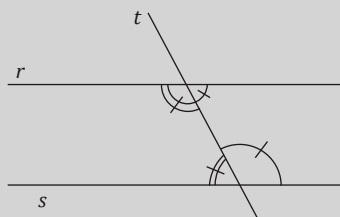
Gli angoli ugualmente segnati sono fra loro *alterni interni*:



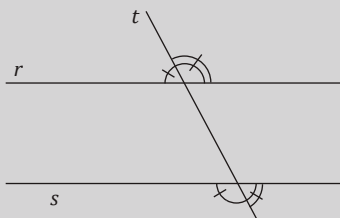
Gli angoli ugualmente segnati sono fra loro *alterni esterni*:



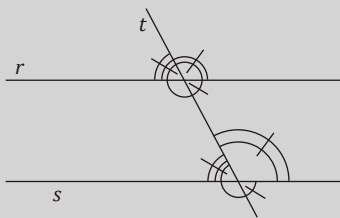
Gli angoli ugualmente segnati sono fra loro *coniugati interni*:



Gli angoli ugualmente segnati sono fra loro *coniugati esterni*:



Gli angoli ugualmente segnati sono fra loro *corrispondenti*:



Se le rette tagliate dalla trasversale sono fra loro parallele vale l'importante teorema seguente.

TEOREMA 5 • Condizione necessaria e sufficiente affinché due rette siano parallele è che, tagliate da una trasversale, formino almeno una coppia di angoli:

- alterni, interni o esterni, fra loro isometrici;
- coniugati, interni o esterni, fra loro supplementari;
- corrispondenti fra loro isometrici.

Il precedente teorema vale solo se imponiamo il seguente postulato, che risulta caratteristico della geometria euclidea.

POSTULATO 6 (Quinto postulato della geometria euclidea) • Dato un punto P e una retta r che non lo contiene, esiste ed è unica la retta passante per P e parallela a r .

La *misura di un angolo*, che consiste nell'assegnare allo stesso un numero reale, può essere eseguita usando diversi metodi. Il metodo sessagesimale consiste nel dividere l'angolo giro in 360 parti uguali, ciascuna delle quali si chiama **angolo grado** e si indica con il simbolo $^\circ$. Ogni angolo grado viene suddiviso in 60 parti uguali, che si chiamano **angoli primi** e si indicano con il simbolo $'$. Infine, ogni angolo primo è diviso in 60 parti uguali, che si chiamano **angoli secondi** e si indicano con il simbolo $''$; se serve si può continuare a dividere usando i numeri decimali.

In questo modo:

- l'angolo piatto misura 180° , perché metà di un angolo giro;
- l'angolo retto misura 90° , perché quarta parte di un angolo giro;
- un angolo che è la sedicesima parte di un angolo giro misura $22,5^\circ$, cioè $22^\circ 30'$;
- un angolo che è la sessantaquattresima parte di un angolo giro misura $5,625^\circ$, cioè $5^\circ 37,5'$ o meglio $5^\circ 37' 30''$;
- un angolo che è l'undicesima parte di un angolo giro misura $32,72^\circ$ (con il simbolo soprasssegnato indichiamo il periodo), cioè $32^\circ 43' 38,1818\dots''$.

Ancora:

- gli angoli acuti misurano meno di 90° ;
- gli angoli ottusi misurano più di 90° .

2 La circonferenza e il cerchio

Passiamo ad un altro importante ente geometrico.

Dato un punto C e un numero reale r , il luogo geometrico dei punti del piano la cui distanza da C è:

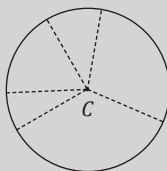
- uguale ad r si dice **circonferenza** di **centro** C e **raggio** r ;
- minore o uguale ad r si dice **cerchio** di **centro** C e **raggio** r .

Diciamo **corda** di una circonferenza Γ , un segmento i cui estremi sono due punti distinti di Γ . La corda passante per il centro si chiama **diametro**.

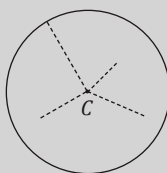
Diciamo **arco** di una circonferenza Γ e avente per estremi i punti A e B di Γ , l'insieme dei punti di Γ compresi tra A e B . Un arco di estremi i punti A e B si dice anche che *sottende* la corda AB .

ESEMPIO 13

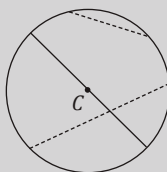
In figura abbiamo una *circonferenza*, in cui abbiamo tracciato alcuni *raggi*:



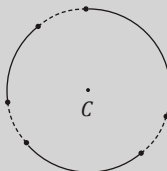
In figura abbiamo un *cerchio*, in cui abbiamo unito il *centro* con alcuni suoi *punti*:



In figura abbiamo alcune *corde* di una *circonferenza*, in particolare quella non tratteggiata è un *diametro*:



In figura abbiamo alcuni *archi*:



Valgono i seguenti *teoremi*.

TEOREMA 6 • Un diametro di una circonferenza Γ non è mai minore di una qualunque altra corda di Γ .

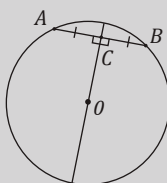
TEOREMA 7 • Per tre punti non allineati passa una e una sola circonferenza.

TEOREMA 8 • Condizione necessaria e sufficiente affinché una perpendicolare a una corda di una circonferenza passi per il centro della circonferenza è che essa sia l'asse del segmento.

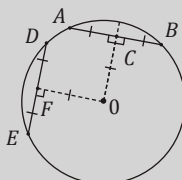
TEOREMA 9 • Condizione necessaria e sufficiente affinché due corde di una stessa circonferenza o di circonferenze isometriche siano isometriche è che abbiano la stessa distanza dal centro.

ESEMPIO 14

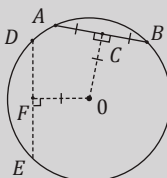
In figura mostriamo una *corda* e il proprio *asse* che passa per il centro:



In figura *corde isometriche* distano «ugualmente» dal centro:



Corde non isometriche, invece, hanno «diversa» distanza dal centro:



TEOREMA 10 • Il rapporto fra la lunghezza di una circonferenza e il proprio diametro è costante.

Il precedente rapporto viene indicato con il simbolo π e ha un valore approssimato di circa 3,1415.

Quindi, come immediato corollario del precedente risultato, abbiamo:

COROLLARIO • La lunghezza di una circonferenza di raggio r è $2\pi r$.

Adesso consideriamo le parti del cerchio.

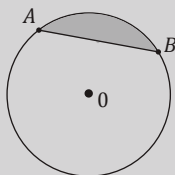
Diciamo **segmento circolare** di base una corda AB di un cerchio Γ , la parte di Γ delimitata dall'arco AB e dalla corda da esso sottesa.

Diciamo **segmento circolare a due basi** la parte di piano delimitata da due corde di una stessa circonferenza fra di loro parallele.

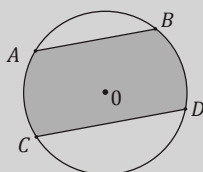
Chiamiamo **settore circolare** la parte di cerchio delimitata da due raggi.

ESEMPIO 15

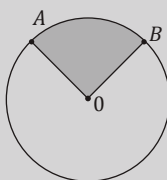
In figura abbiamo i due *segmenti circolari* che insistono sui due archi $\widehat{AB}$:



Un *segmento circolare a due basi*:



Un *settore circolare*:



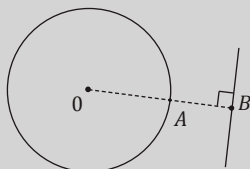
Per stabilire le reciproche posizioni di una retta e di una circonferenza, consideriamo il seguente *teorema*.

TEOREMA 11 • Date una retta r e una circonferenza Γ di centro O e raggio di misura R , indichiamo con d la distanza di r da O , ossia la misura del segmento perpendicolare condotto da O a r . Se:

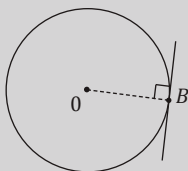
- $d > R$, allora r e Γ non hanno punti in comune, ossia sono esterne;
- $d = R$, allora r e Γ hanno un solo punto in comune, ossia la retta è tangente alla circonferenza;
- $d < R$, allora r e Γ hanno due punti in comune, ossia sono secanti.

ESEMPIO 16

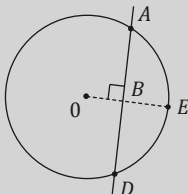
Ecco un esempio di retta e circonferenza *esterne*:



Un esempio di retta *tangente* ad una circonferenza:



Un esempio di retta e circonferenza *secanti*:



Concludiamo il paragrafo con altre nozioni sulla circonferenza.

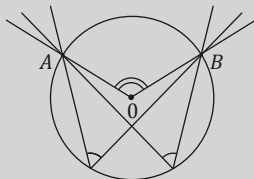
Diciamo **angolo al centro** di una circonferenza Γ un angolo il cui vertice coincide con il centro di Γ e i cui lati sono i prolungamenti di due diametri.

Diciamo **angolo alla circonferenza** Γ un angolo il cui vertice è un punto di Γ e i cui lati sono due rette secanti la circonferenza oppure una secante e l'altra tangente.

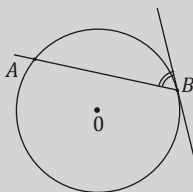
In entrambi i casi l'arco determinato dai lati dell'angolo si dice che è *sotteso* dall'angolo o anche che è l'arco su cui *insiste* l'angolo.

ESEMPIO 17

In figura mostriamo un *angolo al centro* e due *angoli alla circonferenza* che *insistono* sullo stesso arco:



Ecco un *angolo alla circonferenza* formato da una *secante* e una *tangente*:



Valgono i seguenti teoremi.

TEOREMA 12 • In una stessa circonferenza o in circonferenze isometriche, se le corde AB e CD sono isometriche allora:

- gli archi $\widehat{AB}$ e $\widehat{CD}$ sono isometrici;
- gli angoli al centro che insistono sugli archi $\widehat{AB}$ e $\widehat{CD}$ sono isometrici;
- i settori circolari determinati dagli angoli al centro che insistono su $\widehat{AB}$ e $\widehat{CD}$ sono isometrici;
- i segmenti circolari di basi AB e CD sono isometrici.

TEOREMA 13 • Un angolo al centro è isometrico al doppio della misura di uno qualsiasi degli angoli alla circonferenza che insistono sul suo stesso arco.

TEOREMA 14 • Dato un punto P su una circonferenza, l'angolo formato dalla retta diametrale passante per P e dalla tangente alla circonferenza in P è retto.

In una circonferenza un arco è perfettamente determinato dall'angolo al centro che esso sottende, quindi vi è un'ovvia relazione fra tale angolo e la misura dell'arco.

TEOREMA 15 • In una circonferenza di raggio r , un arco che intercetta un angolo al centro che misura α° ha una lunghezza pari a:

$$\frac{2\pi r \cdot \alpha^\circ}{360^\circ}$$

Così, ovviamente, la semicirconferenza, che determina un angolo al centro di 180° , misura:

$$\frac{2\pi r \cdot 180^\circ}{360^\circ} = \pi r$$

3 I poligoni

Prima di trattare i triangoli occorre definire alcuni concetti fondamentali.

Un insieme di tre o più segmenti a due a due fra loro consecutivi ma non adiacenti si chiama **poligonale**; gli eventuali estremi di un segmento che non risultano comuni a un altro segmento vengono detti **estremi liberi**.

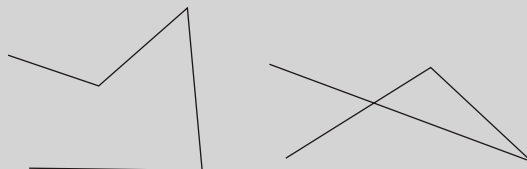
Una poligonale priva di estremi liberi si dice **poligonale chiusa**, diversamente viene chiamata **poligonale aperta**. L'insieme formato dai punti del piano delimitati da una poligonale chiusa e dai punti della poligonale si dice **poligono**, i segmenti che lo compongono si chiamano **lati**, gli estremi dei lati si dicono **vertici**.

Una poligonale in cui due segmenti qualsiasi hanno in comune al massimo un estremo viene detta **poligonale semplice**, diversamente viene detta **poligonale intrecciata**.

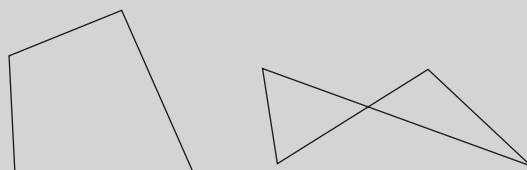
Una figura si dice **convessa** se, ogniqualevolta scegliamo due suoi punti qualsiasi, l'intero segmento che li congiunge è formato solo da punti appartenenti alla figura stessa, diversamente la figura si dice **concava**.

ESEMPIO 18

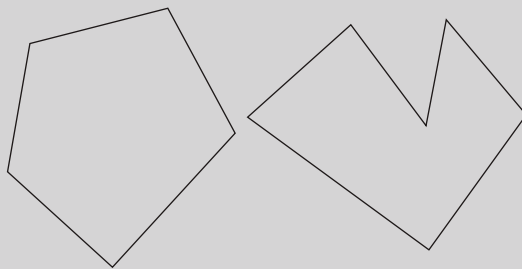
Ecco un esempio grafico di una *poligonale aperta semplice* e di una *intrecciata*:



Un esempio di una *poligonale chiusa semplice* e di una *intrecciata*:



Infine, un *poligono convesso* e uno *concavo*:



4 I triangoli

Concentriamo la nostra attenzione sui poligoni più semplici, quelli formati da tre lati e tre angoli e che perciò vengono chiamati *triangoli* (ovviamente se hanno tre lati determinano tre angoli).

In geometria esistono 3 criteri che permettono di costruire un triangolo conoscendo alcuni suoi elementi, lati e/o angoli.

TEOREMA 16 (Criterio di isometria LAL dei triangoli) • Due triangoli che hanno ordinatamente isometrici due lati e l'angolo tra essi compreso sono fra loro isometrici.

TEOREMA 17 (Criterio di isometria ALA dei triangoli) • Due triangoli che hanno ordinatamente isometrici due angoli e il lato a essi adiacente sono fra loro isometrici.

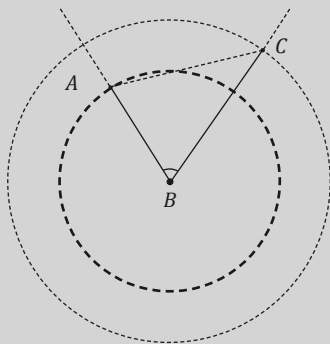
TEOREMA 18 (Criterio di isometria LLL dei triangoli) • Due triangoli che hanno ordinatamente isometrici i tre lati sono fra loro isometrici.

ESEMPIO 19

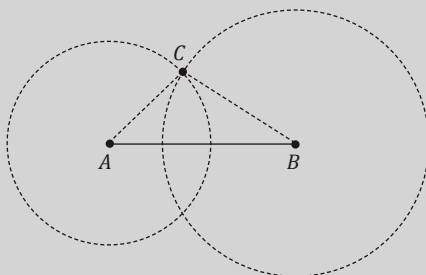
Per costruire un triangolo si usano riga e compasso.

Per esempio, grazie al *criterio LAL*, costruiamo un triangolo nel modo seguente.

Una volta che conosciamo l'angolo lo tracciamo; quindi, usando il compasso, stacciamo sui suoi lati dei segmenti lunghi quanto i lati noti del triangolo, determinando i punti *A* e *C*. A questo punto basta unire *A* con *C* per ottenere l'unico triangolo di lati *AB* e *BC* e angolo $\widehat{ABC}$:



Per il *criterio LLL*, invece, tracciamo uno dei tre lati, AB ; quindi, con il compasso tracciamo le circonferenze che hanno centro, rispettivamente, in A e B , e raggi uguali agli altri due lati noti. Se tali circonferenze si incontrano, abbiamo determinato il terzo vertice e costruito il triangolo:



L'esempio precedente ci suggerisce di enunciare il seguente teorema.

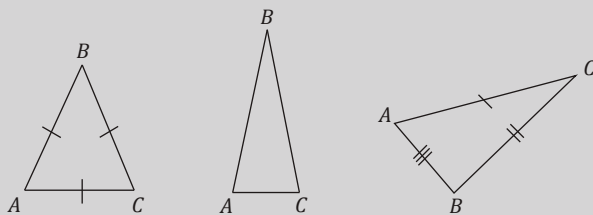
TEOREMA 19 (Disuguaglianza triangolare) • Tre segmenti possono rappresentare i lati di uno stesso triangolo solo se ciascuno di essi è minore della somma degli altri due e il valore assoluto della differenza di due di essi è minore del terzo.

Possiamo classificare i triangoli in base alle reciproche relazioni fra i **lati**. Infatti, un triangolo che ha:

- 3 lati isometrici si dice **equilatero**;
- almeno 2 lati isometrici si dice **isoscele**;
- 3 lati di misura diversa si dice **scaleno**.

ESEMPIO 20

Le figure di seguito riportate rappresentano, rispettivamente, un *triangolo equilatero*, un *triangolo isoscele* e un *triangolo scaleno*:



In un triangolo isoscele non equilatero, i lati isometrici vengono chiamati *lati obliqui*, mentre l'altro lato viene detto *base*. Infine, l'angolo opposto alla base si dice *angolo al vertice*.

Vale il seguente teorema.

TEOREMA 20 • In un triangolo isoscele gli angoli adiacenti alla base sono isometrici.

COROLLARIO • Un triangolo equilatero ha tutti gli angoli isometrici.

Consideriamo alcuni importanti segmenti di un triangolo.

Diciamo **altezza** relativa a un lato di un triangolo il segmento che ha per estremi il vertice opposto al lato e il piede della retta perpendicolare condotta da quel vertice alla retta a cui appartiene il lato.

Diciamo **mediana** relativa a un lato di un triangolo il segmento che ha per estremi il punto medio del lato e il vertice a esso opposto.

Diciamo **bisettrice** di un angolo interno di un triangolo il segmento che ha per estremi il vertice dell'angolo e l'intersezione della semiretta che divide l'angolo a metà con il lato opposto all'angolo.

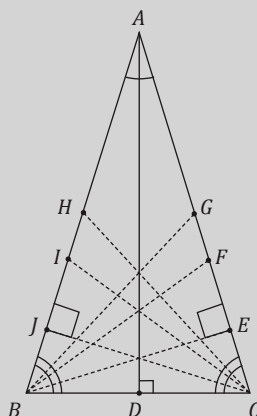
Ogni triangolo ha 3 mediane, 3 bisettrici e 3 altezze.

TEOREMA 21 • *In un triangolo isoscele la bisettrice relativa all'angolo al vertice, la mediana e l'altezza relativa alla base sono lo stesso segmento.*

COROLLARIO • *In un triangolo equilatero le altezze relative a un lato coincidono con le mediane relative allo stesso lato e con le bisettrici dell'angolo opposto a quel lato.*

ESEMPIO 21

Nel triangolo isoscele in figura abbiamo segnato le altezze, le mediane e le bisettrici:



Come si vede, quelli relativi ai lati obliqui sono tutti segmenti distinti, mentre quelli relativi alla base coincidono tutti nell'unico segmento AD.

Nell'esempio precedente abbiamo notato che bisettrici, altezze e mediane si incontrano in un punto. Ciò è valido per qualsiasi triangolo.

TEOREMA 22 • *Le bisettrici degli angoli interni di un triangolo si incontrano in uno stesso punto, che si chiama «incentro» perché è il centro della circonferenza inscritta nel triangolo.*

TEOREMA 23 • *Gli assi dei lati di un triangolo si incontrano in uno stesso punto, che si chiama «circo-centro» perché è il centro della circonferenza circoscritta al triangolo.*

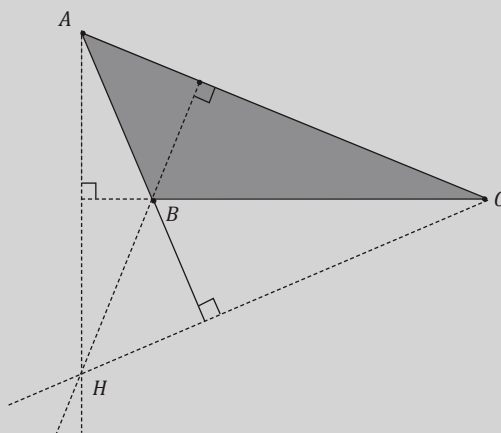
TEOREMA 24 • *Le mediane dei lati di un triangolo si incontrano in uno stesso punto, che si chiama «baricentro», il quale divide ciascuna mediana in due parti, in modo che quella che contiene il vertice sia il doppio dell'altra.*

TEOREMA 25 • *Le altezze dei lati di un triangolo si incontrano in uno stesso punto, che si chiama ortocentro.*

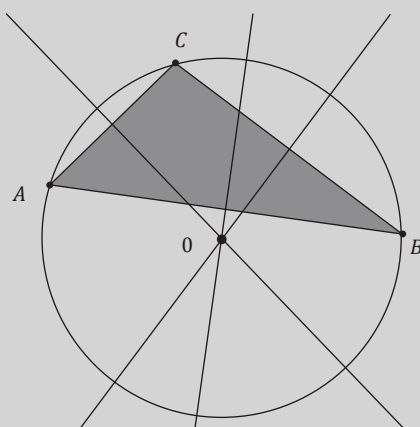
ESEMPIO 22

Il *baricentro* di un triangolo, così come l'*incentro*, sono sempre punti interni al triangolo. Non è così, invece, per l'*ortocentro* e per il *circocentro*, come mostrato nelle figure seguenti.

Nella figura seguente, l'*ortocentro* è indicato con H :



Nella figura seguente, il *circocentro* è indicato con O :



Il precedente esempio ci convince a classificare i triangoli anche in base agli angoli. Enunciamo prima un teorema.

TEOREMA 26 • La somma degli angoli interni di un triangolo è pari a un angolo piatto.

COROLLARIO • In un triangolo un angolo esterno è isometrico alla somma dei due angoli interni non adiacenti a esso.

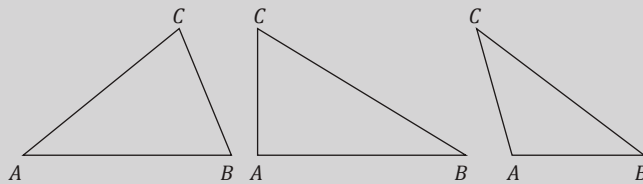
Un triangolo che ha:

- tutti gli angoli interni acuti si dice **acutangolo**;
- un angolo interno retto si dice **rettangolo**;
- un angolo interno ottuso si dice **ottusangolo**.

Il lato maggiore di un triangolo rettangolo si chiama **ipotenusa**, gli altri due lati si chiamano **cateti**.

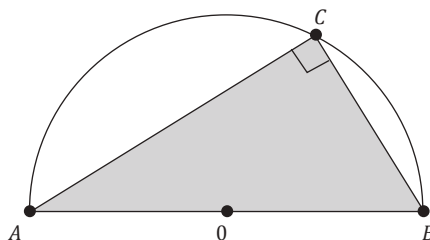
ESEMPIO 23

Le figure di seguito riportate rappresentano, rispettivamente, un *triangolo acutangolo*, un *triangolo rettangolo* e un *triangolo ottusangolo*:



Come immediato corollario del teorema sugli angoli al centro e quelli alla circonferenza abbiamo il seguente.

COROLLARIO • Un triangolo inscritto in una semicirconferenza, ossia tale che il suo lato maggiore coincida con un diametro, è rettangolo nel terzo vertice, quindi l'ipotenusa coincide con il diametro.



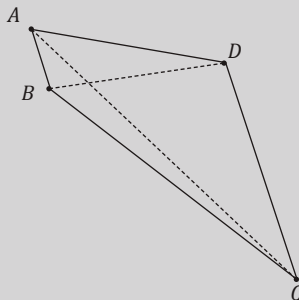
5 Alcuni quadrilateri

Consideriamo i poligoni con più di 3 lati, per i quali possiamo considerare un nuovo ente geometrico.

Un segmento che unisce due vertici non appartenenti a uno stesso lato di un poligono convesso si chiama **diagonale** del poligono.

ESEMPIO 24

Ecco un quadrilatero con le sue *diagonali*:



Un quadrilatero ha sempre 2 diagonali.

In generale, vale il seguente **teorema**.

TEOREMA 27 • Dato un poligono di n vertici, il numero di diagonali del poligono è:

$$d = \frac{n(n-3)}{2}$$

Fra i quadrilateri ne distinguiamo alcuni.

a) Parallelogramma

Un quadrilatero con i lati opposti paralleli si chiama **parallelogramma**.

Per i parallelogrammi valgono diverse proprietà.

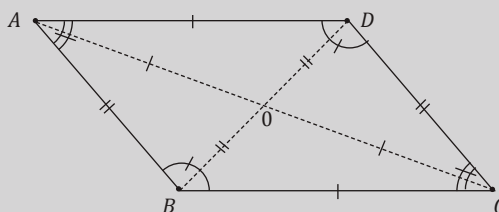
TEOREMA 28 • Condizione necessaria e sufficiente affinché un quadrilatero sia un parallelogramma è che valga almeno una di queste proprietà:

- i lati opposti siano isometrici;
- gli angoli opposti siano isometrici;
- gli angoli adiacenti a ciascun lato siano supplementari;
- le diagonali si incontrino nel loro punto medio.

Un quadrilatero convesso con due lati opposti isometrici e paralleli è un **parallelogramma**.

ESEMPIO 25

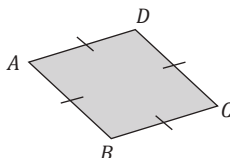
Ecco un *parallelogramma* in cui segniamo con uguali simboli enti fra loro isometrici:



Distinguiamo alcuni particolari parallelogrammi.

Un parallelogramma si chiama:

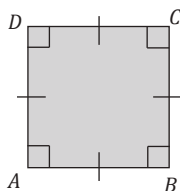
- **rombo** o **losanga**, se è equilatero:



- **rettangolo**, se è equiangolo:



— **quadrato**, se è equilatero ed equiangolo:



Per i precedenti parallelogrammi valgono le ulteriori seguenti proprietà.

TEOREMA 29 • Condizione necessaria e sufficiente affinché un parallelogramma sia un rombo è che valga almeno una delle seguenti proprietà:

- le diagonali siano perpendicolari;
- le diagonali siano bisettrici degli angoli relativi ai vertici che uniscono.

TEOREMA 30 • Condizione necessaria e sufficiente affinché un parallelogramma sia un rettangolo è che abbia le diagonali isometriche.

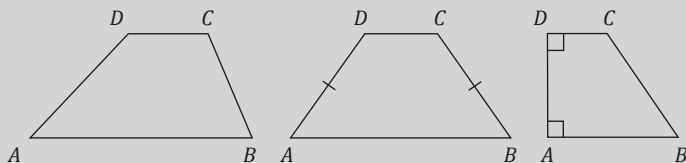
b) Trapezio

Un quadrilatero con solo due lati opposti paralleli si chiama **trapezio**. I lati non paralleli si chiamano **lati obliqui**; i lati fra loro paralleli si chiamano **base maggiore** e **base minore**.

Se i lati obliqui sono isometrici, il trapezio si dice **isoscele**. Se un lato è perpendicolare alle basi, il trapezio si dice **rettangolo**. Se i lati obliqui sono disuguali, il trapezio si dice **scaleno**.

ESEMPIO 26

Le figure seguenti riportano, rispettivamente, un generico trapezio, un trapezio isoscele e un trapezio rettangolo:



Valgono le proprietà seguenti.

TEOREMA 31 • In un trapezio isoscele gli angoli adiacenti a ciascuna base sono isometrici.

c) Quadrilateri inscrittibili e circoscrivibili, poligoni regolari

Concludiamo la trattazione dei quadrilateri parlando dei quadrilateri inscrittibili e circoscrivibili a una circonferenza.

TEOREMA 32 • Condizione necessaria e sufficiente affinché un quadrilatero sia inscrittibile in una circonferenza è che abbia gli angoli interni opposti supplementari.

TEOREMA 33 • Condizione necessaria e sufficiente affinché un quadrilatero sia circoscrivibile a una circonferenza è che la somma delle misure dei suoi lati opposti sia costante.

Adesso un teorema sugli angoli.

TEOREMA 34 • *La somma degli angoli interni di un poligono convesso di n lati è pari a $(n - 2)$ angoli piatti. La somma degli angoli esterni è pari a un angolo giro.*

Infine, un'ultima definizione.

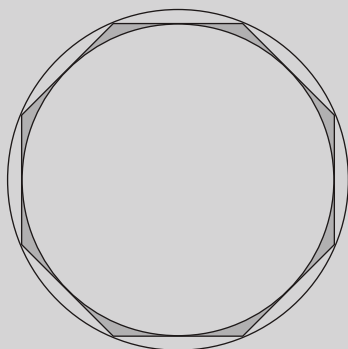
Un poligono equilatero ed equiangolo si dice **poligono regolare**.

Un ultimo *teorema*.

TEOREMA 35 • *Ogni poligono regolare è inscritto e circoscrittibile a una circonferenza.*

ESEMPIO 27

In figura un *ottagono regolare* e le circonferenze rispetto alle quali esso è *inscritto* e *circoscritto*:



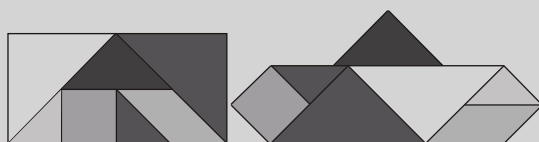
6 Il concetto di area

Il concetto di area di una figura piana viene dato come primitivo. Poniamo qualche definizione e qualche postulato.

Diciamo che due figure piane date sono **equidecomponibili** se esiste una loro suddivisione in un uguale numero di figure che, a due a due, hanno la stessa area.

ESEMPIO 28

Le due figure sono *equidecomponibili* perché formate con gli stessi 7 poligoni:



POSTULATO 7 • *Due figure equidecomponibili hanno la stessa area.*

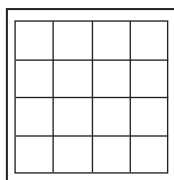
POSTULATO 8 • *Due figure isometriche hanno la stessa area.*

POSTULATO 9 • Data una famiglia F di n figure, non sovrapposte tra loro, e un'altra famiglia F' di n figure, anch'esse non sovrapposte tra loro e che a due a due hanno la stessa area delle figure di F , l'unione delle figure di F rappresenta una figura che ha la stessa area dell'unione delle figure di F' .

POSTULATO 10 • Date due figure con la stessa area, F e F' , le figure che si ottengono da F e F' sottraendo loro due figure della stessa area hanno la stessa area.

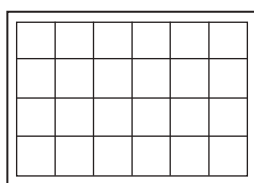
Vediamo alcune formule per il calcolo delle aree dei poligoni.

TEOREMA 36 • L'area di un quadrato di lato ℓ è ℓ^2 .



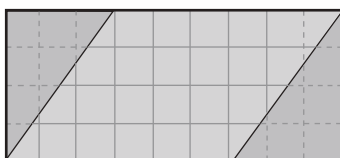
A questo punto è facile generalizzare il risultato all'area di un rettangolo.

TEOREMA 37 • L'area di un rettangolo di lati lunghi a e b è $a \times b$.



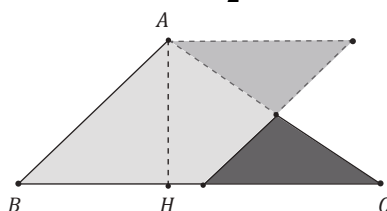
E, più in generale, di un parallelogramma.

TEOREMA 38 • L'area di un parallelogramma in cui un lato è lungo b e la relativa altezza è lunga h è $b \times h$.



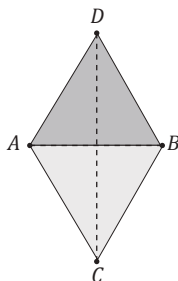
Quindi passiamo ai triangoli.

TEOREMA 39 • Un triangolo ha la stessa area di un parallelogramma in cui un lato è la metà di un lato del triangolo e le altezze relative sono isometriche. Pertanto, se il lato del triangolo è lungo b e la relativa altezza è lunga h , la sua area è $\frac{b \times h}{2}$.



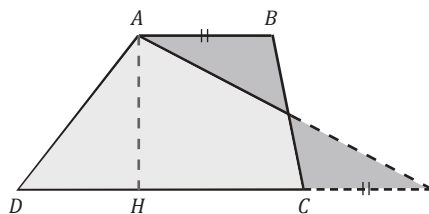
Poiché un rombo è formato da due triangoli isosceli isometrici con la base in comune, facilmente si ottiene il seguente risultato.

TEOREMA 40 • L'area di un rombo le cui diagonali sono lunghe d_1 e d_2 è $\frac{d_1 \times d_2}{2}$.



Infine consideriamo i trapezi.

TEOREMA 41 • Un trapezio ha la stessa area di un triangolo in cui un lato è isometrico alla somma delle basi del trapezio e le altezze relative sono isometriche. Pertanto, se le basi del trapezio sono lunghe b_1 e b_2 e la relativa altezza è lunga h , la sua area è $\frac{b_1 + b_2}{2} \times h$.



Ricordiamo adesso altre relazioni che riguardano le aree.

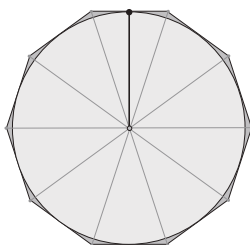
TEOREMA 42 • Il raggio della circonferenza inscritta in un triangolo si ottiene dal rapporto fra la misura dell'area del triangolo e la misura del semiperimetro.

TEOREMA 43 • Il raggio della circonferenza circoscritta a un triangolo si ottiene dal rapporto fra il prodotto delle misure dei lati e il quadruplo della misura dell'area del triangolo.

Tenuto conto del **Teorema 42**, possiamo definire una nuova grandezza per i poligoni regolari, utile a calcolarne l'area.

Il raggio della circonferenza inscritta in un poligono regolare si chiama **apotema** del poligono regolare.

Come mostrato nella figura seguente per un decagono regolare, l'apotema rappresenta l'altezza dei triangoli isosceli fra loro isometrici in cui può dividersi ogni poligono regolare.



In questo modo l'area del poligono è la somma delle aree dei triangoli, che hanno come base uno dei lati del poligono e come altezza relativa l'apotema. Pertanto vale la seguente proprietà.

TEOREMA 44 • *L'area di un poligono regolare si ottiene moltiplicando la misura del suo semiperimetro per quella del suo apotema.*

Sia R il raggio della circonferenza circoscritta al poligono regolare; la tabella seguente riporta le formule per la determinazione del lato, dell'apotema e dell'area di un poligono regolare, noto R .

Poligono regolare	Numero lati	Lato	Apotema	Area
Triangolo	3	$R\sqrt{3}$	$\frac{R}{2}$	$3\frac{\sqrt{3}}{4}R^2$
Quadrato	4	$R\sqrt{2}$	$\frac{\sqrt{2}}{2}R$	$2R^2$
Pentagono convesso	5	$\frac{R}{2}\sqrt{10-2\sqrt{5}}$	$\frac{R}{4}(\sqrt{5}+1)$	$\frac{5}{8}\sqrt{10+2\sqrt{5}}R^2$
Esagono	6	R	$\frac{\sqrt{3}}{2}R$	$3\frac{\sqrt{3}}{2}R^2$
Ottagono convesso	8	$R\sqrt{2-\sqrt{2}}$	$\frac{R}{2}\sqrt{2+\sqrt{2}}$	$2\sqrt{2}R^2$
Decagono convesso	10	$\frac{R}{2}(\sqrt{5}-1)$	$\frac{R}{4}\sqrt{10+2\sqrt{5}}$	$\frac{5}{4}\sqrt{10-2\sqrt{5}}R^2$
Dodecagono convesso	12	$\frac{R}{2}(\sqrt{6}-\sqrt{2})$	$\frac{R}{4}(\sqrt{6}+\sqrt{2})$	$3R^2$

Il cerchio è, in qualche modo, un poligono regolare con infiniti lati; la sua area è pari, quindi, al prodotto tra la misura del suo semiperimetro, cioè la lunghezza della semicirconferenza, e il suo apotema, che evidentemente è lo stesso raggio. Traducendo in formule, si ottiene:

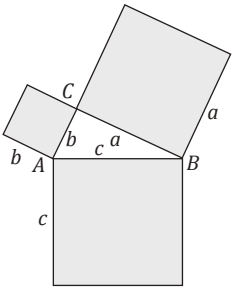
$$\pi r \cdot r = \pi r^2.$$

TEOREMA 45 • *Un settore circolare di una circonferenza di raggio r , che determina un angolo al centro che misura α gradi, ha una misura pari a $\frac{\pi \cdot r^2 \cdot \alpha}{360}$.*

Strettamente legati al problema delle aree sono i seguenti famosi teoremi.

TEOREMA 46 (Teorema di Pitagora) • *In un triangolo rettangolo, il quadrato costruito sull'ipotenusa ha la stessa area della somma dei quadrati costruiti sui cateti.*

Sia ABC un triangolo rettangolo in C :



$\overline{AB} = c$ (ipotenusa)

$\overline{AC} = b$ e $\overline{BC} = a$ (cateti)

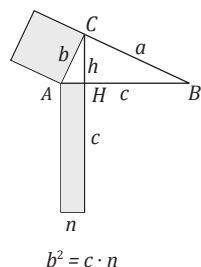
$$c^2 = a^2 + b^2 \Rightarrow \begin{cases} a = \sqrt{c^2 - b^2} \\ b = \sqrt{c^2 - a^2} \\ c = \sqrt{a^2 + b^2} \end{cases}$$

TEOREMA 47 (Inverso del Teorema di Pitagora) • Se in un triangolo ABC il quadrato costruito sul lato maggiore BC ha la stessa area dell'unione dei quadrati costruiti sugli altri due lati, allora il triangolo è rettangolo.

TEOREMA 48 (I Teorema di Euclide) • In un triangolo rettangolo, il quadrato costruito su un cateto ha la stessa area del rettangolo le cui dimensioni sono isometriche all'ipotenusa e alla proiezione di detto cateto sull'ipotenusa.

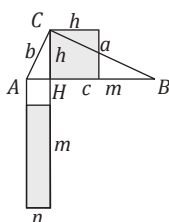
TEOREMA 49 (II Teorema di Euclide) • In un triangolo rettangolo, il quadrato costruito sull'altezza relativa all'ipotenusa ha la stessa area del rettangolo le cui dimensioni sono isometriche alle proiezioni dei cateti sull'ipotenusa.

(I Teorema)



$$b^2 = c \cdot n$$

(II Teorema)



$$h^2 = m \cdot n$$

Sia ABC un triangolo rettangolo in C :

$\overline{AC} = b$ (cateto);

$\overline{AB} = c$ (ipotenusa);

$CH = h$ (altezza relativa ad $\overline{AB}$);

$\overline{AH} = n$ (proiezione di $\overline{AC}$ su $\overline{AB}$);

$\overline{BH} = m$ (proiezione di $\overline{BC}$ su $\overline{AB}$).

TEOREMA 50 (Formola di Erone) • L'area di un triangolo, i cui lati misurano a , b e c , e il cui semiperimetro si indica con p , si ottiene con la seguente formula: $\sqrt{p \cdot (p-a) \cdot (p-b) \cdot (p-c)}$.

7 La similitudine

Dopo avere affrontato i concetti intuitivi di uguaglianza ed area, trattiamo brevemente quello di *somiglianza*.

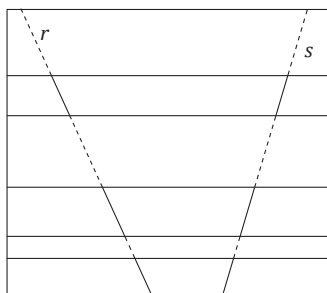
Diciamo che due poligoni sono **simili** fra loro, secondo il rapporto di similitudine $k \neq 0$, se si verificano le seguenti condizioni:

- hanno lo stesso numero di lati;
- esiste una corrispondenza biunivoca fra i loro angoli interni, in modo che due angoli corrispondenti siano fra loro isometrici;
- esiste una corrispondenza biunivoca fra i lati che formano le coppie di angoli corrispondenti isometrici, in modo che il rapporto delle misure dei lati corrispondenti sia sempre uguale a k .

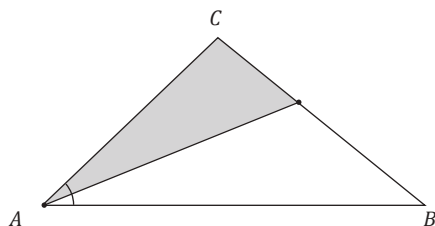
Per indicare che due figure geometriche P e P' sono simili, scriveremo $P \sim P'$.

Sono dimostrabili diversi teoremi relativi alla similitudine. Enunciamo solo i più importanti.

TEOREMA 51 (Teorema di Talete) • Due rette che tagliano un fascio di rette parallele determinano due classi di segmenti fra loro proporzionali.

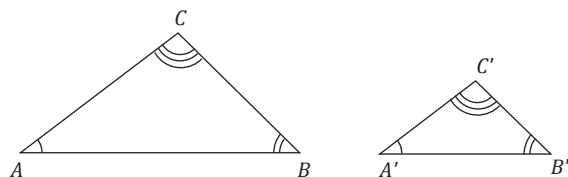


TEOREMA 52 (Teorema della bisettrice) • La bisettrice di un angolo interno di un triangolo divide il lato opposto in parti proporzionali agli altri due lati.

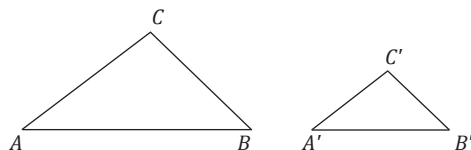


Anche per la similitudine fra triangoli valgono tre criteri.

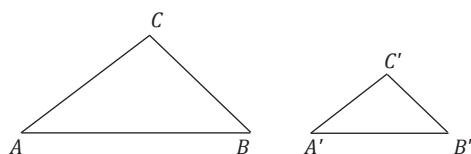
TEOREMA 53 (I criterio di similitudine dei triangoli) • Se due angoli di un triangolo sono isometrici ad altrettanti angoli di un altro triangolo, allora i due triangoli sono simili.



TEOREMA 54 (II criterio di similitudine dei triangoli) • Se due dei tre rapporti fra i lati di due triangoli sono uguali e gli angoli compresi da tali lati sono fra loro isometrici, allora i due triangoli sono simili.



TEOREMA 55 (III criterio di similitudine dei triangoli) • Se in due triangoli può stabilirsi una corrispondenza biunivoca fra i rispettivi lati, in modo che tutti i rapporti delle misure di lati corrispondenti siano uguali, allora i due triangoli sono simili.



Inoltre, valgono le seguenti proprietà.

TEOREMA 56 • In due triangoli simili:

- il rapporto fra due lati corrispondenti è uguale a quello fra le rispettive altezze;
- il rapporto fra i rispettivi perimetri è uguale al rapporto di similitudine;
- il rapporto fra le corrispondenti aree è uguale al quadrato del rapporto di similitudine.

Valgono ancora altre proprietà.

TEOREMA 57 (Teorema delle due corde) • In una circonferenza due corde incidenti sono divise dal loro punto comune in modo che il prodotto delle misure delle due parti dell'una uguaglia il prodotto delle misure delle due parti dell'altra.

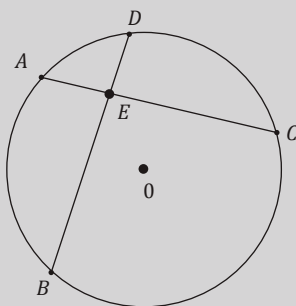
TEOREMA 58 (Teorema delle due secanti) • Condotte da un punto P esterno a una circonferenza Γ due rette a essa secanti, si ottiene che il prodotto delle misure di una secante per la sua parte esterna è uguale al prodotto dell'altra secante per la sua parte esterna.

TEOREMA 59 (Teorema della secante e della tangente) • Condotte per un punto P esterno a una circonferenza Γ una retta r secante e una retta t tangente a Γ , si ottiene che il segmento di tangenza è medio proporzionale fra l'intera secante e la sua parte esterna.

ESEMPIO 29

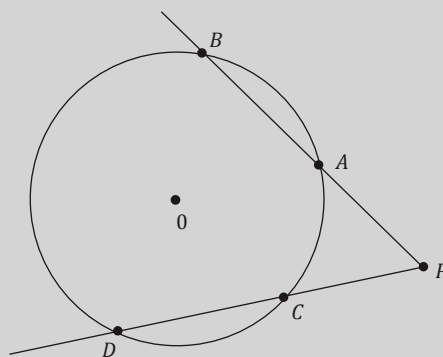
Con riferimento alla figura, per il **teorema delle corde** possiamo scrivere:

$$\overline{AE} \cdot \overline{EC} = \overline{DE} \cdot \overline{EB}$$



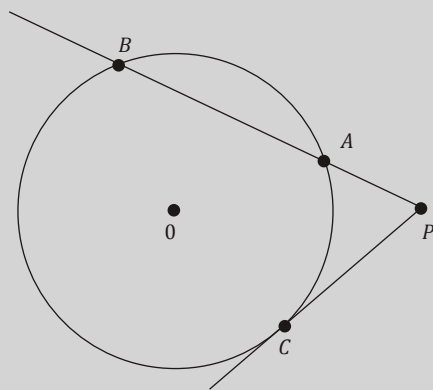
Con riferimento alla figura, per il **teorema delle secanti** possiamo scrivere:

$$\overline{PA} \cdot \overline{PB} = \overline{PC} \cdot \overline{PD}$$



Con riferimento alla figura, per il **teorema della secante e della tangente** possiamo scrivere:

$$\overline{PA} \cdot \overline{PB} = \overline{PC}^2$$



II

La geometria euclidea dello spazio

1 Punti, rette e piani nello spazio

Anche nel trattare la geometria dello spazio enunceremo proprietà e presenteremo oggetti del mondo reale. Intanto cominciamo a dire che esiste un ulteriore ente primitivo da aggiungere a punto, retta e piano, cioè lo spazio. Ovviamente tutte le proprietà della geometria piana continuano a valere anche per quella dello spazio, quando si muove in ambienti piani.

Abbiamo bisogno di aggiungere anche qualche postulato.

POSTULATO 11 • Se una retta e un piano hanno in comune più di un punto, allora tutti i punti della retta sono anche punti del piano e si dice che la retta giace sul piano.

POSTULATO 12 • Per ogni retta passano infiniti piani.

POSTULATO 13 • Dato un piano α , esiste almeno un punto P non appartenente ad α .

Un'altra cosa che possiamo dire è che il piano nella geometria dello spazio si comporta come la retta nella geometria del piano; avremo, quindi, le stesse reciproche posizioni fra due piani viste per le rette nel piano.

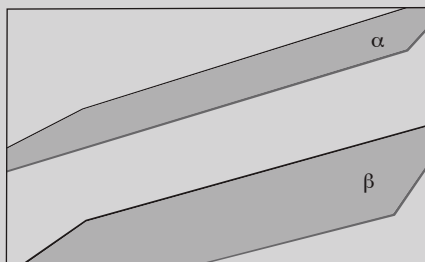
Due piani privi di punti comuni si dicono fra loro **paralleli**.

Due piani aventi 2 punti in comune, e quindi la retta passante per tali punti, si dicono fra di loro **incidenti**.

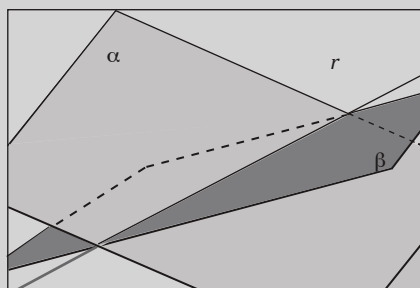
Due piani aventi in comune almeno tre punti non allineati si dicono fra di loro **coincidenti**.

ESEMPIO 30

Come modello di due piani *paralleli* possiamo pensare al pavimento e al tetto di una casa tradizionale, ovviamente immaginando tali oggetti come infiniti:



Come piani *incidenti* possiamo, invece, pensare a due pareti consecutive:



Vediamo i seguenti teoremi.

TEOREMA 60 • Un piano è determinato da:

- 3 punti non complanari;
- una retta e un punto fuori da essa.

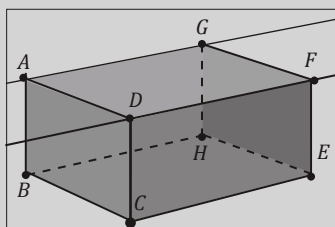
Invece, per le reciproche posizioni fra rette, si aggiunge una quarta possibilità.

Due **rette** che:

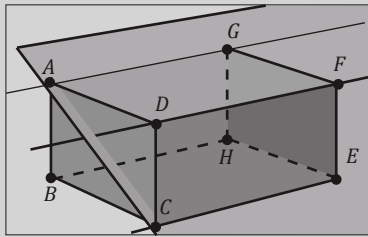
- non hanno punti in comune ma appartengono a uno stesso piano si dicono **parallele**;
- non hanno punti in comune e non appartengono a uno stesso piano si dicono **sgembe**;
- hanno 1 punto in comune si dicono **incidenti**;
- hanno 2 punti in comune si dicono **coincidenti**.

ESEMPIO 31

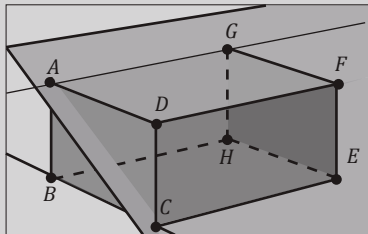
Abbiamo quindi introdotto due tipi di parallelismo, uno più «forte» e uno più «debole». Per capirci consideriamo il seguente modello formato da una scatola, in cui ogni faccia è assimilata a un piano e ogni segmento a una retta:



Le rette indicate sono *parallele* perché non hanno punti in comune e appartengono al «tetto» della scatola. Anche le rette seguenti sono *parallele*, anche se il piano che le contiene all'inizio non era segnato:



Invece le rette seguenti sono *sgembe* perché, come si vede, qualsiasi piano che contenga la retta per AG incontra sempre la retta per BC in un solo punto:



TEOREMA 61 • Esiste un solo piano che contiene due rette incidenti.

TEOREMA 62 • Esiste un solo piano che contiene due rette parallele.

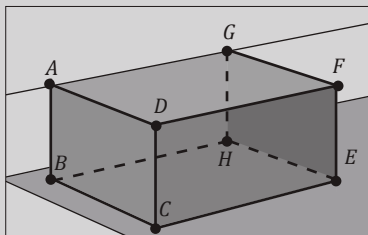
Passiamo alle reciproche posizioni di retta e piano.

Una **retta** e un **piano** che:

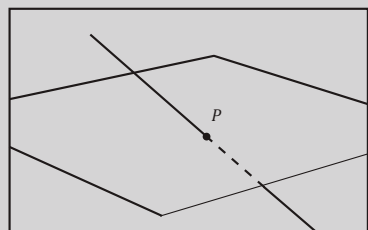
- non hanno punti in comune si dicono **paralleli**;
- hanno 1 punto in comune si dicono **incidenti**.

ESEMPIO 32

La retta segnata e il piano per B, C ed E sono *paralleli*:



La retta e il piano in figura sono *incidenti* nel punto P :



Continuiamo a costruire nozioni riprendendole dalle analoghe della geometria piana.

Un piano suddivide i punti dello spazio in due insiemi di punti non appartenenti al piano; ciascuno di questi insiemi è detto **semispazio**. Il piano che determina un semispazio si dice **origine del semispazio**.

POSTULATO 14 • *Tutti i semispazi sono uguali.*

Possiamo allora costruire l'analogo dell'angolo piano.

Diciamo **angolo diedro**, determinato dai semipiani α e β di origine comune la retta r , l'insieme dei punti dello spazio appartenenti a uno qualunque dei semispazi determinati. La retta r si chiama **spigolo** del diedro.

In analogia con quanto visto nella geometria del piano valgono le seguenti definizioni.

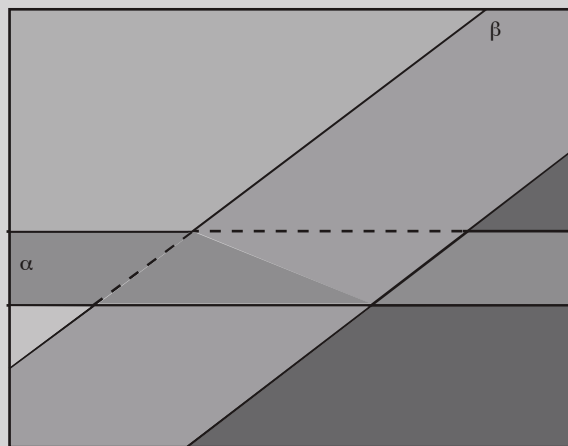
Diciamo **angolo diedro nullo** quello determinato da due semipiani coincidenti.

Diciamo **angolo diedro giro** l'intero spazio.

Diciamo **angolo diedro piatto** quello determinato da due semipiani facenti parte dello stesso piano.

ESEMPIO 33

In figura, i due piani incidenti formano i quattro *angoli diedri* diversamente ombreggiati:



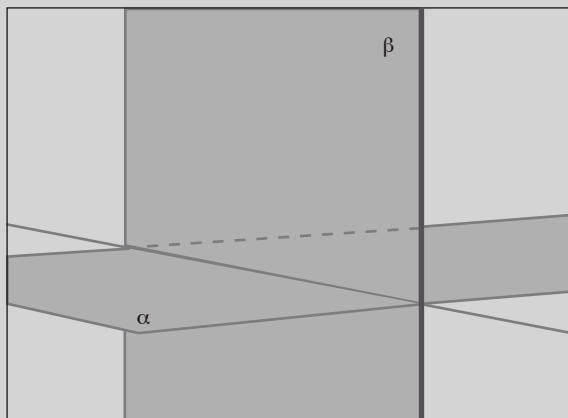
Adesso possiamo fornire i concetti di perpendicolarità nello spazio.

Due **piani** che incontrandosi formano quattro angoli diedri isometrici si dicono fra loro **perpendicolari** e gli angoli diedri così formati si dicono **angoli diedri retti**.

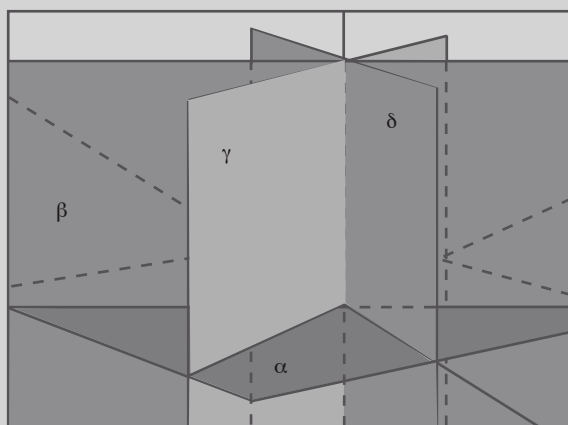
Una **retta** incidente a un piano si dice **perpendicolare al piano** se è perpendicolare a tutte le rette del piano passanti per il punto di incidenza.

ESEMPIO 34

In figura, due piani perpendicolari:



Nella figura successiva, invece, mostriamo perché una retta perpendicolare a un piano deve esserlo a tutte le rette del piano passanti per il punto di incidenza:



Enunciamo altri teoremi.

TEOREMA 63 • Se r è una retta perpendicolare al piano α , e β è un altro piano contenente r , allora β è perpendicolare ad α .

TEOREMA 64 • Se due piani α e β sono perpendicolari e si incontrano lungo la retta r , ogni retta s di α perpendicolare a r è perpendicolare a β .

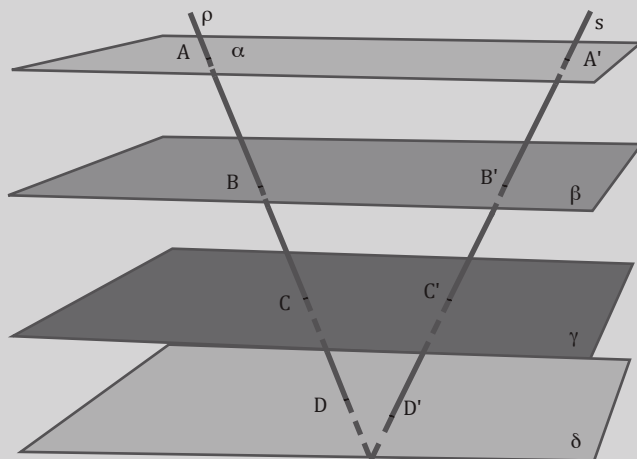
TEOREMA 65 • Due rette perpendicolari allo stesso piano sono fra loro parallele.

TEOREMA 66 • Se r e s sono rette parallele e se r è perpendicolare a un piano α , allora anche s è perpendicolare ad α .

TEOREMA 67 (Teorema di Talete nello spazio) • Due rette trasversali individuano su un fascio di piani paralleli classi di segmenti proporzionali.

ESEMPIO 35

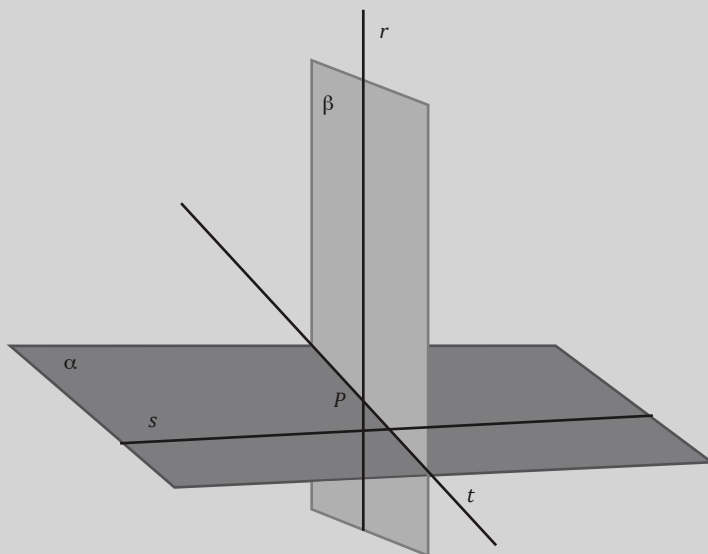
Ecco un'illustrazione grafica del **Teorema di Talete nello spazio**:



TEOREMA 68 (Teorema delle 3 perpendicolari) • Sia r una retta perpendicolare a un piano α in un punto P , sia s una retta di α e sia t la retta di α perpendicolare a s e passante per P ; in tal caso, s è perpendicolare al piano determinato da r e t .

ESEMPIO 36

Ecco un'illustrazione grafica del **Teorema delle 3 perpendicolari**:



2 I poliedri

Adesso vogliamo estendere il concetto di poligono alla geometria nello spazio. Ci rendiamo conto che dobbiamo considerare non segmenti ma poligoni i quali abbiano a due a due dei lati in comune (così come i rispettivi estremi). Il minimo numero di poligoni per racchiudere una parte di spazio è evidentemente quattro.

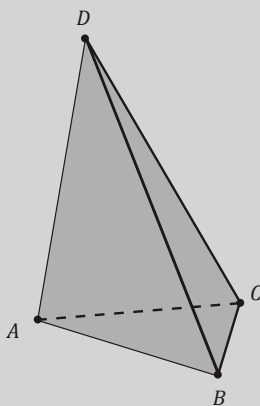
Diciamo **poliedro** la parte di spazio delimitata da quattro o più poligoni aventi i lati a due a due comuni, in modo che non vi siano lati di alcun poligono che non siano comuni a un altro. I poligoni che lo generano si dicono **facce** del poliedro, i loro lati si dicono **spigoli** del poliedro, i loro vertici si dicono **vertici** del poliedro.

a) Alcuni poliedri

A seconda del numero delle sue facce, un poliedro sarà denominato con un vocabolo formato da un *prefisso* riferito a tale numero e dal suffisso *edro* (tetraedro, pentaedro, esaedro, ...).

ESEMPIO 37

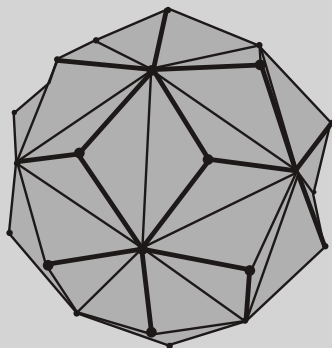
Il poliedro più semplice ha 4 facce che sono triangoli e si chiama *tetraedro*:



I poliedri possono essere *convessi* ma anche *concavi*.

ESEMPIO 38

Questo poliedro *concavo* si chiama icosaedro stellato:



Consideriamo alcuni poliedri particolarmente interessanti.

Diciamo **prisma** un poliedro le cui facce sono parallelogrammi con un solo lato in comune a due a due, e due poligoni di n lati posti su piani paralleli e aventi in comune con ciascun parallelogramma un solo lato e i vertici appartenenti a tale lato. I parallelogrammi si chiamano **facce laterali**, gli altri due poligoni si chiamano **basi** del prisma.

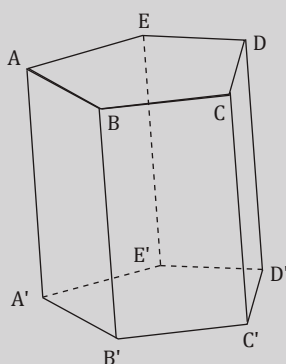
Un prisma che ha tutte le facce formate da parallelogrammi si chiama **parallelepipedo**.

Un prisma i cui spigoli appartenenti alle facce laterali sono perpendicolari ai piani delle basi si chiama **prisma retto**.

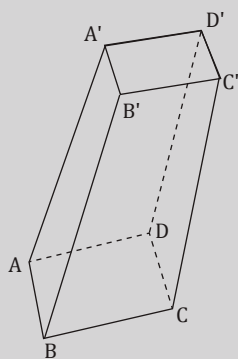
Se i poligoni fra loro paralleli sono poligoni regolari, il prisma si dirà **regolare**.

ESEMPIO 39

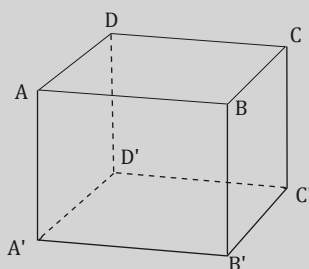
In figura, un *prisma obliquo a basi pentagonali*:



Un *parallelepipedo obliquo*:

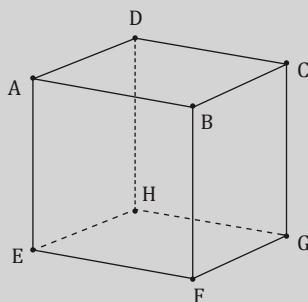


E un *parallelepipedo rettangolo*:



in cui si vede che gli spigoli sono a 4 a 4 isometrici. Un parallelepipedo è perfettamente determinato da 3 numeri, che sono le sue dimensioni.

Un *cubo* è determinato da un solo numero, che è la misura di uno qualsiasi dei suoi 12 spigoli isometrici:



La **superficie di un poliedro** è ovviamente la superficie di tutte le sue facce. Nel caso dei prismi, volendo determinare una formulazione analitica semplice, dobbiamo considerare quelli retti. In questo caso tutte le facce laterali sono rettangoli che hanno una dimensione di uguale misura (l'altezza del prisma), pertanto, la superficie del prisma si ottiene come stabilito dal seguente teorema.

TEOREMA 69 • La superficie di un prisma retto si ottiene moltiplicando la misura del perimetro della base per l'altezza del prisma.

COROLLARIO • La superficie di un parallelepipedo rettangolo di dimensioni a, b e c è $2 \times (ab + ac + bc)$.

COROLLARIO • La superficie di un cubo di lato ℓ è $6 \ell^2$.

Vediamo un'altra interessante famiglia di poliedri.

Si definisce **piramide** un poliedro individuato da una faccia poligonale chiamata **base** e da un **vertice** che non giace sul piano della base e che talora viene chiamato **apice** della piramide. Sono suoi **spigoli** i lati del poligono di base e i segmenti delimitati dall'apice e da ciascuno dei vertici della base. Sono **facce** della piramide la sua base e le facce triangolari (chiamate **facce laterali**) che hanno come vertice il suo apice.

Una piramide in cui il poligono di base è circoscrivibile a una circonferenza e in cui l'altezza condotta dal vertice alla base cade nel centro di questa circonferenza si dice **retta**.

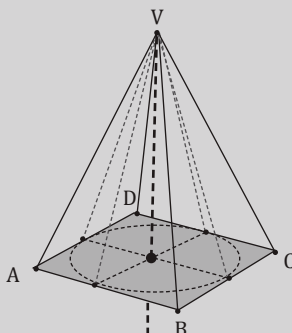
Una piramide retta con base un poligono regolare si dice piramide **regolare**.

Vale il seguente teorema.

TEOREMA 70 • In una piramide retta le congiungenti il vertice con i punti di contatto dei lati della base con la circonferenza inscritta (nella base stessa) sono le altezze delle facce laterali e sono uguali fra loro.

ESEMPIO 40

Ecco una *piramide retta a base quadrangolare*. La figura mostra l'isometria delle altezze di ciascuna faccia laterale:



In virtù del precedente risultato poniamo la seguente definizione.

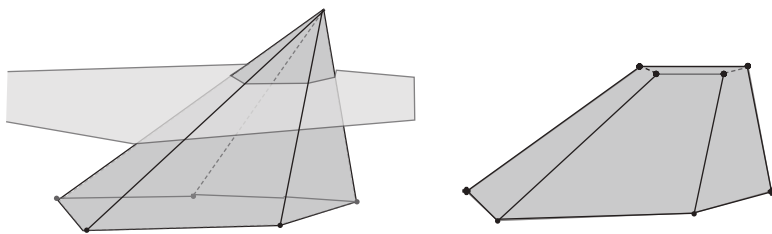
L'altezza isometrica delle facce di una piramide retta si chiama **apotema** della piramide.

Anche nel caso delle piramidi, se vogliamo delle semplici formule, dobbiamo considerare quelle rette.

TEOREMA 71 • *La superficie di una piramide retta si ottiene moltiplicando la misura del semiperimetro della base per l'apotema della piramide e aggiungendovi la superficie della base.*

Talvolta risulta interessante considerare il poliedro che si ottiene sezionando una piramide con un piano parallelo alla sua base.

Il poliedro differenza fra una piramide e la piramide più piccola, che si ottiene sezionando la piramide con un piano parallelo alla base, si chiama **tronco di piramide**.



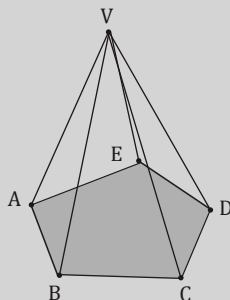
Ovviamente, se la piramide è retta anche il tronco che da essa si ottiene è retto; le facce laterali diventano trapezi, le cui altezze, dette apotema del tronco, sono uguali tra loro. Quindi, anche per tale poliedro vi è una semplice formula per il calcolo della sua superficie.

TEOREMA 72 • *La superficie di un tronco di piramide retta si ottiene moltiplicando la somma delle misure dei semiperimetri delle due basi per l'apotema del tronco di piramide e aggiungendovi le superfici delle basi.*

L'esempio seguente mostra un'interessante relazione fra spigoli, vertici e facce di uno stesso poliedro.

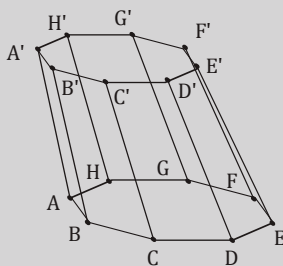
ESEMPIO 41

Consideriamo la seguente *piramide a base pentagonale*:

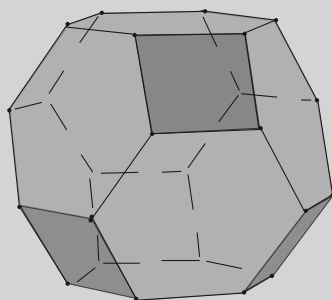


Essa ha 6 facce, 6 vertici e 10 spigoli.

Invece, il seguente *prisma a base ottagonale* ha 10 facce, 16 vertici e 24 spigoli.



Infine, il seguente poliedro è detto *ottaedro troncato* e ha 14 facce, 6 quadrati e 8 esagoni, 24 vertici e 36 spigoli.



In tutti i casi mostrati nell'esempio precedente si nota che il numero delle facce aggiunto a quello dei vertici supera il numero degli spigoli di 2. Questo è un risultato generale.

TEOREMA 73 (Formula di Eulero) • In ogni poliedro convesso vale la seguente formula:

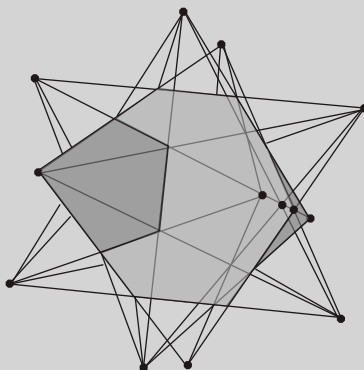
$$F + V = S + 2,$$

con F numero delle facce, V dei vertici e S degli spigoli.

La formula vale anche per poliedri non convessi.

ESEMPIO 42

In figura abbiamo il cosiddetto piccolo dodecaedro stellato, che certamente non è un poliedro convesso:

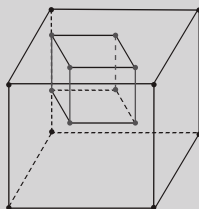


In pratica, su ogni faccia di un poliedro convesso a 12 facce è costruita una piramide. Pertanto, abbiamo $12 \times 5 = 60$ facce. I vertici sono invece $20 + 12 = 32$, cioè i 20 del dodecaedro di partenza e i 12 vertici delle piramidi. Infine, gli spigoli sono quelli delle basi delle piramidi, ossia $\frac{12 \times 5}{2} = 30$, perché ogni spigolo è comune a due piramidi; poi ci sono quelli che da ogni vertice della piramide vanno alla propria base e sono $12 \times 5 = 60$. In totale abbiamo 90 spigoli. In definitiva, la formula di Eulero vale ancora perché si ha: $60 + 32 = 90 + 2$.

Mostriamo che ci sono poliedri non convessi, i quali non verificano la formula di Eulero.

ESEMPIO 43

Il poliedro in figura è formato «bucando» un cubo con un altro cubo:



Questo poliedro ha 11 facce: le 6 del cubo grande, una delle quali è un quadrato con un buco, e solo 5 del cubo piccolo, perché la faccia superiore è bucata. Ha poi 16 vertici, 8 per ciascuno dei due cubi. Infine, ha 24 spigoli, 12 per ciascuno. In pratica in questo modo abbiamo raddoppiato sia il numero dei vertici sia quello degli spigoli di un cubo, ma non abbiamo raddoppiato quello delle facce, dato che appunto ne manca una. Pertanto, la formula di Eulero non vale: $11 + 16 = 24 + 3$.

In virtù del precedente esempio, poniamo la seguente definizione.

Chiamiamo **caratteristica di Eulero** di un poliedro il numero $F + V - S$.

Chiamiamo **poliedri euleriani** quelli che hanno *caratteristica* 2.

b) Poliedri regolari

Adesso vediamo i corrispondenti dei poligoni regolari.

Un poliedro le cui facce sono poligoni regolari dello stesso tipo e tali che in ogni vertice concorrano lo stesso numero di facce si chiama **poliedro regolare**.

Dato che i poligoni regolari sono infiniti, fa una certa sensazione il seguente risultato.

TEOREMA 74 • *Esistono solo 5 poliedri regolari.*

Il risultato precedente è dovuto al fatto che per formare un vertice abbiamo bisogno di almeno 3 poligoni regolari, in modo però che la somma dei loro angoli concorrenti nel vertice sia inferiore a 360° .

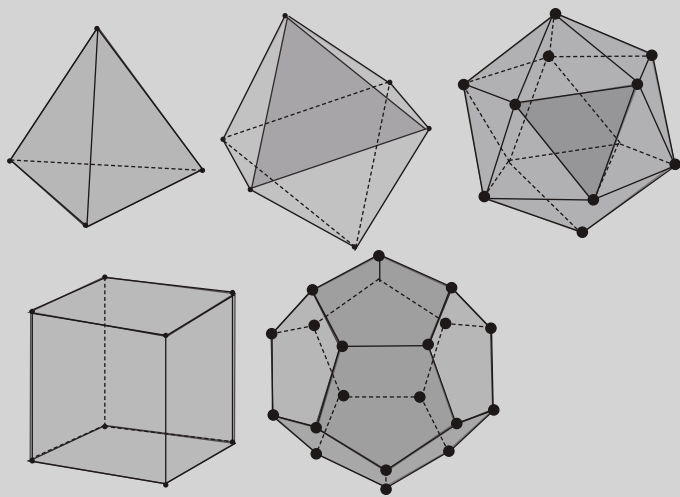
Ma ciò, tenendo conto della misura degli angoli interni di un poligono regolare, può ottenersi solo nei seguenti 5 modi:

- $3 \times 60^\circ = 180^\circ$
- $4 \times 60^\circ = 240^\circ$
- $5 \times 60^\circ = 300^\circ$
- $3 \times 90^\circ = 270^\circ$
- $3 \times 108^\circ = 324^\circ$

Pertanto, si hanno 3 poliedri regolari a faccia triangolare, un poliedro regolare a faccia quadrata e un poliedro regolare a faccia pentagonale.

ESEMPIO 44

Ecco i cinque *poliedri regolari*:



I primi tre sono quelli a facce triangolari (*tetraedro*, *ottaedro*, *icosaedro*), poi abbiamo l'*esaedro* (o *cubo*) a facce quadrate e il *dodecaedro* a facce pentagonali.

Dato che i poliedri regolari sono particolarmente importanti, riassumiamo in una tabella quanti vertici, spigoli e facce hanno.

Poliedro regolare	Facce	Spigoli	Vertici
Tetraedro	4	6	4
Esaedro o Cubo	6	12	8
Ottaedro	8	12	6
Dodecaedro	12	30	20
Icosaedro	20	30	12

Osserviamo che cubo e ottaedro, così come dodecaedro e icosaedro, hanno lo stesso numero di spigoli, mentre il numero delle facce dell'uno è uguale al numero dei vertici dell'altro. Questa particolare proprietà fa sì che, unendo i punti medi delle facce di uno dei due poliedri, si ottiene un poliedro dell'altro tipo, o come si dice, i poliedri sono fra loro **duali**.

Un tetraedro regolare è un solido formato da quattro triangoli equilateri isometrici. Questo significa che possiamo determinare la misura della superficie del tetraedro mediante la misura di uno spigolo.

Sia ℓ la misura dello spigolo di un tetraedro regolare. La misura dell'altezza sarà $\frac{1}{2} \cdot \ell \cdot \sqrt{3}$; l'area di una faccia $\frac{1}{4} \cdot \ell^2 \cdot \sqrt{3}$; l'area laterale $4 \cdot \frac{1}{4} \cdot \ell^2 \cdot \sqrt{3} = \ell^2 \cdot \sqrt{3}$.

Operando in modo analogo, possiamo trovare le superfici degli altri poliedri regolari in funzione della misura del lato, che riassumiamo nella tabella seguente:

Poliedro regolare	Superficie
Tetraedro	$\ell^2 \cdot \sqrt{3}$
Esaedro o cubo	$6\ell^2$
Ottaedro	$2 \cdot \sqrt{3} \cdot \ell^2$
Dodecaedro	$15 \cdot \sqrt{\frac{5+2\sqrt{5}}{5}} \cdot \ell^2$
Icosaedro	$5 \cdot \sqrt{3} \cdot \ell^2$

3 I solidi di rotazione

Oltre ai poliedri ci sono altri interessanti solidi, i cosiddetti *corpi rotondi*. Essi si ottengono dalla rotazione completa di un poligono attorno a una retta, appartenente o no al poligono. Consideriamo i più diffusi solidi di rotazione: il cilindro, il cono e la sfera.

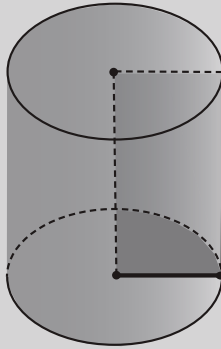
Il solido ottenuto dalla rotazione completa di un rettangolo intorno ad un suo lato si chiama **cilindro**. Il lato attorno a cui si effettua la rotazione si chiama **altezza** del cilindro. Se l'altezza è isometrica al diametro della base, il cilindro si dice **equilatero**.

Il solido ottenuto dalla rotazione completa di un triangolo intorno ad un suo lato si chiama **cono**. Se il triangolo è rettangolo e la rotazione avviene attorno a un cateto, il cono si dice **retto** e l'ipotenusa si chiama **apotema** del cono. Se l'apotema è isometrico al diametro della base, il cono si dice **equilatero**.

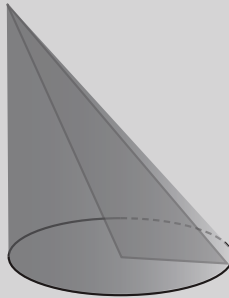
Il solido ottenuto dalla rotazione completa di un semicerchio attorno al proprio diametro si dice **sfera**. Diciamo **superficie sferica**, di centro il punto O e di raggio il segmento di misura r , il luogo geometrico dei punti dello spazio la cui distanza da O misura r .

ESEMPIO 45

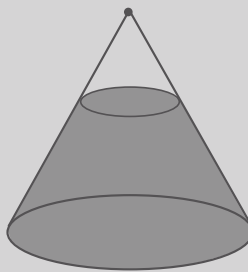
Ecco un esempio grafico di *cilindro*:



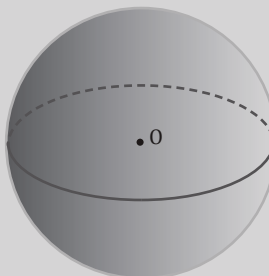
Un esempio di *cono obliquo*:



Un esempio di *tronco di cono*:



Un esempio di *sfera*:



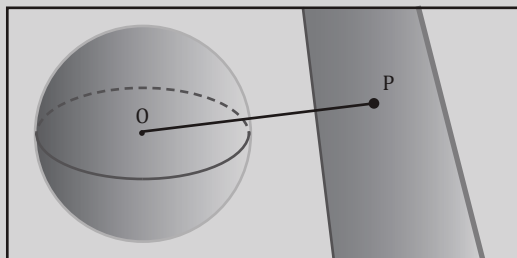
Risulta particolarmente interessante considerare le reciproche posizioni che possono assumere un piano e una sfera.

TEOREMA 75 • Dato un piano Π e una sfera Σ di centro O e raggio R , indichiamo con d la distanza di Π da O , ossia la misura del segmento perpendicolare condotto da O a Π . In tal caso, se:

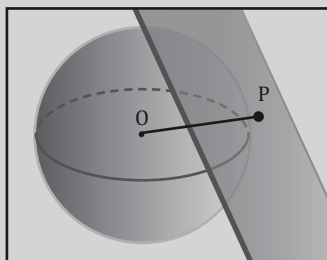
- $d > R$, allora Π e Σ non hanno punti in comune e si dicono **esterni**;
- $d = R$, allora Π e Σ hanno un solo punto in comune e si dicono **tangenti**;
- $d < R$, allora Π e Σ hanno infiniti punti in comune e si dicono **secanti**.

ESEMPIO 46

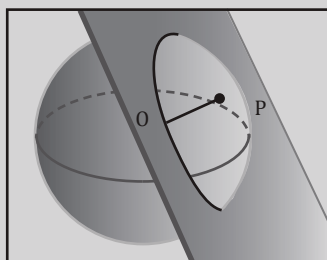
Ecco una sfera e un piano *esterni*:



Una sfera e un piano *tangenti*:



Una sfera e un piano *secanti*:



Vale il seguente **teorema**.

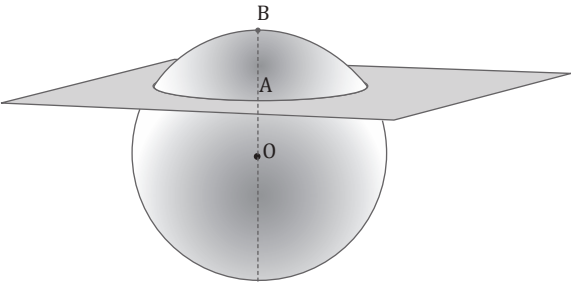
TEOREMA 76 • Un piano Π e una sfera Σ fra loro secanti hanno in comune un cerchio.

Un risultato analogo a quello valido nel piano per la circonferenza vale anche per la sfera.

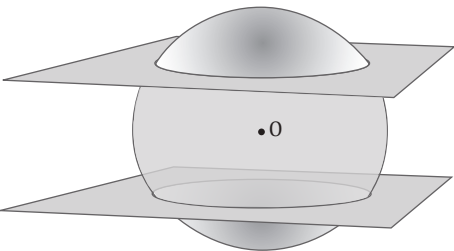
TEOREMA 77 • Per quattro punti non complanari passa una e una sola sfera.

Passiamo ad altre definizioni.

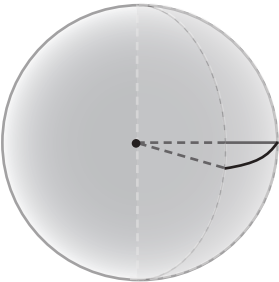
Un piano che seca una superficie sferica la divide in due zone, ciascuna delle quali si chiama **calotta sferica**; il diametro viene diviso in due parti, ciascuno dei quali si chiama **altezza** della rispettiva calotta o segmento.



Due piani fra loro paralleli che secano una superficie sferica la dividono in tre zone, due calotte e una terza che si chiama **zona sferica**; la distanza fra i piani si chiama **altezza**.



La parte di superficie sferica ottenuta mediante due semipiani aventi l'origine in comune con una retta diametrale si chiama **fuso sferico**.



Nella tabella riportiamo le formule per il calcolo delle superfici di alcuni solidi di rotazione.

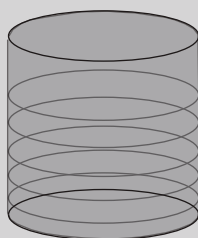
Solido	Superficie	
Sfera	$4\pi r^2$	(r = raggio della sfera)
Cono retto	$\pi r a$	(r = raggio della circonferenza di base del cono; a = apotema del cono)
Tronco di cono	$\pi(r + R)a$	(r e R = raggi delle due basi del tronco di cono; a = apotema del tronco di cono)
Calotta sferica	$2\pi r h$	(r = raggio della circonferenza di base della calotta; h = altezza della calotta)
Fuso sferico che determina un angolo α in radianti	$4r^2 \alpha$	(r = raggio della circonferenza)
Zona sferica	$2\pi r h$	(r = raggio della sfera da cui si ricava la zona sferica; h = altezza della zona sferica)

4 Il concetto di volume

Per trovare il **volume** di un solido, ossia la parte di spazio da esso contenuta, possiamo seguire il procedimento spiegato nel seguente esempio.

ESEMPIO 47

Per determinare il *volume* di un cilindro possiamo pensare che esso sia formato da infinite circonferenze isometriche e concentriche, come mostrato in figura:



Questo ci convince a dire che il volume è la somma delle infinite aree di questi infiniti cerchi. Dato che la somma degli «spessori» dei cerchi è ovviamente l'altezza del cilindro, possiamo dire che il volume del cilindro è dato dal prodotto dall'area di base per l'altezza, cioè in formula $\pi \cdot r^2 \cdot h$.

La procedura precedente può applicarsi anche agli altri tipi di solidi qui considerati, ma ovviamente solo per cilindri e prismi si ottengono facilmente i risultati. In tutti gli altri casi dobbiamo usare il calcolo infinitesimale.

Ci limitiamo a enunciare le formule.

TEOREMA 78 • Il volume:

- del prisma o del cilindro è il prodotto dell'area di base per l'altezza;
- della piramide o del cono è il prodotto di un terzo dell'area di base per l'altezza;
- della sfera è il prodotto di quattro terzi dell'area di base del suo cerchio massimo.

Anche se può sembrare strano, ai volumi delimitati dalle parti della superficie sferica vengono assegnati diversi nomi.

La parte di sfera contenuta da una **calotta sferica** si chiama **segmento sferico a una base**; quella contenuta dalla zona sferica si chiama **segmento sferico a due basi**; quella contenuta dal fuso si chiama **spicchio sferico**.

Nella tabella seguente riportiamo le formule per il calcolo dei volumi di alcuni solidi.

Solido	Volume
Tetraedro regolare	$\frac{\ell^3}{12} \cdot \sqrt{2}$
Esaedro regolare	ℓ^3
Ottaedro regolare	$\frac{\ell^3}{3} \cdot \sqrt{2}$
Dodecaedro regolare	$\frac{15 + 7 \cdot \sqrt{5}}{4} \cdot \ell^3$
Icosaedro regolare	$\frac{15 + 5 \cdot \sqrt{5}}{12} \cdot \ell^3$
Piramide	$\frac{A_b \cdot h}{3}$ (A_b = area di base)
Tronco di piramide	$\frac{(A_b + A_{b'} + \sqrt{A_b \cdot A_{b'}})}{3} \cdot h$ (A_b e $A_{b'}$ = aree delle due basi del tronco di piramide)
Sfera	$\frac{4}{3} \pi \cdot r^3$
Cilindro retto	$\pi r^2 \times h$
Cono retto	$\frac{\pi r^2 \cdot h}{3}$
Tronco di cono	$\pi \frac{r_1^2 + r_2^2 + r_1 \cdot r_2}{3} \cdot h$ (r_1 e r_2 = raggi delle due circonferenze/basi del tronco)
Spicchio sferico che determina un angolo α in radianti	$\frac{2\alpha \cdot r^3}{3}$ (r = raggio della circonferenza)
Segmento sferico a una base	$\frac{\pi \cdot h^2}{3} \cdot (3r - h)$ (r = raggio della circonferenza di base della calotta; h = altezza della calotta)
Segmento sferico a due basi	$\frac{\pi \cdot h}{6} \cdot [3 \cdot (r_1^2 + r_2^2) + h^2]$ (r_1 e r_2 = raggi delle due circonferenze/basi; h = altezza della zona)

I sistemi numerici N, Z, Q, R, C e le strutture algebriche fondamentali

1 Estensione del concetto di numero: dai naturali ai complessi

In questa sede si darà una breve trattazione delle varie estensioni del concetto di numero.

a) Numeri naturali

L'insieme dei *numeri naturali*, indicati con la simbologia $N \{0, 1, 2, 3, \dots\}$, è l'insieme base dei numeri; esso è legato alla *operazione del contare* la quantità degli elementi di un insieme finito, ne rappresenta, cioè, la potenza.

Nell'insieme considerato sono definite:

- la *relazione d'ordine totale* $\leq$;
- le operazioni di *addizione* e di *moltiplicazione*, con le loro *proprietà*.

Dati due numeri naturali a e b , le operazioni di sottrazione e di divisione, indicate, rispettivamente, con $a - b$ e a/b , sono definite solo se, per la prima $a > b$, per la seconda a è un multiplo di b .

I numeri naturali presentano due significati concettualmente eterogenei: quello di *ordinale*, che indica quale posto occupa un dato elemento in una successione ordinata, e quello di *cardinale*, che esprime la quantità.

La prima classificazione dei numeri naturali è fornita dalla divisione per due: un numero naturale si dice *pari* se è divisibile per 2, altrimenti si dice *dispari*. In matematica, dato un numero naturale n , il suo doppio è indicato con $2n$; un numero dispari è indicato con $2n + 1$.

Solo le operazioni di moltiplicazione e di elevamento al quadrato conservano la parità, nel senso che il prodotto di due numeri pari (dispari) è pari (dispari), e il quadrato di un numero pari (dispari) è pari (dispari).

Un'altra tipologia di numeri naturali è quella dei *numeri primi*. Sono tali i numeri interi divisibili solo per 1 e per se stessi. Sono numeri primi i numeri 2, 3, 5, 7, 11 etc.

I numeri naturali o sono numeri primi o sono il prodotto di più numeri primi.

Tra i numeri primi un cenno a parte meritano i *numeri di Mersenne* (1588-1648), che sono del tipo:

$$M_n = 2^n - 1$$

con n intero. Finora il più grande *numero di Mersenne* conosciuto è $2^{1257827} - 1$, che conta 378.632 cifre.

b) Numeri relativi

L'operazione di sottrazione non è sempre possibile nell'insieme N dei naturali, perciò la prima estensione del concetto di numero si verifica con l'introduzione degli interi relativi, indicati con la simbologia $Z = \{\dots, -3, -2, -1, 0, +1, +2, +3, \dots\}$.

Nell'insieme considerato sono definite:

- la *relazione d'ordine totale* $\leq$;
- le operazioni di *addizione* e di *moltiplicazione*, con le loro *proprietà*;
- l'operazione di *sottrazione*, in quanto esiste lo zero e l'*opposto* di ogni elemento.

L'operazione di divisione continua a non essere definita.

c) Numeri razionali

L'estensione del concetto di numero dai relativi Z ai razionali Q rende lecita sempre anche l'operazione di divisione. Questo merita delle puntualizzazioni.

Siano p e q due interi relativi, con q non nullo. Nell'insieme di tutte le frazioni p/q , le due frazioni p/q e p'/q' sono equivalenti se e solo se $pq' = p'q$; per cui ogni numero razionale è rappresentato non solo da una frazione, ma da tutte le altre equivalenti ad essa.

Nell'insieme considerato sono definite:

- la *relazione d'ordine totale* $\leq$;
- le operazioni di *addizione* e di *moltiplicazione*, con le loro *proprietà*;
- l'operazione di *sottrazione*, in quanto esiste lo zero e l'*opposto* di ogni elemento;
- l'operazione di *divisione* di un qualsiasi numero razionale per un altro diverso da zero.

La rappresentazione decimale di un numero razionale $r = \frac{p}{q}$ si ottiene eseguendo la divisione di p per q , ed è del tipo:

$$\pm m, c_1 c_2 \dots c_h \dots$$

dove m è un numero intero non negativo, preceduto dal segno $+$ o $-$, e seguito da una virgola e da una successione di numeri interi $c_1 c_2 \dots c_h \dots$, compresi tra 0 e 9, che si dicono *cifre decimali*. La successione di cifre decimali è *periodica*, nel senso che, a partire da un certo indice in poi, un gruppo di cifre si ripete indefinitamente; tale gruppo è denominato *periodo* ed è indicato sovrapposandogli un trattino, mentre il gruppo di cifre che precede il periodo è denominato *antiperiodo*. Siano $c_1 c_2 \dots c_h$ le cifre costituenti l'antiperiodo, e $c_{h+1} \dots c_{h+i}$ le cifre costituenti il periodo. La rappresentazione del numero razionale, che è denominata *allineamento decimale periodico*, diventa la seguente:

$$\pm m, c_1 c_2 c_h \overline{c_{h+1} \dots c_{h+i}}$$

È ovvio che se il periodo è la sola cifra 0, la rappresentazione decimale del numero razionale è:

$$\pm m, c_1 c_2 \dots c_h \dots$$

d) Numeri reali

L'insieme Q dei numeri razionali presenta due ostacoli, il primo di carattere geometrico: i numeri razionali consentono di misurare solo grandezze commensurabili con una data unità di misura U ; il secondo di carattere algebrico: la risoluzione di equazioni algebriche del tipo $x^2 - 2 = 0$ è impossibile nell'insieme dei razionali.

L'estensione dei numeri dai razionali Q ai reali R elimina questi due ostacoli.

Alla diagonale del quadrato di lato 1 corrisponde il numero reale $\sqrt{2}$ che è la sua misura; a tal proposito, si considerino i numeri:

$$1; 1,4; 1,41; 1,414; \dots$$

$$2; 1,5; 1,42; 1,415; \dots$$

e i loro rispettivi quadrati:

$$1; 1,96; 1,9881; 1,999396; \dots$$

$$4; 2,25; 2,0164; 2,002225; \dots$$

Da ciò si evince che non esiste alcun numero razionale il cui quadrato è 2. Le classi di numeri:

$$A = \{1; 1,4; 1,41; 1,414; \dots\}$$

$$B = \{2; 1,5; 1,42; 1,415; \dots\}$$

sono i valori approssimati per difetto e per eccesso, rispettivamente, della radice quadrata di 2; esse sono *separate*, in quanto ogni numero della prima classe è minore di qualsiasi numero

della seconda classe, e *contigue*, in quanto, scelto un numero positivo ε , per quanto piccolo, è sempre possibile individuare un elemento della prima e un elemento della seconda tali che la loro differenza sia minore di ε .

Le due classi così costruite identificano un *numero irrazionale* avente un numero illimitato di cifre decimali con un allineamento non periodico, maggiore di qualsiasi elemento della prima classe e minore di qualsiasi elemento della seconda classe.

Queste osservazioni conducono a considerare l'insieme R dei numeri reali come costituito da:

- *numeri razionali*, ossia quei numeri reali espressi da allineamenti decimali periodici;
- *numeri irrazionali*, ossia quei numeri reali espressi da allineamenti decimali non periodici.

Per accettare questo nuovo insieme è necessario ricordare che in esso sono definite:

- la *relazione d'ordine totale* $\leq$;
- le operazioni di *addizione* e di *moltiplicazione*, con le loro *proprietà*;
- l'operazione di *sottrazione*, in quanto esiste lo zero e l'*opposto* di ogni elemento;
- l'operazione di *divisione* di un qualsiasi numero razionale per un altro diverso da zero.

Una rappresentazione dei numeri reali è fornita da una retta ordinata su cui si sia fissata un'origine 0 e una unità di misura; tale retta è detta *retta reale*. Ad ogni punto della retta corrisponde uno e un solo numero reale, e viceversa, ad ogni numero reale corrisponde uno e un solo punto della retta. Il segno + o - indica il verso di percorrenza della retta, mentre il numero senza segno ne indica la lunghezza.

e) Numeri complessi

L'ente numerico:

$$a + ib$$

si dice *numero complesso* ed esprime, in forma algebrica, la coppia ordinata (a, b) di numeri reali; in esso a e b sono coefficienti, e precisamente a si dice *parte reale* e b *coefficiente dell'unità immaginaria* i ; inoltre, a e b sono diversi da zero.

L'introduzione dell'insieme dei numeri complessi, indicato con C , è lecita in quanto è possibile sapere quali di essi sono uguali e quali diversi, e in quanto si possono estendere ad essi le regole basilari del calcolo.

Il nuovo insieme numerico è più ampio dell'insieme dei numeri reali, ha la struttura algebrica di *campo* e consente l'operazione di estrazione della radice quadrata di un numero reale negativo. In particolare, quest'ultima operazione può essere eseguita in quanto l'elemento del nuovo insieme numerico C , indicato con i , è tale che $i^2 = -1$.

I *numeri complessi* consentono di dare soluzione ad equazioni del tipo:

$$x^2 = -a \quad \text{con } a \text{ numero reale positivo}$$

I numeri complessi la cui parte reale sia nulla si dicono *immaginari puri*, e il loro quadrato è negativo.

Dato il significato dell'unità immaginaria, si ha:

$$\begin{aligned} i^0 &= 1 \\ i^1 &= i \\ i^2 &= -1 \\ i^3 &= i^2 \cdot i = -i \\ i^4 &= i^2 \cdot i^2 = 1 \\ &\dots\dots\dots \\ i^7 &= i^4 \cdot i^3 = -i \end{aligned}$$

Il *valore assoluto* o *modulo* del numero complesso $a + ib$ è dato da:

$$|a + ib| = \sqrt{a^2 + b^2}$$

Il *numero complesso coniugato* di $a + ib$ è $a - ib$.

2 Numeri algebrici e numeri trascendenti

Un *numero algebrico* è la soluzione dell'*equazione polinomiale*:

$$a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n = 0$$

in cui $a_0 \neq 0$ e $a_1, a_2, \dots, a_n$ sono numeri interi e n è un intero positivo.

Un *numero trascendente* è, invece, un numero che si può esprimere come la soluzione di una qualsiasi equazione polinomiale con coefficienti interi. Esempi tipici di numeri trascendenti sono i numeri π , rapporto costante tra circonferenza e diametro, ed e , base dei logaritmi naturali.

L'insieme dei numeri algebrici è un insieme infinito numerabile, mentre l'insieme dei numeri trascendenti è un insieme infinito non numerabile.

3 Le strutture algebriche

Prima di dare la definizione di *struttura algebrica*, si daranno le definizioni di *legge di composizione interna* e *legge di composizione esterna*.

Si dice *legge di composizione interna* di un insieme A , oppure operazione *binaria* in A , la *relazione*:

$$f: A \times A \rightarrow A$$

generata dalla corrispondenza creata da una certa operazione, che associ ad ogni coppia ordinata di elementi $a, b \in A$ un dato elemento $c \in A$.

Si dice, invece, *legge di composizione esterna* di un insieme A , la *relazione*:

$$f: A \times A \rightarrow B$$

generata dalla corrispondenza creata da una certa operazione, che associ ad ogni coppia ordinata di elementi $a, b \in A$ un dato elemento c appartenente ad un altro insieme B .

A questo punto è possibile dare la definizione di *struttura algebrica*.

Ad un insieme è impressa una struttura algebrica se in esso sono introdotte una o più leggi di composizione.

Si dice *struttura algebrica* S , indicata con:

$$S = (A, T)$$

un insieme A che è dotato di una legge di composizione interna T , tale che ad ogni coppia $a, b \in A$ corrisponda un elemento c di A .

L'insieme A è denominato *sostegno della struttura*, l'elemento c è denominato, invece, *risultato* o *composto* dell'operazione T . È ovvio che ogni struttura può essere dotata di più di una legge di composizione interna.

Nell'insieme N dei naturali si ha:

$$S = (N, +)$$

in quanto l'insieme N è una struttura algebrica dotata dell'operazione di addizione, che costituisce, appunto, una legge di composizione interna.

Sulla base delle definizioni degli insiemi N, Z, Q e indicando con:

$$Q_0 = Q - \{0\}$$

l'insieme dei numeri razionali escluso lo zero, sono strutture algebriche, quindi, le seguenti:

$$\begin{aligned} &(N, +); \quad (N, \cdot) \\ &(Z, +); \quad (Z, -); \quad (Z, \cdot) \\ &(Q, +); \quad (Q, -); \quad (Q, \cdot) \\ &(Q_0, :) \end{aligned}$$

Un elemento h del sostegno A di una struttura $S = (A, T)$ si dice *neutro*, per una legge di composizione interna T definita su A , se:

$$\forall a \in A \Rightarrow a T h = h T a = a$$

Gli elementi neutri delle strutture $(Q, +)$ e $(Q, \cdot)$ sono, rispettivamente, 0 e 1.

Un elemento a' del sostegno A di una struttura $S = (A, T)$, dotata di elemento neutro h , si dice *simmetrico dell'elemento a appartenente ad A* se si ha:

$$a T a' = a' T a = h$$

L'elemento a si dice *simmetrizzabile*.

Tutti gli elementi della struttura $(Z, +)$ sono simmetrizzabili, e l'elemento neutro è 0.

Nell'insieme Q_0 dei numeri razionali relativi, escluso lo zero, tutti gli elementi hanno un reciproco e sono simmetrizzabili all'interno della struttura $(Q_0, \cdot)$; l'elemento neutro è 1:

$$\forall a \in Q_0 \Rightarrow a \cdot \frac{1}{a} = 1$$

4 Le proprietà delle strutture algebriche

La legge di composizione interna T , che opera in una struttura algebrica, può godere delle seguenti proprietà:

— *proprietà commutativa*, se:

$$\forall a, b \in A \Rightarrow a T b = b T a$$

— *proprietà associativa*, se:

$$\forall a, b, c \in A \Rightarrow (a T b) T c = a T (b T c)$$

— *proprietà della cancellazione*: $\forall a, b, c \in A$

$$\text{se } a T b = c T b \Rightarrow a = c \text{ (cancellazione a destra)}$$

$$\text{se } b T a = b T c \Rightarrow a = c \text{ (cancellazione a sinistra)}$$

Secondo i casi, l'elemento b si dice *semplificabile a destra* o *a sinistra*, oppure semplicemente *regolare*, se è semplificabile sia a destra sia a sinistra.

5 Struttura abeliana e struttura regolare

La struttura $S = (A, T)$ si dice *abeliana* se la legge di composizione interna T gode della proprietà commutativa; si dice *regolare* se tutti gli elementi dell'insieme A sono regolari o semplificabili.

La struttura $(N, +)$ è abeliana e regolare; è abeliana ma non regolare la struttura $(N, \cdot)$, perché l'insieme N contiene lo zero, che non è regolare ($a \cdot 0 = b \cdot 0 = 0$ non implica necessariamente che $a = b$).

6 Semigruppı e gruppi

La struttura $S = (A, T)$ si dice *semigruppı* se la legge di composizione interna T gode della proprietà associativa.

Se la legge gode anche della proprietà commutativa, allora il semigruppı si dice *abeliano*.

Sono *semigruppı* le strutture $(N, +)$ e $(N, \cdot)$; essi sono anche abeliani.

Un semigruppı si dice *monoide* se è dotato di elemento neutro.

Le strutture $(Q, +)$ e $(Q, \cdot)$ sono *monoidi*, e i loro elementi neutri sono, rispettivamente, 0 e 1.

Un monoide si dice *gruppı* se, in esso, ogni elemento è simmetrizzabile.

Se la legge di composizione interna di un gruppı è commutativa, il monoide si dice *abeliano*.

La struttura moltiplicativa $(Q_0^+, \cdot)$ è un gruppı abeliano, in quanto nell'insieme Q_0^+ , sostegno della struttura, la moltiplicazione è commutativa, associativa, ha elemento neutro 1 ed ogni elemento è simmetrizzabile (ha un reciproco).

7 L'anello

Sia data la struttura algebrica:

$$S = (A, T, \Omega)$$

dove T e Ω sono due leggi di composizione interne, considerando T come prima legge e Ω come seconda legge; tale struttura si dice *anello* se:

- la struttura (A, T) è un gruppı abeliano;
- la struttura (A, Ω) è un semigruppı;
- la seconda legge (Ω) è distributiva rispetto alla prima sia a destra sia a sinistra.

Un anello si dice *abeliano* quando la legge di composizione interna del semigruppı (A, Ω) gode della proprietà commutativa.

Un anello, inoltre, si dice *unitario* quando il semigruppı (A, Ω) è un monoide.

La struttura $(Z, +, \cdot)$ è un anello abeliano unitario in quanto:

- la struttura $(Z, +)$ è un gruppı abeliano;
- la struttura $(Z, \cdot)$ è un semigruppı;
- la moltiplicazione è distributiva sia a destra sia a sinistra rispetto all'addizione;
- la moltiplicazione è commutativa;
- 1 è l'elemento neutro della moltiplicazione.

8 Il campo

Sia data la struttura algebrica:

$$S = (A, T, \Omega)$$

si indichi con A_0 l'insieme A privo dell'elemento neutro di (A, T) ; tale struttura si dice *campo* se:

- la struttura (A, T) è un gruppı abeliano;
- la struttura (A_0, Ω) è un gruppı abeliano;
- la seconda legge (Ω) è distributiva rispetto alla prima (T) sia a destra sia a sinistra.

È facile dimostrare che la struttura $S = (Q, +, \cdot)$ è un campo.



Espansione Web
Le equazioni diofantee

Il linguaggio dell'algebra lineare e il calcolo vettoriale

1 Le matrici

La *matrice* è un insieme di numeri disposti per righe e per colonne.

In genere, una matrice è contrassegnata da una lettera maiuscola, mentre i numeri sono contrassegnati da una lettera a seguita da un primo pedice, che ne indica la riga, e da un secondo pedice, che ne indica la colonna. Ogni numero si chiama *elemento della matrice*.

Una matrice A di tipo $m \times n$, ossia con m righe e n colonne, si scrive nel modo seguente:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Se una matrice è costituita da una sola riga si chiama *vettore riga*; se, invece, è costituita da una sola colonna si chiama *vettore colonna*.

Se il numero di righe coincide con il numero delle colonne, allora la matrice si dice *matrice quadrata*. Il numero di righe (o di colonne) di una matrice quadrata si definisce *ordine*.

2 Matrici particolari

L'insieme degli elementi di una matrice aventi lo stesso numero di riga e di colonna è definito *diagonale principale*.

Una particolare matrice quadrata è detta *matrice unitaria* o *matrice identità*. In essa gli elementi della diagonale principale sono uguali a 1, mentre tutti gli altri elementi sono uguali a 0. Tale matrice viene indicata con una lettera I seguita da un indice che rappresenta il numero delle righe (o colonne). La matrice unitaria di ordine 4, indicata con I_4 è:

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

La *matrice diagonale* differisce dalla matrice unitaria per avere sulla diagonale principale elementi non tutti unitari.

Una matrice quadrata di ordine n si dice *simmetrica* se:

$$a_{ij} = a_{ji} \quad \forall i \in j$$

Una matrice avente elementi tutti uguali a 0 è detta *matrice nulla*, ed è indicata con 0. Un esempio di matrice nulla è il seguente:

$$0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Una matrice quadrata di ordine n si dice *triangolare superiore* se tutti gli elementi al di sotto della diagonale principale sono nulli; si dice, invece, *triangolare inferiore* se sono nulli tutti gli elementi al di sopra della diagonale principale.

La *trasposta di una matrice* è una matrice ottenuta scambiando le righe con le colonne della matrice stessa. Sia A la matrice, la sua trasposta è indicata con A' :

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \quad A' = \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ a_{13} & a_{23} & a_{33} & a_{43} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}$$

Ovviamente, la trasposta di una trasposta è la matrice originaria.

La trasposta di una matrice identità è la matrice stessa.

3 Le operazioni sulle matrici

Date due matrici A e B , esse sono *uguali* se ciascun elemento di A è uguale al corrispondente (ossia che occupa lo stesso posto) elemento di B . Si considerino le due matrici:

$$A = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix}$$

Esse sono uguali.

a) Somma di due matrici

Date due matrici A e B aventi lo stesso numero di righe e di colonne, si definisce *somma di A e B* la matrice $C = A + B$ avente lo stesso numero di righe e di colonne di A e B e i cui elementi si ottengono operando la somma dei corrispondenti elementi di A e di B . Traducendo in formule, per il generico elemento di posto ij della matrice C , si ha:

$$c_{ij} = a_{ij} + b_{ij}$$

La somma gode delle seguenti due proprietà:

— *proprietà commutativa*, infatti:

$$A + B = B + A$$

— *proprietà associativa*, infatti:

$$A + (B + C) = (A + B) + C$$

b) Differenza di due matrici

La differenza tra due matrici A e B è una matrice avente lo stesso numero di righe e di colonne di A e B , e i cui elementi si ottengono operando la differenza dei corrispondenti elementi di A e di B .

c) Prodotto di due matrici

Per effettuare il prodotto tra due matrici è necessario che il numero delle colonne della prima sia uguale al numero delle righe della seconda.

Date due matrici A (di tipo $m \times n$) e B (di tipo $n \times r$), si definisce *prodotto della matrice A per la matrice B* , la matrice $C = AB$ di tipo $m \times r$ in cui l'elemento di posto ij è la somma dei prodotti degli elementi di uguale posto della riga i -esima di A per la colonna j -esima di B ; per il generico elemento di posto ij , si ha:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}$$

o, in forma compatta:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

La matrice che si ottiene ha un numero di righe uguale a quello della prima matrice e un numero di colonne uguale a quello della seconda.

Nel prodotto tra matrici è fondamentale l'ordine di moltiplicazione, nel senso che, pur essendo possibile il prodotto AB , può non essere possibile il prodotto BA , per la regola del numero di righe e di colonne. Inoltre, anche se il prodotto $B \times A$ fosse possibile, solitamente:

$$AB \neq BA$$

per cui il prodotto tra matrici non è commutativo.

Le due matrici I e O assumono un ruolo simile al corrispondente di 1 e 0 per gli scalari; infatti, si ha:

$$A + O = A$$

$$A - A = O$$

$$A \cdot O = O$$

$$A \cdot I = A$$

A differenza degli scalari, è possibile che il prodotto di due matrici sia una matrice nulla, anche se nessuna delle due è una matrice nulla.

d) Prodotto di una matrice per uno scalare

Data una matrice A e uno scalare λ , il prodotto di λ per A è la matrice λA i cui elementi si ottengono moltiplicando ciascun elemento di A per λ . Si consideri la matrice A definita in precedenza; il prodotto di A per λ è la matrice:

$$\lambda A = \begin{bmatrix} \lambda a_{11} & \lambda a_{12} & \dots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \dots & \lambda a_{2n} \\ \dots & \dots & \dots & \dots \\ \lambda a_{m1} & \lambda a_{m2} & \dots & \lambda a_{mn} \end{bmatrix}$$

4 I determinanti

Il *determinante* è un numero reale associato ad una matrice quadrata.

Per calcolare il determinante di una matrice A (di solito indicato con $|A|$) di ordine n è necessaria la seguente definizione: «si dice *complemento algebrico* (o *cofattore*) dell'elemento a_{ij} della matrice A di ordine n , e si indica con A_{ij} , il determinante della matrice che si ottiene cancellando dalla matrice A la riga i -esima e la colonna j -esima, moltiplicato per il termine $(-1)^{i+j}$, che è positivo o negativo a seconda che $i + j$ sia pari o dispari».

Considerando la matrice quadrata A del terzo ordine:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

il complemento algebrico dell'elemento a_{11} non è altro che il determinante della matrice ottenuta cancellando la prima riga e la prima colonna (in pratica la riga e la colonna che contengono l'elemento a_{11}), moltiplicato per $(-1)^{1+1}$; essendo $i = 1$ e $j = 1$, esso è:

$$A_{11} = (-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$$

Il determinante di una matrice deriva dalla somma dei prodotti tra gli elementi di una riga (o di una colonna) qualunque e i propri complementi algebrici; il determinante della matrice A di ordine n è, dunque:

$$|A| = \sum_{j=1}^n a_{ij} A_{ij} \quad i = 1, 2, \dots, n$$

Nel caso di una *matrice del primo ordine*, il determinante è pari al numero stesso che costituisce la matrice, ossia:

$$|a_{11}| = a_{11}$$

Nel caso di una *matrice del secondo ordine*, risulta invece:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12} a_{21}$$

Per una del *terzo ordine*:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} A_{11} + a_{12} A_{12} + a_{13} A_{13} = a_{11} (a_{22} a_{33} - a_{23} a_{32}) - a_{12} (a_{21} a_{33} - a_{23} a_{31}) + a_{13} (a_{21} a_{32} - a_{22} a_{31}) = a_{11} a_{22} a_{33} - a_{11} a_{23} a_{32} + a_{12} a_{23} a_{31} - a_{12} a_{21} a_{33} + a_{13} a_{21} a_{32} - a_{13} a_{22} a_{31}$$

Se in una matrice A si omettono righe e/o colonne, si ottengono delle submatrici; i determinanti delle submatrici di ordine p si chiamano *minori di ordine p* della matrice A .

Si dice che la matrice A di tipo $m \times n$, con $m \neq n$, ha rango (o caratteristica) r (intero) se tra i minori di ordine r della matrice A ve ne è almeno uno diverso da zero e se tutti i minori di ordine $r + 1$, se esistono, sono nulli.

Il rango di una matrice, in altre parole, è definibile come il massimo ordine dei suoi minori non nulli. Esso non può essere superiore al più piccolo dei numeri m e n .

5 Le proprietà dei determinanti

Le proprietà dei determinanti, fondamentali in quanto, talvolta, rendono più semplice il loro calcolo, sono le seguenti:

- il determinante di una matrice avente una riga (o una colonna) costituita da tutti zero è nullo;
- il determinante di una matrice avente due righe (o colonne) uguali o proporzionali è nullo;
- se in una matrice si scambiano due righe (o colonne) il determinante cambia di segno;
- moltiplicando per λ gli elementi di una riga (o colonna) di una matrice, il determinante risulta moltiplicato anch'esso per λ ;
- il determinante di una matrice è uguale a quello della sua trasposta;
- la somma dei prodotti tra gli elementi di una riga (o colonna) e i complementi algebrici di un'altra riga (o colonna) è nulla;
- il determinante della somma di due matrici non è generalmente uguale alla somma dei determinanti delle due matrici.

6 L'inversa di una matrice

Una matrice quadrata si dice *singolare* se il suo determinante è uguale a zero, *non singolare* se il suo determinante è diverso da zero.

Data una matrice quadrata A di ordine n , si dice *matrice inversa di A* , e si indica con A^{-1} , o con $(a_{ij})^{-1}$, la matrice che soddisfa la relazione:

$$A^{-1}A = AA^{-1} = I_n$$

Per verificare se esiste e, in caso positivo, per vedere come si calcola l'inversa di una matrice, si deve premettere, innanzi tutto, che il determinante della matrice identità, come è facile dimostrare, è uguale a 1; in secondo luogo, si deve considerare la *matrice aggiunta* A^+ di A . Essa si ottiene determinando la trasposta A' di A , e calcolando i complementi algebrici di A' ed è data da:

$$A^+ = \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}$$

L'inversa di una matrice quadrata si ottiene dividendo ciascun elemento della matrice aggiunta per il determinante della matrice originaria, vale a dire:

$$A^{-1} = \frac{A^+}{|A|}$$

Si considerino, ora, alcune proprietà delle matrici inverse:

- l'inversa di una inversa è la matrice originaria:

$$(A^{-1})^{-1} = A$$

- l'inversa di un prodotto è uguale al prodotto delle inverse delle matrici invertite nell'ordine:

$$(AB)^{-1} = B^{-1}A^{-1}$$

- l'inversa di una trasposta è uguale alla trasposta dell'inversa:

$$(A')^{-1} = (A^{-1})'$$

- l'inversa della matrice identità è la stessa matrice identità:

$$I^{-1} = I$$

Moltiplicando primo e secondo membro del sistema espresso in forma matriciale $Ax = b$ per A^{-1} , che è la matrice inversa di A ed esistente solo se $|A| \neq 0$, si ha:

$$x = A^{-1}b$$

La soluzione del sistema si ottiene, quindi, moltiplicando la matrice inversa A^{-1} per il vettore b .

Ricordando che la matrice inversa A^{-1} è data da $A^{-1} = \frac{A^*}{|A|}$, la soluzione può essere riscritta nel modo seguente:

$$\begin{bmatrix} x_1 \\ x_1 \\ \dots \\ x_n \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_1 \\ \dots \\ b_n \end{bmatrix}$$

Per cui la i -esima soluzione è data da:

$$x_i = \frac{1}{|A|} (A_{1i}b_1 + A_{2i}b_2 + \dots + A_{ni}b_n) = \frac{1}{|A|} \sum_{j=1}^n A_{ji}b_j$$

dove il fattore $\sum_{j=1}^n A_{ji}b_j$ è il determinante della matrice A in cui al posto della i -esima colonna si sostituisce il vettore b dei termini noti. Pertanto, *ogni soluzione del sistema si ottiene dal rapporto tra il determinante della matrice in cui al posto dei coefficienti dell'incognita che si sta determinando si sostituiscono i termini noti, e il determinante della matrice dei coefficienti A .*

Il sistema ammette, quindi, la soluzione nulla $(0, 0, \dots, 0)$, se tutti i termini noti sono nulli.

Inoltre, è facile verificare che la soluzione è unica se $|A| \neq 0$, mentre esistono infinite soluzioni se $|A| = 0$; in tal caso il sistema è indeterminato.

La regola di Cramer richiede un numero di operazioni elementari dell'ordine di $(n+2)!$.

9 Il metodo di eliminazione di Gauss

Il metodo che ci si appresta ad esporre prende il nome dal matematico tedesco C.F. Gauss (1777-1855).

È noto l'inconveniente derivante dall'applicazione della *regola di Cramer*: il calcolo della soluzione è troppo laborioso se n è molto elevato. Per questo motivo un algoritmo che si preferisce è quello di Gauss.

Prima di analizzare il *metodo di eliminazione di Gauss* si deve precisare che per sistemi equivalenti si intendono due o più sistemi aventi le stesse soluzioni.

Il *metodo di eliminazione di Gauss* trasforma un dato sistema lineare in un altro equivalente ad esso, la cui matrice dei coefficienti sia *triangolare superiore*, e risolubile con un metodo di risoluzione cosiddetto *all'indietro*; esso consiste nell'eliminazione successiva di tutte le variabili fino a ridursi ad un'equazione in una incognita.

Essenzialmente, tale metodo si articola nei seguenti passi.

- Normalizzazione della prima equazione

Supposto che $a_{11} \neq 0$, si divide la prima equazione per a_{11} , ottenendo un'equazione in cui il coefficiente della variabile x_1 sia uguale a 1.

• Eliminazione della variabile x_1

Tutte le equazioni in cui si presenta x_1 , vale a dire tutte le equazioni in cui $a_{j1} \neq 0$, sono divise per a_{j1} e sottratte alla prima equazione, in modo da ottenere $n-1$ equazioni in cui non si presenta la variabile x_1 ; si ottiene il sistema equivalente al sistema (1):

$$\begin{aligned} x_1 + a'_{12}x_2 + \dots + a'_{1n}x_n &= b'_1 \\ a'_{22}x_2 + \dots + a'_{2n}x_n &= b'_2 \\ &\dots\dots\dots \\ a'_{n2}x_2 + \dots + a'_{nn}x_n &= b'_n \end{aligned} \quad (2)$$

Eseguendo le operazioni di cui ai suddetti passi, sul sistema in $n-1$ equazioni in $n-1$ incognite ottenibile eliminando dal sistema (2) la prima equazione, si perviene ad un sistema equivalente al sistema (2) in cui nelle ultime $n-2$ equazioni non compare più x_2 .

Il procedimento descritto si ripete fino a pervenire ad un sistema equivalente al sistema (1) del tipo:

$$\begin{aligned} x_1 + b_{12}x_2 + \dots + b_{1n}x_n &= c_1 \\ x_2 + \dots + b_{2n}x_n &= c_2 \\ &\dots\dots\dots \\ x_n &= c_n \end{aligned}$$

che si risolve facilmente.

Il metodo di eliminazione di Gauss richiede un numero di operazioni elementari dell'ordine di n^3 .

10 Il teorema di Rouché-Capelli

Sia dato un sistema lineare di m equazioni in n incognite $x_1, x_2, \dots, x_n$ del tipo:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases} \quad \text{con } m \neq n$$

Si consideri il seguente teorema.

TEOREMA DI ROUCHÉ-CAPELLI • Condizione necessaria e sufficiente affinché il sistema lineare di m equazioni in n incognite $Ax = b$ ammetta soluzione è che la matrice A dei coefficienti e la matrice Ab dei coefficienti e dei termini noti abbiano lo stesso rango:

$$\text{rango}(A) = \text{rango}(Ab)$$

dove Ab è uguale a:

$$Ab = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{bmatrix}$$

Sia r il rango comune delle matrici A e Ab . Supponendo che r sia minore o uguale al più piccolo tra i numeri m e n , si prendono r righe e r colonne della matrice A , ottenendo una submatrice di A di

- esiste $0 \in V$ tale che $u + 0 = 0 + u = u$ per ogni $u \in V$;
- esiste $-u \in V$ tale che $u + (-u) = (-u) + u = 0$ per ogni $u \in V$.

L'elemento neutro 0 di V è detto *vettore nullo*.

Pertanto, l'insieme V con l'operazione di addizione costituisce un gruppo abeliano.

2. Un'operazione esterna di *moltiplicazione* tra gli elementi di K e quelli di V , per la quale valgono le seguenti proprietà:

- $\lambda(u + v) = \lambda u + \lambda v$ per ogni $\lambda \in K$ e $u, v \in V$;
- $(\lambda + \alpha)u = \lambda u + \alpha u$ per ogni $\lambda, \alpha \in K$ e $u \in V$;
- $(\lambda\alpha)u = \lambda(\alpha u)$ per ogni $\lambda, \alpha \in K$ e $u \in V$;
- $1u = u$ per ogni $u \in V$.

Ogni elemento dello spazio vettoriale è detto *vettore*.

Un sottoinsieme $W \subseteq V$ è un *sottospazio vettoriale* dello spazio vettoriale V , chiuso rispetto all'addizione e alla moltiplicazione, vale a dire è tale che $u + w \in W$ e $\lambda w \in W$ per ogni $u, w \in W$ e $\lambda \in K$.

13 La combinazione lineare nello spazio vettoriale

Sia V uno spazio vettoriale su un campo K . Siano dati i k vettori $v_1, v_2, \dots, v_k \in V$ e gli scalari (o numeri reali) $\alpha_1, \alpha_2, \dots, \alpha_k \in K$. L'espressione:

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_k v_k$$

o, in forma compatta:

$$\sum_{i=1}^k \alpha_i v_i$$

si definisce *combinazione lineare* dei k vettori di V ed individua un elemento di V .

Il sottospazio vettoriale W generato dai vettori $v_1, v_2, \dots, v_k$ e costituiti dall'insieme di tutte le loro possibili combinazioni lineari è detto *span dei vettori* $v_1, v_2, \dots, v_k$, e si indica con $\text{Span}(W)$.

Se esistono $\alpha_1, \alpha_2, \dots, \alpha_k \in K$, non tutti nulli, l'insieme di vettori $v_1, v_2, \dots, v_k \in V$ dello spazio vettoriale V si dice *linearmente dipendente* se:

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_k v_k = 0$$

e l'uguaglianza appena scritta è detta *relazione di dipendenza lineare* tra $v_1, v_2, \dots, v_k$.

Viceversa, se $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_k v_k = 0$ implica che $\alpha_1, \alpha_2, \dots, \alpha_k = 0$, allora i vettori $v_1, v_2, \dots, v_k$ si dicono *linearmente indipendenti*.

14 Le basi

Un insieme $B = \{v_1, v_2, \dots, v_n\}$ di vettori di uno spazio vettoriale V è una base di V se i vettori $v_1, v_2, \dots, v_n$ rappresentano un sistema di generatori di V e se sono linearmente indipendenti.

Uno spazio vettoriale è detto di *dimensione* finita o infinita a seconda che esista o meno una base.

I vettori:

$$e_1 = (1, 0, 0, \dots, 0); \quad e_2 = (0, 1, 0, \dots, 0), \quad \dots, \quad e_n = (0, 0, 0, \dots, 1)$$

costituiscono la base canonica di R^n .

Ciascuna base di uno spazio vettoriale V è caratterizzata dal numero n di elementi, che costituisce la dimensione dello spazio vettoriale V .

Se $B = \{v_1, v_2, \dots, v_n\}$ è una base di uno spazio vettoriale V e $v \in V$, allora i coefficienti $\alpha_1, \alpha_2, \dots, \alpha_n$, tali che $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$, si dicono *coordinate di v rispetto alla base B* .

Dato un sottoinsieme A di V tale che $A \subseteq V$, si dice che $B \subseteq A$ è un *sottoinsieme massimale in A di vettori linearmente indipendenti* se gli elementi di B sono linearmente indipendenti, e unendo a B un qualsiasi altro elemento di A si ottiene un insieme di vettori linearmente dipendenti.

Da questa definizione discende il seguente teorema.

TEOREMA • Sia $B \subset V$ un sottoinsieme di uno spazio vettoriale V ; allora B è una base se e solo se è un insieme massimale di vettori linearmente indipendenti in un sistema di generatori A . Qualsiasi spazio vettoriale V ammette una base, e se V ha un sistema finito di generatori, allora ammette una base finita.

15 I vettori

Nell'algebra che si andrà a considerare, detta *algebra lineare*, il campo dei numeri reali fornisce non solo gli elementi che costituiscono le componenti di *vettori* ma anche gli *scalari*, ossia i numeri reali che emergono da alcune operazioni come i prodotti dei vettori.

Si dice *vettore* individuato da un segmento orientato non nullo $\overrightarrow{AB}$ l'insieme dei segmenti orientati di lunghezza $|\overrightarrow{AB}|$, e che, se $A \neq B$, sono paralleli ad $\overrightarrow{AB}$ ed hanno il suo stesso verso. Un vettore individuato dal segmento nullo è denominato *vettore nullo*.

Un vettore è generalmente indicato con i simboli $\vec{u}$, $\vec{v}$ etc. Se un vettore $\vec{u}$ è rappresentato da un segmento orientato $\overrightarrow{AB}$, secondo la notazione di *Grassmann* il vettore è espresso come differenza di due punti, vale a dire:

$$\vec{u} = B - A$$

Se gli estremi A e B del segmento coincidono, allora $\vec{u}$ è denominato *vettore nullo* ed è indicato con $\vec{0}$.



Un vettore è individuato da:

- un *modulo*, il quale è un numero reale che indica la lunghezza del segmento, rispetto ad una data unità di misura. Il modulo di un vettore $\vec{u}$ si indica con $|\vec{u}|$ o semplicemente con u ;
- un *verso*, che è quello individuato dalla freccia;
- una *direzione*, che è quella della retta cui appartiene il segmento.

Due vettori $\vec{u}$ e $\vec{v}$ sono uguali se hanno lo stesso modulo, lo stesso verso e la medesima direzione.

Dato un vettore $\vec{v}$ e un punto A del piano, si dice *vettore $\vec{v}$ applicato in A* l'unico segmento orientato $\overrightarrow{AB}$ che individua $\vec{v}$.

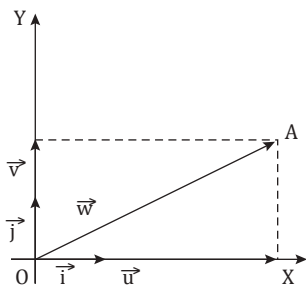
Se due segmenti orientati $\overrightarrow{AB}$ e $\overrightarrow{CD}$ rappresentano uno stesso vettore, allora il quadrilatero $ABCD$ è un parallelogramma.



Si dice *angolo orientato* tra due vettori $\vec{u}$ e $\vec{v}$, l'angolo orientato formato da due semirette aventi la stessa origine, e direzione e verso, rispettivamente, di $\vec{u}$ e $\vec{v}$.

16 Le coordinate cartesiane di vettori

Si dicono *versori degli assi coordinati* i vettori $\vec{i}$ e $\vec{j}$ che, applicati nell'origine O del riferimento cartesiano, danno due segmenti di lunghezza unitaria. Il primo giace sull'asse delle ascisse x ed ha la sua stessa direzione e il suo stesso verso; il secondo giace sull'asse delle ordinate ed ha la sua stessa direzione e il suo stesso verso.



Ogni vettore del tipo $\vec{w} = A - O$ può essere scomposto, in modo unico, in due vettori $\vec{u}$ e $\vec{v}$; il primo ha la direzione dell'asse x , il secondo quella dell'asse y . In formula si ha:

$$\vec{w} = \vec{u} + \vec{v}$$

Indicando con a il modulo di $\vec{u}$ e con b il modulo di $\vec{v}$, si scrive:

$$\vec{u} = a\vec{i} \quad \text{e} \quad \vec{v} = b\vec{j}$$

da cui:

$$\vec{w} = a\vec{i} + b\vec{j}$$

I numeri a e b sono le *componenti* di $\vec{w}$ rispetto al sistema di riferimento cartesiano considerato; siccome, in un sistema di riferimento cartesiano Oxy , le componenti del vettore $\vec{w} = A - O$ corrispondono alle coordinate cartesiane del punto A , allora le componenti del vettore $\vec{w}$ si chiamano anche sue *coordinate cartesiane*.

Precisamente, il numero a è l'ascissa di $\vec{w}$ e il numero b è l'ordinata di $\vec{w}$.

Dati due punti $A(x_1, y_1)$ e $B(x_2, y_2)$, il modulo del vettore $\vec{w} = B - A$ è dato dalla radice quadrata della somma dei quadrati delle sue coordinate; le coordinate del vettore $B - A$ sono date dalla differenza delle corrispondenti coordinate dei punti A e B , per cui si ha:

$$w = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

ESEMPIO 1

Determinare il modulo del vettore avente per estremi:

$$A(2, \sqrt{7}) \quad \text{e} \quad B(-1, 5)$$

Il modulo w del vettore è pari a:

$$w = \sqrt{(-1-2)^2 + (5-\sqrt{7})^2} = \sqrt{9 + 25 + 7 - 10\sqrt{7}} = \sqrt{41 - 10\sqrt{7}}$$

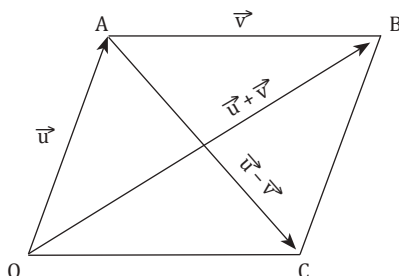
17 Le operazioni sui vettori

a) Somma di due vettori

La somma di due vettori $\vec{u}$ e $\vec{v}$ è il vettore:

$$\vec{w} = \vec{u} + \vec{v}$$

che si ottiene fissando il punto terminale di $\vec{u}$ nel punto iniziale di $\vec{v}$, e congiungendo il punto iniziale di $\vec{u}$ con il punto terminale di $\vec{v}$.



Tale definizione si esprime mediante la *regola del parallelogrammo*, che si enuncia in questo modo:

La somma di due vettori $\vec{u}$ e $\vec{v}$, uscenti da un punto O , è il vettore rappresentato dalla diagonale, uscente sempre da O , del parallelogrammo costruito sui due vettori $\vec{u}$ e $\vec{v}$.

La precedente definizione può essere estesa alla somma di un qualunque numero di vettori; in questo caso prende il nome di *regola della poligonale*, secondo cui:

La somma di più vettori è data dal vettore che unisce un dato punto O con l'ultimo estremo della linea poligonale i cui lati rappresentano i singoli vettori, e sono tracciati l'uno successivamente all'altro, a partire sempre dal punto O .

La somma gode delle seguenti due proprietà:

— *proprietà commutativa*, infatti:

$$\vec{v} + \vec{u} = \vec{u} + \vec{v}$$

— *proprietà associativa*, infatti:

$$(\vec{v} + \vec{u}) + \vec{w} = \vec{v} + (\vec{u} + \vec{w})$$

L'elemento neutro è il vettore nullo $\vec{0}$.

b) Differenza di due vettori

Due vettori si dicono *opposti* quando hanno uguale modulo, stessa direzione e versi contrari.

L'opposto di $\vec{v}$ è $-\vec{v}$, pertanto:

la differenza di due vettori $\vec{u}$ e $\vec{v}$ è data dalla somma di $\vec{u}$ più l'opposto di $\vec{v}$.

$$\vec{u} - \vec{v} = \vec{u} + (-\vec{v})$$

La differenza dei due vettori visti nella figura precedente è rappresentata dall'altra diagonale del parallelogrammo.

c) Prodotto scalare

Dati due vettori $\vec{u}$ e $\vec{v}$, si dice *prodotto scalare* di $\vec{u}$ e $\vec{v}$ il numero reale che si ottiene moltiplicando il prodotto dei moduli dei due vettori per il coseno del loro angolo $\hat{\alpha}$:

$$\vec{u} \cdot \vec{v} = uv \cos \hat{\alpha}$$

Il numero reale che si ottiene sarà positivo, nullo o negativo a seconda che l'angolo tra i due vettori sia, rispettivamente, acuto, retto oppure ottuso.

Siano $\vec{u}$, $\vec{v}$ e $\vec{w}$ tre vettori e sia λ un numero reale. Il prodotto scalare gode delle seguenti proprietà:

— *proprietà commutativa*:

$$\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$$

— *proprietà associativa*:

$$(\lambda \vec{u}) \cdot \vec{v} = \lambda(\vec{u} \cdot \vec{v})$$

— *proprietà distributiva del prodotto scalare rispetto alla somma di due vettori*:

$$(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{w} \cdot \vec{u} + \vec{w} \cdot \vec{v}$$

Inoltre, se $\vec{u}$ e $\vec{v}$ sono due vettori non nulli, ortogonali tra loro, ossia se $\vec{u} \perp \vec{v}$, allora:

$$\vec{u} \cdot \vec{v} = uv \cos 90^\circ = 0$$

Per cui, se dovesse essere:

$$\vec{u} \cdot \vec{v} = 0$$

non essendo nulli $\vec{u}$ e $\vec{v}$, deve essere necessariamente:

$$\cos \hat{\alpha} = 0$$

Da cui segue:

$$\hat{\alpha} = \pm 90^\circ$$

Pertanto i due vettori $\vec{u}$ e $\vec{v}$ sono ortogonali.

— *Il quadrato di un vettore è uguale al quadrato del suo modulo*:

$$\vec{u} \cdot \vec{u} = uu \cos 0^\circ = u^2$$

— *Il prodotto scalare di due versori $\vec{p}$ e $\vec{q}$ è uguale al coseno del loro angolo; si ha*:

$$\vec{p} \cdot \vec{q} = 1 \cdot 1 \cos \hat{\alpha} = \cos \hat{\alpha}$$

— *Il prodotto scalare di un vettore $\vec{u}$ per il proprio versore $\vec{p}$ è uguale al modulo del vettore; si ha*:

$$\vec{u} \cdot \vec{p} = u \cdot 1 \cos 0^\circ = u$$

— *Il prodotto scalare di due vettori $\vec{u}$ e $\vec{v}$, espressi in funzione delle coordinate cartesiane di essi, è dato dalla somma del prodotto delle prime coordinate con il prodotto delle seconde coordinate dei due vettori.*

Diamo dimostrazione dell'enunciato. Siano dati i due vettori:

$$\vec{u} = a\vec{i} + b\vec{j} \quad \text{e} \quad \vec{v} = c\vec{i} + d\vec{j}$$

Scrivendo il loro prodotto scalare, si ha:

$$\vec{u} \cdot \vec{v} = (a\vec{i} + b\vec{j}) \cdot (c\vec{i} + d\vec{j}) = ac\vec{i}^2 + bd\vec{j}^2 + (ad + bc)\vec{i} \cdot \vec{j}$$

Siccome si ha:

$$\vec{i}^2 = 1; \quad \vec{j}^2 = 1; \quad \vec{i} \cdot \vec{j} = 0 \quad (\text{essendo ortogonali tra loro});$$

risulta:

$$\vec{u} \cdot \vec{v} = ac + bd$$

ESEMPIO 2

Determinare a in modo che i due vettori:

$$\vec{a} = 3\vec{i} - 2\vec{j} \quad \text{e} \quad \vec{b} = a\vec{i} + \frac{1}{2}\vec{j}$$

siano perpendicolari.

Due vettori sono perpendicolari se, essendo non nulli, il loro prodotto scalare è uguale a zero. Pertanto, per determinare a , si deve porre la condizione:

$$(3\vec{i} - 2\vec{j}) \cdot \left(a\vec{i} + \frac{1}{2}\vec{j}\right) = 0$$

ossia:

$$3a - 2 \cdot \frac{1}{2} = 0$$

da cui si ricava:

$$a = \frac{1}{3}$$

d) Prodotto vettoriale di due vettori

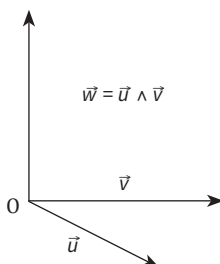
Dati due vettori $\vec{u}$ e $\vec{v}$ si dice *prodotto vettoriale* (o *esterno*) il vettore:

$$\vec{w} = \vec{u} \wedge \vec{v}$$

che ha:

- per modulo un numero che si ottiene moltiplicando il prodotto dei moduli dei due vettori per il valore assoluto del seno del loro angolo;
- per direzione la direzione della perpendicolare al piano determinato dai due vettori;
- per verso quello individuato dal dito medio della mano destra quando il pollice e l'indice della stessa mano, disposti in un piano perpendicolare al medio, indicano rispettivamente il primo e il secondo dei due vettori (questo modo di utilizzare la mano destra è detto *regola della mano destra*).

La rappresentazione grafica del prodotto vettoriale può essere ottenuta in un sistema cartesiano $Oxyz$, in cui i vettori $\vec{i}$, $\vec{j}$ e $\vec{k}$ sono i versori che hanno la stessa direzione e lo stesso verso, rispettivamente, dell'asse x , dell'asse y e dell'asse z .



Il prodotto vettoriale è nullo se:

- i vettori sono paralleli;
- è nullo uno dei due vettori.

Il prodotto vettoriale non gode della *proprietà commutativa*, essendo:

$$\vec{u} \wedge \vec{v} = -\vec{v} \wedge \vec{u}$$

Esso gode della *proprietà distributiva rispetto alla somma dei vettori*, infatti:

$$(\vec{u} + \vec{v}) \wedge \vec{w} = \vec{u} \wedge \vec{w} + \vec{v} \wedge \vec{w}$$

e) Prodotto di un vettore per uno scalare

Sia t un numero reale diverso da zero e $\vec{v}$ un vettore non nullo, si chiama *prodotto del numero reale t per il vettore $\vec{v}$* il vettore $t\vec{v}$ che ha:

- la stessa direzione di $\vec{v}$;
- il verso concorde o contrario a quello di $\vec{v}$, a seconda che il numero t sia positivo o negativo;
- il modulo uguale al prodotto del modulo di $\vec{v}$ per il valore assoluto di t .

Qualsiasi siano i numeri reali λ e α , nonché i vettori $\vec{u}$ e $\vec{v}$, si ha:

- $\lambda(\vec{u} + \vec{v}) = \lambda\vec{u} + \lambda\vec{v}$
- $(\lambda + \alpha)\vec{u} = \lambda\vec{u} + \alpha\vec{u}$
- $(\lambda\alpha)\vec{v} = \lambda(\alpha\vec{v})$
- $1\vec{u} = \vec{u}$

Il metodo delle coordinate per la descrizione di luoghi geometrici

1 La geometria analitica

La geometria analitica è un complesso di procedure che traducono problemi geometrici in problemi analitici, occupandosi della successiva interpretazione geometrica dei risultati ottenuti.

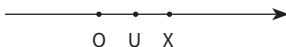
Essa si serve di insiemi numerici ed equazioni per definire ed analizzare costruzioni geometriche.

La geometria analitica studia e rappresenta le proprietà geometriche con il linguaggio dell'algebra, oppure le proprietà algebriche con il linguaggio della geometria (come le equazioni di primo grado con le rette e le equazioni di secondo grado con le coniche...).

Si tratta di due linguaggi astratti, attraverso i quali la geometria analitica costruisce un ponte che funziona in entrambi i versi, arricchendo sia l'algebra sia la geometria.

2 Le coordinate sulla retta

Si consideri una retta e su di essa si fissino un punto O (origine), un verso, che diremo positivo, ed un segmento $\overline{OU}$ come unità di misura.



Fissato un punto X sulla retta, chiamiamo ascissa di X quel numero reale che esprime la misura di OX rispetto a OU .

Tale ascissa sarà positiva se X segue O nel verso stabilito, negativa se X precede O .

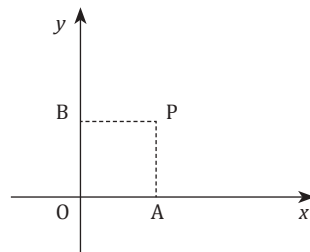
In questo modo, ad ogni punto della retta corrisponde un numero reale X . Viceversa, ad ogni numero reale X corrisponde un punto della retta.

Si dice che *l'insieme dei punti della retta e l'insieme dei numeri reali sono in corrispondenza biunivoca*.

3 Le coordinate cartesiane nel piano

Siano date una retta x orizzontale, sulla quale si sia fissata l'origine O e con una freccia il verso positivo, e una retta y , perpendicolare alla retta x e passante per l'origine O , sulla quale si sia fissato il verso positivo sempre con una freccia. Le due rette così individuate formano un *sistema di assi cartesiani* che dividono il piano in quattro parti dette *quadranti*. Gli assi coordinati si chiamano, rispettivamente, *asse delle ascisse*, che si indica con x , *asse delle ordinate*, che si indica con y . Il punto O è l'*origine degli assi*.

Fissata un'unità di misura per l'asse delle ascisse e una per l'asse delle ordinate (l'unità di misura può essere la stessa), e individuato un qualsiasi punto P del piano, si considerino le proiezioni A su Ox e B su Oy di tale punto.



Le misure dei due segmenti $\overline{OA}$ e $\overline{OB}$ si dicono rispettivamente *ascissa* e *ordinata* del punto P; entrambe prendono il nome di *coordinate* di tale punto.

In generale, ad ogni punto del piano corrisponde una coppia di numeri che sono l'ascissa e l'ordinata del punto.

Il segno delle coordinate di un punto può essere positivo o negativo a seconda del quadrante in cui si trova il suddetto punto; se il punto si trova nel:

- *primo quadrante*, ha ascissa e ordinata entrambe positive;
- *secondo quadrante*, ha ascissa negativa e ordinata positiva;
- *terzo quadrante*, ha ascissa e ordinata entrambe negative;
- *quarto quadrante*, ha ascissa positiva e ordinata negativa.

Inoltre, i punti situati sull'asse delle ascisse (x) hanno ordinata nulla; i punti situati sull'asse delle ordinate (y) hanno ascissa nulla.

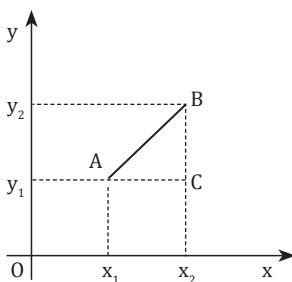
Una coppia di numeri individua un punto del piano, avente come ascissa il primo numero e come ordinata il secondo numero.

4 La distanza di due punti

La misura della distanza di due punti è data dalla radice quadrata della somma dei quadrati delle differenze delle ascisse dei due punti e delle ordinate dei due punti.

Siano dati i due punti:

$$A(x_1, y_1) \quad \text{e} \quad B(x_2, y_2)$$



La distanza $\overline{AB}$ è l'ipotenusa del triangolo rettangolo ABC , il quale ha per cateti $\overline{AC}$ e $\overline{BC}$; dunque, per il *teorema di Pitagora*, si ha:

$$\overline{AB} = \sqrt{\overline{AC}^2 + \overline{BC}^2}$$

da cui:

$$\overline{AB} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

ESEMPIO 1

Determinare la distanza del punto $A(3, 2)$ dal punto $B(1, 1)$.

La distanza del punto A dal punto B è data da:

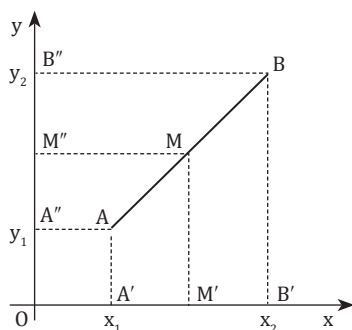
$$d = \sqrt{(3 - 1)^2 + (2 - 1)^2} = \sqrt{5}$$

5 Le coordinate del punto medio di un segmento

L'ascissa e l'ordinata del punto medio M di un segmento $\overline{AB}$ sono uguali, rispettivamente, alla semisomma delle ascisse ed alla semisomma delle ordinate degli estremi A e B del dato segmento.

Si vogliono determinare le coordinate del punto medio M di un segmento avente per estremi:

$$A(x_1, y_1) \quad \text{e} \quad B(x_2, y_2)$$



Le proiezioni M' ed M'' del punto medio sui due assi sono, rispettivamente, i punti medi dei segmenti $\overline{A'B'}$ e $\overline{A''B''}$.

Il punto medio di un segmento ha per ascissa la semisomma delle ascisse degli estremi del segmento e per ordinata la semisomma delle ordinate degli estremi del segmento; pertanto, esso ha coordinate:

$$\overline{OM'} = \frac{x_1 + x_2}{2} \quad \text{e} \quad \overline{OM''} = \frac{y_1 + y_2}{2}$$

ESEMPIO 2

Determinare le coordinate del punto medio del segmento avente per estremi i punti:

$$A(\sqrt{3}, 3) \quad \text{e} \quad B(-\sqrt{3}, 5)$$

Le coordinate del punto medio M del segmento $\overline{AB}$ considerato, si ricavano da:

$$\overline{OM'} = \frac{\sqrt{3} - \sqrt{3}}{2} = 0 \quad \text{e} \quad \overline{OM''} = \frac{3 + 5}{2} = 4$$

Pertanto, il punto medio del segmento $\overline{AB}$ è $M(0, 4)$.

ESEMPIO 3

Determinare la lunghezza della mediana del triangolo avente per vertici $A(1, 2)$, $B(2, 5)$ e $C(3, 0)$, relativa al vertice B .

La *mediana* di un triangolo è il segmento che congiunge un vertice col punto medio del lato opposto; per determinare la lunghezza della mediana, è necessario calcolare il punto medio del lato $\overline{AC}$. Esso è pari a:

$$M(2, 1)$$

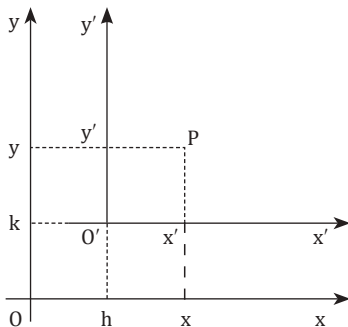
A questo punto bisogna calcolare la distanza tra il vertice $B(2, 5)$ e tale punto:

$$d = \sqrt{(2 - 2)^2 + (5 - 1)^2} = \sqrt{16} = 4$$

È questa, dunque, la lunghezza della mediana.

6 La traslazione d'assi

Si consideri un sistema di coordinate cartesiane Oxy ed un secondo sistema cartesiano $O'x'y'$, parallelo al primo. Siano (h, k) le coordinate di O' rispetto al sistema Oxy .



Sia P un punto di coordinate (x, y) rispetto al sistema Oxy . Chiamate (x', y') le coordinate di P rispetto al sistema $O'x'y'$, si ha:

$$\begin{cases} x' = x - h \\ y' = y - k \end{cases}$$

e in maniera equivalente:

$$\begin{cases} x = x' + h \\ y = y' + k \end{cases}$$

7 La rappresentazione grafica di funzioni

Dati due insiemi X e Y non vuoti, si dice *funzione* o *applicazione* una relazione tra gli elementi di X e gli elementi di Y che associa ad ogni elemento x di X uno e un solo elemento y di Y . Si chiama f l'applicazione di X in Y . In tal caso, per indicare che f è una funzione, si adopera la scrittura:

$$f: X \rightarrow Y$$

L'elemento y di Y che la funzione associa all'elemento x di X si indica con:

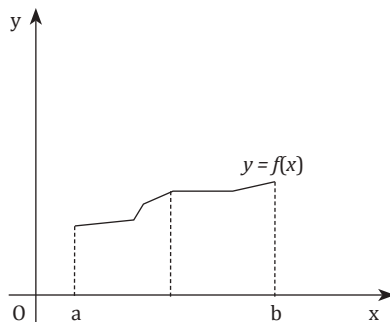
$$y = f(x)$$

in cui $y = f(x) \in Y$ è l'immagine di x in Y .

Il *diagramma* di una funzione $y = f(x)$ è l'insieme dei punti di un piano le cui coordinate soddisfano l'equazione $y = f(x)$.

Fissato un sistema di assi cartesiani ortogonali, si assumano per la x valori compresi in un intervallo $[a, b]$. Attribuendo alla variabile indipendente x il valore x_1 , la variabile dipendente assume il valore $y_1 = f(x_1)$; si individuano, in tal modo, le coordinate x_1 e y_1 di un punto P del piano.

Rappresentando altri punti del piano e congiungendoli, si ottiene un luogo geometrico denominato *diagramma* della funzione $y = f(x)$.

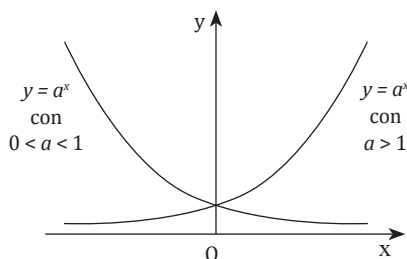


Si consideri la funzione:

$$y = a^x$$

È questa la *funzione esponenziale*, in cui a^x è una potenza di base positiva e diversa da 1; il suo diagramma si chiama *curva esponenziale* ed è sempre situata nel semipiano delle ordinate positive. L'andamento crescente o decrescente di tale curva dipende dal valore assunto da a :

- se $a > 1$, per valori di x negativi e crescenti in valore assoluto, la curva si avvicina indefinitamente all'asse delle ascisse; per valori di x positivi e crescenti in valore assoluto, la curva si allontana dall'asse delle ascisse e tende all'infinito positivo;
- se $0 < a < 1$, per valori di x negativi e crescenti in valore assoluto, la curva si allontana dall'asse delle ascisse; per valori di x positivi e crescenti in valore assoluto, la curva si avvicina indefinitamente all'asse delle ascisse.

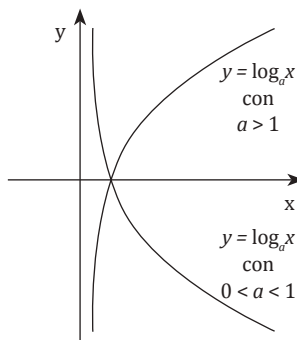


Si consideri, inoltre, la funzione:

$$y = \log_a x$$

È questa la *funzione logaritmica*, il cui diagramma si chiama *curva logaritmica*. Tale curva è sempre situata nel semipiano delle ascisse positive, ma il suo andamento crescente o decrescente dipende dal valore assunto da a :

- se $a > 1$, per valori di y negativi e decrescenti in valore assoluto, la curva si avvicina indefinitamente all'asse delle ordinate; per valori di y positivi e crescenti in valore assoluto, la curva si allontana dall'asse delle ordinate. Inoltre, per $x > 1$ la y è sempre maggiore di zero, mentre per $0 < x < 1$ la y è sempre minore di zero; da ciò si deduce che la curva interseca l'asse x nel punto di ascissa 1;
- se $0 < a < 1$, per valori di y negativi e crescenti in valore assoluto, la curva si allontana dall'asse delle ordinate; per valori di y positivi e crescenti in valore assoluto, la curva si avvicina indefinitamente all'asse delle ordinate. Inoltre, per $x > 1$ la y è sempre minore di zero, mentre per $0 < x < 1$ la y è sempre maggiore di zero.



La funzione logaritmica di base a appena definita è l'inversa della funzione esponenziale di base a .

8 L'equazione generale o implicita della retta

Sia data l'equazione di primo grado:

$$ax + by + c = 0$$

Essa si dice *equazione generale o implicita della retta*.

Risolvendo l'equazione data rispetto a y , si ottiene:

$$y = -\frac{a}{b}x - \frac{c}{b}$$

da cui si evince che l'equazione data ha per diagramma una retta.

A questo punto, è bene distinguere due casi:

— se il coefficiente a è nullo, l'equazione diventa:

$$y = -\frac{c}{b}$$

Il suo diagramma è una retta parallela all'asse delle ascisse, in quanto tutti i suoi punti hanno ordinata costante. Essa rappresenta il luogo geometrico dei punti del piano aventi dall'asse delle ascisse distanza costante $-\frac{c}{b}$;

— se il coefficiente b è nullo, l'equazione diventa:

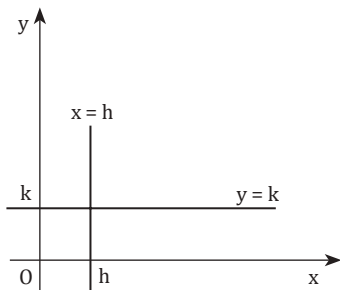
$$x = -\frac{c}{a}$$

il suo diagramma è una retta parallela all'asse delle ordinate, in quanto tutti i suoi punti hanno ascissa costante. Essa rappresenta il luogo geometrico dei punti del piano aventi dall'asse delle ordinate distanza costante $-\frac{c}{a}$.

Quindi, le equazioni del tipo:

$$x = h \quad \text{e} \quad y = k$$

rappresentano, rispettivamente, le equazioni di una retta parallela all'asse delle ordinate e di una retta parallela all'asse delle ascisse.



9 Le rette rispetto all'origine degli assi cartesiani

Si consideri il grafico della funzione:

$$y = mx \quad \text{con } m \text{ costante}$$

I punti di questa curva sono tali che è *costante* il rapporto tra l'ordinata e l'ascissa di ciascuno di essi. Tale rapporto costante è proprio uguale ad $\frac{y}{x} = m$.

Attribuendo alla x i valori $0, 1, 2, 3, \dots$, i corrispondenti valori della y sono $0, m, 2m, 3m, \dots$

Pertanto, congiungendo i punti con l'origine O degli assi, si ottiene il diagramma della funzione.

Il diagramma della funzione $y = mx$ rappresenta, dunque, una retta passante per l'origine, e il coefficiente m , che si chiama *coefficiente angolare*, rappresenta l'inclinazione della retta.

La funzione considerata è la funzione rappresentativa della *proporzionalità diretta*.

Il grafico della funzione:

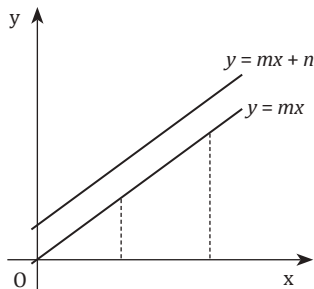
$$y = mx + n$$

è una retta, in cui m è il *coefficiente angolare* ed n è l'*intercetta*. Quest'ultima individua il punto in cui la retta incontra l'asse delle ordinate, quando la variabile x assume valore zero; infatti, si ha:

$$y = m(0) + n = n$$

Il punto in cui la retta incontra l'asse delle ascisse si individua ponendo $y = 0$:

$$x = -\frac{n}{m}$$



La funzione considerata si dice *funzione lineare* perché la variabile x figura al primo grado (in altre parole con l'esponente 1 sottinteso); da ciò si deduce che ogni equazione di primo grado nelle variabili x e y ha per grafico una retta.

10 L'equazione della retta passante per un punto assegnato o per due punti assegnati

Sia data l'*equazione esplicita* della retta:

$$y = mx + n$$

e un punto $P(x_1, y_1)$. La retta data passa per il punto P se e solo se le coordinate del punto P soddisfano l'equazione data, ossia deve essere:

$$y_1 = mx_1 + n$$

Sottraendo dall'equazione della retta questa identità, si ha che al variare di m l'equazione:

$$y - y_1 = m(x - x_1)$$

rappresenta il *fascio proprio di rette* passanti per il punto $P(x_1, y_1)$.

Siano $P(x_1, y_1)$ e $Q(x_2, y_2)$ due punti. Affinché una qualsiasi delle rette del fascio passante per il punto $P(x_1, y_1)$ passi anche per il punto $Q(x_2, y_2)$, è necessario che le coordinate del punto Q soddisfino l'equazione del fascio $y - y_1 = m(x - x_1)$, ossia deve essere:

$$y_2 - y_1 = m(x_2 - x_1)$$

Da cui si ottiene che il valore del coefficiente angolare è:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

L'equazione dell'unica retta passante anche per il punto Q si ottiene sostituendo il valore del coefficiente angolare appena ottenuto nell'equazione del fascio di rette passanti per P ; tale equazione è:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

ESEMPIO 4

Determinare il valore dell'intercetta della retta avente per coefficiente angolare $\frac{1}{2}$ e passante per il punto $P(5, 5)$.

L'equazione della retta passante per il punto $P(5, 5)$ e avente per coefficiente angolare $\frac{1}{2}$ è:

$$y - 5 = \frac{1}{2}(x - 5)$$

Da cui:

$$y = \frac{1}{2}x + \frac{5}{2}$$

L'intercetta è, pertanto, uguale a $\frac{5}{2}$.

11 Rette parallele e rette perpendicolari

Due rette aventi *coefficienti angolari uguali* sono *parallele*; esse hanno equazioni del tipo:

$$y = mx + n \quad \text{e} \quad y = mx + v$$

dove m è il coefficiente angolare comune ad entrambe.

Se le equazioni delle due rette sono date sotto forma implicita, si ha, rispettivamente:

$$ax + by + c = 0 \quad \text{e} \quad a'x + b'y + c' = 0$$

Per determinare la *condizione di parallelismo* delle stesse è necessario, dapprima, porre le equazioni sotto forma esplicita; per cui si ha, rispettivamente:

$$y = -\frac{a}{b}x - \frac{c}{b} \quad \text{e} \quad y = -\frac{a'}{b'}x - \frac{c'}{b'}$$

Pertanto, le due rette sono parallele se e solo se:

$$\frac{a}{b} = \frac{a'}{b'} \quad \text{purché} \quad \frac{c}{b} \neq \frac{c'}{b'}$$

Se anche $\frac{c}{b} = \frac{c'}{b'}$, le due rette coincidono.

Due rette aventi *coefficienti angolari fra loro reciproci e di segno opposto* sono *perpendicolari* (o *ortogonali*); esse hanno equazioni del tipo:

$$y = mx + n \quad \text{e} \quad y = -\frac{1}{m}x + v$$

dove il coefficiente della seconda è il reciproco e di segno opposto di quello della prima. In definitiva, la *condizione di perpendicolarità* di due rette è che il prodotto dei loro coefficienti angolari sia uguale a -1 .

Se le equazioni delle due rette sono date sotto forma implicita, si ha, rispettivamente:

$$ax + by + c = 0 \quad \text{e} \quad a'x + b'y + c' = 0$$

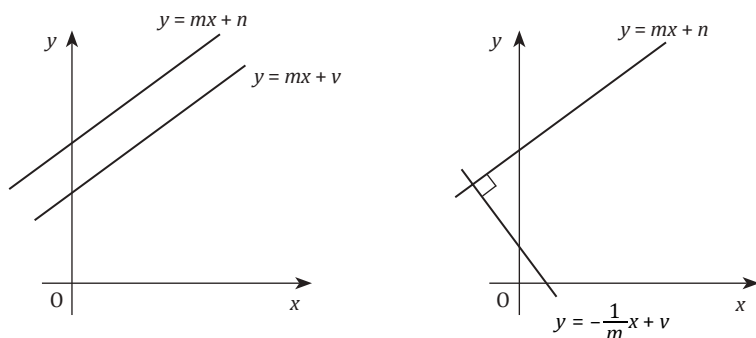
Per determinare la condizione di perpendicolarità delle stesse è necessario, dapprima, porre le equazioni sotto forma esplicita, per cui si ha:

$$y = -\frac{a}{b}x - \frac{c}{b} \quad \text{e} \quad y = -\frac{a'}{b'}x - \frac{c'}{b'}$$

Pertanto, le due rette sono perpendicolari se e solo se:

$$\frac{a}{b} \cdot \frac{a'}{b'} = -1$$

I grafici che mostrano la condizione di parallelismo e di perpendicolarità di due rette sono i seguenti:



ESEMPIO 5

Indicare l'equazione della retta parallela all'asse delle ordinate e passante per il punto dell'asse x , $P(\sqrt{3}, 0)$. La retta di equazione $x = \sqrt{3}$ è il luogo geometrico dei punti aventi tutti ascissa $\sqrt{3}$; essa è, quindi, parallela all'asse delle ordinate.

ESEMPIO 6

Determinare l'equazione della retta che passa per il punto $A(2, 7)$ e parallela alla retta di equazione:

$$y = 2x - 5$$

Sia r la retta cercata e s la retta ad essa parallela di equazione $y = 2x - 5$. Poiché il coefficiente angolare della retta s è 2, anche quello della retta r deve essere 2. La retta passante per $A(2, 7)$ si ottiene da:

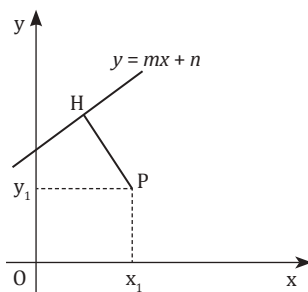
$$y - 7 = 2(x - 2)$$

Di conseguenza, l'equazione cercata è:

$$y = 2x + 3$$

12 La distanza di un punto da una retta

Sia dato un punto $P(x_1, y_1)$, una retta r avente equazione $y = mx + n$ ed un punto H di tale retta. Per trovare la lunghezza del segmento $\overline{PH}$ esprimente la distanza tra il punto P e la retta r , si consideri che la retta passante per P perpendicolare a r ha per coefficiente angolare $-\frac{1}{m}$.



Le coordinate di H si ottengono dal sistema:

$$\begin{cases} y = mx + n \\ y - y_1 = -\frac{1}{m}x + \frac{1}{m}x_1 \end{cases}$$

da cui:

$$H\left(\frac{x_1 + my_1 - mn}{1 + m^2}, \frac{mx_1 + m^2y_1 + n}{1 + m^2}\right)$$

Dalla formula della distanza tra due punti, la lunghezza del segmento $\overline{PH}$ è:

$$d = \sqrt{\left(x_1 - \frac{x_1 + my_1 - mn}{1 + m^2}\right)^2 + \left(y_1 - \frac{mx_1 + m^2y_1 + n}{1 + m^2}\right)^2}$$

Attraverso diversi passaggi si ottiene il modulo della distanza:

$$d = \frac{|mx_1 - y_1 + n|}{\sqrt{1 + m^2}}$$

Volendo calcolare la distanza dell'origine $O(0, 0)$ dalla retta, la formula appena ottenuta diventa:

$$d = \frac{|n|}{\sqrt{1 + m^2}}$$

ESEMPIO 7

Determinare la distanza del punto $P(-2, 3)$ dalla retta di equazione:

$$2x - y + 2 = 0$$

L'equazione della retta data, sotto forma esplicita, è:

$$y = 2x + 2$$

La distanza del punto $P(-2, 3)$ da tale retta è:

$$d = \frac{|2 \cdot (-2) - 3 + 2|}{\sqrt{1 + 2^2}} = \frac{5}{\sqrt{5}} = \sqrt{5}$$

13 Le coniche

Per *conica* si intende il luogo dei punti soddisfacenti, con le loro coordinate cartesiane, una equazione di secondo grado, la quale può contenere tre termini di secondo grado, due termini di primo grado e un termine noto. Essa è del tipo:

$$f(x, y) = ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0$$

Si chiamano coniche la *circonferenza*, l'*ellisse*, l'*iperbole* e la *parabola*.

La denominazione di coniche deriva dal fatto che esse si possono ottenere come intersezioni della superficie di un cono indefinito con un piano che non passa per il vertice del cono.

14 La circonferenza

La *circonferenza* è il luogo geometrico dei punti del piano aventi distanza fissa da un punto, detto *centro*.

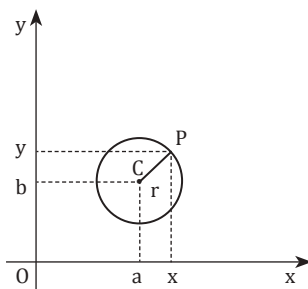
L'equazione della *circonferenza*, avente centro $C(a, b)$, raggio r e passante per un punto $P(x, y)$, si ottiene semplicemente considerando che il punto P appartiene alla circonferenza solo se $\overline{CP} = r$, ossia:

$$\sqrt{(x - a)^2 + (y - b)^2} = r$$

Quadrando si ha:

$$(x - a)^2 + (y - b)^2 = r^2$$

che è l'equazione della circonferenza.



Sviluppando, si ottiene un'equazione di secondo grado che manca del termine in xy , e in cui i coefficienti di x^2 e y^2 sono uguali:

$$x^2 + y^2 - 2ax - 2by + (a^2 + b^2 - r^2) = 0$$

Pertanto, si può affermare che, inversamente, un'equazione di secondo grado del tipo:

$$x^2 + y^2 + \alpha x + \beta y + \gamma = 0$$

rappresenta, in un sistema di assi ortogonali, una circonferenza, in cui $\alpha = -2a$, $\beta = -2b$, $\gamma = a^2 + b^2 - r^2$. Le coordinate del centro e il raggio sono, rispettivamente:

$$C\left(-\frac{\alpha}{2}, -\frac{\beta}{2}\right) \quad r = \frac{1}{2}\sqrt{\alpha^2 + \beta^2 - 4\gamma} \quad \text{con } \alpha^2 + \beta^2 - 4\gamma > 0$$

Se il centro della circonferenza coincide con l'origine O degli assi cartesiani, a e b si annullano, e l'equazione del cerchio diventa:

$$x^2 + y^2 = r^2$$

ESEMPIO 8

Scrivere l'equazione della circonferenza avente centro $C(2, 3)$ e raggio $r = 5$.

Applicando la formula data, l'equazione è la seguente:

$$(x - 2)^2 + (y - 3)^2 = 5^2$$

Quadrando e riducendo i termini simili, essa diventa:

$$x^2 + y^2 - 4x - 6y - 12 = 0$$

a) Mutua posizione di una circonferenza e di una retta

Per conoscere la posizione reciproca di una circonferenza e di una retta basta risolvere il sistema formato dalle equazioni della circonferenza e della retta:

$$\begin{cases} x^2 + y^2 + \alpha x + \beta y + \gamma = 0 \\ y = ax + b \end{cases}$$

L'equazione risolvente, che si ottiene eliminando la y , è un'equazione di secondo grado le cui soluzioni sono diverse a seconda del segno del discriminante; infatti, se:

- $\Delta > 0$ → essa ammette due soluzioni reali e distinte, per cui la retta è *secante*;
- $\Delta = 0$ → essa ammette due soluzioni reali e coincidenti, per cui la retta è *tangente*;
- $\Delta < 0$ → essa ammette due soluzioni complesse coniugate, per cui la retta è *esterna*.

ESEMPIO 9

Determinare le intersezioni della circonferenza avente equazione:

$$x^2 + y^2 - 4x + 8y + 2 = 0$$

con la retta avente equazione $y = 2x + 1$.

Per risolvere il problema si imposta il sistema:

$$\begin{cases} x^2 + y^2 - 4x + 8y + 2 = 0 \\ y = 2x + 1 \end{cases}$$

Applicando il metodo di sostituzione, si ottiene il sistema:

$$\begin{cases} x^2 + 4x^2 + 4x + 1 - 4x + 16x + 8 + 2 = 0 \\ y = 2x + 1 \end{cases}$$

Da cui, l'equazione che consente di risolvere il sistema è:

$$5x^2 + 16x + 11 = 0$$

Essa ha $\Delta > 0$, per cui la retta è secante; le sue radici sono:

$$x_1 = -\frac{11}{5} \quad \text{e} \quad x_2 = -1$$

Sostituendo nelle equazioni date, si ottengono i corrispondenti valori della y :

$$y_1 = -\frac{17}{5} \quad \text{e} \quad y_2 = -1$$

Pertanto, i punti intersezione della retta con la circonferenza sono:

$$A\left(-\frac{11}{5}, -\frac{17}{5}\right) \quad \text{e} \quad B(-1, -1)$$

b) Mutua posizione di due circonferenze

Per conoscere la posizione reciproca di due circonferenze basta risolvere il sistema formato dalle equazioni delle due circonferenze:

$$\begin{cases} x^2 + y^2 + \alpha x + \beta y + \gamma = 0 \\ x^2 + y^2 + \alpha' x + \beta' y + \gamma' = 0 \end{cases}$$

Siccome, sottraendo la seconda equazione dalla prima, si ha:

$$(\alpha - \alpha')x + (\beta - \beta')y + (\gamma - \gamma') = 0$$

e siccome ogni soluzione del sistema dato è anche soluzione dell'equazione di primo grado appena ottenuta, invece di risolvere il sistema di equazioni delle due circonferenze, ci si riconduce al sistema di equazioni del cerchio e della retta.

L'equazione di primo grado, ottenuta sottraendo membro a membro le equazioni delle due circonferenze, si chiama *asse radicale* e rappresenta l'equazione della retta che congiunge i due punti di intersezione delle due circonferenze; tale retta è perpendicolare alla congiungente i centri delle circonferenze.

c) Tangenti ad una circonferenza

Per determinare l'equazione delle tangenti condotte da un punto $P(x_1, y_1)$ ad una circonferenza, bisogna innanzitutto conoscere la posizione reciproca della circonferenza e del punto.

- Se il punto è interno alla circonferenza, ogni retta passante per P interseca la circonferenza in due punti; in tal caso non esistono rette tangenti alla circonferenza passanti per quel punto.
- Se il punto è sulla circonferenza, esiste solo una retta passante per P che è tangente alla circonferenza; essa è perpendicolare al raggio passante per P . L'equazione di tale tangente si determina risolvendo il sistema formato dall'equazione della circonferenza e da quella di una retta generica passante per P ; supponendo che la circonferenza abbia centro nell'origine, si ha:

$$\begin{cases} x^2 + y^2 = r^2 \\ y - y_1 = m(x - x_1) \end{cases}$$

Affinché tale retta sia tangente alla circonferenza è necessario che le soluzioni del sistema siano reali e coincidenti, ovvero che l'equazione di secondo grado, ottenuta eliminando la y dal sistema, abbia discriminante nullo. Ponendo tale discriminante uguale a zero, si ottiene una equazione in m ; sostituendo il valore del coefficiente angolare così ottenuto nell'equazione della retta passante per P , e considerando che $x_1^2 + y_1^2 = r^2$, si ottiene l'equazione della tangente alla circonferenza:

$$x_1 x + y_1 y = r^2$$

- Se il punto è esterno alla circonferenza, si deve risolvere il sistema formato dall'equazione della circonferenza e dall'equazione del fascio di rette passante per P ; le radici dell'equazione di secondo grado, che si ottiene eliminando la y dal sistema, devono essere reali e coincidenti, ossia m deve essere tale da rendere nullo il discriminante di tale equazione. Eguagliando a zero tale discriminante si ottiene un'equazione di secondo grado in m , le cui radici m_1 e m_2 sono i coefficienti angolari delle due tangenti condotte dal punto P esterno al cerchio. In definitiva, le loro equazioni sono:

$$y - y_1 = m_1(x - x_1) \quad \text{e} \quad y - y_1 = m_2(x - x_1)$$

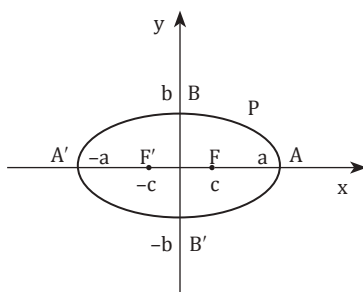
15 L'ellisse

L'ellisse è il luogo geometrico dei punti del piano tali che la somma delle loro distanze dai fuochi, in valore assoluto, è costante ed è uguale all'asse principale.

L'ellisse è una curva chiusa simmetrica rispetto al suo asse principale.

I punti:

- A e A' sono i *vertici principali*;
- F e F' sono i *fuochi*;
- B e B' sono equidistanti dai due fuochi e sono anch'essi dei *vertici*. Il segmento BB' che li unisce è detto *asse minore* dell'ellisse; inoltre, biseca l'*asse principale* $A'A$ nel punto O ed è a esso perpendicolare.



In un sistema di riferimento cartesiano ortogonale, facendo coincidere l'asse principale con l'asse x , l'asse minore con l'asse y , e il centro dell'ellisse con l'origine O , si pone:

$$\overline{A'A} = 2a; \quad \overline{B'B} = 2b; \quad \overline{F'F} = 2c$$

Pertanto, i quattro vertici e i fuochi hanno, rispettivamente, coordinate:

$$A'(-a, 0); A(a, 0)$$

$$B'(0, -b); B(0, b)$$

$$F'(-c, 0); F(c, 0)$$

Sia $P(x, y)$ un punto dell'ellisse. Indicando con i valori assoluti le distanze e in base alla definizione dell'ellisse, si ha:

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a$$

Applicando le regole sui radicali e semplificando, dopo diversi passaggi, si ottiene:

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2)$$

Si dimostra facilmente che $a^2 - c^2 = b^2$; per cui dividendo ambo i membri per a^2b^2 si ottiene l'equazione in *forma canonica* dell'ellisse riferita ai suoi assi:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

in cui a è la lunghezza del semiasse maggiore e b è la lunghezza del semiasse minore.

Pertanto, si può affermare che, inversamente, un'equazione di secondo grado del tipo:

$$\alpha x^2 + \beta y^2 = \gamma \quad (\text{con } \alpha, \beta \text{ e } \gamma \text{ dello stesso segno})$$

rappresenta l'equazione di un'ellisse.

a) Eccentricità dell'ellisse

Essendo $a^2 - c^2 = b^2$, la *semidistanza focale* c è data da:

$$c = \sqrt{a^2 - b^2}$$

Il rapporto tra la semidistanza focale e il semiasse maggiore si può assumere come la misura di quanto l'ellisse, per la sua forma più o meno allungata, differisce dalla circonferenza; tale rapporto si dice *eccentricità dell'ellisse*:

$$e = \frac{c}{a} = \frac{\sqrt{a^2 - b^2}}{a}$$

Da esso si deduce che l'ellisse è tanto più allungata quanto più a è maggiore di b .

Se i fuochi coincidono con i vertici principali si dice che l'ellisse degenera nel diametro maggiore.

Se i fuochi coincidono con il centro della curva, ossia se c è uguale a zero, si è in presenza della circonferenza; in tal caso, l'eccentricità è nulla.

ESEMPIO 10

Determinare semiassi e semidistanza focale dell'ellisse avente equazione:

$$x^2 + 4y^2 = 36$$

L'equazione data, ridotta in *forma canonica*, è:

$$\frac{x^2}{36} + \frac{y^2}{9} = 1$$

Dalla stessa si evince che $a^2 = 36$ e $b^2 = 9$, per cui i semiassi sono $a = 6$ e $b = 3$.

La semidistanza focale è $c = \sqrt{a^2 - b^2} = \sqrt{36 - 9} = \sqrt{27} = 3\sqrt{3}$.

b) Tangenti ad un'ellisse

Per determinare l'equazione della tangente ad un'ellisse in un suo punto $P(x_1, y_1)$ si risolve il sistema tra l'equazione dell'ellisse e quella di una retta generica passante per P .

Si procede, quindi, allo stesso modo visto per la circonferenza, ottenendo il coefficiente angolare m della retta tangente. Sostituendo il valore del coefficiente angolare nell'equazione della retta passante per P , e considerando che $b^2x_1^2 + a^2y_1^2 = a^2b^2$, si ottiene l'equazione della retta tangente all'ellisse nel suo punto $P(x_1, y_1)$:

$$\frac{x_1x}{a^2} + \frac{y_1y}{b^2} = 1$$

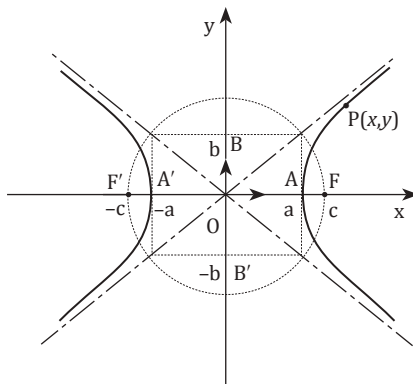
16 L'iperbole

L'iperbole è il luogo geometrico dei punti del piano le cui distanze da due punti fissi, detti fuochi, hanno una differenza costante in valore assoluto.

I punti:

- A e A' sono i *vertici principali* o *reali*; l'iperbole interseca l'asse delle ascisse in questi due punti aventi coordinate $A(a, 0)$ e $A'(-a, 0)$. Inoltre, il segmento $|AA'| = 2a$ si dice *asse trasverso*;
- F e F' sono i *fuochi*. La distanza focale è $FF' = 2c$;

- B e B' sono i *vertici secondari o immaginari*, aventi coordinate $B(0, b)$ e $B'(0, -b)$. Il segmento $|B'B| = 2b$ che li unisce si dice *asse non traverso*.



In un sistema di riferimento cartesiano ortogonale, si fa coincidere l'asse x con la congiungente i due fuochi F e F' , e l'asse y con la perpendicolare a $F'F$ nel punto medio O .

Sia $P(x, y)$ un punto dell'iperbole; indicando con i valori assoluti le distanze, in base alla definizione dell'iperbole, si ha:

$$\left| \sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} \right| = 2a$$

Applicando le regole sui radicali e semplificando, dopo diversi passaggi si ottiene:

$$(c^2 - a^2)x^2 + a^2y^2 = a^2(c^2 - a^2)$$

Si dimostra facilmente che $c^2 - a^2 = b^2$, per cui, dopo aver diviso per a^2b^2 , l'equazione in *forma canonica* dell'iperbole è:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

Pertanto, si può affermare che, inversamente, un'equazione di secondo grado del tipo:

$$\alpha x^2 - \beta y^2 = \gamma$$

rappresenta l'equazione di un'iperbole.

I punti reali dell'iperbole si determinano conducendo per i vertici A e A' due rette parallele all'asse y : la curva assume valori fuori della striscia di piano delimitata dalle due parallele, ossia la curva si compone di due rami distinti. A tale risultato si giunge risolvendo l'equazione dell'iperbole rispetto a y :

$$y = \pm \frac{b}{a} \sqrt{x^2 - a^2}$$

Si vede che essa deve assumere valori per cui $|x| \geq a$.

Ciascun ramo dell'iperbole è simmetrico rispetto all'asse x e si scosta indefinitamente da esso, a partire dai vertici. Infatti:

- attribuendo a x valori da a a ∞ , la y cresce da 0 a ∞ ed assume due valori opposti per ogni valore di x ;
- attribuendo a x valori da $-a$ a $-\infty$, la y cresce da 0 a ∞ ed assume due valori opposti per ogni valore di x .

Ciascun ramo dell'iperbole è simmetrico anche rispetto all'asse y e si scosta indefinitamente da esso.

Infine, l'iperbole è simmetrica rispetto al punto O , che è il centro della curva.

La curva presenta, inoltre, due *asintoti*, ossia esistono due rette a cui la curva si avvicina indefinitamente senza mai raggiungerle; esse hanno, rispettivamente, equazioni:

$$y = \frac{b}{a}x \quad \text{e} \quad y = -\frac{b}{a}x$$

Come per l'ellisse, l'eccentricità è data dal rapporto tra la semidistanza focale e il semiasse; tale rapporto è maggiore di 1. Quanto maggiore è c rispetto ad a tanto più i rami dell'iperbole sono aperti.

ESEMPIO 11

Determinare l'equazione dell'iperbole, riferita ai propri assi, avente per vertici i punti $A(8, 0)$ e $A'(-8, 0)$, e per fuochi $F(10, 0)$ e $F'(-10, 0)$.

L'equazione dell'iperbole riferita ai propri assi è:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

Avendo a e c è possibile ricavare b^2 ; dunque, si ha:

$$b^2 = c^2 - a^2 = 10^2 - 8^2 = 100 - 64 = 36$$

Pertanto, l'equazione dell'iperbole riferita ai propri assi è:

$$\frac{x^2}{64} - \frac{y^2}{36} = 1$$

ossia:

$$9x^2 - 16y^2 = 576$$

ESEMPIO 12

Determinare i vertici dell'iperbole avente equazione:

$$\frac{x^2}{9} - \frac{y^2}{4} = 1$$

I vertici sono i due punti in cui l'iperbole incontra l'asse delle ascisse; per $y = 0$, si ha $x = \pm 3$ e tali punti sono indicati con $A'(-3, 0)$ e $A(3, 0)$.

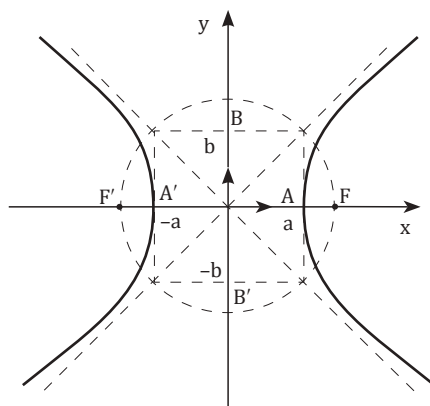
a) Iperbole equilatera

L'iperbole si dice *equilatera* se i suoi assi sono uguali, ossia se $a = b$; la sua equazione, riferita ai propri assi di simmetria, presi come assi coordinati, diventa pertanto:

$$x^2 - y^2 = a^2$$

La semidistanza focale è $c = a\sqrt{2}$, per cui i fuochi hanno coordinate:

$$F(a\sqrt{2}, 0) \quad \text{e} \quad F'(-a\sqrt{2}, 0)$$



Gli asintoti hanno, in tal caso, equazioni:

$$y = x \quad \text{e} \quad y = -x$$

cioè sono le bisettrici degli assi e, quindi, due rette fra loro perpendicolari.

Assumendo come assi coordinati tali asintoti, è facile verificare che la sua equazione diventa $xy = \frac{a^2}{2}$. Risolvendo rispetto a y , si ha:

$$y = \frac{a^2}{2x}$$

Pertanto, si può affermare che, inversamente, un'equazione del tipo:

$$xy = k$$

ossia:

$$y = \frac{k}{x}$$

rappresenta l'equazione di un'iperbole equilatera riferita ai suoi asintoti.

Per tracciare tale diagramma, bisogna considerare che alla x non si può attribuire valore zero; nemmeno la y può essere uguale a zero, per cui la funzione data non interseca gli assi cartesiani.

Da ciò consegue che essa è costituita da due *rami* situati in due quadranti opposti, infatti:

- se $k > 0$, attribuendo a x valori positivi (negativi) anche la y assume valori positivi (negativi); il diagramma sarà contenuto in parte nel primo quadrante degli assi e in parte nel terzo quadrante. Quanto più alto è il valore della x tanto più è basso il valore della y (la x e la y sono inversamente proporzionali); quest'ultima variabile non sarà mai uguale a zero, ma si dice che *tende asintoticamente all'asse positivo delle ascisse*;
- se $k < 0$, attribuendo a x valori positivi (negativi) la y assume valori negativi (positivi); il diagramma sarà contenuto in parte nel secondo quadrante e in parte nel quarto quadrante.

b) Tangenti ad un'iperbole

Le equazioni delle tangenti ad un'iperbole in un punto $P(x_1, y_1)$ sono diverse a seconda dell'iperbole, infatti:

- l'equazione della tangente all'iperbole $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ nel punto $P(x_1, y_1)$ è:

$$\frac{x_1 x}{a^2} - \frac{y_1 y}{b^2} = 1$$

- l'equazione della tangente all'iperbole $x^2 - y^2 = a^2$ nel punto $P(x_1, y_1)$ è:

$$x_1 x - y_1 y = a^2$$

- l'equazione della tangente all'iperbole $xy = k$ nel punto $P(x_1, y_1)$ è:

$$y_1 x + x_1 y = 2k$$

ESEMPIO 13

Determinare l'equazione della tangente all'iperbole equilatera $xy = -2$, nel punto $P(-1, 2)$.

Applicando la formula dell'equazione della tangente all'iperbole equilatera si ottiene:

$$2x - y = -4$$

ossia:

$$y = 2x + 4$$

Allo stesso risultato si sarebbe giunti impostando il sistema tra l'equazione dell'iperbole e quella di una retta generica passante per P :

$$\begin{cases} xy = -2 \\ y - 2 = m(x + 1) \end{cases}$$

Sostituendo nella prima equazione, il valore della y ottenuto dalla seconda equazione, semplificando e isolando i termini simili, si sarebbe ottenuta l'equazione di secondo grado in x :

$$mx^2 + (m + 2)x + 2 = 0$$

che, per quanto esposto, deve avere soluzioni reali e coincidenti o, il che è lo stesso, m deve assumere un valore tale che il discriminante sia nullo:

$$\Delta = (m + 2)^2 - 4 \cdot m \cdot 2 = m^2 - 4m + 4 = 0$$

Si tratta di un'equazione di secondo grado in m , avente per radice $m = 2$; sostituendo il valore del coefficiente angolare appena ottenuto nell'equazione della generica retta passante per P , si sarebbe ottenuta la tangente dell'iperbole equilatera nel punto $P(-1, 2)$:

$$y = 2x + 4$$

17 La parabola

La parabola è il luogo geometrico dei punti del piano equidistanti da un punto fisso, detto fuoco, e da una retta fissa non passante per il fuoco, detta direttrice.

La parabola è una curva aperta costituita da un solo ramo, con un solo fuoco e un solo asse di simmetria, che è l'*asse della parabola*; quest'ultimo passa per il fuoco ed è perpendicolare alla direttrice. Il punto di incontro tra la parabola e il suo asse di simmetria è detto *vertice* della parabola.

Siano F il fuoco, d la direttrice, p (*parametro della parabola*) la misura della distanza $\overline{DF}$ tra F e d . Ricaviamo ora le equazioni delle parabole simmetriche, rispettivamente, all'asse delle y e all'asse delle x .

a) Equazione della parabola simmetrica rispetto all'asse delle y

In questo contesto, si assume:

- come asse delle ascisse la tangente nel vertice della parabola;
- come asse delle ordinate l'asse della parabola.

Come si può verificare dalla figura, l'asse delle x è, altresì, perpendicolare al segmento DF nel suo punto di mezzo (che corrisponde all'origine O degli assi cartesiani).

Il fuoco ha per coordinate $F\left(0, \frac{p}{2}\right)$, mentre la direttrice ha equazione $y = -\frac{p}{2}$.

Si consideri ora un generico punto $P(x, y)$ della parabola e lo si sottoponga alla condizione di essere equidistante da F e da d . Considerando che $H\left(x, -\frac{p}{2}\right)$ è il piede della perpendicolare calata dal punto P sulla direttrice, si ha:

$$\sqrt{x^2 + \left(y - \frac{p}{2}\right)^2} = \sqrt{\left(y + \frac{p}{2}\right)^2}$$

da cui, quadrando, sviluppando, semplificando e risolvendo rispetto a y , si ha:

$$y = \frac{1}{2p}x^2$$

Ponendo $\frac{1}{2p} = a$, si ha:

$$y = ax^2$$

che è l'equazione canonica della parabola nel riferimento scelto.

La parabola *volge la concavità verso l'alto* se a è positivo; *volge la concavità verso il basso* se a è negativo.

b) Equazione della parabola simmetrica rispetto all'asse delle x

In questo contesto, si assume:

- come asse delle ascisse l'asse della parabola;
- come asse delle ordinate la perpendicolare all'asse delle ascisse nel vertice della parabola (cioè la tangente nel suo vertice).

Il fuoco ha per coordinate $F\left(\frac{p}{2}, 0\right)$, mentre la direttrice ha equazione $x = -\frac{p}{2}$.

Sia $P(x, y)$ un generico punto della parabola e $H\left(-\frac{p}{2}, y\right)$ il piede della perpendicolare calata dal punto P sulla direttrice; si ha:

$$\sqrt{\left(x - \frac{p}{2}\right)^2 + y^2} = \sqrt{\left(x + \frac{p}{2}\right)^2}$$

da cui, quadrando, sviluppando e semplificando, si ha:

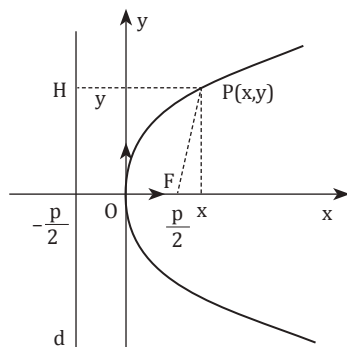
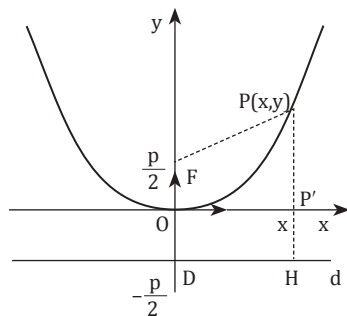
$$y^2 = 2px$$

che è l'equazione canonica della parabola nel riferimento scelto.

Da tale equazione si ha:

$$y = \pm\sqrt{2px}$$

da cui si deduce che, effettivamente, la parabola è costituita da un solo ramo, in quanto essa è situata nel semipiano delle ascisse positive per $p > 0$; è, invece, situata nel semipiano delle ascisse negative per $p < 0$.



ESEMPIO 14

Determinare l'equazione della parabola avente per fuoco $F(3, 0)$ e per vertice $V(2, 0)$.

Dai dati del problema si evince che la parabola è simmetrica rispetto all'asse x ; inoltre, essendo il vertice della parabola equidistante dal fuoco e dalla direttrice, e siccome $\overline{VF} = \sqrt{(2-3)^2} = 1$, la direttrice non può che avere equazione $x = 1$. Pertanto, dato un generico punto $P(x, y)$ della curva, per definizione di parabola, deve essere:

$$\sqrt{(x-3)^2 + y^2} = \sqrt{(x-1)^2}$$

da cui, quadrando e semplificando, si ottiene l'equazione cercata:

$$y^2 = 4x - 8$$

c) Parabola e funzione di secondo grado

Dimostriamo ora che, rispetto ad un sistema di assi cartesiani ortogonali Oxy , l'equazione:

$$y = ax^2 + bx + c$$

con a, b, c costanti arbitrarie, la prima delle quali diversa da zero, rappresenta ancora una parabola.

L'equazione equivale alla seguente:

$$y + \frac{b^2}{4a} - c = ax^2 + bx + \frac{b^2}{4a}$$

che a sua volta, posto $\Delta = b^2 - 4ac$, equivale alla:

$$y + \frac{\Delta}{4a} = a \left(x + \frac{b}{2a} \right)^2$$

Ponendo:

$$x' = x + \frac{b}{2a} \quad \text{e} \quad y' = y + \frac{\Delta}{4a}$$

si ha una traslazione degli assi e l'origine del nuovo sistema ha per coordinate:

$$O' \left(-\frac{b}{2a}; -\frac{\Delta}{4a} \right)$$

L'equazione data diventa:

$$y' = ax'^2$$

È noto che la curva rappresentativa di questa equazione è una parabola. Tale parabola ha l'asse di simmetria parallelo all'asse delle y , e vertice, fuoco ed equazione della direttrice, rispettivamente, uguali a:

$$V \left(-\frac{b}{2a}; -\frac{\Delta}{4a} \right) \quad F \left(-\frac{b}{2a}; \frac{1-\Delta}{4a} \right) \quad y = -\frac{1+\Delta}{4a} \quad (\text{con } \Delta = b^2 - 4ac)$$

Ovviamente, anche se l'equazione di secondo grado $y = ax^2 + bx + c$ è incompleta, essa rappresenta pur sempre una parabola.

d) Mutua posizione di una retta e di una parabola

Sia data una parabola di equazione:

$$y = ax^2 + bx + c$$

e una retta di equazione:

$$y = mx + n$$

Per determinare la posizione reciproca della retta e della parabola si risolve il sistema:

$$\begin{cases} y = mx + n \\ y = ax^2 + bx + c \end{cases}$$

Sostituendo nella seconda equazione il valore della y rilevato dalla prima, si ottiene l'equazione di secondo grado:

$$ax^2 + (b - m)x + c - n = 0$$

La posizione reciproca della retta e della parabola date dipende dal valore assunto dal discriminante di questa equazione; precisamente, se:

- $\Delta > 0 \rightarrow$ le radici dell'equazione data sono reali e distinte, per cui la retta è *secante* alla parabola;
- $\Delta = 0 \rightarrow$ le radici dell'equazione data sono reali e coincidenti, per cui la retta è *tangente* alla parabola;
- $\Delta < 0 \rightarrow$ l'equazione non ammette radici, per cui la retta è *esterna* rispetto alla parabola.

ESEMPIO 15

Determinare la posizione reciproca della parabola di equazione:

$$y = \frac{1}{2}x^2 - 8$$

e della retta di equazione $y = 2x - 10$.

Per determinare la posizione reciproca della retta e della parabola, si risolve il sistema:

$$\begin{cases} y = 2x - 10 \\ y = \frac{1}{2}x^2 - 8 \end{cases}$$

da cui, sostituendo ad y , nella seconda equazione, l'espressione $2x - 10$ ricavata dalla prima, si ottiene (dopo aver ridotto i termini simili ed eliminato il denominatore) l'equazione di secondo grado:

$$x^2 - 4x + 4 = 0$$

Siccome $\Delta = 0$, l'equazione ammette le due radici reali e coincidenti:

$$x_1 = x_2 = 2$$

I corrispondenti valori della y sono:

$$y_1 = y_2 = -6$$

Pertanto, la retta è tangente alla parabola nel punto $A(2, -6)$.

e) Tangenti ad una parabola

Le equazioni delle tangenti ad una parabola in un punto $P(x_1, y_1)$ sono diverse a seconda della parabola, infatti:

- l'equazione della tangente alla parabola $y = ax^2$ nel punto $P(x_1, y_1)$ è:

$$y + y_1 = 2ax_1x$$

- l'equazione della tangente alla parabola $y^2 = 2px$ nel punto $P(x_1, y_1)$ è:

$$y_1 y = p(x + x_1)$$

- l'equazione della tangente alla parabola $y = ax^2 + bx + c$ nel punto $P(x_1, y_1)$ è:

$$y + y_1 = 2ax_1x + b(x + x_1) + 2c$$

18 L'equazione parametrica e cartesiana di un piano

Le equazioni parametriche del piano passante per il punto $P(x_0, y_0, z_0)$ e parallelo ai vettori $u = (u_1, u_2, u_3)$ e $v = (v_1, v_2, v_3)$ sono date da:

$$\begin{cases} x = x_0 + u_1 \cdot s + v_1 \cdot t \\ y = y_0 + u_2 \cdot s + v_2 \cdot t \\ z = z_0 + u_3 \cdot s + v_3 \cdot t \end{cases}$$

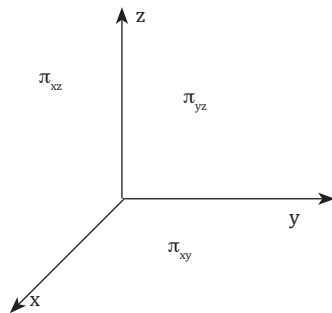
dove s e t sono i parametri. Ad ogni coppia di valori assegnata ad s e t corrisponde un punto del piano (analogamente alle equazioni parametriche di una retta nel piano). Esistono *infinite rappresentazioni parametriche* di un piano. Per passare dalla rappresentazione parametrica di un piano alla sua (unica) *rappresentazione cartesiana*, si ricava un parametro da un'equazione (es. s dalla prima equazione) che si sostituisce nelle altre due. Si ricava l'altro parametro da una delle due equazioni (in cui si è sostituito s) e si sostituisce nella terza, ottenendo un'equazione del tipo:

$$ax + by + cz + d = 0$$

che è l'equazione cartesiana di un piano nello spazio.

• Particolari piani nello spazio

- 1) Il piano π_{xy} ha equazione $z = 0$.
- 2) Il piano π_{yz} ha equazione $x = 0$.
- 3) Il piano π_{xz} ha equazione $y = 0$.
- 4) Un piano parallelo al piano π_{xy} ha equazione del tipo $z = k$.
- 5) Un piano parallelo al piano π_{yz} ha equazione del tipo $x = k$.
- 6) Un piano parallelo al piano π_{xz} ha equazione del tipo $y = k$.
- 7) Un piano del tipo $ax + by + d = 0$ è un piano ortogonale al piano di equazione $z = 0$ (piano π_{xy}).
- 8) Un piano del tipo $by + cz + d = 0$ è un piano ortogonale al piano di equazione $x = 0$ (piano π_{yz}).
- 9) Un piano del tipo $ax + cz + d = 0$ è un piano ortogonale al piano di equazione $y = 0$ (piano π_{xz}).



19 L'equazione cartesiana di un piano passante per tre punti (non allineati)

Un primo modo di procedere (ma poco agevole) è quello di seguito descritto. L'equazione cartesiana di un piano è $ax + by + cz + d = 0$, la quale può anche essere scritta (se $c \neq 0$) nel seguente modo:

$$\frac{a}{c}x + \frac{b}{c}y + z + \frac{d}{c} = 0 \quad \Leftrightarrow \quad z = -\frac{a}{c}x - \frac{b}{c}y - \frac{d}{c}$$

e cioè:

$$z = mx + ny + q$$

dove:

$$m = -\frac{a}{c}; \quad n = -\frac{b}{c}; \quad q = -\frac{d}{c}$$

L'equazione del piano è univocamente determinata allorché sono determinati i coefficienti m , n e q . Imponendo il passaggio del piano per i punti:

$$P_1(x_1, y_1, z_1) \quad P_2(x_2, y_2, z_2) \quad P_3(x_3, y_3, z_3)$$

si ottiene il sistema di tre equazioni nelle tre incognite m , n e q :

$$\begin{cases} z_1 = mx_1 + ny_1 + q \\ z_2 = mx_2 + ny_2 + q \\ z_3 = mx_3 + ny_3 + q \end{cases}$$

Ricavati m , n e q (il sistema ammetterà un'unica terna di soluzioni m , n , q se i tre punti non sono allineati), si sostituiscono tali valori nell'equazione $z = mx + ny + q$ ottenendo l'equazione ricercata.

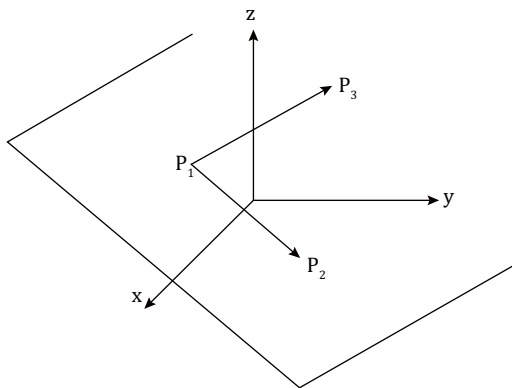
Un modo più elegante (apparentemente più complesso, ma in realtà con calcoli più agevoli) è quello descritto a seguire.

Siano $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$ e $P_3(x_3, y_3, z_3)$ i tre punti non allineati. Si consideri (senza mancare di generalità) il primo punto P_1 e si considerino i vettori:

$$P_2 - P_1 = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$$

$$P_3 - P_1 = (x_3 - x_1, y_3 - y_1, z_3 - z_1)$$

L'equazione del piano p ricercato passa per il punto P_1 ed è parallelo ai vettori $P_2 - P_1$ e $P_3 - P_1$. Quindi segue il procedimento visto precedentemente.



20 Tre vettori complanari

Assegnati tre vettori $s = (s_1, s_2, s_3)$, $u = (u_1, u_2, u_3)$ e $v = (v_1, v_2, v_3)$, condizione necessaria e sufficiente affinché siano *complanari* è che il *determinante* della matrice:

$$\begin{pmatrix} s_1 & s_2 & s_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix} \quad \text{sia uguale a zero}$$

21 Quattro punti complanari

Assegnati quattro punti $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, $P_3(x_3, y_3, z_3)$ e $P_4(x_4, y_4, z_4)$, si fissi arbitrariamente un punto (es. il primo) e si considerino i vettori:

$$P_2 - P_1; P_3 - P_1; P_4 - P_1$$

Condizione necessaria e sufficiente affinché i *quattro punti siano complanari* è che i tre vettori appena scritti siano complanari (caso precedente).

22 Piani e vettori paralleli

Assegnati due piani:

$$\pi_1 \quad a_1x + b_1y + c_1z + d_1 = 0$$

$$\pi_2 \quad a_2x + b_2y + c_2z + d_2 = 0$$

condizione necessaria e sufficiente affinché i *due piani siano paralleli* è che la matrice:

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}$$

abbia rango 1, oppure, equivalentemente, che esista uno scalare λ tale che:

$$a_1 = \lambda \cdot a_2$$

$$b_1 = \lambda \cdot b_2$$

$$c_1 = \lambda \cdot c_2$$

Assegnato un piano $ax + by + cz + d = 0$ ed un vettore $u = (u_1, u_2, u_3)$, la condizione affinché il vettore sia parallelo al piano è che risulti:

$$a \cdot u_1 + b \cdot u_2 + c \cdot u_3 = 0$$

23 Le rette nello spazio euclideo: equazioni parametriche di una retta

Assegnato un punto $P(x_1, y_1, z_1)$ ed un vettore $u = (u_1, u_2, u_3)$, esiste un'unica retta passante per P ed avente la stessa direzione del vettore u (che prende il nome di *vettore direzionale*). Una *rappresentazione parametrica di tale retta* è:

$$\begin{cases} x = x_1 + u_1 \cdot t \\ y = y_1 + u_2 \cdot t \\ z = z_1 + u_3 \cdot t \end{cases}$$

Se si ricava t da una delle tre equazioni e si sostituisce nelle altre due, si ottiene un'equazione *cartesiana* della stessa retta che si presenta come intersezione di due piani.

24 La direzione di una retta espressa in forma cartesiana

Sia assegnata una retta nello spazio mediante una sua *rappresentazione cartesiana* del tipo:

$$\begin{cases} a_1x + b_1y + c_1z + d_1 = 0 \\ a_2x + b_2y + c_2z + d_2 = 0 \end{cases}$$

La *direzione* di tale retta è fornita dal vettore $u = (l, m, n)$, le cui componenti sono i determinanti dei minori del secondo ordine estratti dalla matrice:

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}$$

cioè:

$$l = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} \quad m = - \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} \quad n = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

l, m ed n sono anche detti *parametri direttori* della retta.

25 I fasci di rette nel piano euclideo

Si è visto che l'equazione di una retta nel piano è $ax + by + c = 0$, dove a e b (coefficienti delle incognite) prendono il nome di numeri direttori della retta e c di termine noto.

Assegnate due rette, $r: a_1x + b_1y + c_1 = 0$ ed $s: a_2x + b_2y + c_2 = 0$, si ottiene che esse sono parallele se è nullo il determinante della matrice dei numeri direttori:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = 0 \quad \text{condizione di parallelismo}$$

Sviluppando il determinante si ha: $a_1 \cdot b_2 - a_2 \cdot b_1 = 0 \Leftrightarrow a_1 \cdot b_2 = a_2 \cdot b_1 \Leftrightarrow a_1/b_1 = a_2/b_2$, da cui la nota condizione $m_1 = m_2$.

Due rette, $r: a_1x + b_1y + c_1 = 0$ e $s: a_2x + b_2y + c_2 = 0$, sono *parallele* se hanno i *numeri direttori proporzionali*, cioè se esiste uno scalare k tale che:

$$(a_1, b_1) = k(a_2, b_2) \quad \text{o, equivalentemente,} \quad \begin{cases} a_1 = ka_2 \\ b_1 = kb_2 \end{cases}$$

Un'altra espressione cartesiana equivalente di una retta nel piano è:

$$y = mx + q$$

dove m è il coefficiente angolare e q l'ordinata del punto d'intersezione con l'asse y .

Inoltre, $m = -a/b$ e $q = -c/b$.

L'equazione del *fascio di rette* passante per un punto $P(x_0, y_0)$ è:

$$(y - y_0) = m(x - x_0).$$

Tale espressione risulta essere utile quando è noto il punto P . In alcuni casi si conoscono le equazioni di due rette incidenti e si vuole conoscere l'equazione del fascio di rette passante per il punto di incidenza (si potrebbe calcolare il punto d'intersezione e quindi applicare l'equazione precedente). Esiste una strada più semplice per individuare tale fascio.

Siano $r: a_1x + b_1y + c_1 = 0$ ed $s: a_2x + b_2y + c_2 = 0$ due rette non parallele.

L'equazione del *fascio proprio* di rette passante per il punto d'intersezione delle due rette precedenti è:

$$\lambda(a_1x + b_1y + c_1) + \mu(a_2x + b_2y + c_2) = 0$$

Al variare di λ e μ si ottengono tutte le rette del fascio.

È facile verificare che tale espressione (più generale) «assorbe» quella precedente; infatti, se P ha coordinate (x_0, y_0) , due rette che passano per questo punto sono le rette di equazioni:

$$x = x_0, \text{ da cui } x - x_0 = 0$$

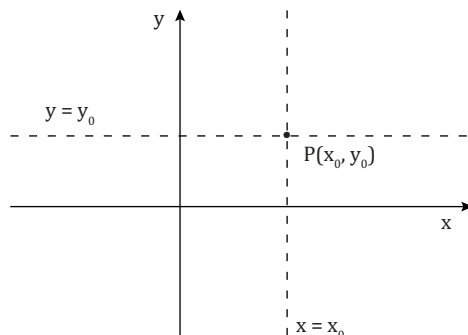
$$y = y_0, \text{ da cui } y - y_0 = 0$$

Quindi l'equazione del fascio diventa in tal caso:

$$\lambda(x - x_0) + \mu(y - y_0) = 0$$

$$\text{da cui: } y - y_0 = -\frac{\lambda}{\mu}(x - x_0)$$

$$\text{e quindi: } (y - y_0) = m(x - x_0).$$



Per *fascio improprio di rette* si intende, invece, un fascio di *rette parallele*.

26 I fasci di piani nello spazio euclideo

Analogamente al caso del fascio di rette nel piano, si parla di fascio di piani nello spazio.

Un *fascio di piani improprio* è un fascio di piani paralleli; un *fascio di piani proprio* è un fascio di centro una retta.

Assegnata una retta nello spazio si può passare alla sua rappresentazione cartesiana, espressa come intersezione di due piani, che è del tipo:

$$\begin{cases} a_1x + b_1y + c_1z + d_1 = 0 \\ a_2x + b_2y + c_2z + d_2 = 0 \end{cases}$$

Il fascio di piani passante per la retta ha equazione:

$$\lambda(a_1x + b_1y + c_1z + d_1) + \mu(a_2x + b_2y + c_2z + d_2) = 0$$

Il fascio di piani paralleli (fascio improprio) al piano di equazione $ax + by + cz + d = 0$ ha equazione:

$$ax + by + cz + k = 0$$

27 La distanza fra due punti nello spazio

Assegnati due punti:

$$P_1(x_1, y_1, z_1) \quad \text{e} \quad P_2(x_2, y_2, z_2)$$

la distanza fra essi è:

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

28 Il punto medio di un segmento nello spazio

Siano $P_1(x_1, y_1, z_1)$ e $P_2(x_2, y_2, z_2)$ gli estremi del segmento.

Il punto medio M avrà coordinate:

$$M\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}, \frac{z_1 + z_2}{2}\right)$$

29 Le superfici nello spazio

Si definisce *equazione di una superficie nello spazio* (nel sistema di coordinate di riferimento scelto nello spazio euclideo) un'equazione a tre variabili che sia soddisfatta dalle coordinate di ciascun punto appartenente a questa superficie e non sia soddisfatta dalle coordinate di alcun altro punto che non appartenga alla superficie.

In numerosi problemi fisici l'equazione di una superficie ha un'importanza primaria, mentre la superficie stessa (cioè l'insieme dei punti) è considerata come un elemento secondario.

Una superficie nello spazio ha un'equazione del tipo:

$$f(x, y, z) = 0$$

Una curva nello spazio (retta, circonferenza, ellisse, iperbole etc.) è in genere espressa come intersezione di due superfici:

$$\begin{cases} f(x, y, z) = 0 \\ g(x, y, z) = 0 \end{cases}$$

La geometria analitica, ed in particolare la geometria analitica dello spazio euclideo (tridimensionale), ha per oggetto principale lo studio delle superfici definite dalle equazioni algebriche rispetto a coordinate cartesiane ortogonali. Tali equazioni hanno le forme:

$$(1) \quad Ax + By + Cz + D = 0$$

$$(2) \quad Ax^2 + By^2 + Cz^2 + 2Dxy + 2Exz + 2Fyz + 2Gx + 2Hy + 2Kz + L = 0$$

Se si eliminano i termini contenenti la variabile z , si ottengono equazioni algebriche di primo e secondo grado nelle variabili x, y . Tali equazioni rappresentano, nel piano, le curve di cui siamo già a conoscenza (rette, circonferenze, ellissi etc.).

La (1) è detta *equazione generale di primo grado* (a condizione che A, B e C non siano contemporaneamente nulli). La (2) è detta *equazione generale di secondo grado*. Le teorie si estendono a equazioni di ordine superiore a due.

Una superficie definita in un sistema di coordinate cartesiane ortogonali da un'equazione algebrica di grado n si chiama *superficie algebrica di ordine n* .

Un piano, ad esempio, è una superficie algebrica del *primo ordine*.

Le quadriche (sfera, ellissoide, iperboloide etc.) sono superfici algebriche del *secondo ordine*, in cui $A, B, C, \dots, L$ sono i coefficienti delle equazioni (numeri assegnati se l'equazione è nota).

a) Equazione della sfera

La sfera di centro $C(\alpha, \beta, \gamma)$ e raggio r ha equazione:

$$(x - \alpha)^2 + (y - \beta)^2 + (z - \gamma)^2 = r^2$$

b) Equazione di una superficie cilindrica che ha per asse l'asse z

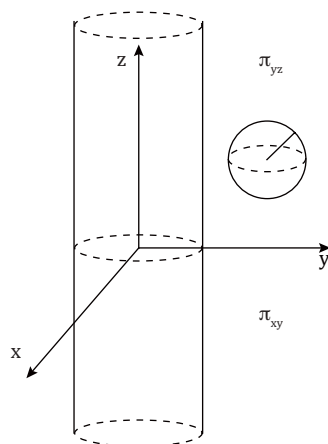
L'equazione di una superficie cilindrica che ha per asse un asse del sistema cartesiano di riferimento (es. asse z) è:

$$x^2 + y^2 = r^2$$

La *direttrice* di tale cilindro è la curva d'intersezione fra la superficie cilindrica ed il piano (superficie) π_{xy} . Quindi la direttrice è la circonferenza di equazione:

$$\begin{cases} x^2 + y^2 = r^2 \\ z = 0 \end{cases}$$

Il cilindro è una superficie ottenuta dalla rotazione di una retta (detta *generatrice*) intorno ad un asse.



c) Equazione segmentaria del piano

Si è visto che l'equazione di un piano è del tipo:

$$ax + by + cz + d = 0$$

dove d è il termine noto e la terna (a, b, c) individua la giacitura del piano.

Se dividiamo tutto per $-d$ otteniamo:

$$\frac{a}{-d}x + \frac{b}{-d}y + \frac{c}{-d}z = 1$$

Posto $-a/d = 1/A$,

$-b/d = 1/B$,

$-c/d = 1/C$,

si ha l'*equazione segmentaria di un piano*:

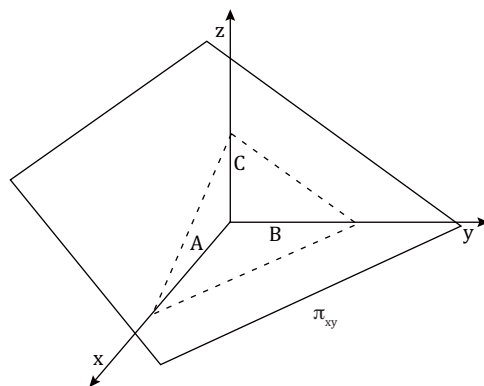
$$\frac{x}{A} + \frac{y}{B} + \frac{z}{C} = 1$$

dove A, B e C rappresentano, geometricamente e a meno del segno, le lunghezze dei segmenti tagliati dal piano sugli assi coordinati.

Analogamente nel piano, l'equazione:

$$\frac{x}{A} + \frac{y}{B} = 1$$

è l'*equazione segmentaria della retta*.



d) Equazione canonica di un ellissoide

Si chiama *ellissoide* una superficie definita in un sistema di coordinate cartesiane ortogonali dall'equazione:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

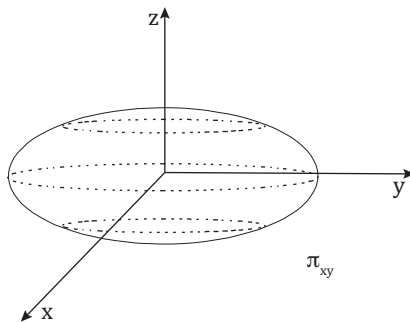
Si pensi alla superficie esterna di un pallone da baseball oppure di un uovo «un po' più simmetrico».

Se si **interseca l'ellissoide** con un piano parallelo al piano coordinato π_{xy} , cioè un piano di equazione $z = k$, la curva generata dalla sezione è definita dal sistema:

$$\begin{cases} \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \\ z = k \end{cases}$$

da cui:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{c^2 - k^2}{c^2}$$



Se $|k| < |c|$ tale equazione rappresenta **un'ellisse**, i cui semiassi sono:

$$s_1 = a\sqrt{\frac{c^2 - k^2}{c^2}}; \quad s_2 = b\sqrt{\frac{c^2 - k^2}{c^2}}$$

Tali semiassi hanno valore massimo quando $k=0$, cioè quando il piano sezionante l'ellissoide è il piano coordinato $z=0$. In tali condizioni, i semiassi dell'ellisse individuata dall'intersezione sono:

$$s_1 = a \quad \text{e} \quad s_2 = b$$

Quando il piano sezionante assume equazione $z = \pm c$, l'ellisse data dall'intersezione degenera in un punto.

Se il piano ha equazione $z = k$ con $|k| > c$, l'ellisse data dall'intersezione è un'ellisse immaginaria.

Analoghe conclusioni si traggono se il piano sezionante è parallelo ad uno degli altri due piani coordinati ($y = k$ e $x = k$).

Le grandezze a , b e c sono dette *assi dell'ellissoide*. Se $a \neq b \neq c$, l'ellissoide è detto *a tre assi*.

Se $a = b \neq c$, le sezioni reali dell'ellissoide con i piani paralleli al piano π_{xy} sono circonferenze di asse z . In tal caso l'ellissoide può essere visto come una superficie di rivoluzione, generata dalla rotazione di un'ellisse intorno ad un semiasse.

Analogo discorso può farsi nei casi in cui $a \neq b = c$ oppure $b \neq a = c$.

Se $a = b = c$ l'ellissoide degenera in una *sfera* di raggio a e centro nell'origine.

e) Equazione canonica di un iperboloide a una falda e a due falde

L'equazione di un *iperboloide a una falda* è:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$

L'iperboloide a una falda può essere pensato come un tubo a sezione in generale ellittica che si apre verso le due estremità, a guisa di collo d'anfora.

È facile dimostrare che le *intersezioni* dell'iperboloide con piani paralleli al piano coordinato π_{xy} , cioè un piano di equazione $z=k$, generano delle curve *ellittiche*.

In particolare se $a=b$, tali ellissi diventano circonferenze; in questo caso l'iperboloide a una falda può essere visto come la rotazione di un ramo di iperbole attorno all'asse z .

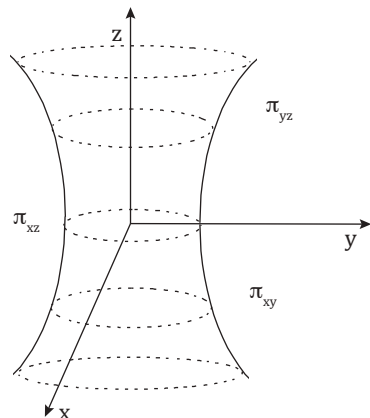
L'equazione di un *iperboloide a due falde* è:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$$

Tale iperboloide *interseca* l'asse z nei punti $(0, 0, +c)$ e $(0, 0, -c)$.

Si può facilmente provare che le *intersezioni* di tale iperboloide con piani del tipo $z=k$, con $|k| > c$, sono ellissi. I semiassi di tali *ellissi* crescono al crescere di k .

Le *intersezioni* dell'iperboloide a due falde, con piani del tipo $y=k$ e $x=k$, sono iperboli.



f) Paraboloide ellittico

L'equazione canonica di un *paraboloide ellittico* è:

$$2z = \frac{x^2}{m} + \frac{y^2}{n}$$

con m e n numeri reali positivi.

L'intersezione del paraboloide ellittico con il piano $y=0$ è:

$$\begin{cases} 2z = \frac{x^2}{m} + \frac{y^2}{n} \\ y = 0 \end{cases} \quad \text{da cui} \quad x^2 = 2zm$$

che nel piano xz rappresenta una *parabola* ascendente, simmetrica rispetto all'asse z e avente per vertice l'origine delle coordinate.

Analogo discorso può farsi sezionando il paraboloide ellittico con il piano $x=0$; la sezione è, anche stavolta, una parabola.

Si ottengono delle parabole anche nel caso in cui il paraboloide venga sezionato con piani paralleli al piano xz (ossia di equazione $y=k$) o con piani paralleli al piano yz (ossia di equazione $x=k$).

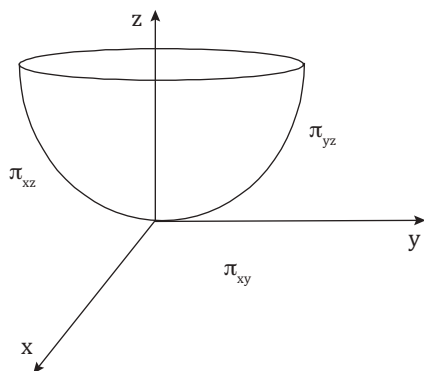
Le intersezioni del paraboloide ellittico con piani paralleli al piano $z=0$, cioè con piani del tipo $z=k$ con $k > 0$, sono:

$$\begin{cases} 2z = \frac{x^2}{m} + \frac{y^2}{n} \\ z = k \end{cases} \quad \text{da cui} \quad \frac{x^2}{m} + \frac{y^2}{n} = 2k$$

che rappresenta un'ellisse di semiassi:

$$s_1 = \sqrt{2km} \quad s_2 = \sqrt{2kn}$$

All'aumentare di k crescono i semiassi.



g) Paraboloide iperbolico (o a sella)

L'equazione canonica del *paraboloide iperbolico* è:

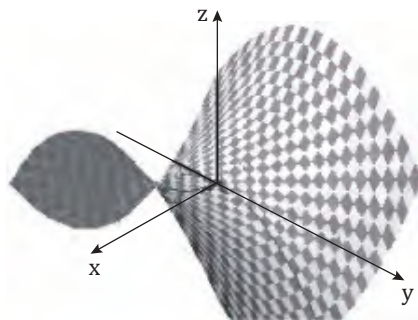
$$2z = \frac{x^2}{m} - \frac{y^2}{n}$$

con m e n positivi.

L'intersezione del paraboloide iperbolico con il piano $y = 0$ è:

$$\begin{cases} 2z = \frac{x^2}{m} - \frac{y^2}{n} \\ y = 0 \end{cases} \quad \text{da cui} \quad x^2 = 2zm$$

che nel piano xz rappresenta una *parabola* ascendente, simmetrica rispetto all'asse z , avente per vertice l'origine delle coordinate.



1 Nozioni di base

Il termine *algoritmo* deriva dal nome di un arabo e precisamente del matematico Muhammad ibn Mûsa al-Khuwârizmî (IX sec.), autore del trattato di algebra in lingua araba *Libro dell'addizione e della sottrazione secondo il calcolo degli indiani*.

Sono trascorsi undici secoli dalla pubblicazione dell'opera e il termine *algoritmo* è stato oggetto di trasformazione, a causa sia della logica matematica sia dell'avvento degli elaboratori elettronici.

Esistono diverse definizioni del termine *algoritmo*. Si può affermare che, *in generale, per algoritmo si intende una successione finita ed ordinata di istruzioni che fanno giungere alla soluzione di un problema; in particolare, per algoritmo di una data operazione si intende il procedimento che si attua per eseguire il calcolo.*

• Proprietà degli algoritmi

1. **FINITEZZA:** eseguendo le istruzioni, un algoritmo, in ogni caso, deve terminare dopo un numero finito di passi.
2. **NON AMBIGUITÀ:** devono essere definiti in modo univoco i passi successivi da seguire per ottenere il risultato; per cui, dire che un algoritmo deve essere *non ambiguo* vuol dire che il significato di ciascuna istruzione deve essere sempre lo stesso qualunque sia l'operatore che esegue l'algoritmo.
3. **GENERALITÀ:** ogni algoritmo deve fornire la soluzione per tutti i problemi appartenenti ad una data classe. Un algoritmo deve essere applicabile a qualunque insieme di dati appartenenti ad un insieme denominato *insieme di definizione o dominio dell'algoritmo*.
4. **ESEGUIBILITÀ:** ogni istruzione deve poter essere eseguita in un tempo finito, da una persona che usi mezzi manuali (carta e penna).
5. **FUNZIONALITÀ:** deve fornire almeno un risultato o dato in uscita.

Esempi di algoritmi sono: il procedimento per il calcolo del prodotto dei primi n numeri naturali; il procedimento di *ricerca dicotomica* di un dato elemento in una successione ordinata, dove per *ricerca dicotomica* si intende un algoritmo in cui, ad ogni passo, la successione è divisa in due parti più o meno uguali ed è deciso, sulla base di un confronto tra l'elemento dato e quello situato a metà della successione, di continuare usando solo una delle due parti, a meno che non si sia già trovata la coincidenza con l'elemento.

ESEMPIO 1

Si consideri uno schema di algoritmo per calcolare il prodotto dei primi n numeri naturali.

Passo 1 - Si distinguono le due variabili A e I , assegnando loro i valori iniziali $A = 1$ e $I = 1$. I rappresenta la *variabile contatore*. Essa tiene conto della successione degli n numeri e, per tale motivo, assumerà valori compresi tra 1 ed n . A è la *variabile prodotto*. Il suo valore sarà pari al prodotto dei primi I numeri (con $1 \leq I \leq n$).

Passo 2 - Si attribuisce ad I il valore $I + 1$.

Passo 3 - Si attribuisce ad A il valore $A \cdot I$.

Passo 4 - Se $I < n$ si ritorna al passo 2; se, invece, $I \geq n$ si avanza al passo 5.

Passo 5 - Il procedimento ha termine e il valore cercato è quello assunto da A .

2 La traduzione di un algoritmo in linguaggio di programmazione

Per risolvere un qualsiasi problema con l'ausilio di un calcolatore è necessario individuare un algoritmo.

In una prima formulazione, ogni problema è espresso in *linguaggio naturale*, che non fornisce alcun metodo per calcolare il risultato in base ai dati di partenza, e che può essere una descrizione dubbia.

Pertanto, per risolvere un problema con l'ausilio di un calcolatore e per automatizzare la soluzione del problema, si devono sempre individuare una serie di regole che, sistematicamente eseguite, consentono di ottenere il risultato del problema o, il che è lo stesso, si deve individuare il procedimento risolutivo del problema.

In secondo luogo, si deve individuare un modo di rappresentare l'algoritmo, le informazioni di cui si dispone e quelle usate dall'algoritmo nel procedimento risolutivo in un linguaggio comprensibile per il calcolatore; si chiamano, rispettivamente, *dati di ingresso* le informazioni di cui si dispone, *programma* la rappresentazione dell'algoritmo e, infine, *dati di uscita* le rappresentazioni date dal calcolatore a seguito del processo di calcolo. *Il dato è la rappresentazione di una informazione, codificata secondo determinate convenzioni, in modo da poter essere trattata da un sistema di elaborazione.*

I *linguaggi di programmazione* sono appositi formalismi mediante i quali è possibile descrivere algoritmi in una forma che sia direttamente comprensibile dall'elaboratore; essi usano parole e regole per la costruzione delle istruzioni, sicuramente più dure di quelle utilizzate nel linguaggio naturale. I linguaggi di programmazione adoperano solo un numero limitato di istruzioni di base, ognuna avente un ruolo definito.

ESEMPIO 2

Si consideri un semplice programma, chiamato *cerchio*, che legge il raggio di un cerchio, ne calcola l'area e ne visualizza il valore, scritto nel linguaggio di programmazione Pascal.

```
PROGRAM cerchio;
(* AREA DI UN CERCHIO *)
VAR area, raggio : real;
BEGIN
  read (raggio);
  area := 3.14159*sqr (raggio);
  writeln (area)
END.
```

Si considerino, ora, le seguenti osservazioni sul programma mostrato.

- Le parole scritte in maiuscolo sono *parole chiave*, aventi un significato predefinito.
- L'intestazione del programma è contenuta nella prima riga ed è *cerchio*.
- Nella seconda riga è scritto un commento che descrive lo scopo del programma.
- Nella terza riga del programma, i valori numerici del raggio e dell'area sono rappresentati dai nomi simbolici *raggio* e *area*, che sono le *variabili*.
- Le istruzioni del programma sono nelle tre righe comprese tra BEGIN ed END. Attraverso esse è inserito nel calcolatore un valore del raggio ed è calcolato il valore dell'area, il quale è successivamente visualizzato.
- Nella sesta riga, il nome *sqr* è una *funzione standard*, utilizzata per calcolare il quadrato del raggio.

3 Le strutture elementari per la descrizione degli algoritmi

Le strutture elementari utilizzate per la descrizione degli algoritmi sono le seguenti:

- L'*istruzione*, che è il tipo di algoritmo più semplice e che non è ulteriormente scomponibile rispetto alle capacità dell'esecutore.
- La *sequenza di istruzioni* con risultati intermedi.
- La *struttura condizionale*, la quale si adotta quando si introduce un controllo su una condizione relativa al risultato di un calcolo.
- La *struttura iterativa*, che permette di ripetere le istruzioni il numero di volte necessario al verificarsi di una certa condizione. Se la condizione è numerica, allora si parla di *iterazione enumerativa*, altrimenti si parla di *iterazione non enumerativa*.

4 La rappresentazione grafica degli algoritmi

Esistono diversi metodi grafici per la rappresentazione degli algoritmi, quali i *diagrammi di flusso* e gli *alberi algoritmici*.

I *diagrammi di flusso* consentono la descrizione degli algoritmi che, a un primo livello di dettaglio, individuano blocchi i quali mostrano operazioni come *risoluzione di una equazione*, *lettura di dati* etc. I blocchi sono i seguenti.

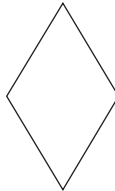
Elaborazione



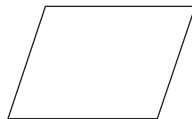
Inizio e fine



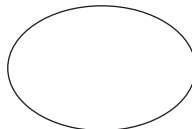
Selezione



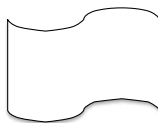
Dati di ingresso e uscita



Connettore

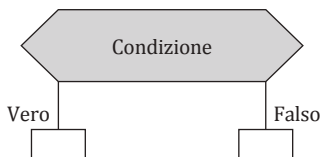


Stampa



Un *albero algoritmico* è una struttura matematica contraddistinta da un insieme di nodi e di rami, e avente un nodo particolare, detto *radice*. Nella struttura ad albero, le operazioni più semplici in cui si scompone una operazione complessa sono specificate ad un livello inferiore rispetto a quello dell'operazione complessa. Le operazioni di uno stesso livello sono eseguite da sinistra verso destra, secondo una priorità temporale.

Di seguito è indicato il simbolo che consente di rappresentare una struttura di base degli algoritmi, e precisamente la struttura condizionale.



5 Il controllo della correttezza

Per provare la correttezza o la scorrettezza di un algoritmo esistono due tipi di metodi:

- *metodi basati sui dati di test*, che consistono nell'eseguire l'algoritmo e nella successiva osservazione del suo comportamento. Tali metodi tendono a ridurre al minimo il numero dei test da effettuare in un algoritmo;
- *metodi formali*, che non eseguono l'algoritmo in quanto ne provano la correttezza su ragionamenti e deduzioni sul testo dello stesso.

6 La complessità degli algoritmi

Per *complessità di un algoritmo* si intende la *misura della gravosità attribuita alle possibili esecuzioni di un algoritmo*; la complessità valuta, dunque, l'efficienza di un algoritmo ed è misurata in termini di risorse che devono essere utilizzate per la soluzione dello stesso.

Si distingue tra:

- *complessità strutturale*, che fa riferimento alla gravosità del lavoro di programmazione;
- *complessità spaziale*, che fa riferimento all'estensione della memoria da usare per l'esecuzione di un algoritmo;
- *complessità temporale*, che fa riferimento al tempo necessario all'esecuzione di un algoritmo.

L'analisi di complessità è espressa come funzione dei dati cui un algoritmo fa riferimento.

Nel valutare la complessità spaziale e temporale di un algoritmo, non si intende calcolare esattamente il tempo e lo spazio fisico di memoria occupata, ma effettuare una loro valutazione approssimata. Per questo motivo si considerano le istruzioni più eseguite dell'algoritmo, disinteressandosi delle altre, e si assume, pertanto, come misura di efficienza di un algoritmo, il numero delle istruzioni più eseguite; tale numero è espresso in funzione della dimensione dei dati di ingresso.

In generale, si dice che un algoritmo presenta *complessità polinomiale* se la funzione esprime la sua complessità è proporzionale a un polinomio di grado n (in cui n è il numero di dati).

7 Cenni sulla computabilità e sulla tesi di Church

Prima di esporre la *tesi di Church* è necessario definire il concetto di *computabilità*, legato a quello di algoritmo. Una funzione f si dice *computabile* se e solo se esiste una procedura, defi-

nibile in un numero finito di passi, attraverso cui si ottiene il valore $f(W)$ della funzione per ogni possibile valore del dominio W .

L'ingegnere inglese Alan Mathison Turing (1912-1954) ha reso chiaro il concetto di computabilità, ideando un automa, detto *macchina di Turing*, in grado di eseguire il calcolo di una funzione (a valori e argomenti naturali) in corrispondenza di ogni valore dell'argomento.

Il matematico americano Alonzo Church (1903-1995) ha compiuto numerose ricerche in tal senso, esponendo una tesi secondo la quale *una funzione è computabile se e solo se è ricorsiva* [1]. Egli ha mostrato, quindi, che i concetti di *computabilità* e di *ricorsività* coincidono.

Si ricordi che il concetto di computabilità, applicato, invece che a funzioni, a proposizioni o predicati, prende il nome di *decidibilità*.

^[1] Una funzione matematica è definita *ricorsivamente* quando nella sua definizione compare un riferimento (chiamata) a se stessa. Esempi di problemi ricorsivi sono: la somma dei primi n numeri naturali, la ricerca di un elemento in una sequenza di interi oppure il calcolo del fattoriale di un numero naturale n .

Elementi di trigonometria e funzioni trigonometriche

1 La circonferenza trigonometrica

Per *circonferenza trigonometrica* (o *goniometrica*) si intende la circonferenza di raggio unitario.

Si consideri un sistema di riferimento cartesiano con origine nel centro di tale circonferenza. Sia P un punto appartenente alla circonferenza e si indichino con P' e P'' le proiezioni ortogonali di P sugli assi coordinati. Indichiamo con α l'angolo individuato dall'asse x e dal raggio corrispondente al punto P .

Si definisce *seno dell'angolo* α il rapporto fra i segmenti PP' e OP :

$$\text{sen} \alpha = PP' / OP$$

Il segmento PP' è da considerarsi orientato, per cui esso sarà positivo o negativo a seconda che il punto P si trovi al di sopra o al di sotto dell'asse x . Il segmento $OP = 1$ perché raggio della circonferenza unitaria. Segue che il $\text{sen} \alpha$ si può confondere con il segmento PP' e risulta positivo nel primo e secondo quadrante, mentre è negativo nel terzo e quarto quadrante.

Si definisce *coseno dell'angolo* α il rapporto fra i segmenti PP'' e OP :

$$\text{cos} \alpha = PP'' / OP$$

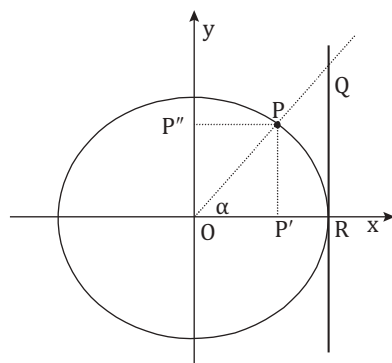
Il segmento PP'' è da considerarsi orientato, per cui esso sarà positivo o negativo a seconda che il punto P si trovi a destra o a sinistra dell'asse y . Visto che $OP = 1$, perché raggio della circonferenza unitaria, si deduce che il $\text{cos} \alpha$ si può confondere con il segmento PP'' e risulta positivo nel primo e quarto quadrante, mentre è negativo nel secondo e terzo quadrante.

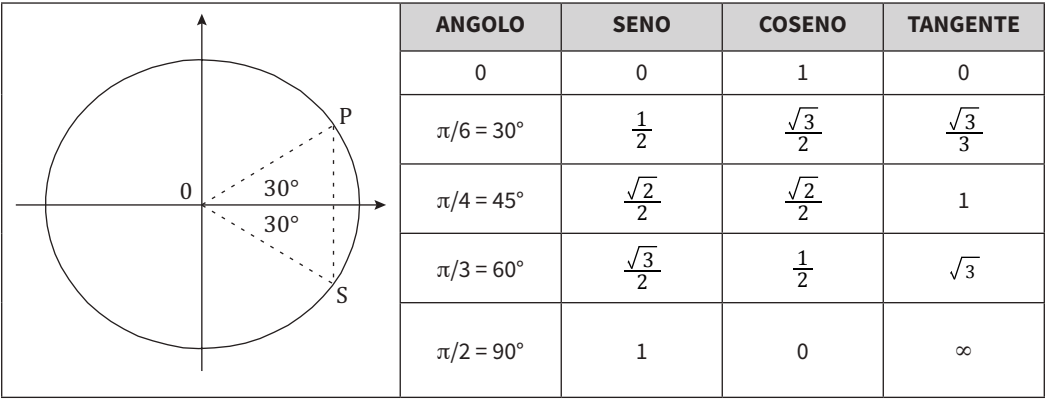
Si definisce *tangente di* α il rapporto fra il $\text{sen} \alpha$ ed il $\text{cos} \alpha$:

$$\text{tg} \alpha = \text{sen} \alpha / \text{cos} \alpha$$

Geometricamente la $\text{tg} \alpha$ coincide con la lunghezza del segmento orientato QR , quindi sarà positiva nel primo e terzo quadrante, mentre risulterà negativa nel secondo e quarto quadrante. Quando α cresce da 0 a $\pi/2$, escluso $\pi/2$, la $\text{tg} \alpha$ assume, sempre crescendo, tutti i valori reali positivi, mentre per $\alpha = \pi/2$ non esiste il valore della tangente; in altre parole, si può dire che quando α cresce da 0 a $\pi/2$, la $\text{tg} \alpha$ cresce da 0 a $+\infty$. Analogamente, quando α si avvicina a $\pi/2$, per valori maggiori di $\pi/2$, la $\text{tg} \alpha$ assume, decrescendo, valori negativi «piccoli quanto si vuole» (in pratica la tangente tende a $-\infty$).

I valori assunti dalle *funzioni trigonometriche* (o *goniometriche*, o *circolari*) calcolate in angoli notevoli sono riportati in tabella. Tali valori sono facilmente deducibili da considerazioni geometriche. Ad esempio, il seno di 30° è pari alla metà del lato del triangolo equilatero OPS di lato $OP = 1$.





2 La relazione fondamentale

$\text{sen}^2\alpha + \text{cos}^2\alpha = 1$

Tale relazione si giustifica applicando il teorema di Pitagora al triangolo OPP' .

3 Le relazioni fra le funzioni trigonometriche

Queste relazioni si ottengono sfruttando la relazione fondamentale. Esse sono utili per la risoluzione di equazioni o disequazioni trigonometriche.

	senα	cosα	tgα
senα	senα	$\pm\sqrt{1-\text{sen}^2\alpha}$	$\frac{\text{sen}\alpha}{\pm\sqrt{1-\text{sen}^2\alpha}}$
cosα	$\pm\sqrt{1-\text{cos}^2\alpha}$	cosα	$\frac{\pm\sqrt{1-\text{cos}^2\alpha}}{\text{cos}\alpha}$
tgα	$\frac{\text{tg}\alpha}{\pm\sqrt{1+\text{tg}^2\alpha}}$	$\frac{1}{\pm\sqrt{1+\text{tg}^2\alpha}}$	tgα

4 Le formule di addizione, sottrazione e duplicazione

Formule di addizione e sottrazione	$\begin{cases} \text{sen}(\alpha \pm \beta) = \text{sen}\alpha\text{cos}\beta \pm \text{cos}\alpha\text{sen}\beta \\ \text{cos}(\alpha \pm \beta) = \text{cos}\alpha\text{cos}\beta \mp \text{sen}\alpha\text{sen}\beta \\ \text{tg}(\alpha \pm \beta) = \frac{\text{tg}\alpha \pm \text{tg}\beta}{1 \mp \text{tg}\alpha\text{tg}\beta} \end{cases}$

Formule di duplicazione	$\begin{cases} \text{sen}2\alpha = 2\text{sen}\alpha\text{cos}\alpha \\ \text{cos}2\alpha = \text{cos}^2\alpha - \text{sen}^2\alpha = 1 - 2\text{sen}^2\alpha = 2\text{cos}^2\alpha - 1 \\ \text{tg}2\alpha = \frac{2\text{tg}\alpha}{1 - \text{tg}^2\alpha} \end{cases}$

5 Le relazioni trigonometriche applicate ai triangoli rettangoli

La trigonometria permette di individuare semplici legami fra gli angoli ed i lati di un poligono; in particolare per i triangoli rettangoli valgono le seguenti relazioni:

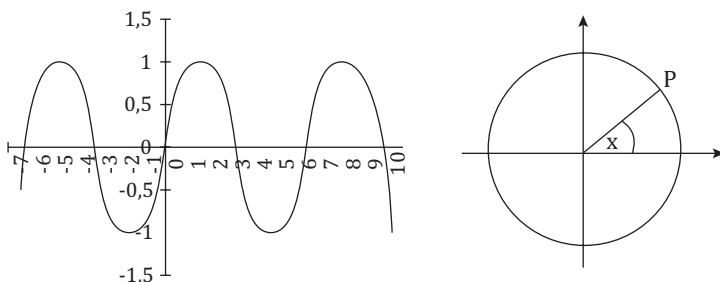
- la misura di un cateto è uguale a quella dell'ipotenusa moltiplicata per il seno dell'angolo opposto al cateto;
- la misura di un cateto è uguale a quella dell'ipotenusa moltiplicata per il coseno dell'angolo acuto adiacente al cateto stesso;
- la misura di un cateto è uguale alla misura dell'altro cateto per la tangente dell'angolo opposto al cateto cercato.

6 Le funzioni goniometriche

Le funzioni goniometriche appartengono alla classe delle funzioni elementari. In particolare, vediamo i comportamenti delle seguenti funzioni: $y = \operatorname{sen} x$, $y = \operatorname{cos} x$, $y = \operatorname{tg} x$, $y = \operatorname{arcsen} x$, $y = \operatorname{arccos} x$, $y = \operatorname{arctg} x$.

7 La funzione $y = \operatorname{sen}(x)$ e la sua inversa $y = \operatorname{arcsen}(x)$

Tale funzione, nota come *funzione sinusoidale*, ha come campo di esistenza tutto l'insieme $\mathfrak{R}$: $X = \mathfrak{R}$.



I punti d'intersezione con l'asse y sono dati da:

$$\begin{cases} y = \operatorname{sen}(x) \\ x = 0 \end{cases}; \quad P(0,0)$$

I punti d'intersezione con l'asse x sono dati da:

$$\begin{cases} y = \operatorname{sen}(x) \\ y = 0 \end{cases}; \quad P(k\pi, 0) \quad \text{con } k \in \mathbb{Z}$$

La funzione $y = \operatorname{sen}(x)$ è una funzione limitata, perché il codominio è compreso fra -1 ed 1:

$$-1 \leq \operatorname{sen}(x) \leq 1 \quad \forall x \in \mathfrak{R}.$$

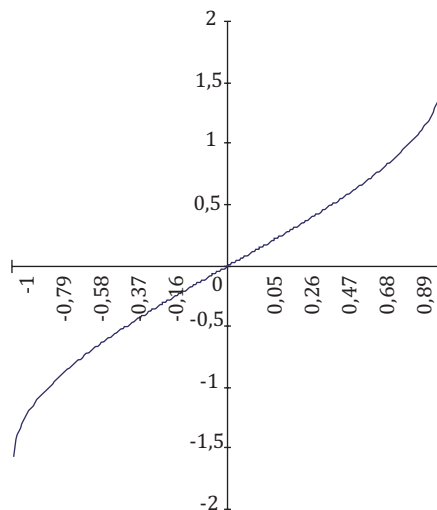
La funzione $y = \sin(x)$ è una funzione periodica, di periodo 2π . Ciò vuol dire che è possibile studiare la funzione in un intervallo di ampiezza 2π (ad es. $[0, 2\pi]$). Il comportamento si ripete periodicamente in tutto l'insieme di esistenza.

La funzione $y = \sin(x)$, nel suo intero insieme di esistenza, non è né crescente né decrescente. Se conveniamo, però, di far variare x soltanto nell'intervallo chiuso $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, in questo intervallo la funzione $y = \sin(x)$ è sempre crescente e assume tutti i valori che vanno da -1 a 1 . Quindi nell'intervallo $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ la funzione si può invertire; la sua funzione inversa si dice *arcoseno* e si indica con la scrittura:

$$y = \arcsin(x)$$

Il campo di esistenza di tale funzione è l'intervallo $X = [-1, 1]$; il codominio è:

$$Y = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$



8 La funzione $y = \cos(x)$ e la sua inversa $y = \arccos(x)$

Tale funzione, nota come *funzione cosinusoidale*, ha come campo di esistenza tutto l'insieme $\mathfrak{R}$: $X = \mathfrak{R}$.

I punti d'intersezione con l'asse y sono dati da:

$$\begin{cases} y = \cos(x) \\ x = 0 \end{cases}; \quad P(0, 1)$$

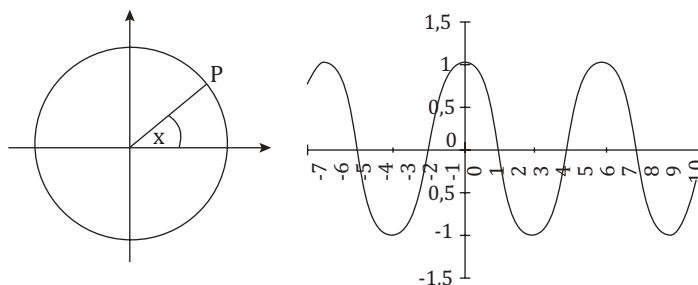
I punti d'intersezione con l'asse x sono dati da:

$$\begin{cases} y = \cos(x) \\ y = 0 \end{cases}; \quad P(\pi/2 + k\pi, 0) \quad \text{con } k \in \mathbb{Z}$$

La funzione $y = \cos(x)$ è una funzione limitata, perché il codominio è compreso fra -1 ed 1 :

$$-1 \leq \cos(x) \leq 1 \quad \forall x \in \mathfrak{R}.$$

La funzione $y = \cos(x)$ è una funzione periodica e di periodo 2π : ciò vuol dire che è possibile studiare la funzione in un intervallo di ampiezza 2π (ad es. $[0, 2\pi]$); il comportamento si ripete periodicamente in tutto l'insieme di esistenza.

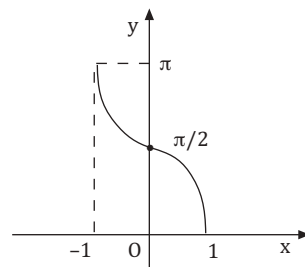


La funzione $y = \cos(x)$ è sempre decrescente nell'intervallo chiuso $[0, \pi]$ e decresce dal valore 1 al valore -1 . Da ciò segue l'esistenza della funzione inversa, detta *arcocoseno*, che si indica con la scrittura:

$$y = \arccos(x)$$

Il campo di esistenza di tale funzione è l'intervallo $X = [-1, 1]$; il codominio è:

$$Y = [0, \pi].$$



9 La funzione $y = \operatorname{tg}(x)$ e la sua inversa $y = \operatorname{arctg}(x)$

Tale funzione (periodica e di periodo π) ha come campo di esistenza l'insieme:

$$X = \mathbb{R} - \{\pi/2 + k\pi \text{ con } k \in \mathbb{Z}\}$$

I punti d'intersezione con l'asse x sono dati da:

$$\begin{cases} y = \operatorname{tg}(x); \\ y = 0 \end{cases}; \quad P(k\pi, 0) \text{ con } k \in \mathbb{Z}$$

I punti $x = \pi/2 + k\pi$ sono di discontinuità. Essendo d'accumulazione per l'insieme di definizione della funzione tangente, ha senso considerarne il limite:

$$\lim_{x \rightarrow (\pi/2 + k\pi)^-} \operatorname{tg}(x) = \infty \quad \lim_{x \rightarrow (\pi/2 + k\pi)^+} \operatorname{tg}(x) = -\infty$$

Quindi la funzione tangente è dotata di *asintoti verticali* di equazione:

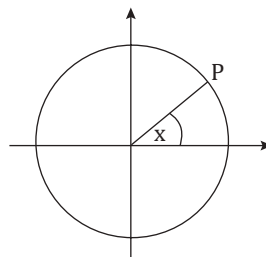
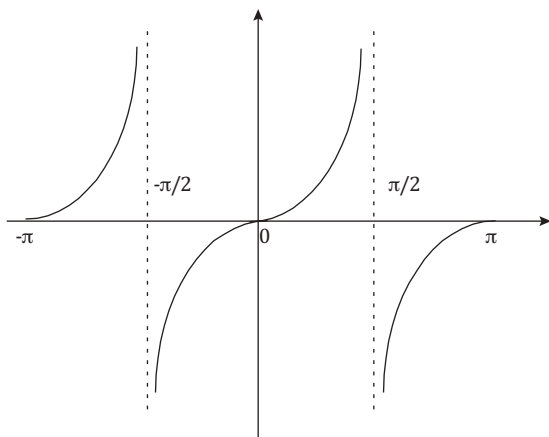
$$\pi/2 + k\pi \quad \text{con } k \in \mathbb{Z}$$

La funzione $y = \operatorname{tg}(x)$ è una funzione illimitata, perché il codominio è compreso fra $-\infty$ ed ∞ .

Essa è anche periodica di periodo π : ciò vuol dire che è possibile studiare la funzione in un intervallo di ampiezza π (ad es. $[0, \pi]$); il comportamento si ripete periodicamente in tutto l'insieme di esistenza.

La funzione tangente è invertibile in ogni intervallo del tipo:

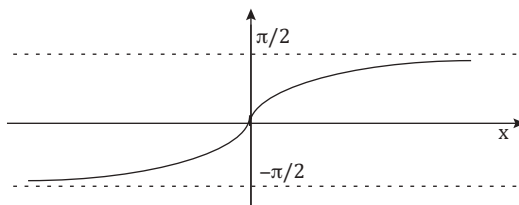
$$]-\pi/2 + k\pi, +\pi/2 + k\pi[\quad \text{con } k \in \mathbb{Z}$$



La funzione inversa $y = \arctg(x)$ è limitata perché il codominio è l'insieme aperto:

$$Y =]-\pi/2, +\pi/2[$$

Il campo di esistenza è, invece, tutto l'insieme $\mathfrak{R}$.



Le funzioni reali di una variabile reale

1 Le funzioni

Il concetto di funzione, secondo la definizione data da Dirichlet, è il seguente: si dice che *una variabile dipendente y è funzione di una variabile indipendente x , secondo la legge f , e si scrive:*

$$y = f(x)$$

quando, assegnato a x un valore qualsiasi, ed effettuando su di esso le operazioni indicate dalla legge di corrispondenza f , si ottiene un unico e finito valore per la y .

La legge f può essere di natura qualsiasi e, inoltre, si dice che la y è funzione analitica della x , in quanto le due variabili sono legate da una interdipendenza espressa da operazioni analitiche.

Le funzioni analitiche si distinguono in:

- *algebriche*, se il legame che intercede tra le due variabili è di natura algebrica ed è rappresentato da un polinomio.

A loro volta le funzioni algebriche si distinguono in *razionali* e *irrazionali*, a seconda che le operazioni che si effettuano sulla variabile indipendente non prevedano o prevedano l'elevazione a potenza con esponente frazionario;

- *trascendenti*, come le funzioni logaritmiche, esponenziali e trigonometriche.

Tali funzioni sono dette anche *funzioni elementari* e, in quanto tali, si distinguono dalle *funzioni composte*, le quali presuppongono un complesso di operazioni che legano le due variabili.

2 Intervallo e intorno

Dati due numeri qualsiasi a e b , con $a < b$, l'insieme di tutti i numeri reali compresi tra a e b si dice:

- *intervallo chiuso*, se a e b sono inclusi, e spesso lo si indica in questo modo:

$$[a, b]$$

In tale caso a e b si dicono, rispettivamente, *minimo* e *massimo* dell'intervallo;

- *intervallo aperto*, se a e b sono esclusi, e spesso lo si indica in questo modo:

$$]a, b[$$

In tale caso a e b si dicono, rispettivamente, *estremo sinistro* e *estremo destro* dell'intervallo.

Dato un numero reale x , si dice *intorno completo* di x un intervallo $]a, b[$ che contiene x .

In particolare:

- l'*intorno sinistro* di x è l'intervallo $]a, x[$ in cui x è l'estremo destro;
- l'*intorno destro* di x è l'intervallo $]x, b[$ in cui x è l'estremo sinistro.

3 Il campo di esistenza di una funzione

Siano A e B due insiemi di numeri reali ed $y = f(x)$ una funzione di A in B . L'insieme A dei valori x , per i quali esiste il corrispondente valore della y , si dice *insieme* (o *campo*) *di esistenza*, o

insieme (o campo) di definizione, o anche *dominio* della funzione. Non è detto, però, che ogni elemento di B debba essere immagine di un elemento di A , cioè vi possono essere elementi di B che non sono immagine di alcun elemento di A . Ne segue che, in generale, l'insieme delle immagini sarà un sottoinsieme proprio di B , che si chiama *immagine di A in B mediante la f* o anche *codominio* della f .

Il campo di esistenza di una funzione analitica dipende dalle operazioni che si devono eseguire sulla variabile indipendente x per ottenere la variabile dipendente y ; ovviamente tali operazioni devono essere sempre possibili nel campo dei numeri reali.

In generale, il campo di esistenza di una:

— *funzione razionale intera* del tipo:

$$y = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$$

è tutto il campo R dei numeri reali;

— *funzione razionale fratta* del tipo:

$$y = \frac{f(x)}{g(x)}$$

è il campo R dei numeri reali, da cui si devono escludere i valori che annullano il denominatore $g(x)$ per i quali la divisione perderebbe significato.

ESEMPIO 1

Determinare il campo di esistenza della funzione:

$$y = \cos 5x - \cos 2x$$

Si tratta di una funzione trigonometrica avente per argomento un numero reale qualsiasi, pertanto il campo di esistenza è $]-\infty, \infty[$.

ESEMPIO 2

Determinare il campo di esistenza della funzione:

$$y = \log \frac{3x+6}{x-1}$$

Si tratta di una funzione logaritmica, per cui l'argomento deve essere maggiore di 0, ossia deve essere:

$$\frac{3x+6}{x-1} > 0 \Rightarrow \begin{cases} 3x+6 > 0 \Rightarrow x > -2 \\ x-1 > 0 \Rightarrow x > 1 \end{cases} \quad \text{oppure} \quad \begin{cases} 3x+6 < 0 \Rightarrow x < -2 \\ x-1 < 0 \Rightarrow x < 1 \end{cases}$$

Pertanto, il campo di esistenza della funzione considerata è $]-\infty, -2[\cup]1, +\infty[$.

4 Le funzioni limitate

Una funzione $y = f(x)$, definita in un dato intervallo $[a, b]$, si dice ivi *limitata* se, per ogni valore di x appartenente al suddetto intervallo, esiste un numero P positivo tale che:

$$|f(x)| \leq P$$

La funzione è:

- *limitata superiormente* se, nell'intervallo $[a, b]$, esiste un punto in cui la funzione assume valore M che è non minore dei valori assunti negli altri punti;
- *limitata inferiormente* se, nell'intervallo $[a, b]$, esiste un punto in cui la funzione assume valore m che è non maggiore dei valori assunti negli altri punti.

5 Le funzioni crescenti e quelle decrescenti

Sia data una funzione $y = f(x)$ e si considerino due punti qualsiasi x_1 e x_2 di un dato intervallo $[a, b]$. La funzione si dice:

- *non decrescente* se $x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2)$;
- *costante* se $x_1 < x_2 \Rightarrow f(x_1) = f(x_2)$;
- *non crescente* se $x_1 < x_2 \Rightarrow f(x_1) \geq f(x_2)$;
- *strettamente crescente* se $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$;
- *strettamente decrescente* se $x_1 < x_2 \Rightarrow f(x_1) > f(x_2)$.

Si dicono *monotone* le funzioni crescenti, decrescenti, non decrescenti o non crescenti, ossia le funzioni che variano sempre in uno stesso verso.

6 Le funzioni composte e quelle inverse

Sia data la funzione:

$$y = f(z)$$

dove z non è variabile indipendente, ma a sua volta funzione ($z = g(x)$) della variabile indipendente x . In tal caso, si ha che la funzione:

$$y = f[g(x)]$$

si dice *funzione composta* di f e di g .

Inoltre, data la funzione:

$$y = f(x)$$

si dice che f è *invertibile* in un intervallo $[a, b]$ chiuso se, ad ogni valore della x in $[a, b]$, corrisponde uno e un solo valore di y in $[a', b']$, dove a' e b' sono il minimo e il massimo della funzione nell'intervallo $[a, b]$, e viceversa ad ogni valore di y in $[a', b']$ corrisponde uno e un solo valore di x in $[a, b]$. La funzione è, pertanto, invertibile nell'intervallo $[a, b]$ se è continua in $[a, b]$ ed è sempre crescente o sempre decrescente in detto intervallo. La *funzione inversa* si indica in questo modo:

$$x = g(y) = g[f(x)]$$

7 Le funzioni elementari

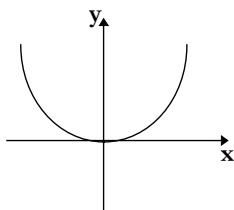
La funzione costante $[f(x) = k \quad \forall x \in \mathfrak{R} \text{ dove } k \in \mathfrak{R}]$, la funzione lineare $[f(x) = ax + b \quad \forall x \in \mathfrak{R} \text{ dove } a, b \in \mathfrak{R}]$, le funzioni circolari $[f(x) = \sin(x), f(x) = \cos(x) \quad \forall x \in \mathfrak{R}; f(x) = \operatorname{tg}(x) \quad \forall x \in \mathfrak{R} - \{\pi/2 + k\pi, \text{ con } k \in \mathbb{Z}\}]$ sono funzioni elementari.

a) Funzione potenza

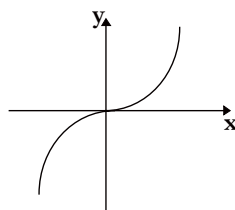
La *funzione potenza di esponente intero non negativo* n è definita in tutto $\mathfrak{R}$.

$$f(x) = x^n \quad x \in \mathfrak{R} \quad \text{ed} \quad n \in \mathbb{N}$$

- se n è pari, la funzione è *pari*, cioè è simmetrica rispetto all'asse y . In altre parole, $f(-x) = f(x) \quad \forall x \in \mathfrak{R}$;
- se n è dispari, la funzione è *dispari*, cioè è simmetrica rispetto all'origine. In altre parole, $f(-x) = -f(x) \quad \forall x \in \mathfrak{R}$.



n pari: $f(x) = f(-x)$
 f è decrescente fra $]-\infty, 0]$
 f è crescente fra $[0, +\infty[$



n dispari: $f(x) = -f(-x)$
 f è crescente fra $]-\infty, +\infty[$

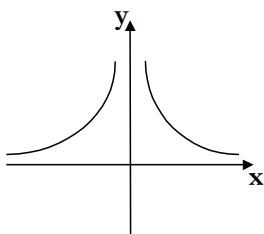
La funzione potenza di esponente intero negativo n è definita in tutto $\mathbb{R}$ escluso $x = 0$.

$$f(x) = x^n \quad x \in \mathbb{R} - \{0\} \quad \text{ed } n \in \mathbb{Z} - \mathbb{N}$$

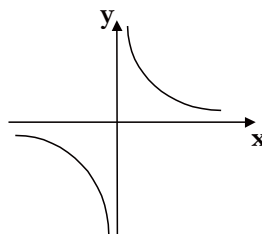
Bisogna tener presente che dire « n negativo» equivale a considerare la funzione:

$$f(x) = 1/x^n \quad x \in \mathbb{R} - \{0\} \quad \text{ed } n \in \mathbb{N}$$

- se n è pari, la funzione è pari, quindi simmetrica rispetto all'asse y ;
- se n è dispari, la funzione è dispari, quindi simmetrica rispetto all'origine.



n pari: $f(x) = f(-x)$
 f è crescente fra $]-\infty, 0]$
 f è decrescente fra $[0, +\infty[$



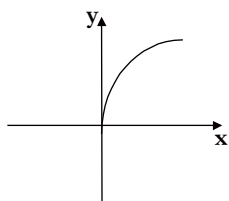
n dispari: $f(x) = -f(-x)$
 f è decrescente in $\mathbb{R} - \{0\}$

b) Funzione radice

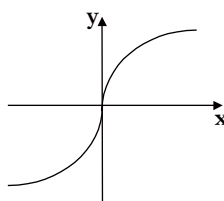
La funzione radice ennesima è la funzione potenza di esponente $1/n$.

$$f(x) = x^{\frac{1}{n}} \quad \text{oppure} \quad f(x) = \sqrt[n]{x}$$

- se n è pari, la funzione è definita solo in $[0, +\infty[$;
- se n è dispari, la funzione è dispari ed è definita in tutto $\mathbb{R}$.



n pari: f è crescente
 fra $[0, +\infty[$



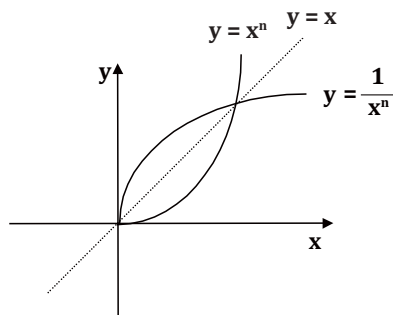
n dispari: $f(x) = -f(-x)$
 f è crescente fra $]-\infty, +\infty[$

La funzione radice ennesima $f(x) = x^{\frac{1}{n}}$ con n pari è l'inversa della restrizione nel campo $[0, +\infty[$ della funzione potenza con esponente n pari.

La funzione radice ennesima $f(x) = x^{\frac{1}{n}}$ con n dispari è l'inversa della funzione potenza con esponente n dispari.

La funzione potenza ennesima e la funzione radice ennesima sono simmetriche rispetto alla bisettrice del I e del III quadrante.

Il punto d'intersezione è $P = (1, 1)$.



Si analizzi ora la funzione potenza avente come esponente un numero reale α :

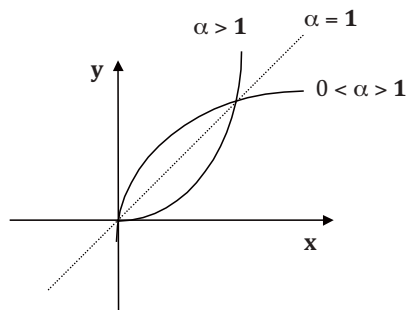
$$f(x) = x^\alpha$$

Si distinguono due situazioni.

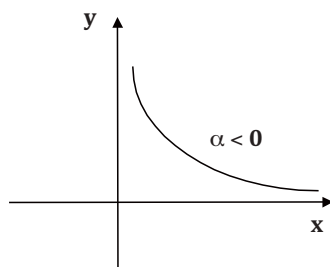
1. Se $\alpha > 0$, il dominio $X = [0, +\infty[$ e il codominio $Y = [0, +\infty[$, la funzione è strettamente crescente. In particolare si distinguono i casi in cui:

- $0 < \alpha < 1$
- $\alpha = 1$
- $\alpha > 1$

La funzione $f(x) = x^\alpha$ è dotata di minimo $m = 0$; inoltre, $\forall \alpha > 0$, $f(1) = 1$.



2. Se $\alpha < 0$, il dominio $X =]0, +\infty[$ e il codominio $Y =]0, +\infty[$, la funzione è strettamente decrescente ed è dotata di estremo inferiore ($e' = \inf x^\alpha = 0$). L'estremo superiore è, invece, infinito ($e'' = \sup x^\alpha = \infty$).



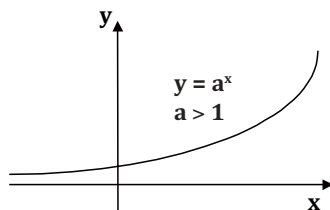
c) Funzione esponenziale di base a

La funzione esponenziale di base un numero reale positivo a :

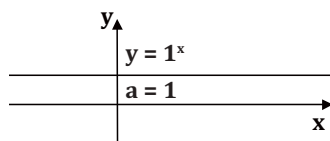
$$f(x) = a^x \quad x \in \mathbb{R}$$

Si distinguono tre casi:

1. se $a > 1$
 - la funzione è strettamente crescente
 - $\sup a^x = +\infty$
 - $\inf a^x = 0$

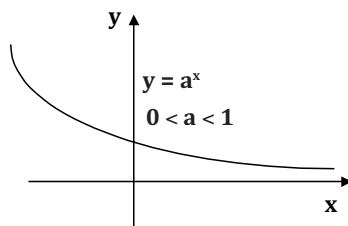


2. se $a = 1$
 - la funzione si riduce ad una funzione costante:
- $$f(x) = a$$



3. se $0 < a < 1$

- la funzione è strettamente decrescente
- $\sup a^x = +\infty$
- $\inf a^x = 0$



d) Funzione logaritmo in base a

$$f(x) = \log_a x;$$

dominio $X =]0, +\infty[;$

codominio $Y =]-\infty, +\infty[.$

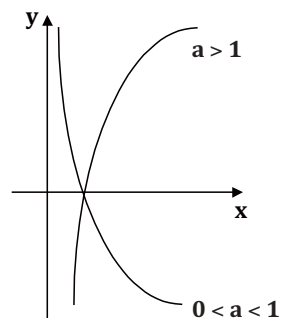
Si distinguono due casi:

1. se $a > 1$

- la funzione è strettamente crescente;

2. se $0 < a < 1$

- la funzione è strettamente decrescente.



La funzione logaritmo è la funzione inversa della funzione esponenziale.

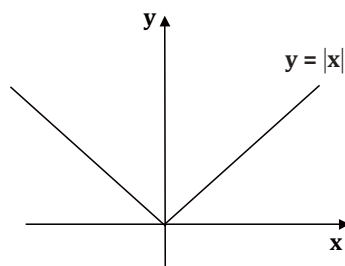
e) Funzione valore assoluto

$$y = |x|$$

dominio $X = \mathbb{R};$

codominio $Y = [0, +\infty[.$

$$y = |x| \Rightarrow \begin{cases} y = x & \text{per } x \geq 0 \\ y = -x & \text{per } x < 0 \end{cases}$$



Si noti che:

- f è strettamente crescente per $x \geq 0$;
- f è strettamente decrescente per $x < 0$;
- f è pari.

8 I limiti di funzioni

Il *punto limite* o di *accumulazione* x_0 del campo d'esistenza di una funzione è quel punto tale che, in ogni suo intorno, per quanto piccolo, cadono sempre infiniti punti del campo di esistenza; tale punto può anche non appartenere al campo di esistenza considerato.

Si dice che l è il limite della funzione $y = f(x)$ per x tendente ad x_0 e si scrive:

$$\lim_{x \rightarrow x_0} f(x) = l$$

quando in corrispondenza di un numero positivo ε , fissato a piacere, è possibile determinare un *intorno completo* di x_0 [1], tale che, per ogni x di tale intorno, escluso eventualmente x_0 , risulti soddisfatta la disequazione:

$$|f(x) - l| < \varepsilon$$

In tal caso si dice anche che $y = f(x)$ si avvicina a l mentre x si avvicina a x_0 .

Una funzione $y = f(x)$, per x tendente a x_0 , si dice:

- *convergente*, se ha per limite un valore finito l ;
- *divergente*, se ha per limite l'infinito positivo o negativo;
- *indeterminata*, se non ha limite.

Il *criterio generale di convergenza di Cauchy* afferma che *condizione necessaria e sufficiente affinché una funzione $y = f(x)$ ammetta limite finito per x tendente ad x_0 è che per ogni numero positivo ε , piccolo a piacere, si possa trovare un intorno di x_0 tale che per due suoi punti x_1 e x_2 distinti da x_0 , si abbia:*

$$|f(x_1) - f(x_2)| < \varepsilon$$

9 I limiti destro e sinistro

Spesso occorre precisare il modo di tendere della x al valore x_0 : infatti la x può tendere a tale valore sia da sinistra che da destra.

Si dice che l è il limite sinistro (destro) della funzione $y = f(x)$ per x tendente da sinistra (destra) ad un valore x_0 , e si scrive, rispettivamente:

$$\lim_{x \rightarrow x_0^-} f(x) = l \quad \text{e} \quad \lim_{x \rightarrow x_0^+} f(x) = l$$

se, scelto un numero positivo ε , si può determinare, nel campo di esistenza della funzione, rispettivamente, un intorno:

$$(x_0 - \delta, x_0) \quad \text{e} \quad (x_0, x_0 + \delta)$$

tale che per ogni x appartenente a questo intorno, diverso da x_0 , si abbia:

$$|f(x) - l| < \varepsilon$$

Per un dato intorno, il limite destro può essere diverso dal limite sinistro e, inoltre, può anche esistere uno solo dei due limiti.

Si può dimostrare che se i limiti destro, l_1 , e sinistro, l_2 , esistono e sono uguali, cioè:

$$l_1 = l_2 = l$$

allora l è il limite della funzione in esame (per $x \rightarrow x_0$). Viceversa, se una funzione ha per limite l , per $x \rightarrow x_0$, allora i limiti destro e sinistro esistono e sono uguali.

10 Funzioni, limiti e infinito

Talvolta, al tendere di x ad x_0 , la funzione $y = f(x)$ aumenta o diminuisce illimitatamente. In tal caso si dice che la funzione ha per limite l'infinito (positivo o negativo), per x tendente ad x_0 , e si scrive:

$$\lim_{x \rightarrow x_0} f(x) = \infty$$

[1] Si chiama *intorno completo* di un numero reale (o di un punto) c un qualsiasi intervallo aperto che contenga c . Se si designa con δ il raggio dell'intorno, allora tale intorno si indica con $(c - \delta; c + \delta)$. Esso risulta essere l'insieme degli $x \in \mathbb{R}$ tali che $|x - c| < \delta$, con $\delta > 0$.

se, scelto un numero positivo P , comunque grande, si può determinare, nel campo di esistenza della funzione, un intorno di x_0 tale che per ogni x appartenente a tale intorno, diverso da x_0 , si abbia:

$$|f(x)| > P$$

Al tendere di x stesso all'infinito, la funzione $y = f(x)$ può avere per limite un numero finito l o può avere per limite l'infinito.

Si dice che la funzione ha per limite l per x tendente all'infinito, e si scrive:

$$\lim_{x \rightarrow \infty} f(x) = l$$

se, scelto un numero positivo ε , si può determinare un numero positivo P tale che per ogni x verificante la condizione:

$$|x| > P$$

si abbia:

$$|f(x) - l| < \varepsilon$$

Specificamente, se quest'ultima relazione è verificata solo per $x < -P$ o solo per $x > P$, si afferma che esistono, rispettivamente, i seguenti limiti:

$$\lim_{x \rightarrow -\infty} f(x) = l \quad \text{e} \quad \lim_{x \rightarrow +\infty} f(x) = l$$

Al tendere di x all'infinito, la funzione può avere per limite l'infinito; a questo punto, scelto un numero positivo P , comunque grande, si può determinare un numero positivo N tale che:

— se per $x > N$ si ha che $f(x) < -P$, oppure $f(x) > P$, allora si dice che esistono, rispettivamente, i seguenti limiti:

$$\lim_{x \rightarrow +\infty} f(x) = -\infty \quad \text{e} \quad \lim_{x \rightarrow +\infty} f(x) = +\infty$$

— se per $x < -N$ si ha che $f(x) < -P$, oppure $f(x) > P$, allora si dice che esistono, rispettivamente, i seguenti limiti:

$$\lim_{x \rightarrow -\infty} f(x) = -\infty \quad \text{e} \quad \lim_{x \rightarrow -\infty} f(x) = +\infty$$

11 I teoremi sui limiti di funzioni

TEOREMA SULL'UNICITÀ DEL LIMITE • Se esiste il limite della funzione $y = f(x)$, per x tendente ad x_0 , tale limite è unico.

Il teorema si dimostra per assurdo. Si supponga che, per $x \rightarrow x_0$, la funzione ammetta due limiti finiti l_1 ed l_2 , e sia $l_1 < l_2$. Allora, per definizione di limite, per ogni $\varepsilon > 0$, scelto ad arbitrio, devono esistere due intorni di x_0 (chiamiamoli H_1 e H_2), per ogni x dei quali, escluso al più x_0 , si abbia rispettivamente:

$$l - \varepsilon < f(x) < l + \varepsilon \quad \forall x \in H_1 (x \neq x_0) \quad \text{e} \quad m - \varepsilon < f(x) < m + \varepsilon \quad \forall x \in H_2 (x \neq x_0)$$

Nell'intorno $H = H_1 \cap H_2$, le relazioni soprascritte devono valere simultaneamente. Per ε abbastanza piccolo perché risulti $l + \varepsilon < m - \varepsilon$, dovrebbe risultare:

$$f(x) < l + \varepsilon < m - \varepsilon < f(x)$$

il che è assurdo. Ciò prova l'unicità del limite.

TEOREMA DEL CONFRONTO • Se $f(x)$, $g(x)$ e $z(x)$ sono tre funzioni definite nello stesso intervallo, eccettuato al più un punto x_0 di questo, e se per ogni x risulta $f(x) \leq g(x) \leq z(x)$, e se, inoltre, è $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} z(x) = l$, si conclude che $\lim_{x \rightarrow x_0} g(x) = l$.

Per definizione di limite, si ha:

$$|f(x) - l| < \varepsilon \quad \text{e} \quad |z(x) - l| < \varepsilon$$

ovvero:

$$l - \varepsilon < f(x) < l + \varepsilon \quad \text{e} \quad l - \varepsilon < z(x) < l + \varepsilon$$

cioè:

$$l - \varepsilon < f(x) \leq z(x) < l + \varepsilon$$

Siccome, per definizione, si ha:

$$f(x) \leq g(x) \leq z(x)$$

deve essere anche:

$$l - \varepsilon < f(x) \leq g(x) \leq z(x) < l + \varepsilon$$

ossia:

$$l - \varepsilon < g(x) < l + \varepsilon$$

e quindi:

$$|g(x) - l| < \varepsilon$$

12 Le operazioni sui limiti di funzioni

Siano date due funzioni $f(x)$ e $g(x)$, definite in un intervallo comune e tali che:

$$\lim_{x \rightarrow x_0} f(x) = l_1 \quad \text{e} \quad \lim_{x \rightarrow x_0} g(x) = l_2$$

Si considerino, adesso, le seguenti operazioni.

ADDIZIONE • Il limite di una somma di funzioni convergenti è uguale alla somma dei limiti delle singole funzioni, cioè:

$$\lim_{x \rightarrow x_0} [f(x) + g(x)] = \lim_{x \rightarrow x_0} f(x) + \lim_{x \rightarrow x_0} g(x) = l_1 + l_2$$

SOTTRAZIONE • Il limite della differenza di funzioni convergenti è uguale alla differenza dei limiti delle singole funzioni, ovvero:

$$\lim_{x \rightarrow x_0} [f(x) - g(x)] = \lim_{x \rightarrow x_0} f(x) - \lim_{x \rightarrow x_0} g(x) = l_1 - l_2$$

MULTIPLICAZIONE • Il limite del prodotto di funzioni convergenti è uguale al prodotto dei limiti delle singole funzioni, vale a dire:

$$\lim_{x \rightarrow x_0} [f(x) \cdot g(x)] = \lim_{x \rightarrow x_0} f(x) \cdot \lim_{x \rightarrow x_0} g(x) = l_1 \cdot l_2$$

Tale limite può essere:

- ∞ , se una delle funzioni ha per limite ∞ e il modulo dell'altra funzione, in un dato intorno di x_0 , è maggiore di una quantità positiva fissa, o comunque se il limite di quest'altra funzione non è nullo;
- zero, se una delle due funzioni ha per limite zero e l'altra un limite diverso da zero.

In particolare, *il limite del prodotto di una costante per una funzione è uguale al prodotto della costante per il limite della funzione*, ossia:

$$\lim_{x \rightarrow x_0} k \cdot f(x) = k \cdot \lim_{x \rightarrow x_0} f(x) = k \cdot l_1$$

DIVISIONE • *Il limite del rapporto di funzioni convergenti è uguale al rapporto dei limiti dei termini, sempre che il limite del denominatore sia diverso da zero*. Traducendo in formule, si ha:

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow x_0} f(x)}{\lim_{x \rightarrow x_0} g(x)} = \frac{l_1}{l_2}$$

RECIPROCO • *Il limite del reciproco di una funzione è uguale al reciproco del limite (finito e non nullo) della funzione data*. Traducendo in formule, si ha:

$$\lim_{x \rightarrow x_0} \frac{1}{f(x)} = \frac{1}{\lim_{x \rightarrow x_0} f(x)} = \frac{1}{l_1}$$

Tale limite è:

- $\pm\infty$, se la funzione tende a zero per valori positivi o negativi, rispettivamente;
- zero, se il limite della funzione $f(x)$ è l'infinito.

POTENZA • *Il limite della potenza di una funzione è uguale alla potenza del limite della funzione*. Traducendo in formule, si ha:

$$\lim_{x \rightarrow x_0} [f(x)]^n = [\lim_{x \rightarrow x_0} f(x)]^n$$

Tali operazioni non possono essere effettuate indistintamente, in quanto talvolta possono dare luogo ad una delle seguenti *forme indeterminate*: $\infty - \infty$, $0 \cdot \infty$, $\frac{0}{0}$, $\frac{\infty}{\infty}$, 0^0 , 1^∞ , ∞^0 .

13 Il confronto di infinitesimi e di infiniti

Secondo il limite cui tendono, le funzioni si chiamano:

- *infinitesimi* per $x \rightarrow x_0$ (o per $x \rightarrow \infty$) quando risulta: $\lim_{x \rightarrow x_0} f(x) = 0$ oppure $\lim_{x \rightarrow \infty} f(x) = 0$;
- *infiniti* per $x \rightarrow x_0$ (o per $x \rightarrow \infty$) quando risulta: $\lim_{x \rightarrow x_0} f(x) = \infty$ oppure $\lim_{x \rightarrow \infty} f(x) = \infty$.

Può accadere che, calcolando il limite del rapporto di due funzioni $f(x)$ e $g(x)$, ci si trovi di fronte al caso in cui entrambe abbiano per limite l'infinito o per limite lo zero, ossia si presenti una delle seguenti *forme indeterminate*:

$$\frac{0}{0} \quad \text{oppure} \quad \frac{\infty}{\infty}$$

In tali situazioni, nulla si può dire sul valore del limite del rapporto, tuttavia, applicando diversi artifici o il *Teorema di L'Hôpital* (che vedremo nel capitolo seguente), è possibile calcolare il limite del rapporto in questione.

Si consideri la forma indeterminata $\frac{0}{0}$. A seconda che il limite del rapporto sia uguale a zero, a infinito o a una quantità fissa diversa da zero, si dice, rispettivamente, che l'infinitesimo del numeratore è di *ordine superiore*, *inferiore* o *dello stesso ordine* dell'infinitesimo del denominatore; ciò per sottolineare il fatto che $f(x)$ è più, meno o ugualmente veloce di $g(x)$ nel tendere a zero.

Si consideri, ora, la forma indeterminata $\frac{\infty}{\infty}$. A seconda che il limite del rapporto sia uguale a infinito, a zero o a una quantità fissa diversa da zero, si dice, rispettivamente, che l'infinito del numeratore è di *ordine superiore*, *inferiore* o *dello stesso ordine* dell'infinito del denominatore; ciò per sottolineare il fatto che $f(x)$ è più, meno o ugualmente veloce di $g(x)$ nel tendere a infinito.

14 Le funzioni continue

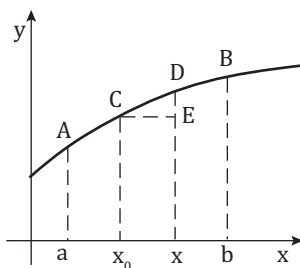
Una funzione $y = f(x)$, definita in un intervallo $[a, b]$, si dice *continua in un punto* x_0 di detto intervallo se il valore $f(x_0)$ che assume in x_0 è il limite a cui tende la funzione per $x \rightarrow x_0$. Traducendo in formula, si ha:

$$\lim_{x \rightarrow x_0} f(x) = f(x_0)$$

Se detta condizione è verificata solo in un intorno sinistro $(x_0 - \delta, x_0)$ o solo in un intorno destro $(x_0, x_0 + \delta)$ di x_0 , si dice che *la funzione è continua in x_0 solo a sinistra o solo a destra*.

Inoltre, si dice che *una funzione è continua in un intervallo* quando è continua in ogni punto dell'intervallo.

Considerando la figura, si dice che la funzione $y = f(x)$, definita nell'intervallo $[a, b]$, è continua in x_0 quando ad ogni piccola variazione di x dal valore x_0 corrisponde una piccola variazione $\overline{ED}$ di $f(x)$ da $f(x_0)$.



15 Le funzioni discontinue

Una funzione $y = f(x)$, definita in un intervallo (a, b) , si dice *discontinua in un punto* di detto intervallo se il valore $f(x_0)$ che assume in x_0 è diverso dal limite a cui tende la funzione per $x \rightarrow x_0$. Traducendo in formula si ha:

$$\lim_{x \rightarrow x_0} f(x) \neq f(x_0)$$

In tal caso, si dice che la funzione presenta in x_0 un *punto di discontinuità* o *punto singolare*.

I punti di discontinuità si classificano in tre specie.

- **Punti di discontinuità di prima specie**

Si dice che nel punto x_0 la funzione $y = f(x)$ ha una *discontinuità di prima specie* se in tale punto esistono finiti il limite destro e quello sinistro, ed essi sono diversi tra loro; la differenza tra i due limiti si dice *salto* della funzione.

• Punti di discontinuità di seconda specie

Si dice che nel punto x_0 la funzione $y = f(x)$ ha una *discontinuità di seconda specie* se in tale punto uno almeno dei due limiti destro e sinistro non esiste, oppure uno almeno dei due limiti destro e sinistro è l'infinito.

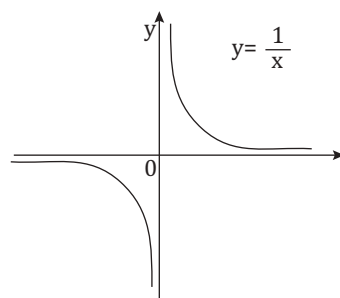
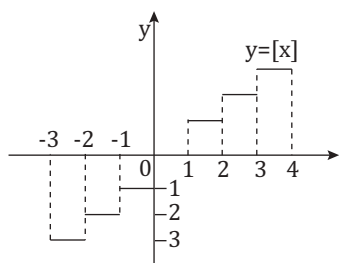
• Punti di discontinuità eliminabile o di terza specie

Si dice che nel punto x_0 la funzione $y = f(x)$ ha una *discontinuità di terza specie* se in tale punto esiste ed è finito il limite della funzione per $x \rightarrow x_0$, ma è diverso dal valore $f(x_0)$ che può anche non esistere.

In questo caso, si dice che la discontinuità è *eliminabile*, in quanto si assume come valore della funzione in x_0 il valore del limite per $x \rightarrow x_0$; si pone cioè: $f(x_0) = \lim_{x \rightarrow x_0} f(x)$.

L'esempio classico di *discontinuità di terza specie* è dato dalla funzione $y = \frac{\sin x}{x}$; il valore della funzione, nel punto $x = 0$, non esiste, in quanto assume il valore indeterminato $\frac{0}{0}$. In tale punto, però, esiste finito il limite, che è pari ad 1. La funzione, pertanto, è prolungabile per continuità in 0, punto in cui assume valore 1.

Nelle figure seguenti sono indicati due diversi tipi di funzioni che presentano, rispettivamente, *discontinuità di prima e di seconda specie*.



La prima è la funzione $y = [x]$, definita dalla legge:

$$y = \begin{cases} x, & \text{se } x \text{ è un numero intero} \\ \text{all'intero precedente la } x, & \text{se } x \text{ non è intero} \end{cases}$$

La funzione ha *discontinuità di prima specie* in tutti i punti aventi per ascissa un numero intero, in cui il salto è uguale a 1.

La seconda è la funzione $y = \frac{1}{x}$, che presenta una *discontinuità di seconda specie* nel punto avente per ascissa 0, infatti:

$$\lim_{x \rightarrow 0^-} \frac{1}{x} = -\infty \quad \text{e} \quad \lim_{x \rightarrow 0^+} \frac{1}{x} = +\infty$$

16 I teoremi sulle funzioni continue

I. TEOREMA • La somma, la differenza e il prodotto di più funzioni continue in un punto x_0 sono funzioni continue.

- II. TEOREMA DELLE FUNZIONI ELEMENTARI** • *Tutte le funzioni elementari, e cioè le funzioni razionali e irrazionali, goniometriche, esponenziali e logaritmiche, sono continue in tutti i punti del campo di esistenza in cui sono definite.*
- III. TEOREMA DELLA PERMANENZA DEL SEGNO** • *Se una funzione $f(x)$ è continua in x_0 ed è $f(x_0) \neq 0$, in un intorno convenientemente piccolo di x_0 la funzione conserva il segno di $f(x_0)$.*
- IV. TEOREMA DI WEIERSTRASS** • *Se una funzione $f(x)$ è continua in un intervallo $[a, b]$, estremi inclusi, ammette in esso un massimo e un minimo.*
- V. TEOREMA DELL'ESISTENZA DEGLI ZERI** • *Se una funzione $f(x)$ è continua in un intervallo $[a, b]$, estremi inclusi, e negli estremi a e b assume valori di segno opposto, esiste almeno un punto interno ad $[a, b]$ in cui la funzione si annulla.*
- VI. TEOREMA DI BOLZANO** • *Se una funzione $f(x)$ è continua in un intervallo $[a, b]$, estremi inclusi, assume in esso ogni valore compreso tra il suo massimo e il suo minimo.*

Ci limitiamo a citare, ora, la seguente, importante proprietà.

PROPRIETÀ • *Se una funzione $f(x)$ è continua in un intervallo $[a, b]$, estremi inclusi, è in esso limitata.*



Espansione Web

Le funzioni polinomiali

Il calcolo differenziale per funzioni di una variabile reale

1 La derivata

Sia $y = f(x)$ una funzione definita in un intervallo $[a, b]$, e siano x e $x + h$ due punti di detto intervallo; si definisce:

— *incremento della variabile indipendente* la quantità:

$$\Delta x = h$$

— *incremento della variabile dipendente* la variazione positiva della funzione dal punto avente ascissa x al punto avente ascissa $x + h$:

$$\Delta y = f(x + h) - f(x)$$

— *rapporto incrementale* il rapporto tra l'incremento della funzione e l'incremento della variabile indipendente:

$$\frac{\Delta y}{\Delta x} = \frac{f(x + h) - f(x)}{h}$$

Si dice *derivata della funzione* $y = f(x)$ il limite, se esiste ed è finito, del rapporto incrementale, per Δx che tende a zero:

$$f'(x) = y' = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{h}$$

2 Le derivate destra e sinistra

È possibile che, pur non esistendo il limite per $h \rightarrow 0$ del rapporto incrementale di cui al precedente paragrafo, esistano, e siano finiti, il limite destro e/o il limite sinistro. Si parla in tal caso, rispettivamente, di *derivata destra* e *derivata sinistra*, e si indicano in questo modo:

$$f'(x)^+ = \lim_{h \rightarrow 0^+} \frac{f(x + h) - f(x)}{h} \quad \text{e} \quad f'(x)^- = \lim_{h \rightarrow 0^-} \frac{f(x + h) - f(x)}{h}$$

Una funzione si dice *derivabile in tutto l'intervallo in cui è definita* se ammette la derivata in ogni punto di detto intervallo.

Inoltre, *ogni funzione* $f(x)$ *derivabile in un punto* x *è ivi continua*; non è vero il contrario, ossia non tutte le funzioni continue in un intervallo sono derivabili in ogni punto dell'intervallo. Infatti, la derivata destra potrebbe essere diversa dalla derivata sinistra in punti che prendono il nome di *punti angolosi*.

La funzione:

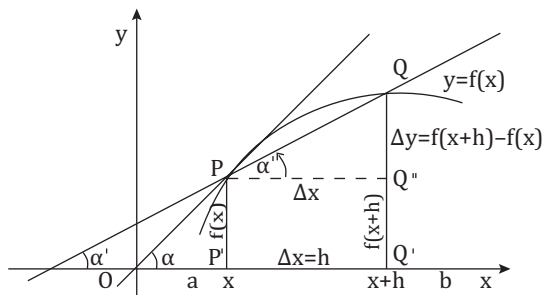
$$y = |x - 5|$$

è continua nel punto $x = 5$, tuttavia non è ivi derivabile; infatti, la derivata destra e sinistra sono, rispettivamente, uguali a $+1$ e -1 .

3 Il significato geometrico della derivata

Il diagramma della funzione $y = f(x)$ sia rappresentato dalla curva tracciata nella figura seguente.

La derivata $f'(x)$ è il coefficiente angolare della retta tangente alla curva $y = f(x)$ nel punto di ascissa x .



Il rapporto incrementale:

$$\frac{\Delta y}{\Delta x} = \frac{f(x+h) - f(x)}{h} = \frac{\overline{PQ'}}{\overline{PQ''}} = \operatorname{tg} \alpha'$$

essendo il triangolo $PQ''Q'$ rettangolo in Q'' , rappresenta la tangente trigonometrica dell'angolo α' individuato dall'asse delle ascisse e dalla retta passante per i punti P e Q , aventi ascissa, rispettivamente, x e $x+h$.

Se h tende a zero, il punto Q tende al punto P , per cui la secante PQ ruota intorno a P e tende alla posizione limite della tangente alla curva nel punto P . Di conseguenza, l'angolo α' tende ad assumere il valore α , e il coefficiente angolare della retta secante tende al coefficiente angolare della retta tangente in P . Pertanto, derivata $f'(x)$ è proprio il coefficiente angolare della retta tangente alla curva nel punto di ascissa x :

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \operatorname{tg} \alpha$$

Se $f'(x) = 0$ si ha che $\operatorname{tg} \alpha = 0$, per cui la tangente nel punto di ascissa x è parallela all'asse x .

Se esistono e sono diverse tra loro le derivate destra e sinistra, la curva ammette in P due tangenti diverse.

4 Il differenziale

La derivata di una funzione $y = f(x)$ in un dato intervallo è:

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$$

Posto:

$$\varepsilon = \frac{\Delta y}{\Delta x} - f'(x) \quad (1)$$

si ha:

$$\lim_{\Delta x \rightarrow 0} \varepsilon = 0$$

cioè ε è un infinitesimo per $\Delta x \rightarrow 0$.

Dalla (1) si ricava:

$$\Delta y = f'(x)\Delta x + \varepsilon\Delta x$$

Poiché, in generale, è $f'(x) \neq 0$, il prodotto $f'(x)\Delta x$ è, per ogni x costante e Δx variabile, un infinitesimo dello stesso ordine di Δx , per $\Delta x \rightarrow 0$. Invece, $\varepsilon\Delta x$ è sempre un infinitesimo d'ordine superiore rispetto a Δx , perché è:

$$\lim_{\Delta x \rightarrow 0} \frac{\varepsilon\Delta x}{\Delta x} = \lim_{\Delta x \rightarrow 0} \varepsilon = 0$$

L'incremento della funzione Δy è dato, quindi, dalla somma di una parte principale $f'(x)\Delta x$ e di un infinitesimo $\varepsilon\Delta x$.

La parte principale si chiama *differenziale della funzione* $y = f(x)$ ed è uguale al prodotto della derivata della funzione per l'incremento arbitrario della variabile indipendente; si indica in questo modo:

$$dy = f'(x)\Delta x$$

Se $f(x) = x$, si ha $f'(x) = 1$, e quindi:

$$dx = 1 \cdot \Delta x = \Delta x$$

Pertanto, la derivata può essere vista come rapporto di differenziali; infatti, si ha:

$$\frac{dy}{dx} = f'(x)$$

Graficamente, il differenziale di una funzione non è altro che l'incremento dell'ordinata del punto $P[x, f(x)]$ della tangente al diagramma della funzione (nel punto P stesso), quando si passa da tale punto a quello di ascissa $x + \Delta x$.

5 Le regole di derivazione

Siano date due funzioni $f(x)$ e $g(x)$, definite in un intervallo comune e derivabili in quei punti dell'intervallo cui si fa riferimento.

I. TEOREMA • La derivata di una somma di funzioni è uguale alla somma delle derivate delle singole funzioni. In formule, si ha:

$$y = f(x) + g(x) \Rightarrow y' = f'(x) + g'(x)$$

II. TEOREMA • La derivata di una differenza di funzioni è uguale alla differenza delle derivate delle singole funzioni. In formule, si ha:

$$y = f(x) - g(x) \Rightarrow y' = f'(x) - g'(x)$$

III. TEOREMA • La derivata di un prodotto di due funzioni è uguale alla somma dei prodotti di ciascuna funzione per la derivata dell'altra. In formule, si ha:

$$y = f(x) \cdot g(x) \Rightarrow y' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

In particolare, la derivata del prodotto di una costante per una funzione è uguale al prodotto della costante per la derivata della funzione. In formule, si ha:

$$y = kf(x) \Rightarrow y' = kf'(x)$$

IV. TEOREMA • La derivata di un prodotto di più di due funzioni è uguale alla somma dei prodotti della derivata di ciascuna funzione per tutte le altre funzioni non derivate. Date, per esempio, tre funzioni, $f(x)$, $g(x)$ e $z(x)$, si ha:

$$y = f(x) \cdot g(x) \cdot z(x) \Rightarrow y' = f'(x) \cdot g(x) \cdot z(x) + f(x) \cdot g'(x) \cdot z(x) + f(x) \cdot g(x) \cdot z'(x)$$

V. TEOREMA • La derivata del quoziente di due funzioni è uguale a una frazione avente al denominatore il quadrato del denominatore e al numeratore la differenza fra il prodotto del denominatore per la derivata del numeratore e il prodotto del numeratore per la derivata del denominatore. In formule, si ha:

$$y = \frac{f(x)}{g(x)} \Rightarrow y' = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$$

VI. TEOREMA • La derivata della potenza n^{ma} di una funzione $f(x)$ è uguale al prodotto di $n[f(x)]^{n-1}$ per la derivata della funzione. In formule, si ha:

$$y = [f(x)]^n \Rightarrow y' = n[f(x)]^{n-1}f'(x)$$

Tale regola vale, oltre che per le potenze ad esponente intero positivo n , anche per le potenze aventi come esponente un numero reale qualunque.

ESEMPIO 1

Calcolare la derivata della funzione:

$$y = \sin x \cdot \cos x$$

Ricordando la formula di duplicazione del coseno, la derivata prima della funzione data è:

$$y' = \cos^2 x - \sin^2 x = \cos 2x$$

6 Le derivate di funzioni composte e di funzioni inverse

Sia $z = g(x)$ una funzione reale definita in un intervallo I e derivabile nel punto $x_0 \in I$, e sia $y = f(z)$ una funzione reale definita in un intervallo I' , tale che $g(I) \subseteq I'$, e derivabile nel punto $z_0 = g(x_0)$. Allora la funzione composta $y = f[g(x)]$ è derivabile in x_0 , e si ha:

$$y'(x_0) = f'(z_0) \cdot g'(x_0)$$

Volendo indicare la derivata in un punto generico, possiamo scrivere:

$$y'(x) = f'(z) \cdot g'(x)$$

In pratica, per derivare una funzione di funzione, questa si deve anzitutto derivare rispetto alla variabile da cui dipende direttamente e poi moltiplicare per la derivata di quest'ultima rispetto alla variabile indipendente. Questo teorema si estende anche al caso in cui le funzioni componenti siano più di due.

È noto che la funzione $y = f(x)$ si dice *invertibile* in un intervallo chiuso $[a, b]$ se ad ogni valore della x corrisponde uno e un solo valore di y in $[a', b']$, dove a' e b' sono il minimo e il massimo della funzione nell'intervallo $[a, b]$, e, viceversa, se ad ogni valore di y in $[a', b']$ corrisponde uno e un solo valore di x in $[a, b]$. In altre parole, la funzione è invertibile se è continua in detto intervallo.

Se la funzione $y = f(x)$ è derivabile, con derivata non nulla nell'intervallo $[a, b]$, la sua *funzione inversa* $x = g(y)$ è derivabile, con derivata non nulla, in $[a', b']$, e la sua derivata, per ogni coppia di valori corrispondenti x e y , è uguale al reciproco della derivata della funzione data.

In formule, si ha:

$$g'(y) = \frac{1}{f'(x)}$$

ESEMPIO 2

Calcolare la derivata prima della funzione:

$$y = e^{\sqrt{\sin x}}$$

Applicando la regola di derivazione di funzioni composte e le regole di derivazione delle funzioni componenti la funzione data, si ha che la derivata in questione è uguale a:

$$y' = e^{\sqrt{\sin x}} \cdot D\sqrt{\sin x} = e^{\sqrt{\sin x}} \cdot \frac{1}{2\sqrt{\sin x}} \cdot D\sin x = e^{\sqrt{\sin x}} \cdot \frac{\cos x}{2\sqrt{\sin x}}$$

7 Le derivate di ordine superiore

Se una funzione $f(x)$ è derivabile in un intervallo (a, b) e la funzione $f'(x)$, che rappresenta la derivata di $f(x)$ in (a, b) , è a sua volta derivabile in (a, b) , si dice *derivata seconda* di $f(x)$ in (a, b) la derivata di $f'(x)$. Essa si indica con uno dei seguenti simboli:

$$y'', \quad f''(x), \quad D^2f(x)$$

La derivata della derivata seconda è detta *derivata terza* e così via. Più generalmente, così proseguendo, si giunge al concetto di derivata n -esima:

$$D^{(n)}f(x) = D[D^{(n-1)}f(x)]$$

Le derivate di ordine superiore al primo si calcolano applicando in maniera replicata le regole di derivazione precedentemente trattate.

8 I teoremi sulle derivate

TEOREMA DI LAGRANGE O DEL VALORE MEDIO • Se una funzione $y = f(x)$ è continua in un intervallo chiuso $[a, b]$ e derivabile nei punti interni di tale intervallo, allora esiste almeno un punto x_0 interno ad $[a, b]$, tale che:

$$\frac{f(b) - f(a)}{b - a} = f'(x_0)$$

Geometricamente, il teorema appena enunciato si traduce nell'esistenza, nell'arco di linea rappresentante la funzione, di un punto nel quale la tangente è parallela alla corda che congiunge i punti estremi dell'arco.

COROLLARI

- I. Se in tutto l'intervallo $[a, b]$ si ha $f'(x) = 0$, allora la funzione è costante nell'intervallo $[a, b]$.
- II. Due funzioni $f(x)$ e $g(x)$, derivabili e aventi uguale derivata in tutti i punti dell'intervallo $[a, b]$, differiscono tra loro di una costante.
- III. Se in un intervallo $[a, b]$ la derivata $f'(x)$ è sempre positiva (negativa), allora la funzione è, in tale intervallo, crescente (decrescente).

TEOREMA DI ROLLE • Se una funzione $y = f(x)$, continua in un intervallo chiuso $[a, b]$ e derivabile nei punti interni di tale intervallo, assume valori uguali agli estremi dell'intervallo, esiste almeno un punto x_0 interno ad $[a, b]$, tale che $f'(x_0) = 0$.

Geometricamente, il teorema appena enunciato si traduce nell'esistenza, nell'arco di linea rappresentante la funzione, di un punto nel quale la tangente è parallela all'asse delle x .

TEOREMA DI CAUCHY O DEGLI INCREMENTI FINITI • Se $f(x)$ e $g(x)$ sono funzioni continue in un intervallo chiuso $[a, b]$, derivabili nei punti interni di tale intervallo, e la derivata $g'(x)$ non si annulla mai in tale intervallo, esiste almeno un punto x_0 , interno ad $[a, b]$, tale che il rapporto degli incrementi finiti delle due funzioni è uguale al rapporto dei valori delle loro derivate nel punto x_0 . In formule, si ha:

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(x_0)}{g'(x_0)}$$

9 Il Teorema di L'Hôpital

Il Teorema di L'Hôpital si enuncia nel seguente modo:

Siano $f(x)$ e $g(x)$ due funzioni con le seguenti proprietà:

- entrambe derivabili in un intorno di x_0 , escluso al più x_0 , con x_0 numero reale qualsiasi oppure $x_0 = \infty$;
- con $g'(x)$ mai nulla in detto intorno;
- per $x \rightarrow x_0$ entrambe tendono a 0 o a ∞ , per cui il limite del rapporto di $f(x)$ e $g(x)$ assume una delle forme indeterminate $\frac{0}{0}$ oppure $\frac{\infty}{\infty}$.

In tali ipotesi, se esiste il $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$, sussiste la relazione:

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$$

Qualora anche il limite del rapporto delle derivate prime si presenti sotto forma indeterminata, ma per $f'(x)$ e $g'(x)$ valgano le stesse proprietà di $f(x)$ e $g(x)$, si può provare a calcolare il $\lim_{x \rightarrow x_0} \frac{f''(x)}{g''(x)}$, dove $f''(x)$ e $g''(x)$ sono le derivate seconde di $f(x)$ e $g(x)$.

La regola enunciata si applica, ovviamente, anche in presenza del solo limite destro o del solo limite sinistro.

Talvolta anche i limiti che si presentano sotto una delle forme indeterminate $+\infty - \infty$, $0 \cdot \infty$, 0^0 , 1^∞ e ∞^0 possono valutarsi attraverso la loro trasformazione nelle due forme indeterminate considerate, e calcolarsi, quindi, mediante l'applicazione del Teorema di L'Hôpital.

10 Le relazioni tra derivate e funzioni crescenti o decrescenti

Riprendiamo ora un concetto già esaminato, quello della crescita o della decrescenza di una funzione $y = f(x)$. Questa volta, però, lo esaminiamo alla luce dello studio delle derivate.

Sia data una funzione $y = f(x)$, definita in un dato intervallo $[a, b]$, e sia x_0 un punto di detto intervallo. Per ogni h positivo e sufficientemente piccolo, la funzione si dice:

- *crescente in x_0* , se $f(x_0 - h) < f(x_0) < f(x_0 + h)$;
- *costante in x_0* , se $f(x_0 - h) = f(x_0) = f(x_0 + h)$;
- *decrescente in x_0* , se $f(x_0 - h) > f(x_0) > f(x_0 + h)$.

Inoltre, la funzione $y = f(x)$ è:

- *crescente* in ogni punto dell'intervallo $[a, b]$, se la sua derivata $f'(x)$ è sempre positiva in tale intervallo;
- *costante* in ogni punto dell'intervallo $[a, b]$, se la sua derivata $f'(x)$ è sempre uguale a zero in tale intervallo;
- *decrescente* in ogni punto dell'intervallo $[a, b]$, se la sua derivata $f'(x)$ è sempre negativa in tale intervallo.

Da ciò discende la seguente regola: *per determinare gli intervalli in cui una funzione è crescente, costante o decrescente si deve discutere la funzione $f'(x)$, vale a dire bisogna calcolarne gli zeri e definire gli intervalli in cui essa è, rispettivamente, positiva, nulla o negativa.*

11 Massimi e minimi

Sia data una funzione $y = f(x)$, definita in un intervallo $[a, b]$, e sia x_0 un punto di detto intervallo. Se in un intorno sufficientemente piccolo di x_0 i valori della funzione sono tutti non maggiori di $f(x_0)$, si dice che la funzione presenta, nel punto x_0 , un *massimo relativo*.

Il punto x_0 è, invece, un punto di *minimo relativo* se in un intorno sufficientemente piccolo di x_0 i valori della funzione sono tutti non minori di $f(x_0)$.

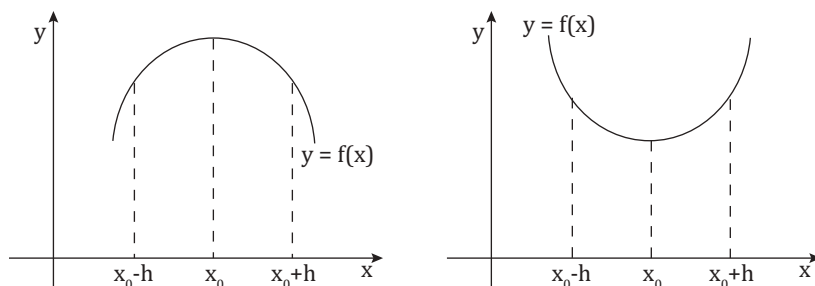
Al fine di determinare i punti di massimo e minimo relativo di una funzione $y = f(x)$, valgono i seguenti teoremi:

PRIMO TEOREMA • Se il punto x_0 interno all'intervallo $[a, b]$, è un punto di massimo o di minimo relativo per la funzione $y = f(x)$, e se in tale punto la funzione è derivabile, si ha necessariamente $f'(x_0) = 0$.

Il teorema inverso non vale, ossia in un punto la derivata prima della funzione può essere nulla senza che in quel punto si abbia un massimo o un minimo relativo. Geometricamente ciò vuol dire che in quel punto la tangente alla curva è parallela all'asse x ; il punto in questione si chiama *flesso a tangente orizzontale*.

SECONDO TEOREMA • Se la funzione $y = f(x)$ è finita e continua in un intervallo $[a, b]$, come pure lo sono le sue derivate successive, e se $f'(x_0) = 0$, si dice che il punto P di ascissa x_0 è:

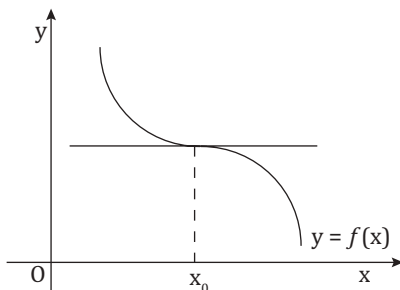
- un punto di massimo relativo proprio, se $f''(x_0) < 0$;
- un punto di minimo relativo proprio, se $f''(x_0) > 0$;
- un punto di flesso orizzontale, se $f''(x_0) = 0$ e $f'''(x_0) \neq 0$.



In generale, se una funzione $y = f(x)$ è finita e continua in un intervallo $[a, b]$, come pure lo sono le sue derivate successive, e se in un punto x_0 di detto intervallo sono nulle tutte le derivate successive fino a quella di ordine $n - 1$, mentre è diversa da zero la derivata di ordine n , si ha:

- se n è pari, in x_0 la funzione presenta un massimo relativo o un minimo relativo a seconda che la derivata di ordine n sia negativa o positiva;

- se n è dispari, in x_0 la funzione presenta un punto di flesso orizzontale crescente o decrescente a seconda che la derivata di ordine n sia positiva o negativa.

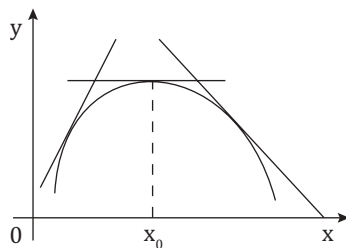
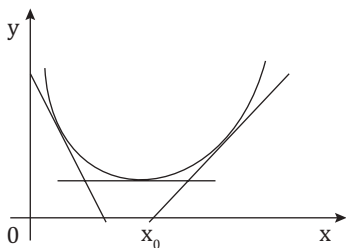


I punti di *massimo assoluto* o di *minimo assoluto* di una funzione definita e continua in un intervallo $[a, b]$ e ivi derivabile si trovano calcolando le ordinate dei punti di massimo e minimo relativi, nonché confrontando queste con le ordinate degli estremi dell'intervallo $[a, b]$; il massimo assoluto e il minimo assoluto sono, rispettivamente, il più grande e il più piccolo di tali valori.

12 Le concavità di una curva

Sia data una funzione $y=f(x)$, definita in un intervallo $[a, b]$ insieme alle sue derivate prima seconda e terza. Si dice che in un punto P , di ascissa x_0 , la curva volge la concavità verso l'alto o verso il basso se, dato un intorno completo di x_0 , per ogni x di tale intorno diverso da x_0 le ordinate dei punti della curva sono, rispettivamente, maggiori o minori di quelle della tangente in P , corrispondenti alle medesime ascisse. In x_0 la curva ha concavità verso l'alto o verso il basso a seconda che $f''(x_0) > 0$ o $f''(x_0) < 0$.

Si dice che in un punto F la curva ha un flesso o un punto di inflessione quando in esso cambia il verso della sua concavità e la tangente in detto punto attraversa la curva, e se $f''(x_0) = 0$ e $f'''(x_0) \neq 0$.



In generale, se una funzione $y=f(x)$ è finita e continua in un intervallo $[a, b]$, come pure lo sono le sue derivate successive, e se in un punto x_0 di detto intervallo sono nulle tutte le derivate successive fino a quella di ordine $n-1$, mentre è diversa da zero la derivata di ordine n , si ha:

- se n è pari, in x_0 la curva volge la concavità verso l'alto o verso il basso a seconda che la derivata di ordine n sia positiva o negativa;
- se n è dispari, in x_0 la curva ha un flesso.

13 Gli asintoti

Una retta r è detta *asintoto* della curva $y=f(x)$ se la distanza di un punto $P(x, y)$ di questa curva dalla retta tende a zero quando il punto indicato si allontana indefinitamente sulla curva, rispetto all'origine delle coordinate (cioè quando almeno una delle coordinate del punto considerato tende all'infinito).

Gli asintoti ad una curva si chiamano:

- *verticali*, se sono paralleli all'asse y ;
- *orizzontali*, se sono paralleli all'asse x ;
- *obliqui*, gli altri.

La curva di equazione $y = f(x)$ ha per:

- *asintoto verticale* la retta di equazione $x = x_0$ se:

$$\lim_{x \rightarrow x_0} f(x) = +\infty \quad \text{oppure} \quad \lim_{x \rightarrow x_0} f(x) = -\infty \quad (\text{o entrambi})$$

- *asintoto orizzontale*:

- la retta di equazione $y = y_0$ se:

$$\lim_{x \rightarrow +\infty} f(x) = y_0 \quad \text{oppure} \quad \lim_{x \rightarrow -\infty} f(x) = y_0$$

- le rette di equazione $y = h$ e $y = k$ se:

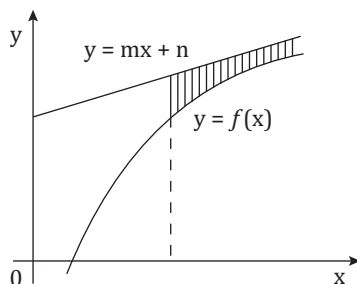
$$\lim_{x \rightarrow +\infty} f(x) = h \quad \text{oppure} \quad \lim_{x \rightarrow -\infty} f(x) = k$$

- *asintoto obliquo* la retta di equazione $y = mx + n$ se:

$$\lim_{x \rightarrow +\infty} [mx + n - f(x)] = 0 \quad \text{oppure} \quad \lim_{x \rightarrow -\infty} [mx + n - f(x)] = 0$$

Il coefficiente angolare m e l'intercetta n della retta asintotica sono, rispettivamente:

$$m = \lim_{x \rightarrow \infty} \frac{f(x)}{x} \quad \text{e} \quad n = \lim_{x \rightarrow \infty} [f(x) - mx]$$



14 Lo studio del grafico di una funzione

Per studiare le principali proprietà e descrivere una curva avente equazione $y = f(x)$ è utile procedere secondo il seguente ordine:

- 1) determinare gli intervalli in cui essa è definita e continua, e quindi il suo comportamento agli estremi degli intervalli;
- 2) determinare gli eventuali punti di intersezione della curva con gli assi coordinati;
- 3) ricercare gli eventuali asintoti;
- 4) ricercare le eventuali simmetrie rispetto all'asse x o all'asse y , oppure rispetto all'origine degli assi;
- 5) determinare gli intervalli in cui la funzione è positiva o negativa;
- 6) determinare gli intervalli in cui è crescente o decrescente;
- 7) determinare i massimi e i minimi;
- 8) determinare gli intervalli in cui volge la concavità verso l'alto o verso il basso;
- 9) determinare i punti di flesso.

Calcolo integrale per funzioni di una variabile reale ed elementi di teoria della misura

1 L'integrale indefinito

Premettendo che per *primitiva* di una funzione $f(x)$, nell'intervallo $[a, b]$, si intende una funzione $F(x)$ – derivabile in ogni punto di questo intervallo – la cui derivata è $f(x)$, si dice *integrale indefinito di una funzione continua reale $f(x)$* la totalità delle funzioni primitive della $f(x)$, e si indica in questo modo:

$$\int f(x) dx$$

Se $F(x)$ è un integrale indefinito di $f(x)$ lo sarà anche $F(x) + c$, dove c è una costante cui si attribuisce qualunque valore reale, e si ha:

$$\int f(x) dx = F(x) + c$$

Inoltre:

- l'integrale indefinito del prodotto di una costante k per una funzione continua è uguale al prodotto della costante per l'integrale indefinito della funzione. In formule, si ha:

$$\int kf(x) dx = k \int f(x) dx$$

- l'integrale indefinito di una somma di funzioni continue è uguale alla somma degli integrali indefiniti delle singole funzioni. Siano $f(x)$ e $g(x)$ due funzioni; in tal caso, si ha:

$$\int [f(x) + g(x)] dx = \int f(x) dx + \int g(x) dx$$

L'operazione di integrazione indefinita si manifesta, quindi, come l'operazione inversa della derivazione; ma mentre l'operazione di derivazione non si applica a tutte le funzioni continue, l'operazione di integrazione può essere applicata a tutte le funzioni continue.

ESEMPIO 1

Calcolare l'integrale indefinito:

$$\int \frac{\sin \sqrt{x}}{\sqrt{x}} dx$$

La funzione in esame è una funzione composta da una funzione trigonometrica e da una funzione irrazionale; pertanto, l'integrale è di facile soluzione considerando che:

$$y = \cos f(x) \Rightarrow y' = -\sin f(x) \cdot f'(x)$$

in cui:

$$f(x) = \sqrt{x} \Rightarrow f'(x) = \frac{1}{2\sqrt{x}}$$

Pertanto:

$$\int \frac{\operatorname{sen} \sqrt{x}}{\sqrt{x}} dx = -2 \cos \sqrt{x} + c$$

ESEMPIO 2

Calcolare l'integrale indefinito:

$$\int \frac{x-2}{x+2} dx$$

L'integrale si risolve applicando alcuni semplici artifici matematici. Esso, infatti, può essere riscritto nella seguente maniera:

$$\int \frac{x-2}{x+2} dx = \int \frac{x-2+2-2}{x+2} dx = \int \frac{x+2}{x+2} dx - \int \frac{4}{x+2} dx = \int dx - 4 \int \frac{1}{x+2} dx = x - 4 \log |x+2| + c$$

ESEMPIO 3

Calcolare l'integrale indefinito:

$$\int \frac{1}{x^2 + x + 1} dx$$

L'integrale dato si risolve considerando che:

$$x^2 + x + 1 = x^2 + x + \frac{1}{4} - \frac{1}{4} + 1 = \frac{3}{4} + \left(x + \frac{1}{2}\right)^2$$

Pertanto, si ha:

$$\begin{aligned} \int \frac{1}{x^2 + x + 1} dx &= \int \frac{1}{\frac{3}{4} + \left(x + \frac{1}{2}\right)^2} dx = \int \frac{1}{\frac{3}{4} + \left(\frac{2x+1}{2}\right)^2} dx = \frac{4}{3} \int \frac{1}{1 + \left(\frac{2x+1}{\sqrt{3}}\right)^2} dx = \\ &= \frac{4}{3} \cdot \frac{\sqrt{3}}{2} \operatorname{arctg}\left(\frac{2x+1}{\sqrt{3}}\right) + c = \frac{2}{\sqrt{3}} \operatorname{arctg}\left(\frac{2x+1}{\sqrt{3}}\right) + c \end{aligned}$$

Nei tre paragrafi successivi saranno trattati i **tre metodi di integrazione**.

2 L'integrazione per sostituzione

Il *metodo di integrazione per sostituzione* consiste nel trasformare la variabile di integrazione x , di un integrale indefinito $\int f(x) dx$, in un'altra variabile, in modo da ottenere una nuova funzione di più agevole integrazione.

Sia $\varphi(t)$ una funzione di t derivabile e invertibile; si pone, quindi:

$$x = \varphi(t)$$

[e diciamo $t = h(x)$ la funzione inversa] il cui differenziale è:

$$dx = \varphi'(t) dt$$

Pertanto, si ha:

$$\int f(x) dx = \int f[\varphi(t)] \varphi'(t) dt$$

Una volta calcolato l'integrale in questo modo ottenuto, si sostituisce alla t il suo valore $t = h(x)$: la funzione che ne risulta è l'integrale indefinito proposto.

ESEMPIO 4

Calcolare l'integrale indefinito:

$$\int \frac{e^x}{1 - e^{2x}} dx$$

L'integrale si risolve applicando il metodo di sostituzione della variabile. Si pone:

$$e^x = t \Rightarrow x = \log t \Rightarrow dx = \frac{1}{t} dt$$

Pertanto, l'integrale considerato si trasforma nell'integrale di facile soluzione:

$$\int \frac{t}{1 - t^2} \frac{1}{t} dt = \int \frac{1}{1 - t^2} dt = \frac{1}{2} \log \left| \frac{1+t}{1-t} \right| + c$$

Per cui:

$$\int \frac{e^x}{1 - e^{2x}} dx = \frac{1}{2} \log \left| \frac{1 + e^x}{1 - e^x} \right| + c$$

3 L'integrazione per decomposizione

L'integrale indefinito di una funzione continua non può essere determinato ricorrendo ad un metodo generale. Spesso si adoperano degli artifici che consentono di riportare gli integrali dati a degli integrali indefiniti immediati. Si ricorre, quindi, al *metodo di integrazione per decomposizione*, che consiste nel decomporre, appunto, la funzione data nella somma algebrica di funzioni, di cui siano noti o, in ogni caso, di facile calcolo gli integrali indefiniti.

Data la frazione $\frac{f(x)}{g(x)}$, si possono presentare diversi casi:

- *il denominatore presenta radici reali distinte*. Siano $x_1, x_2, \dots, x_n$ le radici del denominatore $g(x)$ e siano $A, B, \dots, N$ i coefficienti da determinare; si pone:

$$\frac{f(x)}{g(x)} = \frac{A}{x - x_1} + \frac{B}{x - x_2} + \dots + \frac{N}{x - x_n}$$

Per determinare tali coefficienti è sufficiente liberarsi dei denominatori e applicare il *principio di identità dei polinomi*;

- *il denominatore presenta radici complesse*. La scomposizione si esegue normalmente, solo che nella frazione elementare, la quale presenta un fattore del tipo $(x^2 + p)$, dove p è un numero intero positivo, si pone al numeratore un binomio del tipo $Ax + B$. Si procede poi normalmente, liberandosi dei denominatori;
- *la frazione ha il numeratore di grado più alto di quello del denominatore*. Per scomporre tali frazioni si esegue la divisione. Siano $q(x)$ il quoziente ed $r(x)$ il resto della divisione. Si pone, quindi:

$$\frac{f(x)}{g(x)} = q(x) + \frac{r(x)}{g(x)}$$

- *il denominatore presenta radici anche multiple*. Se il denominatore $g(x)$ ha per radice x_1 di grado n , si pone:

$$\frac{f(x)}{g(x)} = \frac{A}{(x - x_1)^n} + \frac{B}{(x - x_1)^{n-1}} + \dots + \frac{N}{x - x_1}$$

Qualora le radici multiple del denominatore fossero più di una, la scomposizione sarebbe analoga. Infatti, la frazione si scomporrebbe in un numero di frazioni pari alla molteplicità delle radici del polinomio al denominatore.

ESEMPIO 5

Calcolare l'integrale indefinito:

$$\int \frac{1}{2x^2 - 5x + 3} dx$$

Il polinomio al denominatore della frazione è scomponibile nel seguente modo:

$$2x^2 - 5x + 3 = (2x - 3)(x - 1)$$

Pertanto, la funzione razionale in esame può essere decomposta in frazioni elementari. Si ha:

$$\frac{1}{(2x - 3)(x - 1)} = \frac{A}{2x - 3} + \frac{B}{x - 1}$$

in cui A e B sono costanti da determinare. Liberando dal denominatore, si ha:

$$1 = (x - 1)A + (2x - 3)B = (A + 2B)x - A - 3B$$

Per cui deve, necessariamente, essere:

$$\begin{cases} A + 2B = 0 \\ -A - 3B = 1 \end{cases}$$

da cui si ricava facilmente che $A = 2$ e $B = -1$.

Sostituendo ad A e B i valori appena ottenuti, si ha:

$$\frac{1}{(2x - 3)(x - 1)} = \frac{2}{2x - 3} - \frac{1}{x - 1}$$

L'integrale si risolve, così, facilmente:

$$\int \frac{1}{2x^2 - 5x + 3} dx = \int \frac{1}{(2x - 3)(x - 1)} dx = \int \frac{2}{2x - 3} dx - \int \frac{1}{x - 1} dx = \log \left| \frac{2x - 3}{x - 1} \right| + c$$

ESEMPIO 6

Calcolare l'integrale indefinito:

$$\int \frac{9x^3 + x - 3}{9x^2 + 1} dx$$

Siccome nella funzione integranda il numeratore è di grado più alto del denominatore, si effettua la divisione e si ha:

$$\frac{9x^3 + x - 3}{9x^2 + 1} = x - \frac{3}{9x^2 + 1}$$

Considerando che $9x^2 + 1 = 1 + (3x)^2$, l'integrale dato si risolve facilmente:

$$\int \frac{9x^3 + x - 3}{9x^2 + 1} dx = \int x dx - \int \frac{3}{1 + (3x)^2} dx = \frac{1}{2}x^2 - \arctg 3x + c$$

4 L'integrazione per parti

Siano f e g due funzioni della x , continue e dotate di derivata continua. Si ha:

$$d(fg) = fdg + gdf, \quad \text{ossia: } d(fg) = fg'dx + gf'dx$$

e quindi:

$$fg'dx = d(fg) - gf'dx$$

Integrando ambo i membri di quest'ultima uguaglianza, si ottiene:

$$\int f(x)g'(x)dx = f(x)g(x) - \int g(x)f'(x)dx$$

In questa formula è contenuta la *formula di integrazione per parti*. Essa è applicata tutte le volte che di uno dei due fattori, $g'x$ nel nostro caso, si riesce a determinare una primitiva $g(x)$.

La regola di integrazione per parti si può enunciare nel seguente modo: *l'integrazione del prodotto di un fattore finito f per un fattore differenziale $dg = g'dx$ è uguale al fattore finito per l'integrale del fattore differenziale meno l'integrale del prodotto dell'integrale trovato per il differenziale del fattore finito.*

ESEMPIO 7

Calcolare l'integrale indefinito:

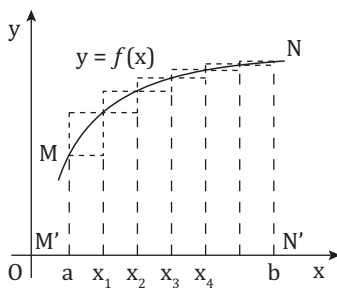
$$\int x \log x \, dx$$

L'integrale dato si risolve con il metodo di integrazione per parti. Assumendo come fattore differenziale $x dx = d \frac{x^2}{2}$, si ha:

$$\int x \log x \, dx = \frac{x^2}{2} \log x - \int \frac{x^2}{2} \cdot \frac{1}{x} dx = \frac{x^2}{2} \log x - \frac{x^2}{4} + c = \frac{x^2}{2} \left(\log x - \frac{1}{2} \right) + c$$

5 L'integrale definito

Sia data una funzione $y = f(x)$, definita e continua, derivabile e a derivata continua in un intervallo $[a, b]$; la si ponga, inoltre, positiva in $[a, b]$. Il concetto di *integrale definito* è illustrato esaminando l'area limitata dalla curva di equazione $y = f(x)$, dall'asse delle x e dalle due parallele all'asse y condotte per gli estremi a e b .



Si divida l'intervallo $[a, b]$ in n parti mediante i punti di ascissa $a, x_1, x_2, \dots, b$ e siano:

- $h_1, h_2, h_3, \dots, h_n$ le ampiezze di ciascun intervallo;
- $m_1, m_2, m_3, \dots, m_n$ i valori minimi che la funzione assume in ciascuno di questi intervalli;
- $M_1, M_2, M_3, \dots, M_n$ i valori massimi che la funzione assume in ciascuno di questi intervalli.

L'area del trapezio mistilineo $MNN'M'$ è compresa tra le somme:

$$s'(h) = m_1 h_1 + m_2 h_2 + \dots + m_n h_n = \sum_{i=1}^n m_i h_i$$

$$s''(h) = M_1 h_1 + M_2 h_2 + \dots + M_n h_n = \sum_{i=1}^n M_i h_i$$

le quali indicano, rispettivamente, la somma delle aree dei rettangoli *inscritti*, contenuti nel trapezoide, e la somma delle aree dei rettangoli *circoscritti*, contenenti il trapezoide. Entrambe variano al variare delle singole lunghezze di ciascun intervallo.

Si dimostra che:

$$\lim_{h \rightarrow 0} s'(h) = \lim_{h \rightarrow 0} s''(h)$$

Tale limite è proprio uguale all'estremo superiore dei valori assunti da $s'(h)$ e all'estremo inferiore dei valori assunti da $s''(h)$.

Si operi una nuova suddivisione dell'intervallo $[a, b]$ e siano Δx_i e $f(\chi_i)$, rispettivamente, la lunghezza di uno degli intervallini in cui si è suddiviso l'intervallo $[a, b]$ e il valore che la funzione assume in un punto χ_i dell'intervallo Δx_i ; al tendere di Δx a zero, la somma:

$$\sum_{i=1}^n f(\chi_i) \Delta x_i$$

tende al limite comune di $s'(h)$ e $s''(h)$, e si ha:

$$\int_a^b f(x) dx = \lim_{\Delta x \rightarrow 0} \sum_{i=1}^n f(\chi_i) \Delta x_i$$

Il limite dato è uguale all'*integrale definito* della funzione continua $f(x)$ estesa all'intervallo $[a, b]$, dove a e b sono detti *estremi di integrazione* (rispettivamente *inferiore* e *superiore*) e indicano gli estremi entro cui varia la x (detta *variabile di integrazione*). L'integrale definito, essendo un numero, non dipende assolutamente dalla variabile x .

Si può affermare, quindi, che l'*integrale definito* è la *somma di infiniti termini infinitesimi*, rappresentati da tanti rettangoli di altezza $f(x)$ e base infinitesima dx .

Prima di esporre le proprietà dell'integrale definito, premettiamo le seguenti definizioni:

- se si *invertono gli estremi di integrazione*, il valore dell'*integrale cambia segno*. Infatti, posto $a < b$, si ha:

$$\int_a^b f(x) dx = - \int_b^a f(x) dx$$

- se l'*estremo superiore d'integrazione* è uguale all'*estremo inferiore*, il valore dell'*integrale* è *nullo* e si ha:

$$\int_a^a f(x) dx = 0$$

6 Le proprietà dell'integrale definito

- L'*integrale definito di un prodotto di una costante per una funzione* è uguale al prodotto della costante per l'*integrale della funzione*. In formule, si ha:

$$\int_a^b k f(x) dx = k \int_a^b f(x) dx$$

- Siano $f(x)$ e $g(x)$ due funzioni integrabili in $[a, b]$; l'*integrale della loro somma* è uguale alla *somma degli integrali delle singole funzioni*. In formule, si ha:

$$\int_a^b [f(x) + g(x)] dx = \int_a^b f(x) dx + \int_a^b g(x) dx$$

— Sia c un punto dell'intervallo $[a, b]$ in cui la funzione è continua; si ha:

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

Si consideri, ora, l'enunciato del teorema che segue.

TEOREMA DELLA MEDIA • L'integrale definito di una funzione continua $f(x)$ è uguale all'ampiezza $(b - a)$ dell'intervallo di integrazione, moltiplicata per il valore che la funzione integranda assume in un punto x_0 del suddetto intervallo. In formule, si ha:

$$\int_a^b f(x)dx = (b - a)f(x_0)$$

7 La relazione tra integrale indefinito e integrale definito

Sia data una funzione $y = f(x)$, limitata e integrabile in un intervallo $[a, b]$, e sia x un punto di detto intervallo. La funzione data è integrabile nell'intervallo $[a, x]$ e, indicando con t la variabile di integrazione, per non confonderla con la variabile indipendente, si ha:

$$F(x) = \int_a^x f(t)dt$$

Quindi $F(x)$ è una funzione di x nell'intervallo $[a, b]$ ed è detta *funzione integrale* di $f(x)$.

Da quanto detto dimostriamo il successivo, importante teorema.

TEOREMA FONDAMENTALE DEL CALCOLO INTEGRALE O TEOREMA DI TORRICELLI • Se la funzione integranda $f(x)$ è continua, esiste la derivata della funzione integrale $F(x)$ nel punto x ed è uguale al valore che la funzione integranda assume nello stesso punto, ossia:

$$F'(x) = f(x)$$

L'incremento della $F(x)$ relativo al punto x e all'incremento h è:

$$F(x + h) - F(x) = \int_a^{x+h} f(t)dt - \int_a^x f(t)dt$$

ossia, in base alle proprietà degli integrali definiti, è:

$$F(x + h) - F(x) = \int_a^x f(t)dt + \int_x^{x+h} f(t)dt - \int_a^x f(t)dt$$

cioè:

$$F(x + h) - F(x) = \int_x^{x+h} f(t)dt$$

Sia x_0 un conveniente punto dell'intervallo $[x, x + h]$. Applicando il *teorema della media*, si ha:

$$F(x + h) - F(x) = hf(x_0)$$

da cui:

$$\frac{F(x + h) - F(x)}{h} = f(x_0)$$

Se si fa tendere h a zero, anche x_0 tende a x , per cui si ha:

$$\lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = \lim_{x_0 \rightarrow x} f(x_0) = f(x)$$

da cui si ricava quanto si voleva dimostrare, ossia:

$$F'(x) = f(x)$$

Il teorema appena dimostrato consente il calcolo dell'integrale definito di una funzione per mezzo dell'integrale indefinito della funzione stessa. Infatti, se $G(x)$ è una qualsiasi primitiva della funzione $f(x)$, si ha:

$$G'(x) = f(x)$$

Siccome $F(x)$ è la funzione integrale di $f(x)$, vuol dire che $F(x)$ e $G(x)$ differiscono tra loro per una costante; cioè:

$$G(x) = F(x) + c$$

ossia, in base al teorema appena dimostrato:

$$G(x) = \int_a^x f(t) dt + c$$

in cui, ponendo $x = a$ e considerando che $\int_a^a f(t) dt = 0$, si ha:

$$G(a) = c$$

Pertanto, si ha:

$$G(x) = \int_a^x f(t) dt + G(a)$$

da cui si ricava:

$$\int_a^x f(t) dt = G(x) - G(a)$$

Per $x = b$, si ha quindi:

$$\int_a^b f(t) dt = [G(x)]_a^b = G(b) - G(a)$$

Da quanto esposto, scaturisce la regola seguente.

REGOLA • Una volta calcolata una primitiva $G(x)$ di una funzione $f(x)$, l'integrale definito tra a e b della $f(x)$ è uguale alla differenza tra il valore assunto da $G(x)$ nell'estremo superiore b e il valore assunto da $G(x)$ nell'estremo inferiore a .

8 Le aree di superfici

Geometricamente, l'integrale definito di una funzione f continua in $[a, b]$ rappresenta l'area di un trapezoide, delimitato dalla curva di equazione $y = f(x)$, dall'asse x e dalle parallele all'asse y di equazioni $x = a$ e $x = b$. Tuttavia, nel calcolare tale area, la semplice applicazione di questo risultato non è sufficiente ma bisogna considerare diverse situazioni.

- La curva può non giacere nel primo o nel secondo quadrante, ma nel terzo o nel quarto. In tali casi la funzione assume valori negativi, per cui è negativo anche l'integrale definito, del quale bisogna considerare il valore assoluto.
- La curva può giacere in parte nella regione delle ordinate positive e in parte in quella delle ordinate negative. In tale caso si sommano i valori assoluti degli integrali relativi ai due intervalli in cui la funzione è, rispettivamente, positiva e negativa.
- L'area di superficie che interessa è quella limitata da due curve e da rette parallele all'asse y ; bisogna fare attenzione a verificare, nel caso concreto, quale curva limiti superiormente o inferiormente l'area interessata.

ESEMPIO 8

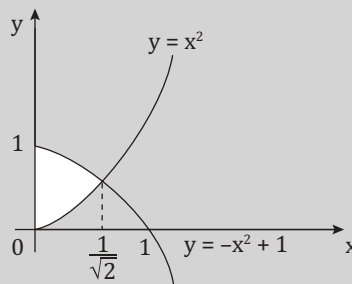
Calcolare l'area della regione di piano limitata dalle due parabole aventi equazioni:

$$y = x^2 \quad \text{e} \quad y = -x^2 + 1$$

nel semipiano delle ascisse positive.

Per calcolare l'area delimitata dalle due curve è necessario trovare le ascisse dei punti di intersezione delle due parabole. Pertanto, risolvendo il sistema formato dalle due equazioni, si ha: $x = \pm \frac{1}{\sqrt{2}}$. Siccome interessa solo l'area nel semipiano delle ascisse positive, l'intervallo al quale bisogna applicare il calcolo integrale è $\left[0, \frac{1}{\sqrt{2}}\right]$; dunque, si ha:

$$\begin{aligned} S &= \int_0^{\frac{1}{\sqrt{2}}} [(-x^2 + 1) - x^2] dx = \int_0^{\frac{1}{\sqrt{2}}} (1 - 2x^2) dx = \left[x - \frac{2}{3} x^3 \right]_0^{\frac{1}{\sqrt{2}}} = \\ &= \frac{1}{\sqrt{2}} - \frac{2}{3(\sqrt{2})^3} = \frac{1}{\sqrt{2}} - \frac{2}{6\sqrt{2}} = \frac{\sqrt{2}}{3} \end{aligned}$$

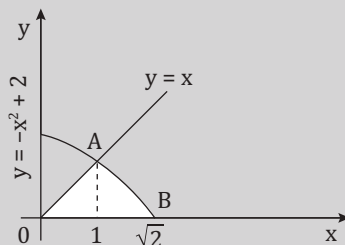
**ESEMPIO 9**

Calcolare l'area della regione di piano limitata dalla parabola di equazione $y = -x^2 + 2$, dalla retta di equazione $y = x$ (bisettrice del primo e del terzo quadrante) e dall'asse x , nel semipiano delle ascisse positive.

Per calcolare l'area della regione di piano considerata è necessario determinare, nel semipiano delle ascisse positive, il punto di intersezione della parabola con la retta data e l'ascissa del punto in cui la parabola interseca l'asse x . Il primo si ottiene risolvendo il sistema formato dalle due equazioni ed è dato dal punto $A(1, 1)$; la seconda si ottiene ponendo, nell'equazione della parabola, $y = 0$, ed è uguale a $\sqrt{2}$. Da quest'ultima ascissa si ottiene il punto $B(\sqrt{2}, 0)$.

Pertanto, l'intervallo al quale bisogna applicare il calcolo integrale è $[0, \sqrt{2}]$; inoltre la regione risulta essere divisa in due parti, di cui la prima è limitata superiormente dalla retta, dall'origine al punto A , e la seconda dalla parabola da A a B . Di conseguenza, l'area cercata è data da:

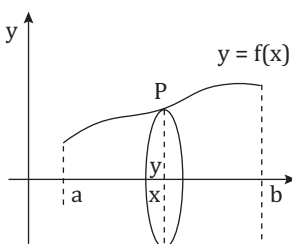
$$\begin{aligned} S &= \int_0^1 x dx + \int_1^{\sqrt{2}} (-x^2 + 2) dx = \left[\frac{1}{2} x^2 \right]_0^1 + \left[-\frac{1}{3} x^3 + 2x \right]_1^{\sqrt{2}} = \\ &= \frac{1}{2} + \left(-\frac{2\sqrt{2}}{3} + 2\sqrt{2} + \frac{1}{3} - 2 \right) = \frac{1}{2} + \frac{4\sqrt{2}}{3} - \frac{5}{3} = \frac{4\sqrt{2}}{3} - \frac{7}{6} \end{aligned}$$



9 I volumi dei solidi di rotazione

Per determinare il volume di un solido di rotazione attraverso l'applicazione del calcolo integrale, si applica un ragionamento analogo a quello visto per la determinazione dell'area di un trapezoide.

Sia data una funzione $y = f(x)$ continua in un intervallo $[a, b]$ e si supponga che essa sia non negativa e crescente in tutto $[a, b]$. Si consideri il trapezoide delimitato dalla curva di equazione $y = f(x)$, dall'asse delle x e dalle due parallele all'asse y condotte per gli estremi a e b . Ruotando di un giro completo tale trapezoide intorno all'asse x , detto *asse di rotazione*, si ottiene un solido di rotazione.



Ogni punto del trapezoide, durante la rotazione, genera una circonferenza, per cui la sezione del solido con un piano perpendicolare all'asse di rotazione, in un punto qualunque dell'intervallo $[a, b]$, è un cerchio avente per raggio l'ordinata della curva nel punto considerato. Di conseguenza, l'area di una qualunque sezione del solido intorno all'asse x è data da:

$$S = \pi [f(x)]^2$$

Si dimostra che il *volume del solido di rotazione* intorno all'asse delle ascisse è dato da:

$$V = \pi \int_a^b [f(x)]^2 dx$$

Pertanto, nota l'equazione della curva $f(x)$, è possibile calcolare il volume del solido di rotazione applicando la formula appena data. Si considerino, ora, alcuni solidi di rotazione.

— Il volume del cono generato dalla rotazione, attorno all'asse x , della retta avente equazione:

$$y = \frac{r}{h}x$$

è dato da:

$$V = \pi \int_0^h \frac{r^2}{h^2} x^2 dx = \pi \left[\frac{r^2}{h^2} \frac{x^3}{3} \right]_0^h = \pi \frac{r^2}{h^2} \frac{h^3}{3} = \frac{1}{3} \pi r^2 h$$

— Il volume della sfera generata dalla rotazione, attorno all'asse x , della semicirconferenza avente equazione:

$$y = \sqrt{r^2 - x^2}$$

è dato da:

$$V = \pi \int_{-r}^r (r^2 - x^2) dx = \pi \left[r^2 x - \frac{x^3}{3} \right]_{-r}^r = \pi \left[r^3 - \frac{r^3}{3} + r^3 - \frac{r^3}{3} \right] = \frac{4}{3} \pi r^3$$

- Il volume dell'ellissoide allungato generato dalla rotazione, attorno al suo asse maggiore, dell'ellisse avente equazione:

$$y = \frac{b}{a} \sqrt{a^2 - x^2}$$

è dato da:

$$V = \pi \int_{-a}^a \frac{b^2}{a^2} (a^2 - x^2) dx = \pi \frac{b^2}{a^2} \left[a^2 x - \frac{x^3}{3} \right]_{-a}^a = \pi \frac{b^2}{a^2} \left[a^3 - \frac{a^3}{3} + a^3 - \frac{a^3}{3} \right] = \frac{4}{3} \pi ab^2$$

- Il volume del paraboloide generato dalla rotazione, attorno all'asse x , del ramo di parabola avente equazione:

$$y = \sqrt{2px}$$

è dato da:

$$V = \pi \int_0^a 2px dx = \pi \left[2p \frac{x^2}{2} \right]_0^a = \pi pa^2$$

ESEMPIO 10

Calcolare il volume del solido generato dalla rotazione completa, intorno all'asse x , della superficie limitata dalla curva di equazione:

$$y = \sqrt{\operatorname{tg} x}$$

nell'intervallo $\left[0, \frac{\pi}{4}\right]$.

Il volume del solido di rotazione considerato è dato da:

$$V = \pi \int_0^{\frac{\pi}{4}} (\sqrt{\operatorname{tg} x})^2 dx = \pi \int_0^{\frac{\pi}{4}} \operatorname{tg} x dx = \pi \left[-\log \cos x \right]_0^{\frac{\pi}{4}} = \pi \left(-\log \frac{\sqrt{2}}{2} \right) = \pi \left(-\log 2^{\frac{1}{2}} \right) = \frac{\pi}{2} \log 2$$

ESEMPIO 11

Calcolare il volume del solido generato dalla rotazione completa, intorno all'asse x , della superficie limitata dalla parabola di equazione:

$$y = -x^2 + 4$$

La curva di equazione data interseca l'asse delle ascisse nei punti -2 e $+2$; pertanto, il volume del solido di rotazione è dato da:

$$\begin{aligned} V &= \pi \int_{-2}^2 (-x^2 + 4)^2 dx = \pi \int_{-2}^2 (x^4 - 8x^2 + 16) dx = \pi \left[\frac{1}{5} x^5 - \frac{8}{3} x^3 + 16x \right]_{-2}^2 = \\ &= \pi \left\{ \left(\frac{32}{5} - \frac{64}{3} + 32 \right) - \left(-\frac{32}{5} + \frac{64}{3} - 32 \right) \right\} = \pi \frac{512}{15} \end{aligned}$$

Successioni, serie numeriche ed equazioni differenziali

1 Insiemi e limiti

Come è noto, il concetto matematico di *insieme* è un concetto primitivo, ossia non definibile. Esso indica ogni raggruppamento di oggetti, persone, numeri o cose aventi una proprietà caratteristica comune.

Si dice che il punto c è un *punto limite* o di *accumulazione* per l'insieme I quando, in un intorno di c , per quanto piccolo esso sia, vi cadono infiniti punti di I . Tutti i punti dell'insieme che non sono punti limite si dicono *punti isolati*.

Si consideri l'insieme:

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{n}, \dots$$

Esso ha per estremo superiore il numero 1, che è anche il massimo valore dell'insieme, e per estremo inferiore lo zero, che non appartiene all'insieme ma è un suo punto di accumulazione, in quanto in un intorno di 0 cadono infiniti elementi dell'insieme.

A quanto detto segue un importante teorema.

TEOREMA DI BOLZANO-WEIERSTRASS • Se in un intervallo $[a, b]$ cadono infiniti punti di un insieme I , esiste in esso almeno un punto limite (o di accumulazione) di I .

2 Le successioni

Una *successione* è un insieme di infiniti numeri reali, che si indicano con:

$$a_1, a_2, a_3, \dots, a_n, \dots$$

e che sono in corrispondenza con l'insieme dei numeri naturali.

Il termine a_n si dice *termine generale* della successione, la quale è, talvolta, indicata semplicemente con a_n .

Una successione si dice:

- *crescente*, se $a_1 < a_2 < a_3 < \dots < a_n < \dots$;
- *decrescente*, se $a_1 > a_2 > a_3 > \dots > a_n > \dots$;
- *costante*, se $a_1 = a_2 = a_3 = \dots = a_n = \dots$.

3 I limiti di successioni

Si dice che la successione:

$$a_1, a_2, a_3, \dots, a_n, \dots$$

ha per limite l se, dato un certo numero positivo ε , piccolo a piacere, si può determinare un termine a_n della successione tale che, per esso e per i suoi successivi, si abbia:

$$|l - a_n| < \varepsilon$$

e si scrive:

$$\lim_{n \rightarrow \infty} a_n = l$$

LIMITE ZERO. Si dice che la *successione* ha limite zero e si scrive:

$$\lim_{n \rightarrow \infty} a_n = 0$$

se, dato un certo numero positivo ε , piccolo a piacere, è possibile trovare un elemento a_n tale che per esso e per i suoi successivi si abbia:

$$|a_n| < \varepsilon$$

LIMITE INFINITO. Si dice che la *successione* ha per limite l'infinito positivo (o negativo) e si scrive:

$$\lim_{n \rightarrow \infty} a_n = \pm \infty$$

se i suoi termini (almeno da un certo punto) sono tutti positivi (o negativi) e, scelto un numero P , grande a piacere, è possibile trovare un elemento a_n tale che esso e i suoi successivi siano maggiori, in valore assoluto, di P ; ossia deve risultare:

$$|a_n| > P$$

Una successione si dice:

- *convergente*, se ha per limite un valore finito l ;
- *divergente*, se ha per limite l'infinito positivo o negativo;
- *indeterminata*, se non ha limite.

Il **criterio generale di convergenza di Cauchy** afferma che *condizione necessaria e sufficiente affinché una successione*

$$a_1, a_2, a_3, \dots, a_n, \dots$$

sia convergente è che la differenza, in valore assoluto, tra un suo termine a_n e un successivo a_{n+s} sia più piccola di un dato numero ε , piccolo a piacere, ossia si abbia:

$$|a_{n+s} - a_n| < \varepsilon$$

4 I teoremi sui limiti di successioni

Il **teorema sull'unicità del limite** vale anche per le successioni. Esso si dimostra per assurdo, allo stesso modo del corrispondente teorema sui limiti di funzioni, pertanto per una sua dimostrazione si rimanda a quanto già esposto in precedenza.

Siano date due successioni:

$$a_1, a_2, a_3, \dots, a_n, \dots \quad \text{e} \quad b_1, b_2, b_3, \dots, b_n, \dots$$

Siano:

$$\lim_{n \rightarrow \infty} a_n = l_1 \quad \text{e} \quad \lim_{n \rightarrow \infty} b_n = l_2$$

i limiti rispettivi delle due successioni; si considerino, ora, i teoremi seguenti.

TEOREMA DELLA SOMMA • Il limite di una successione somma di due successioni convergenti è uguale alla somma dei limiti delle singole successioni; per le successioni a_n e b_n si ha:

$$\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n = l_1 + l_2$$

TEOREMA DELLA DIFFERENZA • Il limite di una successione differenza di due successioni convergenti è uguale alla differenza dei limiti delle singole successioni; per le successioni a_n e b_n , si ha:

$$\lim_{n \rightarrow \infty} (a_n - b_n) = \lim_{n \rightarrow \infty} a_n - \lim_{n \rightarrow \infty} b_n = l_1 - l_2$$

TEOREMA DEL PRODOTTO • Il limite di una successione prodotto di successioni convergenti è uguale al prodotto dei limiti delle singole successioni; per le successioni a_n e b_n , si ha:

$$\lim_{n \rightarrow \infty} (a_n \cdot b_n) = \lim_{n \rightarrow \infty} a_n \cdot \lim_{n \rightarrow \infty} b_n = l_1 \cdot l_2$$

TEOREMA DEL QUOZIENTE • Il limite di una successione quoziente di due successioni tendenti rispettivamente a l_1 e l_2 (con $l_2 \neq 0$) è uguale al quoziente dei limiti delle due successioni; per le successioni a_n e b_n , si ha:

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n} = \frac{l_1}{l_2}$$

TEOREMA DEGLI INVERSI O DEI RECIPROCI • Il limite del reciproco di una successione è uguale al reciproco del limite (finito e non nullo) della successione data; per la successione a_n , si ha:

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} = \frac{1}{l_1}$$

Inoltre, se la successione a_n tende al limite 0, la successione degli inversi tende al limite ∞ e viceversa.

5 Le serie numeriche

Sia data una successione di numeri reali:

$$a_1, a_2, \dots, a_n, \dots$$

Si dice *serie associata alla successione* il seguente simbolo:

$$a_1 + a_2 + \dots + a_n + \dots \quad \left(\text{o anche: } \sum_{n=1}^{\infty} a_n \right)$$

in cui i numeri $a_1, a_2, \dots$ si dicono *termini della serie*.

Si ponga:

$$\begin{aligned} S_1 &= a_1; \\ S_2 &= a_1 + a_2; \\ &\dots\dots\dots \\ S_n &= a_1 + a_2 + \dots + a_n. \end{aligned}$$

Le somme $S_1, S_2, \dots, S_n$ sono denominate *ridotte* o *somme parziali della serie*.

S_n è una funzione della variabile intera e positiva n , per cui ha senso calcolarne il limite al tendere di n a $+\infty$.

Se tale limite esiste ed è finito, vale a dire se:

$$\lim_{n \rightarrow \infty} S_n = S$$

si dice che la serie è *convergente*.

Il numero S si dice *valore* o *somma* (di infiniti termini) della serie e si indica:

$$S = a_1 + a_2 + \dots + a_n + \dots$$

Se tale limite esiste ed è infinito, vale a dire se:

$$\lim_{n \rightarrow \infty} S_n = +\infty \quad \text{e} \quad \lim_{n \rightarrow \infty} S_n = -\infty$$

si dice che la serie è *divergente*.

Una serie si dice *indeterminata* quando non è né convergente né divergente.

ESEMPIO 1

La serie dei numeri interi positivi:

$$1 + 2 + 3 + \dots + n + \dots$$

è divergente, infatti:

$$S_n = 1 + 2 + 3 + \dots + n + \dots = \frac{1+n}{2} n$$

da cui:

$$\lim_{n \rightarrow \infty} S_n = +\infty$$

Data la serie:

$$a_1 + a_2 + \dots + a_n + \dots$$

si considerino i seguenti due criteri.

CRITERIO DI CAUCHY O DELLA RADICE • La serie è assolutamente convergente se esiste il limite:

$$\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = l$$

ed è $l < 1$; non è convergente se è $l > 1$; nulla si può affermare se è $l = 1$.

CRITERIO DI D'ALEMBERT O DEL RAPPORTO • Se tutti i termini della serie sono diversi da zero ed esiste il limite:

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = l$$

la serie è assolutamente convergente se è $l < 1$; non è convergente se è $l > 1$; nulla si può affermare se è $l = 1$.

6 Le serie geometriche

Si dice *serie geometrica*, di ragione q , la serie:

$$1 + q + q^2 + \dots + q^n + \dots$$

dove q è un numero reale qualunque.

Sia:

$$S_n = 1 + q + q^2 + \dots + q^{n-1}$$

Se $q \neq 1$, si ha:

$$S_n = \frac{1 - q^n}{1 - q}$$

Si distinguano i seguenti quattro casi.

- Se $-1 < q < +1$, allora:

$$\lim_{n \rightarrow \infty} q^n = 0$$

per cui:

$$\lim_{n \rightarrow \infty} S_n = \frac{1}{1-q}$$

La serie geometrica, in tal caso, è convergente.

- Se $q < -1$ oppure $q > +1$, allora:

$$\lim_{n \rightarrow \infty} q^n = \infty$$

per cui:

$$\lim_{n \rightarrow \infty} S_n = \infty$$

La serie geometrica è, dunque, divergente.

- Se $q = +1$, la serie geometrica è tale che:

$$S_n = n$$

In tal caso la serie è divergente, essendo:

$$\lim_{n \rightarrow \infty} S_n = \infty$$

- Se $q = -1$, la serie geometrica diventa:

$$1 - 1 + 1 - 1 + \dots$$

Essa è indeterminata, essendo:

$$\begin{aligned} S_n &= 0 \text{ se } n \text{ è pari} \\ S_n &= 1 \text{ se } n \text{ è dispari} \end{aligned}$$

Come si può notare, non esiste alcun limite di S_n al tendere di n a ∞ .

7 Le serie di funzioni

Sia data una successione di funzioni:

$$f_1(x), f_2(x), \dots, f_n(x), \dots$$

definite in un intervallo $[a, b]$, e si consideri la *serie di funzioni*:

$$f_1(x) + f_2(x) + \dots + f_n(x) + \dots$$

La serie considerata diviene la serie numerica:

$$f_1(x_0) + f_2(x_0) + \dots + f_n(x_0) + \dots$$

fissando un valore x_0 all'interno dell'intervallo $[a, b]$, in modo che ogni funzione assuma un valore determinato.

Si dice che *la serie di funzioni è convergente, divergente o indeterminata nel punto x_0 se risulta, rispettivamente, convergente, divergente o indeterminata la serie numerica formata con i valori che le funzioni della serie assumono per $x = x_0$.*

8 Le serie di potenze

Fra le serie di funzioni, hanno particolare importanza le *serie di potenze*, le quali hanno la seguente forma:

$$a_0 + a_1x + a_2x^2 + \dots + a_nx^n + \dots$$

dove i coefficienti $a_0, a_1, a_2, \dots$ sono numeri reali e x è una variabile reale.

ESEMPIO 2

Ad esempio, sono serie di potenze le seguenti:

$$\frac{1}{2}x + \frac{2}{3}x^2 + \frac{3}{4}x^3 + \dots + \frac{n}{n+1}x^n + \dots; \quad x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots + \frac{1}{n!}x^n + \dots$$

Il *campo di convergenza* di una serie di potenze è l'insieme dei valori per cui la serie converge.

TEOREMA • Se una serie di potenze converge solo per qualche valore della x , esiste, allora, un numero positivo R tale che la serie è assolutamente convergente per ogni valore della x interno all'intervallo $(-R, R)$, e non converge per i valori esterni a detto intervallo.

Tale intervallo è denominato *intervallo di convergenza della serie di potenze* ed R è il *raggio di convergenza*.

Se una serie di potenze converge solo per $x = 0$, si dice che il suo raggio di convergenza è nullo; se, invece, la serie di potenze converge per qualunque valore della x , si dice che il suo raggio di convergenza è infinito.

TEOREMA • Sia data la serie di potenze:

$$a_0 + a_1x + a_2x^2 + \dots + a_nx^n + \dots$$

Se i suoi coefficienti sono tutti diversi da zero e se esiste il limite:

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \omega$$

allora il raggio di convergenza è nullo se $\omega = +\infty$; è più infinito se $\omega = 0$; è uguale a $1/\omega$ in tutti gli altri casi.

9 La serie di Fourier

Se una funzione $f(x)$ è definita nell'intervallo $[0, 2\pi]$, è possibile esprimerla come somma di una serie del tipo:

$$\begin{aligned} & \frac{1}{2}a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \dots + (a_n \cos nx + b_n \sin nx) + \dots = \\ & = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \end{aligned}$$

detta *serie trigonometrica*. I coefficienti $a_0, a_1, a_2, \dots, a_n, \dots, b_1, b_2, \dots, b_n, \dots$ sono numeri reali e x varia nell'intervallo $[0, 2\pi]$. Le funzioni $(a_1 \cos x + b_1 \sin x), (a_2 \cos 2x + b_2 \sin 2x) \dots$ sono periodiche di periodo 2π .

Supponendo che la funzione sia integrabile nell'intervallo $[0, 2\pi]$, i coefficienti a_n e b_n si determinano ricorrendo alle cosiddette *formule di Eulero-Fourier*:

$$\begin{aligned} a_0 &= \frac{1}{\pi} \int_0^{2\pi} f(x) dx \\ a_1 &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos x dx & b_1 &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin x dx \\ a_2 &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos 2x dx & b_2 &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin 2x dx \\ &\dots\dots\dots \\ a_n &= \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx dx & b_n &= \frac{1}{\pi} \int_0^{2\pi} f(x) \sin nx dx \end{aligned}$$

La serie trigonometrica ottenuta con dette formule si chiama *serie di Fourier relativa alla funzione*, che a sua volta è detta *sviluppatibile in serie di Fourier*.

10 Le equazioni differenziali

Si chiama *equazione differenziale ordinaria, di ordine n* , un'equazione nella quale figura come incognita una funzione y della sola variabile reale x , che stabilisca un legame fra la variabile x , la funzione y e le prime n derivate $y', y'', \dots, y^{(n)}$ di questa.

In generale, un'equazione differenziale di ordine n si presenta nella forma:

$$f(x, y, y', y'', \dots, y^{(n)}) = 0$$

La *soluzione* o l'*integrale* dell'equazione differenziale è una funzione y della variabile x che, con le sue derivate (fino a quella di ordine n), soddisfi l'equazione differenziale.

Un'equazione differenziale si dice:

- *lineare*, se la funzione è un polinomio di primo grado nelle $n + 1$ variabili;
- *di forma normale*, se è risolta rispetto alla derivata di ordine massimo.

L'equazione differenziale più semplice è l'equazione differenziale del primo ordine:

$$y' = f(x)$$

che può scriversi anche sotto la forma:

$$\frac{dy}{dx} = f(x)$$

da cui:

$$dy = f(x) dx$$

Ricorrendo al calcolo integrale, si ha:

$$\int dy = \int f(x) dx$$

vale a dire:

$$y = F(x) + c$$

che è l'*integrale generale* dell'equazione differenziale data, se $F(x)$ è una primitiva della funzione $f(x)$, e se tale funzione è continua, eccetto (al più) un numero dato di punti, nell'intervallo in cui è definita; c è, invece, una costante arbitraria.

Analogamente, si dimostra che la soluzione o l'integrale dell'equazione differenziale del secondo ordine:

$$y'' = f(x)$$

è:

$$y = F_1(x) + cx + c_1$$

dove c e c_1 sono costanti arbitrarie.

Dagli esempi fatti si evince che l'integrazione di un'equazione differenziale ordinaria ammette, quando possibile, infinite soluzioni, ognuna delle quali dipende dai valori delle costanti arbitrarie.

La soluzione si dice *integrale generale* se contiene il numero massimo di costanti arbitrarie, mentre si dicono *integrali particolari* quelle soluzioni che si ottengono assegnando valori specifici alle costanti. Le soluzioni si dicono *integrali singolari* se esse non si ottengono dando valori specifici alle costanti.

11 Tipi di equazioni differenziali

a) Equazioni differenziali a variabili separate

Le equazioni differenziali del tipo:

$$y' = \frac{f(x)}{g(y)}$$

sono equazioni differenziali a variabili separate e il cui integrale generale è:

$$G(y) = F(x) + c$$

in cui $G(y)$ e $F(x)$ sono due primitive, rispettivamente, di $g(y)$ e di $f(x)$.

b) Equazioni differenziali lineari del primo ordine

Le equazioni differenziali del primo ordine si dicono *lineari* quando la funzione incognita y e la sua derivata prima y' non figurano con esponente maggiore di uno.

Esse potranno, quindi, scriversi sotto la forma:

$$y' = f(x)y + g(x)$$

dove $f(x)$ e $g(x)$ sono funzioni continue nell'intervallo di definizione.

L'equazione differenziale si dice *omogenea* se è $g(x) = 0$.

Se è $g(x) \neq 0$, allora si dimostra che l'integrale generale dell'equazione differenziale lineare data è:

$$y = \left[\int g(x) e^{-\int f(x) dx} + c \right] e^{\int f(x) dx}$$

in cui $e^{\int f(x) dx}$ è denominato *fattore integrante*.

c) Equazioni differenziali lineari omogenee del secondo ordine a coefficienti costanti

Si dice *equazione differenziale lineare omogenea del secondo ordine a coefficienti costanti* un'equazione che determini un legame tra una variabile indipendente x , una funzione y della variabile x , nonché la derivata del primo e del secondo ordine della funzione suddetta, e che sia del tipo:

$$y'' + ay' + by = 0$$

dove a e b sono costanti.

L'equazione:

$$r^2 + ar + b = 0$$

si dice *equazione caratteristica* dell'equazione differenziale data.

L'integrale generale dell'equazione differenziale è diverso a seconda delle radici ρ_1 e ρ_2 dell'equazione caratteristica. Infatti, se esse sono:

— *reali e distinte*, l'integrale generale è:

$$y = c_1 e^{\rho_1 x} + c_2 e^{\rho_2 x}$$

— *reali, coincidenti* e uguali a $\rho_1 = \rho_2 = \rho$, l'integrale generale si scrive:

$$y = c_1 e^{\rho x} + c_2 x e^{\rho x}$$

— *complesse, coniugate* e uguali a $\rho_1 = \alpha + i\beta$ e $\rho_2 = \alpha - i\beta$, l'integrale generale, sulla base anche delle formule di Eulero, è:

$$y = c_1 e^{\alpha x} \cos \beta x + c_2 e^{\alpha x} \sin \beta x$$

I processi di approssimazione e di stima degli errori

1 Introduzione

Il *Calcolo numerico* è una parte fondamentale della Matematica Applicata: esso si occupa della predisposizione di *modelli matematici* rappresentativi di dati fenomeni.

Un modello matematico può essere definito come una rappresentazione formale delle conoscenze relative ad un fenomeno; con la sua costruzione non si realizza un'esatta riproduzione del fenomeno studiato, ma se ne fornisce una versione semplificata.

Un modello matematico consiste, tipicamente, in un insieme di equazioni e disequazioni che legano le variabili rilevanti del fenomeno allo studio, e le cui soluzioni si determinano con metodi numerici che si servono degli strumenti di calcolo elettronico.

2 Gli errori

Di un numero reale, che si presenta in un calcolo, si considera solo un numero finito di cifre decimali. Sia a un numero reale e a^* un suo valore approssimato. La differenza:

$$a - a^*$$

si dice *errore*, il cui valore assoluto:

$$\varepsilon = |a - a^*|$$

è detto *errore assoluto* di a^* nell'approssimare a .

Si dice *errore relativo* di a^* la quantità:

$$\varepsilon_r = \frac{|a - a^*|}{|a|}$$

Il numero positivo:

$$\frac{|a|}{|a - a^*|}$$

si dice *grado di precisione* dell'approssimazione mediante a^* del numero reale $a \neq a^*$.

Se il numero reale a appartiene ad un intervallo aperto del tipo:

$$]a_1, a_2[$$

che prende il nome di *intervallo di indeterminazione*, allora le quantità:

$$a - a_1 \quad \text{e} \quad a - a_2$$

sono gli errori che si riferiscono al numero reale a e ai suoi valori approssimati, per difetto (a_1) e per eccesso (a_2).

Le quantità:

$$a - a_1 \quad \text{e} \quad a_2 - a$$

sono i corrispondenti errori assoluti.

a) Tipi di errori

L'errore appena considerato deriva dalla discrepanza esistente tra un modello matematico, che costituisce la semplificazione di un fenomeno, e la realtà. Esso è denominato *errore nei dati iniziali*.

Esistono, inoltre, i seguenti tipi di errori.

• Errore di troncamento

Questo tipo di errore si presenta ogniqualvolta la soluzione di un problema scaturisce da una successione infinita di operazioni e si è pertanto obbligati ad interrompere il calcolo dopo un numero finito di passi.

Questo si verifica quando si sostituisce ad una serie una sua somma parziale. Sia S la somma di una serie e S_n la successione delle sue somme parziali. Dà luogo a un errore di troncamento l'uguaglianza approssimata:

$$S \cong S_n$$

Si consideri, ora, un teorema sulle serie numeriche che si attua nel calcolo con troncamento.

TEOREMA SULLE SERIE ALTERNATE • Se a_n è una successione decrescente di numeri reali non negativi tale che:

$$\lim_{n \rightarrow \infty} a_n = 0$$

allora la serie alternata:

$$a_1 + (-a_2) + a_3 + (-a_4) + \dots + (-1)^{n-1} a_n + \dots$$

è convergente. Sia S la sua somma; per ogni $r \in \mathbb{N}^*$, si ha:

$$\left| S - \sum_{n=1}^r (-1)^{n-1} a_n \right| \leq a_{r+1}$$

Pertanto, l'errore che si commette sostituendo alla somma della serie alternata quella dei suoi primi r termini è inferiore o uguale al termine tralasciato.

• Errore di arrotondamento

Questo tipo di errore deriva dall'arrotondamento che viene eseguito sui numeri che intervengono in un calcolo. Un tipo di arrotondamento può essere quello che aggiunge 5 alla prima cifra decimale che si vuole tralasciare, e che tronca, quindi, il numero ottenuto.

Il numero trascendente $e = 2,718281828\dots$, base dei logaritmi neperiani, può essere arrotondato alla quarta cifra decimale e il risultato che si ottiene è $e^* = 2,7183$.

Siano a e a^* , rispettivamente, un numero reale e il suo valore arrotondato alla n -sima cifra decimale. Si dimostra che:

$$|a - a^*| \leq \frac{1}{2} \cdot 10^{-n}$$

Pertanto, si può scrivere:

$$a = a^* \pm \frac{1}{2} \cdot 10^{-n}$$

ESEMPIO 1

Il valore arrotondato del numero trascendente $e = 2,718281828\dots$, alla quarta cifra decimale, è $e^* = 2,7183$. Quindi si ha:

$$e = 2,7183 \pm \frac{1}{2} \cdot 10^{-4}$$

b) Propagazione degli errori

Siccome gli *errori di troncamento e di arrotondamento* possono avere influsso sulla precisione dei risultati di un calcolo, sono stati approntati alcuni metodi atti a quantificare tali effetti.

In questo contesto saranno considerati solo gli effetti degli errori sulle operazioni di *addizione* e di *moltiplicazione*. Precisamente, si stabiliranno delle relazioni tra l'approssimazione dei dati e l'approssimazione dei risultati che si ottengono eseguendo sui dati stessi le operazioni di addizione e di moltiplicazione.

A tal fine, siano a e b due numeri reali, e siano a^* e b^* due valori approssimati, rispettivamente, di a e di b .

- **Propagazione degli errori nell'addizione**

Si considerino le conseguenze del calcolo approssimato della somma:

$$a + b$$

Si indichino con ε_1 e con ε_2 gli errori, rispettivamente, di a^* e di b^* . Si possono, pertanto, scrivere le due uguaglianze:

$$\begin{aligned} a &= a^* + \varepsilon_1 \\ b &= b^* + \varepsilon_2 \end{aligned}$$

Sommandole membro a membro, si ha:

$$a + b = a^* + b^* + (\varepsilon_1 + \varepsilon_2)$$

ossia:

$$(a + b) - (a^* + b^*) = \varepsilon_1 + \varepsilon_2$$

Da ciò discende che $\varepsilon_1 + \varepsilon_2$ è l'errore di $a^* + b^*$ nell'approssimare la somma $a + b$; in altre parole, *l'errore della somma è uguale alla somma degli errori degli addendi*.

Siccome:

$$|\varepsilon_1 + \varepsilon_2| \leq |\varepsilon_1| + |\varepsilon_2|$$

si ha:

$$|(a + b) - (a^* + b^*)| \leq |\varepsilon_1| + |\varepsilon_2|$$

ossia *l'errore assoluto della somma è minore o uguale alla somma degli errori assoluti dei singoli addendi*.

Se ε_1 e ε_2 sono gli errori dovuti all'arrotondamento, rispettivamente, di a alle prime i cifre decimali e di b alle prime j cifre decimali, si ha:

$$-\frac{1}{2} \cdot 10^{-i} \leq \varepsilon_1 \leq \frac{1}{2} \cdot 10^{-i}$$

$$-\frac{1}{2} \cdot 10^{-j} \leq \varepsilon_2 \leq \frac{1}{2} \cdot 10^{-j}$$

cioè, ricorrendo ai valori assoluti:

$$|\varepsilon_1| \leq \frac{1}{2} \cdot 10^{-i}$$

$$|\varepsilon_2| \leq \frac{1}{2} \cdot 10^{-j}$$

da cui:

$$|\varepsilon_1 + \varepsilon_2| \leq \frac{1}{2} \cdot 10^{-i} + \frac{1}{2} \cdot 10^{-j}$$

Si può concludere affermando che l'errore $\varepsilon_1 + \varepsilon_2$ di $a^* + b^*$ nell'approssimazione di $a + b$ è compreso tra gli estremi:

$$-\frac{1}{2}(10^{-i} + 10^{-j}) \quad \text{e} \quad \frac{1}{2}(10^{-i} + 10^{-j})$$

In particolare, se $i = j$, gli estremi saranno:

$$-10^{-i} \quad \text{e} \quad 10^{-i}$$

- Propagazione degli errori nella moltiplicazione

Si considerino le conseguenze del calcolo approssimato del prodotto:

$$a \cdot b$$

Sapendo che ε_1 e ε_2 sono, rispettivamente, l'errore di a^* e di b^* , dalle uguaglianze:

$$\begin{aligned} a &= a^* + \varepsilon_1 \\ b &= b^* + \varepsilon_2 \end{aligned}$$

si ha:

$$a \cdot b = (a^* + \varepsilon_1) \cdot (b^* + \varepsilon_2) = a^*b^* + a^*\varepsilon_2 + b^*\varepsilon_1 + \varepsilon_1\varepsilon_2$$

Pertanto, l'errore di a^*b^* nell'approssimazione di $a \cdot b$ è uguale a:

$$|(a \cdot b) - (a^* \cdot b^*)| = |a^*\varepsilon_2 + b^*\varepsilon_1 + \varepsilon_1\varepsilon_2|$$

Siccome il prodotto $\varepsilon_1 \cdot \varepsilon_2$ può essere trascurato, si ha che l'errore di a^*b^* è approssimativamente uguale a:

$$|\varepsilon| = |a^*\varepsilon_2 + b^*\varepsilon_1|$$

Essendo:

$$|a^*\varepsilon_2 + b^*\varepsilon_1| \leq |a^*\varepsilon_2| + |b^*\varepsilon_1| = |a^*| \cdot |\varepsilon_2| + |b^*| \cdot |\varepsilon_1|$$

si ha:

$$|\varepsilon| \leq |a^*| \cdot |\varepsilon_2| + |b^*| \cdot |\varepsilon_1|$$

Se ε_1 e ε_2 sono gli errori dovuti all'arrotondamento, rispettivamente, di a alle prime i cifre decimali e di b alle prime j cifre decimali, si ha:

$$|\varepsilon| = |a^*\varepsilon_2 + b^*\varepsilon_1| \leq |a^*| \cdot \frac{1}{2} \cdot 10^{-j} + |b^*| \cdot \frac{1}{2} \cdot 10^{-i}$$

In particolare, se $i = j$, si ha:

$$|\varepsilon| = |a^*\varepsilon_2 + b^*\varepsilon_1| \leq (|a^*| + |b^*|) \frac{1}{2} \cdot 10^{-i}$$

3 L'interpolazione

L'*interpolazione* è il procedimento mediante il quale si sostituisce alla successione di valori disponibili di una funzione $f(x)$ una successione più completa di dati, che saranno, tutti o parte, dati teorici in grado di fornire informazioni sull'andamento della funzione in corrispondenza di valori della x all'interno del campo di osservazione.

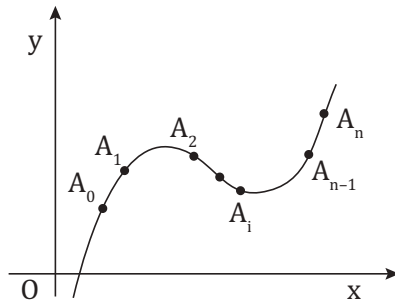
L'interpolazione mira ad ottenere i valori assunti dalla funzione in corrispondenza di valori della x diversi da quelli dati; a tale procedimento si ricorre quando, pur conoscendo i valori che la funzione assume in corrispondenza di valori dati della x , non si conosce l'espressione analitica della funzione stessa, o quando, pur conoscendola, essa è troppo complessa.

Siano dati i punti $A_0, A_1, A_2, \dots, A_n$, aventi per coordinate, rispettivamente, $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, con $x_0 < x_1 < x_2 < \dots < x_n$. Con l'interpolazione si mira a costruire un polinomio di grado minore o uguale a n , il cui grafico passi per i punti A_i con $i = 0, 1, 2, \dots, n$, ossia un polinomio tale che:

$$p(x_i) = y_i \quad i = 0, 1, 2, \dots, n$$

Il polinomio $p(x)$, in pratica, approssima la funzione $f(x)$ in un dato intorno dei valori noti della x , detti *punti di interpolazione*.

L'interpolazione mediante i valori di un polinomio $p(x)$ è denominata *interpolazione polinomiale*, e il polinomio è denominato *polinomio interpolatore* della funzione tra i valori della x , o *polinomio interpolante* f tra i valori noti della x . Per il *principio di identità dei polinomi*, $p(x)$ è sicuramente unico.



a) Formula di interpolazione di Lagrange

La *formula di interpolazione di Lagrange* è la seguente:

$$p(x) = \sum_{i=0}^n y_i \prod_{j \neq i} \frac{(x - x_j)}{(x_i - x_j)}$$

Il procedimento di interpolazione è *lineare* quando il polinomio interpolatore della funzione è di primo grado, ossia quando nella formula data $i = 1$. Si ha, invece, *interpolazione quadratica* quando il polinomio è di secondo grado, ossia quando nella formula data $i = 2$; e così via.

È possibile ricavare una misura dell'errore che si commette ponendo:

$$p(x) = f(x)$$

in un procedimento di interpolazione. A tal proposito, si consideri una funzione $y = f(x)$, dotata nell'intervallo $[a, b]$ di derivate continue fino a quella di ordine $n + 1$, e sia M un maggiorante della derivata $f^{(n+1)}(x)$; si dimostra che:

$$|f(x) - p(x)| \leq M \frac{(b-a)^{n+1}}{(n+1)!}$$

dove $p(x)$ è il polinomio interpolatore della funzione f espresso dalla formula di Lagrange.

b) Formula di interpolazione di Newton

Prima di dare la formula di interpolazione di Newton, è essenziale introdurre il concetto di *differenze divise* della funzione $f(x)$, mediante le uguaglianze:

$$\begin{aligned} f(x_i) &= y_i \\ f[x_i, x_{i+1}] &= \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \\ f[x_i, x_{i+1}, x_{i+2}] &= \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i} \\ &\dots\dots\dots \\ f[x_i, x_{i+1}, \dots, x_{i+h}] &= \frac{f[x_{i+1}, \dots, x_{i+h}] - f[x_i, \dots, x_{i+h-1}]}{x_{i+h} - x_i} \end{aligned}$$

Tali uguaglianze sono dette *differenze divise*, rispettivamente, di ordine zero, uno, due etc., relative ai punti (x_i, y_i) .

Il polinomio interpolatore di Lagrange, relativo ai differenti $n + 1$ punti (x_i, y_i) , può essere espresso in termini di differenze divise della funzione f mediante la formula:

$$\begin{aligned} p(x) = & f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots + \\ & + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1)(x - x_{n-1}) \end{aligned}$$

detta *formula di interpolazione di Newton*.

4 La risoluzione approssimata di equazioni

Si considereranno, in questo paragrafo, alcuni metodi per la risoluzione numerica di equazioni del tipo:

$$f(x) = 0$$

dove $f(x)$ non è una funzione lineare.

a) Metodo di bisezione

Sia data l'equazione:

$$f(x) = 0 \tag{1}$$

dove la funzione $f(x)$ è continua in un intervallo $[a, b]$ e $f(a) \cdot f(b) < 0$.

Per cercare la radice dell'equazione (1), che appartiene all'intervallo $[a, b]$, si divide tale intervallo in due parti uguali.

Se $f\left(\frac{a+b}{2}\right) = 0$, allora $\frac{a+b}{2}$ è una radice dell'equazione.

Se $f\left(\frac{a+b}{2}\right) \neq 0$, si consideri quello dei due intervalli $\left[a, \frac{a+b}{2}\right]$ oppure $\left[\frac{a+b}{2}, b\right]$ alle

estremità del quale la funzione $f(x)$ assume valori di segno opposto. Il nuovo intervallo dimezzato $[a_1, b_1]$ viene ancora suddiviso in due parti, sulle quali si ripete il procedimento sopra

esposto. Si ottiene così, a un certo punto, o una radice esatta dell'equazione (1) oppure una successione di intervalli $[a_1, b_1], [a_2, b_2], \dots, [a_n, b_n], \dots$, tali che per ogni n :

$$f(a_n) \cdot f(b_n) < 0 \quad (2)$$

e inoltre:

$$b_n - a_n = \frac{b-a}{2^n}, \text{ con } \lim_{n \rightarrow \infty} (b_n - a_n) = 0$$

Pertanto, per l'*assioma della continuità*, gli intervalli $[a_1, b_1], \dots, [a_n, b_n], \dots$ hanno uno e un solo elemento comune c , e risulta:

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = c$$

Nella disuguaglianza (2) si passi, ora, al limite per $n \rightarrow +\infty$. Poiché $f(x)$ è continua, si ha: $[f(c)]^2 \leq 0$. Ne segue che $f(c) = 0$, cioè c è una radice dell'equazione (1) ed è evidente che:

$$0 \leq c - a_n \leq \frac{b-a}{2^n}$$

b) Metodo delle iterate successive

Con questo metodo si trasforma l'equazione $f(x) = 0$ in un'equazione equivalente del tipo:

$$x = g(x) \quad (3)$$

in cui $g(x)$ può essere scelta in infiniti modi, ponendo, ad esempio, $g(x) = x \pm f(x)$.

Una volta determinato, con metodi grafici per dirne una, un valore approssimato (anche largamente) di una soluzione dell'equazione data, lo si indichi con x_0 e lo si porti al secondo membro dell'equazione (3). Si ottiene, dunque, un certo numero:

$$x_1 = g(x_0)$$

Si sostituisce tale valore, x_1 , sempre al secondo membro della (3), ricavando un nuovo numero:

$$x_2 = g(x_1)$$

e così di seguito. Si ottiene, in tal modo, una successione $x_0, x_1, \dots, x_n, \dots$ di numeri, in modo che x_n si ricava da x_{n-1} , mediante la formula:

$$x_n = g(x_{n-1}) \quad (4)$$

Se questa successione è convergente, cioè se esiste:

$$c = \lim_{n \rightarrow +\infty} x_n$$

allora, passando al limite nell'equazione (4) e supponendo la $g(x)$ continua, si ha:

$$\lim_{n \rightarrow +\infty} x_n = g\left(\lim_{n \rightarrow +\infty} x_{n-1}\right), \text{ da cui } c = g(c)$$

Ciò prova che il numero c è una radice dell'equazione (3) e tale radice si calcola, mediante la (4), con l'approssimazione voluta.

Un metodo iterativo è convergente, in generale, se soddisfa il teorema seguente.

TEOREMA • Se la funzione $g(x)$ è definita e derivabile su tutto l'intervallo $[a, b]$, e se esiste un numero r tale che:

$$|g'(x)| \leq r < 1$$

su tutto $[a, b]$, allora:

- il processo iterativo $x_n = g(x_{n-1})$ converge indipendentemente dal valore iniziale $x_0 \in [a, b]$;
- il valore limite $c = \lim x_n$ è l'unica radice dell'equazione $x = g(x)$, sull'intervallo $[a, b]$.

c) Metodo di Newton

Sia $f(x)$ una funzione reale definita nell'intervallo $[a, b]$ e ivi derivabile due volte. Inoltre, in $[a, b]$ la $f''(x)$ sia sempre positiva o sempre negativa e la f assuma, agli estremi di $[a, b]$, valori di segno opposto, cioè sia:

$$f(a) \cdot f(b) < 0$$

Si vuole determinare un valore approssimativo dell'unica radice c che l'equazione $f(x) = 0$ ammette in $[a, b]$.

Supponiamo $f''(x) > 0$ e che in $a = x_0$ la f sia positiva. La tangente al diagramma della f nel punto A di ascissa $x_0 = a$ è:

$$y = f(x_0) + f'(x_0)(x - x_0)$$

Essa taglia l'asse x nel punto di ascissa:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Si ha $x_1 > x_0$, perché $f(x_0) > 0$ per ipotesi e perché risulta sicuramente $f'(x_0) < 0$.

Inoltre $f(x_1) > 0$ perché, essendo $f''(x) > 0$, il grafico è al di sopra della tangente in A. Ne segue che $x_1 < c$ e, quindi, la radice c risulta interna all'intervallo $[x_1, b]$.

Si applichi a tale intervallo lo stesso procedimento, considerando la tangente in $A_1[x_1, f(x_1)]$ e la sua intersezione con l'asse x . Ciò conduce al punto di ascissa:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \quad \text{con } x_1 < x_2 < c < b.$$

Si prosegua nell'intervallo $[x_2, b]$, ottenendo:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} \quad \text{con } x_1 < x_2 < x_3 < c < b.$$

e così di seguito.

Si ottiene, dunque, la successione crescente $x_1, x_2, \dots, x_n, \dots$ di numeri minori della radice c , in modo che x_n si ricava da x_{n-1} mediante la formula:

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$$

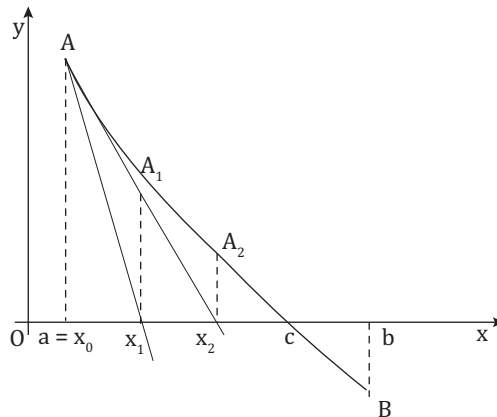
Tale successione converge a c , ovvero risulta:

$$\lim_{n \rightarrow +\infty} x_n = c$$

Il *metodo di Newton* o *metodo delle tangenti*, appena analizzato, è applicato sovente per determinare le soluzioni di un'equazione del tipo:

$$x^n = c$$

Graficamente, il procedimento consiste nel sostituire alla curva $y = f(x)$ la tangente nel punto $[x_n, f(x_n)]$ alla curva medesima.



5 L'integrazione numerica

Sia data una funzione f continua e definita in un intervallo $[a, b]$, e sia F una primitiva di f . In tal caso, l'integrale definito di f sull'intervallo considerato è:

$$\int_a^b f(x) dx = [F(x)]_a^b$$

L'integrale non è sempre calcolabile, o perché non esiste alcun metodo analitico per integrare la funzione, oppure perché, pur esistendo, esso è alquanto complesso. In tali situazioni, vengono in nostro soccorso alcuni metodi di integrazione numerica approssimata, i quali saranno esposti brevemente qui di seguito.

a) Metodo dei rettangoli

Per illustrare questo metodo si consideri una funzione f continua e definita in un intervallo $[a, b]$. Effettuiamo una scomposizione di $[a, b]$ nei seguenti n intervalli parziali:

$$[a_0, a_1], [a_1, a_2], \dots, [a_i, a_{i+1}], \dots, [a_{n-1}, a_n] \quad \text{con } a_0 = a \text{ e } a_n = b$$

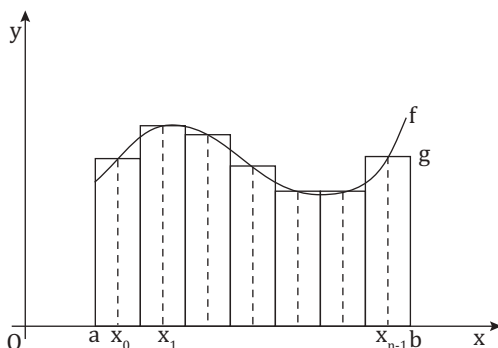
ognuno di ampiezza pari a:

$$h = \frac{b-a}{n}$$

e i cui punti medi (crescenti) sono:

$$x_0, x_1, \dots, x_i, \dots, x_{n-1}$$

La funzione è approssimata, in ognuno di questi intervalli parziali, da una funzione costante $f(x_i)$, dove x_i è il punto medio del generico intervallo i -esimo.



La somma delle aree degli n rettangoli rappresentati in figura è uguale all'integrale della funzione costante g (a scalino) che approssima la funzione f , vale a dire:

$$\int_a^b g(x)dx = h[f(x_0) + f(x_1) + f(x_2) + \dots + f(x_{n-1})]$$

Assegnando a h il suo valore, si può scrivere l'uguaglianza approssimata:

$$\int_a^b f(x)dx = \frac{b-a}{n}[f(x_0) + f(x_1) + f(x_2) + \dots + f(x_{n-1})] = R_n$$

che è la *formula dei rettangoli*.

La *maggiorazione dell'errore* dovuta all'approssimazione appena fatta è data da:

$$\left| \int_a^b f(x)dx - R_n \right| \leq M \frac{(b-a)^2}{n}$$

dove $M > 0$ è tale che:

$$|f'(x)| \leq M \quad \forall x \in [a, b]$$

b) Metodo dei trapezi

Anche per illustrare questo metodo si deve considerare una funzione f continua e definita in un intervallo $[a, b]$, del quale si effettua una scomposizione in n intervalli parziali:

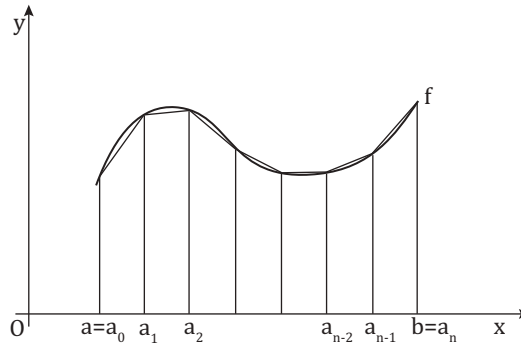
$$[a_0, a_1], [a_1, a_2], \dots, [a_p, a_{i+1}], \dots, [a_{n-1}, a_n] \quad \text{con } a_0 = a \text{ e } a_n = b$$

ognuno di ampiezza pari a:

$$h = \frac{b-a}{n}$$

La funzione è approssimata, in ognuno degli intervalli parziali $[a_i, a_{i+1}]$, dal polinomio di Lagrange di primo grado, relativo ai punti a_i, a_{i+1} . Si ottiene, pertanto, la *formula del trapezio* relativa all'intervallo $[a_i, a_{i+1}]$:

$$\int_{a_i}^{a_{i+1}} f(x)dx = \frac{a_{i+1} - a_i}{2} [f(a_i) + f(a_{i+1})]$$



Siccome:

$$\int_a^b f(x) dx = \sum_{i=0}^{n-1} \int_{a_i}^{a_{i+1}} f(x) dx$$

e considerando che $a_{i+1} - a_i = \frac{b-a}{n}$, si ottiene l'approssimazione della funzione f con una funzione lineare a tratti, la quale, graficamente, è rappresentata da n segmenti. Dunque, l'uguaglianza approssimata:

$$\int_a^b f(x) dx \approx \frac{b-a}{2n} \left[f(a) + 2 \sum_{i=1}^{n-1} f(a_i) + f(b) \right] = T_n$$

è la *formula dei trapezi*.

La *maggiorazione dell'errore*, dovuta all'approssimazione appena fatta, è data da:

$$\left| \int_a^b f(x) dx - T_n \right| \leq \frac{M}{2} \frac{(b-a)^3}{n^2}$$

dove $M > 0$ è tale che:

$$|f''(x)| \leq M \quad \forall x \in [a, b]$$

c) Metodo di Simpson

Anche per illustrare il *metodo di Simpson* si considera una funzione f continua e definita in un intervallo $[a, b]$, del quale si effettua una scomposizione in n intervalli parziali:

$$[a_0, a_1], [a_1, a_2], \dots, [a_{n-2}, a_{n-1}], [a_{n-1}, a_n] \quad \text{con } a_0 = a \text{ e } a_n = b$$

ognuno di ampiezza pari a:

$$h = \frac{b-a}{n}$$

e i cui punti medi (crescenti) sono:

$$x_0, x_1, \dots, x_p, \dots, x_{n-1}$$

La funzione è approssimata, in ognuno degli intervalli parziali $[a_i, a_{i+1}]$, dal polinomio di Lagrange di primo grado, relativo ai punti a_i, x_i, a_{i+1} . In tal caso, si ottiene la *formula di Simpson* relativa all'intervallo $[a_i, a_{i+1}]$:

$$\int_{a_i}^{a_{i+1}} f(x) dx = \frac{a_{i+1} - a_i}{6} [f(a_i) + 4f(x_i) + f(a_{i+1})]$$

Con le notazioni fatte per il metodo precedente, si ottiene la formula:

$$\int_a^b f(x) dx = \frac{b-a}{6n} \sum_{i=0}^{n-1} [f(a_i) + 4f(x_i) + f(a_{i+1})]$$

vale a dire l'uguaglianza approssimata:

$$\int_a^b f(x) dx = \frac{b-a}{6n} \left[f(a) + 2 \sum_{i=1}^{n-1} f(a_i) + 4 \sum_{i=0}^{n-1} f(x_i) + f(b) \right] = S_n$$

che è la *formula di Simpson*.

La *maggiorazione dell'errore* dovuta all'approssimazione appena fatta è data da:

$$\left| \int_a^b f(x) dx - S_n \right| \leq \frac{M}{6} \frac{(b-a)^4}{n^2}$$

dove $M > 0$ è tale che:

$$|f'''(x)| \leq M \quad \forall x \in [a, b]$$

ESEMPIO 2

Si consideri la funzione:

$$f(x) = x^2 \quad x \in [0, 1]$$

Il valore esatto dell'integrale della funzione tra 0 e 1 è:

$$\int_0^1 f(x) dx = \int_0^1 x^2 dx = \frac{1}{3}$$

Considerando che il punto medio dell'intervallo $[0, 1]$ è 0,5, la formula di Simpson per $n = 1$ è:

$$\int_{a_i}^{a_{i+1}} f(x) dx = \frac{1}{6} [f(0) + 4f(0,5) + f(1)] = \frac{1}{6} [0 + 4 \cdot 0,25 + 1] = \frac{1}{6} \cdot 2 = \frac{1}{3}$$

il che equivale a dire che la formula di Simpson fornisce il valore esatto dell'integrale.

1 Indagine statistica e tabelle

Un'**indagine statistica** è un'operazione condotta, mediante l'osservazione, su elementi indefiniti di un determinato collettivo, con l'obiettivo di distinguerli e classificarli secondo le **modalità** di uno o più caratteri.

L'**unità statistica** è la componente elementare del collettivo; è su di essa che si acquisiscono le informazioni.

Nell'ambito delle unità statistiche si distinguono:

- *unità semplici*, ad esempio una singola persona o una singola abitazione;
- *unità composte*, che sono insiemi di *unità semplici* simili ed esistenti anche a prescindere dall'unità composta. Un esempio può essere rappresentato dalla famiglia, intesa come insieme dei suoi componenti, oppure da un edificio, inteso come insieme di abitazioni;
- *unità complesse*, che sono insiemi di unità semplici diverse, atte però a caratterizzarle nella loro totalità. Un esempio può essere il processo di produzione di prodotti assemblati, effettuato da un'impresa che si occupa non solo dell'assemblaggio delle singole componenti, ma anche della loro produzione.

I **caratteri** sono gli aspetti del fenomeno oggetto di rilevazione. A loro volta i caratteri statistici si distinguono in **qualitativi** (come il genere di attività, la professione esercitata etc.) e **quantitativi** (come il reddito, la produzione, l'età etc.).

L'indagine statistica si articola nelle seguenti **fasi**:

- *identificazione del collettivo statistico*, ovvero l'operazione di definizione delle unità statistiche che lo costituiscono;
- *rilevazione*, che è l'operazione mediante la quale si acquisiscono le modalità di uno o più caratteri del collettivo statistico; essa può riguardare l'intera popolazione oggetto di osservazione oppure può essere *per campione*, ossia riguardare un sottoinsieme della popolazione;
- *elaborazione*, ossia l'operazione di classificazione (in tabelle e grafici) e sintesi dei dati risultanti dallo spoglio;
- *interpretazione*, la quale, sulla base delle conoscenze sul fenomeno oggetto di studio, chiarisce i risultati acquisiti.

Le **tabelle statistiche** emergono dalle operazioni di spoglio dei risultati di indagini statistiche, attraverso la classificazione dei dati rilevati in base alle modalità (o manifestazioni) dei caratteri.

Esse possono essere *semplici* o *multiple* (o a più entrate), a seconda se riportano, rispettivamente, le informazioni statistiche su un fenomeno collettivo in relazione a un solo carattere, o a più di un carattere combinando ciascuna modalità di un carattere con le modalità dell'uno o degli altri caratteri.

Una tabella statistica si definisce mediante qualificazioni non determinabili numericamente e/o mediante numeri, che sono:

- **intensità**, se mostrano la misura o la grandezza di un carattere (come il peso di una persona, l'ammontare degli investimenti di un'azienda etc.);

- **frequenze**, se mostrano il numero di volte in cui una modalità del carattere si presenta nelle unità statistiche (come il numero di iscritti alle liste di leva in uno specifico anno, il numero di iscritti ai licei scientifici in un dato anno scolastico etc.).

A loro volta le frequenze si distinguono in:

- *frequenze assolute*, le quali indicano il numero di unità di un collettivo che presenta una data modalità di un carattere;
- *frequenze relative*, che si ottengono dividendo ogni frequenza assoluta per il totale delle stesse;
- *frequenze percentuali*, le quali si ottengono moltiplicando per 100 ognuna delle frequenze relative, e indicano la frazione delle osservazioni che presentano una data modalità del carattere;
- *frequenze cumulate*, le quali indicano le frequenze delle osservazioni che hanno un valore del carattere minore (oppure minore o uguale) ad una prestabilita modalità.

In una tabella statistica semplice si presentano due sole colonne: una *colonna madre*, che indica le modalità del carattere; e una *seconda colonna*, tracciata a destra della precedente, in cui, in corrispondenza di ogni modalità, si indica l'intensità o la frequenza accertata.

La **distribuzione statistica** è l'insieme delle determinazioni del carattere e delle rispettive frequenze. Essa è denominata *variabile statistica* se il carattere è quantitativo; è chiamata, invece, *mutabile statistica* se il carattere è qualitativo. Inoltre, una variabile statistica è *discreta* o *continua*, a seconda dell'insieme di dati di riferimento.

Le mutabili statistiche, a loro volta, si distinguono in:

- *rettilinee*, se esiste un ordine naturale o logico delle modalità (come, ad esempio, il numero di operai metalmeccanici per livello);
- *serie storiche* o *temporali* o *cronologiche*, se il principio regolatore è il tempo, il quale è inteso come progressione cronologica; ovviamente, una serie storica è una particolare mutabile rettilinea (come, ad esempio, il numero di autovetture di una data marca vendute in diversi anni);
- *cicliche*, se il tempo è inteso in termini di periodicità, per cui non esistono né una modalità iniziale né una modalità finale (come, ad esempio, le precipitazioni nevose nei diversi mesi di un anno);
- *sconnesse*, se non esistono né un ordine logico né un ordine naturale secondo cui sono disposte le modalità (come, ad esempio, il numero dei voti ottenuti da diversi partiti in una consultazione elettorale).

2 Le distribuzioni statistiche semplici

a) Variabili statistiche

• Variabili statistiche discrete

Sia dato un collettivo di N unità statistiche, delle quali si voglia rilevare un carattere quantitativo discreto, e si indichino con:

- $x_1, x_2, \dots, x_s$ le s modalità del carattere;
- $N_1, N_2, \dots, N_s$ le s frequenze corrispondenti alle modalità.

La tabella statistica descrittiva dei dati è denominata *seriazione statistica discreta*.

Modalità del carattere (x)	Frequenze
x_1	N_1
x_2	N_2
x_3	N_3
...	...
x_s	N_s
Totale	N

ESEMPIO 1

Famiglie residenti in un dato Comune, suddivise per numero di componenti

Numero componenti	Famiglie
1	153
2	225
3	335
4	564
5	346
6	133
7	75
8 e più	49
Totale	1.880

• Variabili statistiche continue

Sia dato un collettivo di N unità statistiche, delle quali si voglia rilevare un carattere quantitativo continuo, e siano s le classi di modalità; si indichino con:

- $x_1 \vdash x_2, x_2 \vdash x_3, \dots, x_s \vdash x_{s+1}$ i valori estremi delle s classi chiuse a sinistra e aperte a destra;
- $N_1, N_2, \dots, N_s$ le s frequenze corrispondenti alle modalità delle s classi.

La tabella statistica descrittiva dei dati è denominata *seriazione statistica continua*.

Classi di modalità del carattere (x)	Frequenze
$x_1 \vdash x_2$	N_1
$x_2 \vdash x_3$	N_2
...	...
$x_i \vdash x_{i+1}$	N_i
...	...
$x_s \vdash x_{s+1}$	N_s
Totale	N

In una variabile statistica continua:

- il *modulo* o *ampiezza* della classe è la differenza tra l'estremo superiore e l'estremo inferiore della classe;
- la *densità di frequenza* è il rapporto tra la frequenza e l'ampiezza di una classe;
- il *valore centrale* è la semisomma degli estremi di una classe.

ESEMPIO 2

Percentuali di un campione di famiglie milanesi per classi di reddito mensile (dati fittizi)

Classi di reddito	Percentuali di famiglie
Fino a € 1.000	26,6
€ 1.000 - € 1.500	19,9
€ 1.500 - € 2.000	17,4
€ 2.000 - € 2.500	18,7
€ 2.500 - € 3.000	10,4
Oltre € 3.000	7
Totale	100

b) Mutabili statistiche

Sia dato un collettivo di N unità statistiche, delle quali si voglia rilevare un carattere qualitativo, e si indichino con:

- $A_1, A_2, \dots, A_s$ le s modalità del carattere;
- $N_1, N_2, \dots, N_s$ le s frequenze corrispondenti alle modalità delle s classi.

La tabella statistica corrispondente è detta *serie di frequenze*. Analogamente, se in luogo delle frequenze si fossero riscontrate le intensità, la tabella avrebbe assunto la denominazione di *serie di intensità*.

Modalità del carattere (x)	Frequenze
A_1	N_1
A_2	N_2
A_3	N_3
...	...
A_s	N_s
Totale	N

ESEMPIO 3

Prenotazioni alberghiere nella Penisola sorrentina secondo i mesi dell'anno (dati fittizi)

Mesi dell'anno	N. prenotazioni
Gennaio	1.230
Febbraio	1.123
Marzo	2.203
Aprile	2.548
Maggio	2.324
Giugno	4.954
Luglio	5.657
Agosto	8.108
Settembre	6.993
Ottobre	3.441
Novembre	3.220
Dicembre	4.088
Totale	45.889

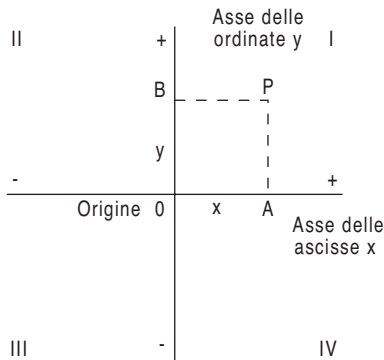
3 Le rappresentazioni grafiche

Molteplici sono le *rappresentazioni grafiche* dei dati statistici. Qui di seguito prenderemo in considerazione solo quelle che si prestano più soddisfacentemente a un'interpretazione dei dati.

a) Riferimento cartesiano ortogonale

Le rappresentazioni grafiche di più largo impiego sfruttano l'elemento lineare, avvalendosi di un riferimento in *coordinate cartesiane ortogonali*; tale sistema è costituito da due rette ortogonali, il cui punto di intersezione (0) è denominato *origine*, e in cui la retta orizzontale è detta *asse delle ascisse*, mentre la retta verticale è chiamata *asse delle ordinate*.

Su entrambi gli assi si fissano un'unità di misura dei segmenti e un orientamento.

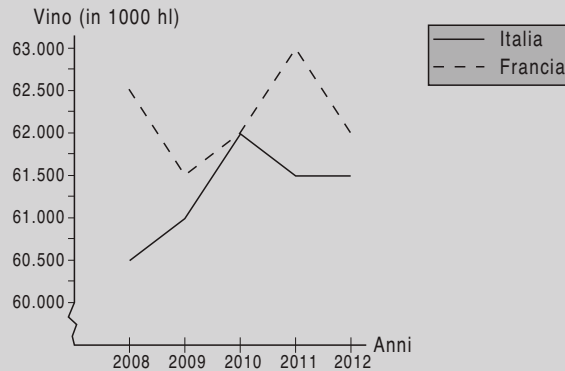


Una volta stabilite le opportune unità di misura per entrambi gli assi, la rappresentazione grafica in questione è particolarmente utile nel caso di *serie temporali*: sull'asse delle ascisse si fissa la relativa unità temporale (giorno, mese, anno...) e sull'asse delle ordinate le modalità

del carattere esaminato, riferite ai diversi tempi. Unendo i differenti punti-immagine tracciati, si ottiene una curva di evidente utilità ai fini dell'interpretazione dei dati statistici.

ESEMPIO 4

Produzione di vino dell'Italia e della Francia (dati fittizi)

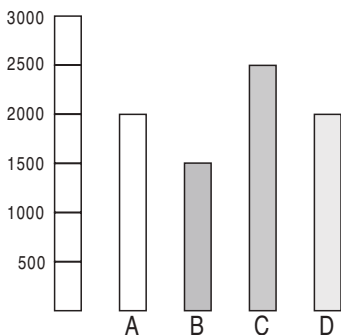


b) Ortogrammi

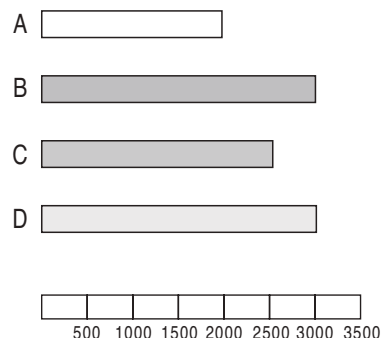
La rappresentazione grafica con *ortogrammi* si basa sulla corrispondenza tra intensità o frequenze e superfici rettangolari, e si attua attraverso:

- rettangoli equidistanziati, di uguale base e aventi altezze uguali o proporzionali alle intensità o frequenze da rappresentare (*ortogramma a colonne*);
- rettangoli equidistanziati, di uguale altezza e aventi basi uguali o proporzionali alle intensità o frequenze da rappresentare (*ortogramma a nastri*).

Le due figure seguenti indicano la rappresentazione di dati mediante, rispettivamente, un ortogramma a colonne e un ortogramma a nastri.



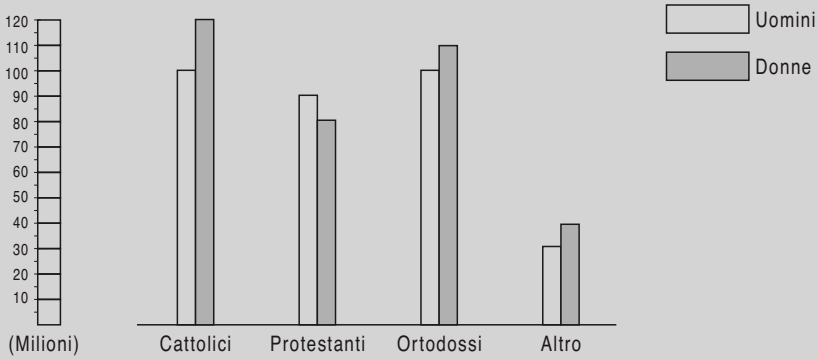
Ortogramma a colonne



Ortogramma a nastri

ESEMPIO 5

Stima della religione professata dai cittadini europei (dati fittizi)

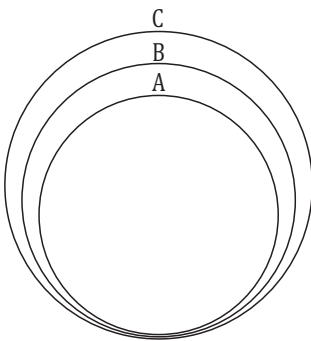


c) Areogrammi per cerchi e per settori circolari

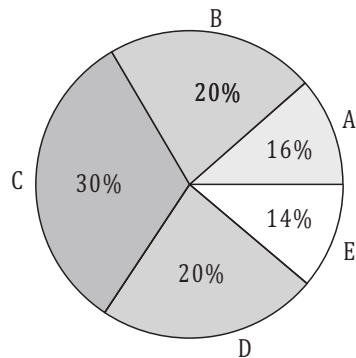
La rappresentazione grafica con *areogrammi* si basa sulla corrispondenza tra intensità o frequenze e superfici di cerchi o di settori circolari.

Negli *areogrammi per cerchi*, le misure delle superfici dei cerchi devono essere proporzionali alle intensità o alle frequenze da rappresentare. Si assume, in tal caso, un cerchio-base di raggio r e area πr^2 che indica l'intensità o la frequenza più bassa; per rappresentare un'intensità o frequenza doppia o pari a k volte quella minima, si ricorre a cerchi aventi, rispettivamente, area doppia o pari a k volte quella minima.

Negli *areogrammi per settori circolari*, l'area del cerchio esprime l'intensità o la frequenza totale (espressa in termini assoluti o percentuali). I settori circolari, invece, esprimono le intensità o le frequenze delle varie modalità del carattere esaminato; essi sono individuati dividendo l'intero angolo giro in parti proporzionali alle intensità o alle frequenze da rappresentare.



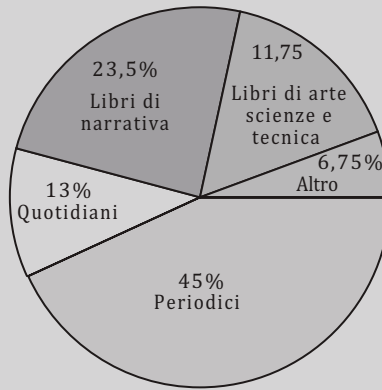
Areogrammi per cerchi



Areogrammi per settori circolari

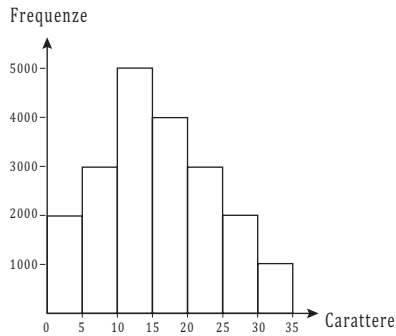
ESEMPIO 6

Lettere preferite in vacanza da un campione di famiglie italiane (dati fittizi)



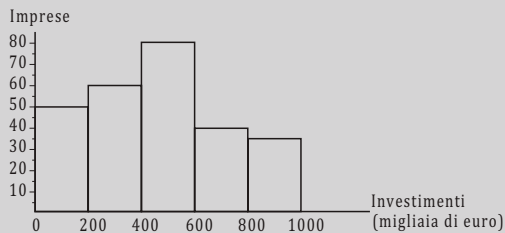
d) Istogrammi

Gli *istogrammi* costituiscono la rappresentazione grafica adeguata delle distribuzioni di frequenze per classi. Per costruire tali grafici si riporta sull'asse delle ascisse, senza soluzione di continuità, l'ampiezza delle successive classi, e sull'asse delle ordinate la frequenza o la densità di frequenza (rapporto tra frequenza e ampiezza di classe) di ciascuna classe.



ESEMPIO 7

Investimenti di un gruppo di imprese appartenenti al settore meccanico



4 Gli indici statistici per variabili quantitative

Gli *indici di posizione* o *medie* sono quantità idonee a dare un'idea d'insieme di un dato collettivo statistico, sostituendosi, pertanto, a tutti gli altri elementi che lo costituiscono.

Essi si distinguono in:

- *medie analitiche*, che si determinano considerando tutti i valori di una data variabile statistica e sono la *media aritmetica*, la *media armonica*, la *media geometrica* etc.;
- *medie lasche*, che si determinano considerando solo dati elementi della distribuzione e sono la *mediana*, la *moda* etc.

Gli *indici di variabilità* sono quantità atte a misurare la tendenza di un fenomeno ad assumere diverse modalità quantitative.

Essi si distinguono in:

- *misure analitiche*, che si suddividono a loro volta in:
 - misure idonee a indicare la *dispersione* dei dati intorno a una media, come gli scostamenti medi dalla media aritmetica, dalla mediana etc.;
 - misure idonee a indicare la *disuguaglianza* dei dati tra loro, come le differenze medie, che si determinano dalle differenze tra ogni intensità e tutte le altre;
- *misure di posizione*, come il *campo di variazione* e la *differenza interquartilica*.

Gli *indici assoluti di variabilità* sono espressi nella stessa unità di misura del fenomeno oggetto di osservazione. Gli *indici relativi di variabilità*, in quanto ottenuti rapportando i corrispondenti indici assoluti ad altri parametri, sono numeri puri, ossia senza alcuna unità fisica; essi sono particolarmente utili nei confronti tra distribuzioni.

Gli *indici di forma* sono quantità atte ad indicare la forma, appunto, della curva rappresentativa delle distribuzioni statistiche.

5 Gli indici di posizione

a) Media aritmetica

Sia dato un collettivo statistico di N unità e una variabile statistica $X = (x_1, x_2, \dots, x_N)$. La *media aritmetica* della variabile X è un indice di posizione e può essere definita come *quella intensità che può essere sostituita ai singoli valori della variabile, in modo che resti invariata l'intensità globale*.

Si distingue tra *media aritmetica semplice* e *media aritmetica ponderata*.

- **Media aritmetica semplice**

Si ottiene rapportando l'intensità globale di un carattere al numero totale dei casi osservati, ossia:

$$M = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$$

Per la definizione di media aritmetica, si ha:

$$\underbrace{M + M + \dots + M}_{N \text{ volte}} = x_1 + x_2 + \dots + x_N$$

Si noti che la media aritmetica semplice di una variabile statistica viene indicata con il simbolo M_s ; la corrispondente media di una variabile casuale è indicata, invece, con la lettera greca μ .

- **Media aritmetica ponderata**

Nel caso in cui le singole modalità della variabile statistica esibiscano frequenze diverse dall'unità, si introduce il concetto di *media aritmetica ponderata*.

Se $N_1, N_2, \dots, N_s$ sono le rispettive frequenze delle modalità $x_1, x_2, \dots, x_s$ della variabile statistica X , la media aritmetica ponderata è:

$$M = \frac{x_1 N_1 + x_2 N_2 + \dots + x_s N_s}{N} = \frac{\sum_{i=1}^s x_i N_i}{N}$$

Se i dati sono raggruppati in classi e se sono note le classi di valori del carattere nonché le rispettive frequenze, si possono presentare due casi nel calcolo della media:

- *è noto anche l'ammontare del carattere per ognuna delle classi.* In tal caso il calcolo della media è semplice in quanto si applica direttamente la formula della media aritmetica, ossia si divide l'ammontare del carattere per il numero dei casi osservati;
- *non è noto l'ammontare del carattere per ognuna delle classi.* In tal caso si ricorre ad un artificio, si suppone cioè che le intensità siano concentrate nel valore centrale di ogni classe.

ESEMPIO 8

I calcoli per la determinazione della media aritmetica sono contenuti nella tabella seguente, in cui x_i indica il generico numero i di componenti del nucleo familiare; N_i il numero di famiglie con i componenti; e il prodotto $x_i N_i$, infine, l'ammontare complessivo di individui costituenti le famiglie con i componenti (la somma di tali prodotti rappresenta l'ammontare totale del carattere). In tale calcolo, si è ipotizzato come ultimo valore del numero di componenti del nucleo familiare la cifra 8.

x_i	N_i	$\bar{x}_i N_i$
1	153	153
2	225	450
3	335	1.005
4	564	2.256
5	346	1.730
6	133	798
7	75	525
8	49	392
Totale	1.880	7.309

Applicando la formula per il calcolo della media aritmetica ponderata, si ottiene il numero medio di componenti della famiglia per il collettivo considerato:

$$M_p = \frac{7.309}{1.880} = 3,88$$

- **Proprietà della media aritmetica**

La media aritmetica di una variabile statistica X è:

- *interna*, vale a dire che il suo valore è sempre minore dell'intensità minima e sempre maggiore dell'intensità massima di una variabile statistica;

- *traslativa*, ovvero che, se ai valori della variabile X si addiziona o si sottrae uno stesso numero a , si ottiene una nuova variabile $X' = X \pm a$, avente media aritmetica uguale alla media aritmetica della variabile X , rispettivamente, aumentata o diminuita del numero a ;
- *omogenea*, ossia che, se i valori della variabile X sono moltiplicati o divisi per uno stesso numero b , si ottiene una nuova variabile $X' = bX$, avente media aritmetica uguale alla media aritmetica della variabile X , rispettivamente, moltiplicata o divisa per il numero b ;
- *associativa*, cioè che, se i valori della variabile X sono suddivisi in due o più insiemi, la media aritmetica della variabile statistica è uguale alla media aritmetica delle medie parziali dei singoli insiemi, ponderate con le numerosità degli insiemi stessi.

Ad ogni variabile statistica X è possibile associare una variabile denominata *scostamento* o *scarto dalla media aritmetica*. Se M è la media aritmetica della variabile X , per la generica modalità x_i , la variabile scarto è data da:

$$e_i = x_i - M$$

Per tale variabile valgono le seguenti proprietà:

- *La somma algebrica degli scarti è uguale a zero.*

Per *dati non raggruppati*, tale proprietà si esprime nel modo seguente:

$$\sum_{i=1}^N (x_i - M) = \sum_{i=1}^N e_i = 0$$

Per *dati raggruppati*, essa si esprime così:

$$\sum_{i=1}^s (x_i - M) N_i = \sum_{i=1}^s e_i N_i = 0$$

- *La somma dei quadrati degli scarti è minima* quando gli scarti sono calcolati dalla media aritmetica M .

Per *dati non raggruppati*, tale proprietà si esprime nel modo seguente:

$$\sum_{i=1}^N (x_i - M)^2 = \sum_{i=1}^N e_i^2 < \sum_{i=1}^N (x_i - A)^2$$

Per *dati raggruppati*, essa si esprime così:

$$\sum_{i=1}^s (x_i - M)^2 N_i = \sum_{i=1}^s e_i^2 N_i < \sum_{i=1}^s (x_i - A)^2 N_i$$

dove A , in entrambi i casi, rappresenta un numero qualunque diverso da M .

b) Media quadratica

Siano dati un collettivo statistico di N unità e una variabile statistica $X = (x_1, x_2, \dots, x_N)$. La *media quadratica* della variabile X è un indice di posizione e può essere definita come *quella intensità che può essere sostituita ai singoli valori della variabile, in modo che resti invariata la somma dei quadrati delle intensità*.

Si distingue tra *media quadratica semplice* e *media quadratica ponderata*.

- **Media quadratica semplice**

Si ottiene dalla radice quadrata del rapporto tra la somma dei quadrati delle intensità e il numero totale dei casi osservati, ossia:

$$M_q = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N}}$$

Per la definizione di media quadratica, si ha:

$$\underbrace{M_q^2 + M_q^2 + \dots + M_q^2}_{N \text{ volte}} = x_1^2 + x_1^2 + \dots + x_N^2$$

- **Media quadratica ponderata**

Con la solita notazione, la media quadratica ponderata è data da:

$$M_q = \sqrt{\frac{\sum_{i=1}^s x_i^2 N_i}{N}}$$

La media quadratica è anch'essa *interna*, *associativa* e *omogenea*, ma *non è traslativa*.

ESEMPIO 9

Si consideri la seguente tabella, in cui le prime due colonne sono uguali a quelle dell'esempio 8. La terza colonna contiene i valori x_i^2 , i quali rappresentano i quadrati dei numeri i , ossia dei componenti del nucleo familiare. La quarta colonna, invece, contiene i valori $x_i^2 N_i$, ossia i prodotti degli x_i^2 per le rispettive frequenze. Come al solito, si è ipotizzato come ultimo valore del numero di componenti del nucleo familiare la cifra 8.

x_i	N_i	x_i^2	$x_i^2 N_i$
1	153	1	153
2	225	4	900
3	335	9	3.015
4	564	16	9.024
5	346	25	8.650
6	133	36	4.788
7	75	49	3.675
8	49	64	3.136
Totale	1.880	—	33.341

Applicando la formula della media quadratica ponderata, si ha:

$$M_q = \sqrt{\frac{33.341}{1.880}} = \sqrt{17,73} = 4,21$$

c) Media armonica

Siano dati il solito collettivo statistico di N unità e una variabile statistica $X = (x_1, x_2, \dots, x_N)$. La *media armonica* della variabile X è un indice di posizione e può essere definita come *quella intensità che può essere sostituita ai singoli valori della variabile, in modo che resti invariata la somma dei reciproci delle intensità*.

Si distingue tra *media armonica semplice* e *media armonica ponderata*.

• Media armonica semplice

Si ottiene rapportando il numero totale dei casi osservati alla somma dei reciproci delle intensità, ossia:

$$M_h = \frac{N}{\sum_{i=1}^N \frac{1}{x_i}}$$

Per la definizione di media armonica, si ha:

$$\underbrace{\frac{1}{M_h} + \frac{1}{M_h} + \dots + \frac{1}{M_h}}_{N \text{ volte}} = \frac{N}{M_h} = \frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_N}$$

• Media armonica ponderata

Con la solita notazione, la media armonica ponderata è data da:

$$M_h = \frac{N}{\sum_{i=1}^s \frac{N_i}{x_i}}$$

La media armonica è anch'essa *interna*, *associativa* e *omogenea*, ma *non è traslativa*.

ESEMPIO 10

Si consideri la seguente tabella, in cui le prime due colonne sono uguali a quelle degli esempi 8 e 9. La terza e quarta colonna contengono, rispettivamente, i valori $1/x_i$ e N_i/x_i , ossia i reciproci dei generici numeri i di componenti del nucleo familiare e i prodotti di tali reciproci per le rispettive frequenze. Si è nuovamente ipotizzato come ultimo valore del numero di componenti del nucleo familiare la cifra 8.

x_i	N_i	$1/x_i$	N_i/x_i
1	153	1	153,00
2	225	0,500	112,50
3	335	0,333	111,55
4	564	0,250	141,00
5	346	0,200	69,20
6	133	0,167	22,21
7	75	0,143	10,72
8	49	0,125	6,12
Totale	1.880	—	626,30

Applicando la formula della media armonica ponderata, si ha:

$$M_h = \frac{1.880}{626,3} = 3$$

d) Media geometrica

Siano dati il solito collettivo statistico di N unità e una variabile statistica $X = (x_1, x_2, \dots, x_N)$. La *media geometrica* della variabile X è un indice di posizione e può essere definita come *quella intensità che può essere sostituita ai singoli valori della variabile, in modo che resti invariato il prodotto delle intensità*.

Si distingue tra *media geometrica semplice* e *media geometrica ponderata*.

- Media geometrica semplice

Si ottiene dalla radice N -esima del prodotto delle intensità, ossia:

$$M_g = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N}$$

Per la definizione di media geometrica, si ha:

$$\underbrace{M_g \cdot M_g \cdot \dots \cdot M_g}_{N \text{ volte}} = M_g^N = x_1 \cdot x_2 \cdot \dots \cdot x_N$$

- Media geometrica ponderata

Con la solita notazione, la media geometrica ponderata è data da:

$$M_g = \sqrt[N]{x_1^{N_1} \cdot x_2^{N_2} \cdot \dots \cdot x_s^{N_s}}$$

In pratica, siccome risulta difficile calcolare la media geometrica, sia semplice sia ponderata, si ricorre spesso ai logaritmi. Infatti, si ha:

$$\log M_g = \frac{1}{N} \sum_{i=1}^N \log x_i \quad \text{e} \quad \log M_g = \frac{1}{N} \sum_{i=1}^s N_i \log x_i$$

Dal logaritmo si risale poi al numero, pervenendo ad un valore approssimato di M_g .

La media geometrica è anch'essa *interna*, *associativa* e *omogenea*, ma *non è traslativa*.

ESEMPIO 11

Si consideri la seguente tabella, in cui le prime due colonne sono uguali a quelle degli esempi precedenti. La terza e la quarta colonna contengono, invece, i valori $\log x_i$ e $N_i \log x_i$, ossia i logaritmi dei generici numeri i di componenti del nucleo familiare e i prodotti di tali logaritmi per le rispettive frequenze. Anche in tale caso si è ipotizzato come ultimo valore del numero di componenti del nucleo familiare la cifra 8.

x_i	N_i	$\log x_i$	$N_i \log x_i$
1	153	0,000	0,000
2	225	0,301	67,725
3	335	0,477	159,795
4	564	0,602	339,528
5	346	0,699	241,854
6	133	0,778	103,474
7	75	0,845	63,375
8	49	0,903	44,247
Totale	1.880	—	1.019,998

Sulla base dei dati che si evincono dalla tabella, il logaritmo della media geometrica è dato da:

$$\log M_g = \frac{1.019,998}{1.880} = 0,542$$

Risalendo dal logaritmo al numero, si ottiene la media geometrica ponderata della distribuzione considerata:

$$M_g = 10^{0,542} = 3,48$$

e) Relazioni tra le medie

Per una data variabile statistica, i valori delle quattro medie considerate (aritmetica, quadratica, armonica e geometrica) stanno nella seguente relazione:

$$M_h < M_g < M < M_q$$

Se le intensità di una variabile statistica sono tutte uguali, ovviamente, la relazione che lega le quattro medie è la seguente:

$$M_h = M_g = M = M_q$$

ESEMPIO 12

Le medie analitiche calcolate per la distribuzione di dati, di cui alla tabella dell'esempio 8, dimostrano l'esistenza della suddetta relazione tra le stesse, infatti:

$$\begin{aligned} M_h &= 3 \\ M_g &= 3,48 \\ M &= 3,88 \\ M_q &= 4,21 \end{aligned}$$

da cui la relazione.

f) Cenni sulla media di somme di potenze

La *media di somme di potenze* è una formula generale comprensiva di numerose medie; essa è data dalla somma delle potenze di ordine s dei termini della distribuzione statistica, divisa per l'analoga somma delle potenze di ordine $s - 1$. Nel caso di dati non raggruppati, la sua espressione è la seguente:

$$M_{s,s-1} = \frac{x_1^s + x_2^s + \dots + x_N^s}{x_1^{s-1} + x_2^{s-1} + \dots + x_N^{s-1}} = \frac{\sum_{i=1}^N x_i^s}{\sum_{i=1}^N x_i^{s-1}}$$

g) Moda

La *moda* o *valore normale* o *norma* è un indice di posizione e può essere definita come *quella modalità del carattere che presenta la frequenza più alta*.

La moda, in alcuni casi, può non esistere, e, anche se esiste, può non essere unica. Una distribuzione che presenta una sola moda si dice *unimodale*; una distribuzione che presenta più di una moda si dice *plurimodale*; infine, una distribuzione che non presenta moda si dice *zeromodale*.

La moda è una media *interna*, *traslativa*, *omogenea*, ma *non associativa*.

Per il calcolo della moda, nel caso di un collettivo che si distribuisce secondo le modalità di un carattere discreto, si possono distinguere diversi casi:

- se il collettivo statistico si presenta con modalità non raggruppate in classi, la moda si calcola considerando, una ad una, le modalità e verificando quale di queste ultime ha una frequenza maggiore;
- se il collettivo statistico si presenta con modalità raggruppate in classi, con classi di uguale ampiezza, la moda cade nella classe avente frequenza più alta;

- se il collettivo statistico si presenta con modalità raggruppate in classi, con classi di diversa ampiezza, si dividono le frequenze delle classi per il numero di valori del carattere contenuto in esse; la moda cade nella classe che presenta il quoziente più alto.

Per il calcolo della moda, nel caso di un collettivo che si distribuisce secondo le modalità di un carattere continuo, si procede in maniera leggermente differente perché non si tiene conto più delle frequenze ma delle densità di frequenza, ossia del rapporto tra frequenza di ciascuna classe e ampiezza della classe stessa; pertanto, la *moda* è *quel valore intorno a cui i casi sono addensati in maggior misura*. Si possono presentare due casi:

- se il collettivo si presenta con modalità raggruppate in classi della stessa ampiezza, la classe modale è quella che presenta, indifferentemente, maggiore frequenza o densità di frequenza;
- se il collettivo si presenta con modalità raggruppate in classi di diversa ampiezza, la classe modale è quella che presenta la densità di frequenza più alta.

ESEMPIO 13

Si consideri la procedura di determinazione della classe modale relativamente alla tabella seguente.

Distribuzione delle imprese presenti in una data area per classi di dipendenti.

Classi di dipendenti	Numero di aziende
10 - 30	21
30 - 50	30
50 - 75	35
75 - 100	30
100 - 150	34
150 - 200	13
200 - 350	75
350 - 600	148
Totale	386

I calcoli per la determinazione della classe modale, tenendo conto del fatto che le classi hanno diversa ampiezza, sono contenuti nella tabella seguente, in cui n_i è il numero di valori del carattere contenuto nella generica classe i -esima e N_i/n_i è il quoziente tra la frequenza della classe i -esima e il numero di valori del carattere di detta classe.

$x_i - x_{i+1}$	N_i	n_i	N_i/n_i
10 - 30	21	21	1,000
31 - 50	30	20	1,500
51 - 75	35	25	1,400
76 - 100	30	25	1,200
101 - 150	34	50	0,680
151 - 200	13	50	0,260
201 - 350	75	150	0,500
351 - 600	148	250	0,592
Totale	386	—	—

Il quoziente massimo tra frequenza di classe e corrispondente numero di valori del carattere della classe è 1,5 e si rileva per la classe «31 - 50 dipendenti», la quale è, pertanto, la classe modale.

Senza tenere conto della diversa ampiezza delle classi, quella modale sarebbe risultata la classe «351 - 600 dipendenti». Infatti, per essa, il numero di aziende è 148, il massimo che si riscontra percorrendo con lo sguardo la colonna delle frequenze.

h) Mediana

Data una variabile statistica, le cui intensità siano in ordine crescente o decrescente, si definisce *mediana* quell'indice di posizione che *corrisponde al valore della variabile statistica che bipartisce la distribuzione*, lasciando la metà dei casi a sinistra e l'altra metà a destra.

Per il calcolo della mediana per dati non raggruppati in classi, si identifica il posto centrale; a tal proposito si possono distinguere due casi:

— i valori osservati sono in numero N dispari. In tal caso, il posto centrale è unico ed è:

$$C = \frac{N+1}{2}$$

La mediana è quella intensità che occupa tale posto centrale;

— i valori osservati sono in numero N pari. In tal caso, sono due i posti centrali:

$$C_1 = \frac{N}{2} \quad \text{e} \quad C_2 = \frac{N}{2} + 1$$

Si assume come mediana la media aritmetica delle intensità che occupano i due posti centrali appena considerati.

Per il calcolo della mediana, nel caso di un collettivo che si distribuisce secondo le modalità di un carattere discreto, con dati raggruppati in classi, il procedimento è il seguente:

1. si calcola il posto centrale, come nel caso di dati non raggruppati in classi;
2. sulla base delle frequenze cumulate, si determina la classe in cui cade la mediana;
3. si determina la differenza tra il posto centrale e la frequenza cumulata fino alla classe che precede quella contenente la mediana;
4. si divide la frequenza della classe mediana per l'ampiezza della stessa classe;
5. per ottenere la mediana si divide la differenza di cui al punto 3 per il quoziente di cui al punto 4.

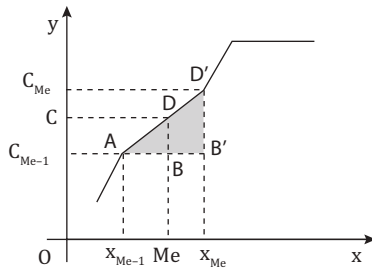
Per il calcolo della mediana, nel caso di un collettivo che si distribuisce secondo le modalità di un carattere continuo, si considera la *spezzata cumulativa* che rappresenta graficamente le frequenze cumulate.

Siano:

- x_{Me-1} l'estremo inferiore della classe mediana;
- x_{Me} l'estremo superiore della classe mediana;
- α_{Me} l'ampiezza della classe mediana;
- N_{Me} la frequenza della classe mediana;
- C la frequenza cumulata coincidente con il posto centrale;
- C_{Me-1} la frequenza cumulata fino alla classe che precede immediatamente la mediana;
- C_{Me} la frequenza cumulata fino alla classe mediana compresa.

La formula per il calcolo della mediana è:

$$Me = x_{Me-1} + \frac{\alpha_{Me}}{N_{Me}}(C - C_{Me-1})$$



Essa si ottiene considerando i triangoli simili ABD e AB'D'.

- **Proprietà della mediana**

Considerando una variabile statistica X e gli scarti corrispondenti $x_i - Me$, valgono le seguenti proprietà:

- *il numero degli scarti di segno positivo è uguale al numero degli scarti di segno negativo;*
- *la somma dei valori assoluti degli scarti dalla mediana è minima.*

La seconda proprietà, scegliendo un qualsiasi numero A , può essere così espressa:

- *per dati non raggruppati:*

$$\sum_i |x_i - Me| \leq \sum_i |x_i - A|$$

- *per le distribuzioni di frequenze:*

$$\sum_i |x_i - Me| N_i \leq \sum_i |x_i - A| N_i$$

ESEMPIO 14

Si consideri la procedura di determinazione della mediana relativamente alla tabella seguente.

Presenze in un villaggio turistico per classi di età

Classi di età	Numero di presenti	Frequenze cumulate
0 - 20	94	94
20 - 30	104	198
30 - 50	132	330
50 - 60	121	451
60 - 80	70	521
Oltre 80	18	539
Totale	539	—

Il numero di osservazioni è dispari, per cui il rango del posto centrale è:

$$C = \frac{539 + 1}{2} = 270$$

ovvero il turista che presenta l'età mediana è il 270° ed è compreso nella terza frequenza. Ciò sta a significare che la classe mediana è la terza, «età tra i 30 e i 50 anni». Per applicare la formula relativa alla determinazione della mediana si deve considerare che:

$$x_{Me-1} = 30, \alpha_{Me} = 20, N_{Me} = 132, C = 270, C_{Me-1} = 198$$

Si ha, dunque:

$$Me = 30 + \frac{20}{132}(270 - 198) = 40,91$$

Pertanto, l'età mediana stimata è 40,91 anni.

i) Percentili

I *percentili* sono indici di posizione e si definiscono come *quelle intensità che ripartono la distribuzione in due parti, lasciando da una parte una data percentuale dei casi e dall'altra la restante percentuale*.

Si distinguono:

- i *terzili*, che sono 2. Il primo terzile (T_1) lascia alla sua sinistra il 33,3% dei casi e alla sua destra il rimanente 66,7%; il secondo terzile (T_2), invece, lascia alla sua sinistra il 66,7% dei casi e alla sua sinistra il rimanente 33,3% dei casi;
- i *quartili*, che sono 3. Il primo quartile (Q_1) lascia alla sua sinistra il 25% dei casi e alla sua destra il restante 75% dei casi; il secondo quartile (Q_2) lascia alla sua sinistra il 50% dei casi e alla sua destra il restante 50% dei casi (coincide, pertanto, con la mediana); infine, il terzo quartile (Q_3) lascia alla sua sinistra il 75% dei casi e alla sua destra il restante 25% dei casi.

Allo stesso modo, si definiscono i *decili* che sono 9, i *centili* che sono 99 etc.

6 Indici di variabilità

a) Campo di variazione

Il *campo di variazione* è un indice di variabilità definito come *la differenza tra l'intensità massima e l'intensità minima di una variabile statistica*, ossia:

$$\omega = X_{\max} - X_{\min}$$

Esistono degli inconvenienti circa l'uso di questa misura di variabilità. Infatti, essa tiene conto solo dei due valori estremi di una distribuzione, non considerando la struttura interna della stessa. Inoltre, è influenzabile da valori estremi anormali e, quando la classe iniziale e quella finale sono aperte, non è determinabile.

b) Differenza interquartilica

La *differenza interquartilica* è un indice di variabilità determinato dalla *differenza tra il terzo e il primo quartile*, ossia:

$$\delta_q = Q_3 - Q_1$$

Nell'intervallo da esso determinato cade il 50% dei casi.

Al pari del campo di variazione, non tiene conto della struttura interna di una distribuzione.

c) Scostamento semplice medio dalla media aritmetica

Siano dati un collettivo statistico di N unità e una variabile statistica X avente media aritmetica M . Lo *scostamento semplice medio dalla media aritmetica* può essere definito come *la media aritmetica, semplice o ponderata, dei valori assoluti degli scarti dalla media aritmetica*.

Per *dati non raggruppati*, lo scostamento semplice medio dalla media aritmetica è dato da:

$$S_M = \frac{\sum_{i=1}^N |x_i - M|}{N}$$

Per *dati raggruppati*, invece, è dato da:

$$S_M = \frac{\sum_{i=1}^s |x_i - M| N_i}{N}$$

Nel caso in cui il collettivo statistico presenti modalità raggruppate in classi, si sostituisce a ciascun valore x_i il valore centrale di ogni classe.

A differenza del campo di variazione e della differenza interquartilica, tale indice di variabilità tiene conto di tutte le intensità di una distribuzione.

d) Scostamento semplice medio dalla mediana

Siano dati un collettivo statistico di N unità e una variabile statistica X avente mediana Me . Lo *scostamento semplice medio dalla mediana* può essere definito come *la media aritmetica, semplice o ponderata, dei valori assoluti degli scarti dalla mediana*.

Per *dati non raggruppati*, lo scostamento semplice medio dalla mediana è dato da:

$$S_{Me} = \frac{\sum_{i=1}^N |x_i - Me|}{N}$$

Per *dati raggruppati*, invece, è dato da:

$$S_{Me} = \frac{\sum_{i=1}^s |x_i - Me| N_i}{N}$$

Anche per lo scostamento semplice medio dalla mediana, nel caso in cui il collettivo statistico presenti modalità raggruppate in classi, si sostituisce a ciascun valore x_i il valore centrale di ogni classe.

e) Scarto quadratico medio

Lo *scarto quadratico medio* (s.q.m.) è un indice di variabilità, e tra questi ultimi è un «indice di dispersione». Esso è definibile come *la media quadratica degli scostamenti tra le singole intensità e la media aritmetica*.

L'espressione dello s.q.m. per *dati non raggruppati* è la seguente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - M)^2}{N}}$$

Per *dati raggruppati*, la sua espressione è, invece, la seguente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^s (x_i - M)^2 N_i}{N}}$$

In presenza di dati raggruppati in classi, si ricorre al modello delle frequenze concentrate nei valori centrali delle classi.

Per evitare che per caratteri continui, ricorrendo ai valori centrali delle classi, si ottenga un valore approssimato dello s.q.m., si fa uso della cosiddetta *correzione di Sheppard*, grazie alla quale la formula diventa:

$$\sigma = \sqrt{\frac{\sum_{i=1}^s (x_i - M)^2 N_i}{N} - \frac{d^2}{12}}$$

dove d è l'ampiezza costante delle classi.

Lo s.q.m. è un indice di variabilità assoluta e viene espresso nella medesima unità di misura rispetto alla quale si esprimono i dati statistici.

Il rapporto:

$$\frac{\sigma}{M}$$

fra lo s.q.m. e la media aritmetica rappresenta un indice di variabilità relativa ed è chiamato *coefficiente di variabilità*. Tale coefficiente è un numero puro e viene generalmente espresso in percentuale, cioè si pone:

$$\text{coeff. di var.} = \frac{\sigma}{M} \cdot 100$$

Spesso vengono divise le singole intensità di una data variabile statistica per lo s.q.m., ottenendo misure standardizzate. Altre volte, invece, si dividono i corrispondenti scarti dalla media per lo s.q.m., ottenendo gli *scarti standardizzati*:

$$s = \frac{X - M}{\sigma}$$

f) Devianza e varianza

La *devianza* è un indice assoluto di variabilità definibile come *la somma, semplice o ponderata, dei quadrati degli scarti tra le singole intensità e la media aritmetica*.

Per dati non raggruppati, l'espressione della devianza è:

$$D(X) = \sum_{i=1}^N (x_i - M)^2$$

Per dati raggruppati, l'espressione della devianza è:

$$D(X) = \sum_{i=1}^s (x_i - M)^2 N_i$$

La *varianza* è il rapporto tra la devianza e la numerosità N del collettivo, ossia:

$$\sigma^2 = \frac{D(X)}{N}$$

La varianza è, dunque, il quadrato dello s.q.m. ed è espressa nel quadrato dell'unità di misura del fenomeno oggetto di studio.

Per evitare che il calcolo della devianza e, conseguentemente, della varianza sia troppo laborioso a causa della presenza degli scarti, si utilizzano le seguenti formule semplificate:

$$D(X) = \sum_{i=1}^s x_i^2 N_i - M^2 N \quad \text{e} \quad \sigma^2 = \frac{\sum_{i=1}^s x_i^2 N_i}{N} - M^2$$

Per cui l'espressione dello scarto quadratico medio è la seguente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^s x_i^2 N_i}{N} - M^2}$$

Considerando la definizione di media quadratica, lo scarto quadratico medio e la varianza possono essere scritti in questo modo:

$$\sigma = \sqrt{M_q^2 - M^2} \quad \text{e} \quad \sigma^2 = M_q^2 - M^2$$

ESEMPIO 15

Calcolare la varianza della distribuzione delle altezze di 125 atleti, riportata nella tabella seguente.

Atleti per classi di altezze

Classi di altezze	Numero di atleti
170 - 175	14
175 - 180	18
180 - 185	28
185 - 190	33
190 - 195	17
195 - 200	15
Totale	125

Per determinare la varianza attraverso la formula semplificata è sufficiente conoscere il quadrato della media quadratica della distribuzione e il corrispondente quadrato della media aritmetica.

La media aritmetica della distribuzione, ricorrendo ai valori centrali delle classi, è:

$$M = \frac{172,5 \cdot 14 + 177,5 \cdot 18 + 182,5 \cdot 28 + 187,5 \cdot 33 + 192,5 \cdot 17 + 197,5 \cdot 15}{125} = 185,14$$

il cui quadrato è:

$$M^2 = 34.276,82$$

Il quadrato della media quadratica è:

$$M_q^2 = \frac{172,5^2 \cdot 14 + 177,5^2 \cdot 18 + 182,5^2 \cdot 28 + 187,5^2 \cdot 33 + 192,5^2 \cdot 17 + 197,5^2 \cdot 15}{125} = 34.331,85$$

La varianza della distribuzione di dati di cui alla tabella è:

$$\sigma^2 = 34.331,85 - 34.276,82 = 55,03$$

g) Differenze medie

• Differenza media con ripetizione

La *differenza media con ripetizione* è un indice assoluto di variabilità che misura la *disuguaglianza media* tra le intensità di una variabile statistica. Essa si ottiene confrontando tutti i termini due a due, non escludendo neppure il confronto di un termine con se stesso; può essere espressa dalla seguente formula:

$$\Delta_R = \frac{\sum_{i,j} |x_i - x_j|}{N^2}$$

Il numero totale delle differenze tra le intensità è N^2 , di cui N sono nulle perché riferite al confronto tra valori uguali ($x_i - x_i = 0$); quelle diverse da zero sono pari a: $N^2 - N = N(N - 1)$.

• Differenza media senza ripetizione

La *differenza media senza ripetizione* è un indice assoluto di variabilità che misura la *disuguaglianza media* tra le intensità di una variabile statistica. Essa si ottiene confrontando tutti i termini due a due, escludendo il confronto di ciascun termine con se stesso; può essere espressa dalla seguente formula:

$$\Delta = \frac{\sum_{i,j} |x_i - x_j|}{N(N - 1)}$$

I confronti fra le intensità sono, questa volta, pari a: $N^2 - N = N(N - 1)$.

Dividendo tra loro la differenza media con ripetizione e la differenza media senza ripetizione, si ottiene la seguente relazione:

$$\Delta_R = \frac{N-1}{N} \Delta$$

h) Indici rapportati al massimo della variabilità

Un modo per ottenere un indice relativo di variabilità è di rapportare l'indice assoluto al corrispondente indice della distribuzione massimante, ossia della distribuzione che presenti massima variabilità. Si ottiene, in questa maniera, un indice che varia tra 0 e 1, e che assume valore 0 in caso di variabilità nulla e valore 1 in caso di variabilità massima.

Il procedimento da seguire è comune a tutti gli indici e si basa sulle seguenti ipotesi:

- le determinazioni della variabile statistica devono essere comprese in un intervallo definito da un valore minimo $x_{(1)}$ e da un valore massimo $x_{(N)}$. Tali valori possono anche non essere i valori minimo e massimo osservati della variabile X ;
- la distribuzione massimante si ottiene supponendo che le N determinazioni assumano solo il valore minimo e il valore massimo, con frequenze pari, rispettivamente, a d e a $N - d$;
- la distribuzione massimante deve avere la stessa media aritmetica della distribuzione osservata, ossia:

$$\frac{x_{(1)}d + x_{(N)}(N - d)}{N} = M$$

La quantità al primo membro dell'uguaglianza rappresenta proprio la media aritmetica della distribuzione massimante.

- Scarto quadratico medio in rapporto al suo massimo

Lo s.q.m. della distribuzione massimante è:

$$\sigma_{max} = \sqrt{\frac{(x_{(1)} - M)^2 d + (x_{(N)} - M)^2 (N - d)}{N}}$$

Attraverso semplici passaggi si giunge alla seguente espressione dello s.q.m. della distribuzione massimante:

$$\sigma_{max} = \sqrt{(M - x_{(1)})(x_{(N)} - M)}$$

L'indice relativo di variabilità *s.q.m. in rapporto al suo massimo* è, pertanto:

$$\frac{\sigma}{\sigma_{max}} = \frac{\sigma}{\sqrt{(M - x_{(1)})(x_{(N)} - M)}}$$

- Differenza media senza ripetizione in rapporto al suo massimo

La differenza media della distribuzione massimante è:

$$\Delta_{max} = \frac{2N(M - x_{(1)})(x_{(N)} - M)}{(N - 1)(x_{(N)} - x_{(1)})}$$

L'indice relativo di variabilità *differenza media senza ripetizione in rapporto al suo massimo* è, pertanto:

$$\frac{\Delta}{\Delta_{max}}$$

i) Indici di concentrazione

Un aspetto della variabilità è la *concentrazione*, la quale ha senso solo per caratteri additivi e trasferibili (come il reddito, il capitale azionario etc.), vale a dire solo se la sua intensità globale o una sua parte possono essere possedute da una o più unità.

Un carattere, viceversa, si dice *equidistribuito* se non è concentrato in assoluto, vale a dire determinate frazioni del suo ammontare appartengono a uguali frazioni di unità statistiche.

Per misurare il *grado di concentrazione* di una variabile statistica X riguardante un collettivo di N unità, si dispongono le sue determinazioni in ordine non decrescente (cioè sia: $0 \leq x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$) e si ragiona nel modo seguente:

- la prima unità osservata costituisce $\frac{1}{N}$ del totale, le prime due costituiscono $\frac{2}{N}$ del totale, e così via, indicando con p_i tali frazioni cumulate, si ha:

$$p_1 = \frac{1}{N}, \quad p_2 = \frac{2}{N}, \quad p_i = \frac{i}{N}, \quad \dots; \quad p_N = \frac{N}{N} = 1$$

- l'intensità della prima unità, $x_{(1)}$, è indicata con s_1 , le intensità delle prime due unità, $x_{(1)} + x_{(2)}$, sono indicate con s_2 , e così via; si ha:

$$s_1 = x_{(1)}; \quad s_2 = x_{(1)} + x_{(2)}; \quad \dots; \quad s_i = x_{(1)} + x_{(2)} + \dots + x_{(i)}; \quad \dots; \quad s_N = x_{(1)} + x_{(2)} + \dots + x_{(N)};$$

- si indica con $q_1 = \frac{S_1}{S_N}$ la frazione dell'intensità della prima unità sull'intensità globale, con $q_2 = \frac{S_2}{S_N}$ la frazione dell'intensità delle prime due unità sull'intensità globale, e così via; si ha:

$$q_1 = \frac{S_1}{S_N}; \quad q_2 = \frac{S_2}{S_N}; \quad \dots; \quad q_i = \frac{S_i}{S_N}; \quad \dots; \quad q_N = \frac{S_N}{S_N} = 1$$

In caso di equidistribuzione, si ha per ogni i :

$$p_i = q_i$$

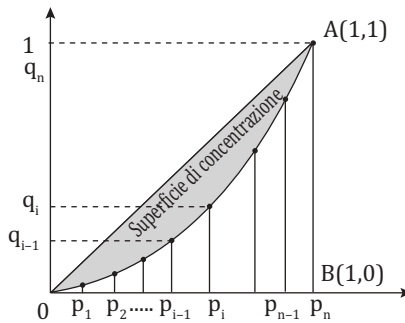
Se non esiste equidistribuzione, necessariamente, si ha:

$$p_1 > q_1; \quad p_2 > q_2; \quad \dots; \quad p_i > q_i; \quad \dots; \quad p_N = q_N$$

Nella situazione di massima concentrazione, si ha:

$$q_0 = q_1 = q_2 = \dots = q_{N-1} = 0; \quad q_N = 1$$

È possibile analizzare i diversi casi di distribuzione del carattere ricorrendo ad un sistema di riferimento cartesiano, vale a dire ad un diagramma i cui punti-immagine sono le coppie (p_i, q_i) — per convenzione si pone $p_0 = q_0 = 0$ — e che è denominato *spezzata di concentrazione*.



La spezzata di concentrazione è compresa in un quadrato avente lato unitario [infatti, il vertice è $A(p_N, q_N) = A(1,1)$]; inoltre, essa ha andamento ascendente, essendo ogni sua ordinata q_i maggiore della precedente q_{i-1} .

Unendo i punti (p_i, q_i) con una linea continua si ottiene la cosiddetta *curva di concentrazione* o *curva di Lorenz*, della quale si considerino le due situazioni limite:

- di *equidistribuzione*, la cui rappresentazione grafica è fornita dal *segmento di equidistribuzione*, vale a dire dalla diagonale OA del quadrato;
- di *massima concentrazione*, la cui rappresentazione grafica è fornita dal triangolo OBA , in quanto tutti i punti della curva ricadono sull'asse delle ascisse, tranne l'ultimo di coordinate $A(1,1)$.

La *superficie di concentrazione* è quella delimitata dalla curva di concentrazione e dal segmento di equidistribuzione, aumenta al crescere della concentrazione ed assume valore massimo 0,5, che equivale all'area del triangolo rettangolo OBA , avente entrambi i cateti uguali a 1.

La misura della concentrazione è fornita dal *rapporto di concentrazione di Gini*:

$$R = \frac{\sum_{i=1}^{N-1} (p_i - q_i)}{\sum_{i=1}^{N-1} p_i}$$

Il rapporto R è un numero puro che assume valore:

- 0 → in caso di equidistribuzione, in quanto tutte le differenze $p_i - q_i$ sono nulle;
- 1 → in caso di massima concentrazione, in quanto tutte le q_i fino alla $(N - 1)$ esima sono nulle.

Tale misura di concentrazione si serve del cosiddetto *metodo dei rettangoli*. Si dimostra, infatti, che la quantità:

$$\frac{1}{N} \sum_{i=1}^{N-1} (p_i - q_i)$$

è un valore approssimato della superficie di concentrazione; essa rappresenta la somma delle aree dei rettangoli di base $\frac{1}{N}$ (che è la lunghezza costante di ogni segmento in cui la base OB del quadrato è suddivisa dai punti aventi ascisse $p_1, p_2, \dots, p_N$) e altezza $p_i - q_i$.

In maniera analoga, si dimostra che la quantità:

$$\frac{1}{N} \sum_{i=1}^{N-1} p_i$$

è un valore approssimato dell'area del triangolo OBA (rappresentante la massima concentrazione).

Pertanto, R può essere visto come il rapporto tra l'area di concentrazione e l'area approssimata di massima concentrazione.

L'area di concentrazione, che è essa stessa una misura di concentrazione, e che varia tra 0 e 1/2, è moltiplicata per 2, per disporre di un indice normalizzato che vari tra 0 e 1, vale a dire:

$$R^* = 2 \text{ area di concentrazione}$$

Con il *metodo dei trapezi* si giunge alla formula per il calcolo di R^* .

L'area di concentrazione S_c è data dalla differenza tra l'area del triangolo OBA (1/2) e l'area residua S_r rispetto al triangolo OBA , vale a dire:

$$S_c = \frac{1}{2} - S_r$$

da cui:

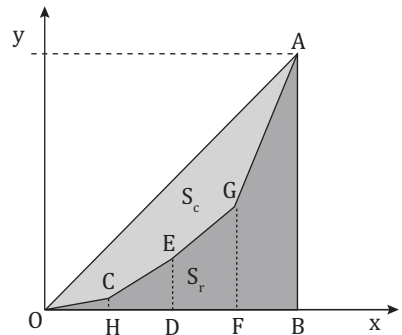
$$R^* = 2 \left(\frac{1}{2} - S_r \right) = 1 - 2S_r$$

L'area residua S_r si determina considerando che essa è uguale all'area del triangolo OHC e dei trapezi successivi $CEDH$, $EGFD$ etc. aventi per basi q_i e q_{i+1} e per altezza $p_{i+1} - p_i$. Siccome si è posto, per convenzione, $p_0 = q_0 = 0$, si ha che il triangolo OHC ha per cateti $p_1 = p_1 - p_0$ e q_1 . Da ciò si deduce che l'area residua, applicando la formula per il calcolo della superficie di un trapezio, è uguale a:

$$S_r = \frac{1}{2} \sum_{i=0}^{N-1} (p_{i+1} - p_i)(q_{i+1} + q_i)$$

Da tale formula si deduce quella di R^* , vale a dire:

$$R^* = 1 - \sum_{i=0}^{N-1} (p_{i+1} - p_i)(q_{i+1} + q_i)$$



In caso di *dati raggruppati*, si possono individuare due situazioni diverse:

- se non si presentano classi di valori, allora il calcolo di R si riconduce a quello della differenza media senza ripetizione, mentre il calcolo di R^* si riconduce a quello della differenza media con ripetizione;
- se si presentano classi di valori, allora conviene utilizzare il *metodo dei trapezi*.

L'indice δ del Gini è un altro indice di concentrazione e, a differenza del rapporto di concentrazione, fornisce informazioni sulla concentrazione del fenomeno, non tenendo conto della superficie compresa tra la spezzata di concentrazione e il segmento di equidistribuzione.

Tale indice si basa sulle quantità $1 - p_i$ e $1 - q_i$. Esse sono sicuramente inferiori all'unità, per cui esiste certamente un δ_i tale che:

$$(1 - q_i)^{\delta_i} = 1 - p_i$$

L'indice si ricava ricorrendo ai logaritmi ed è tanto più grande quanto più alta è la concentrazione.

ESEMPIO 16

Calcolare, attraverso il metodo dei trapezi, la misura della concentrazione dei dati di cui alla tabella seguente, in cui con x_i si indica il reddito mensile dell' i -esima unità (percettore di reddito), con N_i la frequenza (l' i -esimo percettore di reddito), con C_i la frequenza cumulata dei primi i individui, con S_i il reddito cumulato delle prime i unità, con p_i la frazione cumulata dei primi i redditi, e con q_i la frazione cumulata dei primi i redditi.

Distribuzione dei redditi di 4 individui

x_i	N_i	C_i	S_i	p_i	q_i
1.500	1	1	1.500	0,25	0,223
1.650	1	2	3.150	0,50	0,468
1.730	1	3	4.880	0,75	0,726
1.840	1	4	6.725	1,00	1,000

Per la determinazione di R^* si rende necessario ricorrere alla tabella seguente.

I	p_i	$p_{i+1} - p_i$	q_i	$q_{i+1} + q_i$	$(p_{i+1} - p_i)(q_{i+1} + q_i)$
0	0,00	0,25	0,000	0,223	0,056
1	0,25	0,25	0,223	0,691	0,173
2	0,50	0,25	0,468	1,194	0,298
3	0,75	0,25	0,726	1,726	0,431
4	1,00	—	1,000	—	—
Totale					0,958

Applicando la formula, si ha:

$$R^* = 1 - 0,958 = 0,042$$

Il valore ottenuto di R^* evidenzia come vi sia una bassa concentrazione dei redditi tra i 4 individui considerati.

j) Momenti

Il *momento* è un parametro distributivo dato dalla media aritmetica, semplice o ponderata, delle potenze r^{me} delle differenze tra le intensità della variabile statistica e un polo λ .

L'espressione del *momento* per *dati non raggruppati* è:

$${}_{{}_\lambda} m_r = \frac{\sum_{i=1}^N (x_i - \lambda)^r}{N}$$

mentre l'analoga espressione per *dati raggruppati* è:

$${}_{{}_\lambda} m_r = \frac{\sum_{i=1}^s (x_i - \lambda)^r N_i}{N}$$

Si distinguono diversi tipi di momenti:

- *rispetto all'origine*, se il polo λ assume valore 0;
- *centrali*, se il polo λ coincide con la media aritmetica M .

7 Indici di forma

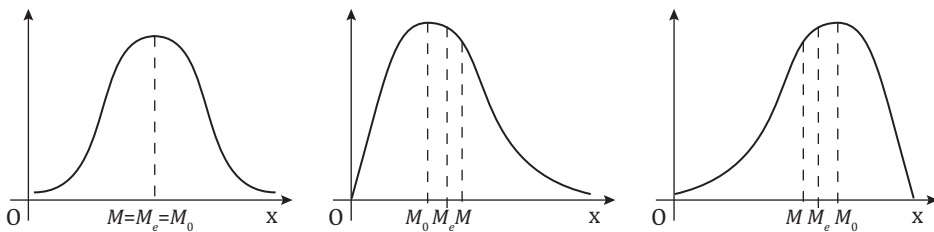
a) Indici di asimmetria

Come è noto, la mediana bipartisce una distribuzione. Graficamente, una distribuzione è *simmetrica* se per il suo diagramma in coordinate cartesiane, il ramo a destra della mediana è sovrapponibile a quello di sinistra attraverso un ribaltamento.

Con le notazioni già viste, si considerino le relazioni tra media aritmetica, mediana e moda, per i diversi tipi di distribuzione; infatti, si parla di:

- *simmetria*, se $M = M_e = M_0$;
- *asimmetria positiva*, se $M_0 < M_e < M$. In tale distribuzione, il ramo destro è più allungato di quello sinistro;
- *asimmetria negativa*, se $M < M_e < M_0$. In tale distribuzione, il ramo sinistro è più allungato di quello destro.

Nelle tre figure seguenti sono riportati esempi, rispettivamente, di una curva che presenta *simmetria*, di una che presenta *asimmetria a destra* e di una che presenta *asimmetria a sinistra*.



Esistono diversi *indici di asimmetria*, ossia indici di forma atti a valutare il grado di asimmetria di distribuzioni statistiche.

Essi si distinguono in *misure assolute* e *misure relative*.

• Misure assolute

Tali misure si ottengono dalle differenze:

$$M - M_e \quad \text{e} \quad M - M_0$$

che sono *nulle* se esiste *simmetria*; *positive* se esiste *asimmetria positiva*; *negative* se esiste *asimmetria negativa*.

Talvolta, come *misura assoluta* viene scelta quella del *momento centrale del terzo ordine*:

$${}_M m_3 = \frac{\sum_{i=1}^s (x_i - M)^3 N_i}{N}$$

che è nullo per distribuzioni simmetriche; positivo per distribuzioni con asimmetria positiva; negativo per distribuzioni con asimmetria negativa.

- **Misure relative**

Tali misure si ottengono dai rapporti:

$$\alpha_1 = \frac{M - M_e}{\sigma} \quad \text{e} \quad \alpha_2 = \frac{M - M_o}{\sigma}$$

A volte, come *misura relativa* viene scelta quella dell'*indice di asimmetria di Fisher*, dato dal momento centrale del terzo ordine rapportato al cubo dello s.q.m.:

$$\alpha_F = \frac{1}{\sigma^3} \frac{\sum_{i=1}^s (x_i - M)^3 N_i}{N}$$

ESEMPIO 17

Calcolare l'indice di asimmetria di Fisher per la distribuzione di dati di cui alla tabella dell'esempio 15.

Della distribuzione considerata non solo è nota la media aritmetica $M = 185,14$, ma anche la varianza; per cui, dalla radice quadrata di quest'ultima, si ottiene lo s.q.m., vale a dire:

$$\sigma = \sqrt{\sigma^2} = \sqrt{55,03} = 7,418$$

L'indice di asimmetria di Fisher è:

$$\begin{aligned} \alpha_F &= \frac{1}{7,418^3} \cdot \left[(172,5 - 185,14)^3 \cdot 14 + (177,5 - 185,14)^3 \cdot 18 + \right. \\ &\quad \left. + (182,5 - 185,14)^3 \cdot 28 + (187,5 - 185,14)^3 \cdot 33 + \right. \\ &\quad \left. + (192,5 - 185,14)^3 \cdot 17 + (197,5 - 185,14)^3 \cdot 15 \right] \cdot \frac{1}{125} = \\ &= \frac{1}{408,188} \cdot [-1280,064] \cdot \frac{1}{125} = -0,025 \end{aligned}$$

La distribuzione presenta lieve asimmetria negativa.

b) Indice di curtosi

La *curtosi* è un indice di forma legato alla maggiore o minore gibbosità di una curva nelle vicinanze del suo massimo e alla maggiore o minore lunghezza delle code, rispetto alla curva Normale, la quale è la curva continua rappresentativa delle distribuzioni che più frequentemente si incontrano in Statistica.

Una curva si dice:

- *leptocurtica* o *ipernormale* se ha la sommità più appuntita della distribuzione normale, con un aspetto snello in prossimità del massimo;
- *mesocurtica* o *normale* se ha le caratteristiche morfologiche di una distribuzione normale;
- *platicurtica* o *iponormale* se ha la sommità più appiattita della distribuzione normale, con un aspetto copioso in prossimità del massimo.

Il grado di curtosi di una distribuzione si misura servendosi dell'*indice di Pearson*, dato dal rapporto tra il momento centrale del quarto ordine e la quarta potenza dello s.q.m.:

$$\gamma_p = \frac{1}{\sigma^4} \frac{\sum_{i=1}^s (x_i - M)^4 N_i}{N} - 3$$

Tale indice assume *valore 0 per distribuzioni normali*, *valori positivi per distribuzioni leptocurtiche* e *valori negativi per distribuzioni platicurtiche*.

8 I rapporti statistici

I *rapporti statistici* sono misure elementari finalizzate al confronto tra i dati stessi. In un rapporto statistico si mettono a confronto due termini, frequenze o quantità, di cui uno almeno è di natura statistica; inoltre, tra i due termini deve sussistere un qualche legame logico. I rapporti così costruiti permettono di confrontare l'intensità di un fenomeno misurato su un collettivo, in tempi o luoghi diversi, e sono largamente impiegati nella descrizione di fenomeni di tipo socio-economico.

Assumono un valore sempre positivo e non dipendono dall'unità di misura, consentendo così un confronto tra fenomeni diversi.

A questo punto è bene dare alcuni cenni sui diversi tipi di rapporti statistici.

- **Rapporti di composizione**

I *rapporti di composizione* mettono in relazione una modalità del fenomeno con il fenomeno nel suo complesso.

- **Rapporti di densità**

I *rapporti di densità* mettono in relazione la frequenza complessiva di un collettivo o l'intensità globale di un dato carattere con una dimensione di spazio o di tempo; essi sono particolarmente utilizzati per esprimere il cosiddetto *grado di affollamento* delle abitazioni.

- **Rapporti di coesistenza**

I *rapporti di coesistenza* si determinano dal quoziente tra le intensità di due fenomeni diversi nello stesso luogo, o tra le intensità di uno stesso fenomeno in due luoghi diversi.

- **Rapporti di durata e di ripetizione**

I *rapporti di durata* si determinano rapportando la consistenza di un dato fenomeno alla parte del fenomeno che si rinnova nell'unità di tempo considerata.

Siccome i flussi di rinnovo non sempre sono uguali tra loro, i rapporti di durata si esprimono rapportando la consistenza media di un fenomeno alla media aritmetica dei flussi di entrata e di uscita. Indicando con C_0 e C_1 la consistenza all'inizio del periodo considerato e la consistenza alla fine del periodo, e sapendo che $C_1 = C_0 + (E - U)$, dove E e U sono, rispettivamente, i flussi in entrata e i flussi in uscita, la *consistenza media* è data da:

$$C = \frac{C_0 + C_1}{2} = \frac{C_0 + (C_0 + E - U)}{2} = \frac{2C_0 + E - U}{2}$$

Il rapporto di durata è, pertanto, dato da:

$$d = \frac{2C_0 + E - U}{E + U}$$

I *rapporti di ripetizione* si determinano, invece, rapportando il flusso di rinnovo di un dato fenomeno alla sua consistenza media.

Mentre i *rapporti di durata* esprimono la *vita media di un elemento di una data popolazione*, i *rapporti di ripetizione* esprimono *quante volte la popolazione considerata si rinnova nell'unità di tempo*.

• Rapporti di derivazione

I *rapporti di derivazione* si determinano rapportando l'intensità o la frequenza di un dato fenomeno all'intensità o frequenza di un altro fenomeno, che costituisce il presupposto del primo e per questo è detto *fenomeno generante*; tali rapporti sono, quindi, utilizzati quando esiste un nesso di causalità tra due fenomeni.

• Variazioni percentuali

Le *variazioni relative* di un dato fenomeno riferite all'unità di tempo si denominano *saggi di incremento o di decremento* e si esprimono generalmente in *percentuali*. Esse si ottengono effettuando il rapporto tra l'incremento assoluto (o il decremento assoluto) verificatosi in un fenomeno in un dato periodo e l'intensità dello stesso fenomeno all'inizio del periodo considerato; moltiplicando la variazione relativa così ottenuta per 100 si ottiene l'incremento (o il decremento) relativo percentuale.

ESEMPIO 18

Determinare l'incremento relativo percentuale di un fenomeno che, dal tempo 0 al tempo 1, passa da un valore di 50 ad un valore di 100.

Il fenomeno passa da un valore iniziale di 50 ad un valore di 100, per cui l'*incremento assoluto* è pari a: $100 - 50 = 50$; l'*incremento relativo*, invece, è pari a: $\frac{100 - 50}{50} = 1$; moltiplicando quest'ultimo valore per 100 si ottiene l'*incremento relativo percentuale*, che ovviamente in questo caso è uguale al 100%.

• Numeri indici

I *numeri indici* consentono di confrontare le intensità di un fenomeno in situazioni temporali e/o spaziali differenti.

I numeri indici vengono distinti in:

- *indici semplici*, utilizzati quando ci si riferisce a fenomeni semplici o elementari;
- *indici sintetici*, utilizzati quando ci si riferisce a fenomeni complessi, costituiti da insiemi di fenomeni semplici della stessa natura;
- *indici composti*, utilizzati quando ci si riferisce a fenomeni complessi, costituiti da insiemi di fenomeni semplici di natura diversa.

Bisogna, inoltre, fare una distinzione tra *numeri indici a base fissa* e *numeri indici a base mobile*.

In una serie di numeri indici a base fissa, ciascuna intensità è rapportata a una sola di esse, assunta come *base*.

In una serie di numeri indici a base mobile, ciascuna intensità è rapportata a quella precedente, assumendo quest'ultima come base; tali indici sono anche detti *concatenati*.

9 Distribuzioni statistiche doppie

Se per ciascun elemento del collettivo statistico si rilevano due o più caratteri si ottengono le *distribuzioni statistiche multiple*.

In questo contesto considereremo solo le *distribuzioni statistiche doppie*, che, come il termine stesso dice, fanno riferimento a due caratteri.

a) Variabili statistiche doppie o distribuzioni bivariate

Sia dato un collettivo di N unità statistiche delle quali si vogliano rilevare due caratteri quantitativi X e Y , e si indichino con:

- $x_1, x_2, \dots, x_p$ le p modalità del carattere X ;
- $y_1, y_2, \dots, y_q$ le q modalità del carattere Y ;
- N_{ij} la frequenza con cui si presenta la coppia di modalità (x_i, y_j) ;
- $N_{i.}$ il totale parziale di colonna, che indica il numero di volte in cui si presenta la modalità x_i del carattere X , a prescindere da come essa sia accoppiata con le modalità del carattere Y ;
- $N_{.j}$ il totale parziale di riga, che indica il numero di volte in cui si presenta la modalità y_j del carattere Y , a prescindere da come essa sia accoppiata con le modalità del carattere X .

La tabella statistica descrittiva dei dati è denominata *tabella statistica a doppia entrata*.

La prima riga, in cui si evidenziano le modalità del carattere X , è denominata *riga madre*; la prima colonna, in cui si evidenziano le modalità del carattere Y , è denominata *colonna madre*.

La parte centrale della tabella è una matrice costituita da q righe e p colonne; in ogni cella sono evidenziate le frequenze corrispondenti alle varie coppie di modalità dei caratteri X e Y .

L'ultima riga e l'ultima colonna sono denominate *riga marginale* e *colonna marginale* e indicano, rispettivamente, i totali parziali di colonna e di riga.

Y \ X	x_1	x_2	...	x_i	...	x_p	Totale
y_1	N_{11}	N_{21}	...	N_{i1}	...	N_{p1}	$N_{.1}$
y_2	N_{12}	N_{22}	...	N_{i2}	...	N_{p2}	$N_{.2}$
...	...	...	...	...	...	...	...
y_j	N_{1j}	N_{2j}	...	N_{ij}	...	N_{pj}	$N_{.j}$
...	...	...	...	...	...	...	...
y_q	N_{1q}	N_{2q}	...	N_{iq}	...	N_{pq}	$N_{.q}$
Totale	$N_{.1}$	$N_{.2}$	...	$N_{.i}$	...	$N_{.p}$	$N_{..}$

b) Mutabili doppie

Sia dato un collettivo di N unità statistiche delle quali si vogliano rilevare due caratteri qualitativi A e B , e si indichino con:

- $A_1, A_2, \dots, A_p$ le p modalità del carattere A ;
- $B_1, B_2, \dots, B_q$ le q modalità del carattere B ;
- N_{ij} la frequenza con cui si presenta la coppia di modalità (A_i, B_j) ;
- $N_{i.}$ il totale parziale di colonna, che indica il numero di volte in cui si presenta la modalità A_i del carattere A , a prescindere da come essa sia accoppiata con le modalità del carattere B ;
- $N_{.j}$ il totale parziale di riga, che indica il numero di volte in cui si presenta la modalità B_j del carattere B , a prescindere da come essa sia accoppiata con le modalità del carattere A .

La tabella statistica a doppia entrata, descrittiva dei dati, è analoga alla corrispondente per *variabili doppie*.

Si dicono *distribuzioni marginali* determinate distribuzioni semplici ottenute associando:

- la riga madre con l'ultima riga delle frequenze marginali;
- la colonna madre con l'ultima colonna delle frequenze marginali.

10 Connessione e concordanza/discordanza

Le relazioni statistiche si distinguono in *generiche* e *specifiche*.

Le prime studiano la *connessione*, cioè accertano che una influenza (unidirezionale o bidirezionale) esista tra due caratteri. Le seconde studiano la *concordanza* o la *discordanza*, ossia stabiliscono in che modo l'influenza tra due caratteri si manifesta.

La connessione sussiste tra fenomeni quantitativi, ma anche tra fenomeni qualitativi, o tra un fenomeno che è qualitativo e un fenomeno che è quantitativo; viceversa, si ha concordanza o discordanza solo tra due caratteri commensurabili, pertanto, o quando entrambi sono quantitativi o quando entrambi sono qualitativi con le stesse modalità.

11 Regressione

Un modello può essere definito come una rappresentazione formale delle conoscenze relative ad un fenomeno. Con la sua costruzione non si realizza una esatta riproduzione del fenomeno studiato ma se ne fornisce una versione semplificata.

I modelli si dividono in due grandi categorie, ciascuna delle quali presuppone assunzioni e conoscenze diverse: *modelli per serie storiche* e *modelli di regressione*.

I modelli per serie storiche basano la loro conoscenza e il loro comportamento sulla storia del fenomeno oggetto di studio.

I modelli di regressione si basano sull'esistenza di una relazione causa-effetto tra il fenomeno osservato e una o più variabili esplicative; essi sono costituiti da una sola o più equazioni lineari (o non lineari) nei parametri.

Bisogna fare un'ulteriore distinzione tra *modelli di regressione semplice* e *multipla*. I primi descrivono come una data variabile (indipendente o esplicativa) «spieghi» un'altra variabile (dipendente); i secondi sono utilizzati quando le variabili indipendenti sono più di una.

a) Regressione lineare semplice

Il *modello di regressione semplice* intende spiegare il legame funzionale esistente tra una variabile x indipendente ed una variabile y dipendente.

Tuttavia, questo modo di procedere presenta degli inconvenienti. Infatti, non è possibile affermare che tra le due variabili x e y esista una perfetta relazione matematica del tipo:

$$y = f(x)$$

in quanto non si dispone di tutte le informazioni relative al fenomeno x e al fenomeno y , ma solo di un campione di osservazioni su di essi. Tale campione è estratto da una data popolazione, che darà luogo a coppie del tipo $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, le quali, solo in termini probabilistici, esprimono la relazione esistente tra le corrispondenti variabili della popolazione. Inoltre, a causa di fenomeni imprevedibili (errori di misurazione, scarti accidentali etc.), non esiste un legame di tipo deterministico tra le variabili. Costituisce, quindi, una semplificazione affermare che il fenomeno y è «spiegato» dal fenomeno x , poiché nella realtà esistono interrelazioni tra variabili che non possono essere compendiate in alcuna scrittura. Per questi motivi, il modello di riferimento sarà del tipo:

$$y = f(x) + \varepsilon$$

La variabile y è una variabile casuale risultante dalla somma di una componente deterministica $f(x)$ e di una componente stocastica ε , dove ε è una variabile casuale che ha funzioni compensative per le discrepanze esistenti tra il modello e la realtà.

La variabile casuale ε è definita *errore* o *scarto* tra le costruzioni teoriche e la realtà osservata; infatti, si ha:

$$\varepsilon = y - f(x)$$

Dopo aver provveduto ad individuare la variabile x che «spiega» il fenomeno in oggetto, si passa alla fase di specificazione del modello, che consiste nella sua rappresentazione formale.

Si assume che la relazione tra le variabili x e y sia lineare nei parametri, in modo da rendere il modello di regressione pari a:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

Tale relazione è valida per la generica coppia di osservazioni (x_i, y_i) . La variabile y è data, dunque, dalla somma di $\beta_0 + \beta_1 x_i$ (parte sistematica) e ε_i (parte stocastica).

Il modello di regressione lineare si basa su diverse ipotesi:

1. $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \dots, n$
2. $E(\varepsilon_i) = 0$
3. $Var(\varepsilon_i) = \sigma^2$
4. $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \text{per } i \neq j$
5. x è una variabile deterministica.

La prima è l'*ipotesi lineare*, la quale specifica la variabile dipendente y , osservata negli n tempi, come una funzione lineare della variabile esplicativa x , cui si aggiunge, in ogni tempo, una manifestazione non osservabile della variabile aleatoria ε .

La seconda è l'*ipotesi di non sistematicità degli errori* e deriva dalla necessità che scarti positivi e negativi si compensino in media 0, il che è lo stesso, che i termini di errore siano mediamente nulli.

La terza e la quarta ipotesi indicano, rispettivamente, che le componenti stocastiche non dipendono dal tempo di rilevazione del fenomeno (*ipotesi di omoschedasticità*) e sono tra loro incorrelate, cioè hanno covarianza nulla.

La quinta ipotesi considera i dati relativi alla variabile indipendente x come non soggetti a deviazioni di natura accidentale.

Lo stadio successivo, relativo alla costruzione di un modello, riguarda la stima dei parametri; è necessario, pertanto, trovare una relazione lineare tra la variabile x e la variabile y , in modo da minimizzare il valore della componente stocastica. Dal punto di vista geometrico è necessario trovare quella retta $y = \beta_0 + \beta_1 x$ che attraversa la nuvola di punti (x_i, y_i) per $i = 1, 2, \dots, n$.

L'equazione generica della retta di regressione è:

$$y = \beta_0 + \beta_1 x$$

dove β_0 è l'intercetta e β_1 è il coefficiente angolare.

Se il calcolo dei due parametri è effettuato con il metodo dei minimi quadrati, proposto da Gauss, si sceglie quella retta per cui si verifica che la somma dei quadrati degli scostamenti tra i valori teorici e quelli osservati del carattere y sia minima o, il che è lo stesso, si minimizza la quantità:

$$S = S(\beta_0, \beta_1) = \sum (y_i - \beta_0 - \beta_1 x_i)^2 = \text{minimo rispetto a } (\beta_0, \beta_1)$$

Uguagliando a zero le derivate parziali rispetto ai parametri, si ha:

$$\begin{cases} \frac{\partial S}{\partial \beta_0} = -2 \sum (y_i - \beta_0 - \beta_1 x_i) = 0 \\ \frac{\partial S}{\partial \beta_1} = -2 \sum (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \end{cases}$$

che, semplificando, si riduce al sistema di due equazioni lineari nelle due incognite β_0 e β_1 :

$$\begin{cases} \beta_0 n + \beta_1 \sum x_i = \sum y_i \\ \beta_0 \sum x_i + \beta_1 \sum x_i^2 = \sum x_i y_i \end{cases}$$

Le relazioni del sistema sono le equazioni normali, la cui soluzione è:

$$\begin{cases} \hat{\beta}_0 = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2} \\ \hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \end{cases}$$

La fase di verifica è volta ad accertare la validità del modello sulla base delle osservazioni disponibili.

b) Regressione lineare multipla

Nella realtà è opportuno analizzare modelli in cui le variabili indipendenti, che «spiegano» una variabile dipendente, sono più di una; si usa a tal fine il *modello di regressione multipla*.

La rappresentazione formale del modello considera la variabile dipendente y come una funzione lineare delle k variabili esplicative $x_1, x_2, \dots, x_k$.

Il modello di regressione multipla sarà:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

Le ipotesi su cui si basa tale modello sono:

— *ipotesi di linearità*:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

Tale ipotesi specifica la variabile dipendente y , osservata negli n tempi, come una funzione lineare tramite i coefficienti β delle variabili x_j (con $j = 1, 2, \dots, k$), cui si aggiunge, in ogni tempo, una manifestazione ε della variabile aleatoria.

Il modello può essere rappresentato in forma compatta nel seguente modo:

$$Y = X\beta + \varepsilon$$

dove Y è il vettore costituito dall'insieme ordinato delle n osservazioni sulla variabile dipendente, X è la matrice ottenuta dall'insieme ordinato delle n osservazioni sulle k variabili esplicative, β è il vettore costituito dai k coefficienti delle variabili esplicative x_j (con $j = 1, 2, \dots, k$), ed ε è il vettore delle manifestazioni non osservabili della componente stocastica di disturbo negli n tempi cui si riferiscono le rilevazioni campionarie;

— *ipotesi di non sistematicità degli errori*:

$$E(\varepsilon_i) = 0 \quad i = 1, 2, \dots, n$$

Tale ipotesi specifica che l'influenza dell'errore su y è, in media, nulla, ossia che il valore atteso della componente stocastica è nullo;

- *ipotesi di omoschedasticità:*

$$\text{Var}(\varepsilon_i) = \sigma^2 \quad i = 1, 2, \dots, n$$

Da tale ipotesi deriva che la variabilità dell'errore è costante;

- *ipotesi di assenza di correlazione nei disturbi:*

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$$

Da questa ipotesi deriva che gli errori relativi ad osservazioni campionarie diverse sono incorrelati;

- *ipotesi di non stocasticità delle variabili esplicative*, la quale indica che le variabili x_j sono deterministiche;
- *ipotesi di non collinearità delle variabili esplicative*, la quale indica che fra due o più delle variabili x_j non esiste una perfetta relazione lineare.

Talvolta i dati empirici violano una o più ipotesi del modello lineare di regressione; ecco il motivo per cui si studia l'adeguatezza del modello di regressione.

Elementi di statistica inferenziale e di calcolo delle probabilità

1 Gli schemi di campionamento

Il *campione* è quella parte di un collettivo statistico sottoposto ad osservazione. L'insieme di campioni di una certa ampiezza, che si possono estrarre da un dato collettivo con una data procedura, è denominato *universo dei campioni*. L'ampiezza del campione è indicata con il simbolo n , mentre l'ampiezza della corrispondente popolazione è indicata con N . In genere, si definiscono *piccoli campioni* quelli per cui $n < 30$.

Esistono diversi schemi di scelta dei campioni, tutti riconducibili a due categorie fondamentali: *campioni casuali* e *campioni ragionati*.

a) Campioni casuali

Si parla di *campioni casuali* quando la scelta delle unità-campione, da una determinata popolazione, è del tutto casuale. Nello scegliere un campione casuale, è fondamentale che le singole unità della popolazione abbiano la stessa probabilità di entrare a far parte del campione.

Esistono diverse procedure di scelta casuale.

— Campionamento semplice

È la procedura di scelta casuale più semplice e consiste nell'assegnare un numero d'ordine progressivo alle unità costituenti il collettivo statistico, impostando una corrispondenza biunivoca con palline aventi numerosità pari a quella del collettivo. Tali palline sono inserite in un'urna, dalla quale si estrae il predeterminato numero n di unità.

Il campionamento considerato può essere effettuato:

- *con ripetizione o bernoullianamente*, se, ad ogni estrazione, la pallina estratta è rimessa nell'urna, per cui la composizione della stessa non si altera, e di conseguenza non si altera nemmeno la probabilità di scelta di una data unità; pertanto, il numero di elementi costituenti l'universo dei campioni è:

$$N \cdot N \cdot \dots \cdot N = N^n$$

- *senza ripetizione o esaustivamente*, se, ad ogni estrazione, la pallina estratta non è rimessa nell'urna, la cui composizione si riduce, di volta in volta, di una unità, e di conseguenza si altera, ad ogni sorteggio, la probabilità di scelta delle unità; di conseguenza, il numero di elementi costituenti l'universo dei campioni è:

$$N \cdot (N - 1) \cdot (N - 2) \cdot \dots \cdot (N - n + 1) = \frac{N!}{(N - n)!}$$

Tale numero è, invece, pari a:

$$\frac{N!}{n!(N - n)!}$$

se l'estrazione è in blocco, vale a dire se le unità sono estratte insieme, e non una dopo l'altra.

I risultati di un'indagine campionaria sono tanto più attendibili quanto maggiore è la numerosità del campione, e tanto meno quanto maggiore è la variabilità del carattere investigato; per ovviare a questo secondo inconveniente, in pratica, si ricorre alla procedura di *campionamento stratificato*. Essa consiste nel raggruppare le unità del collettivo in strati il più possibile omogenei rispetto al carattere oggetto d'osservazione, e nell'estrarre, quindi, un dato numero di unità-campione da ogni strato.

Il campionamento semplice è usato difficilmente a causa dell'impossibilità di enumerare tutte le unità di un collettivo statistico. Per tale motivo, vengono spesso utilizzati altri tipi di campionamento, tra cui si ricorda quello *a grappolo* e quello *areale*.

— *Campionamento a grappolo*

Tale metodo non prevede l'estrazione di singole unità della popolazione, ma di grappoli di individui, cioè agglomerati di unità statistiche. Esso è utilizzato nelle indagini sulla popolazione quando non si hanno a disposizione schedari relativi a singoli individui ma a famiglie.

— *Campionamento areale*

Rientra nel campionamento a grappolo e si basa su una carta geografica o topografica, suddivisa in aree, che si considerano come insiemi di unità statistiche. Le aree-campione sono predeterminate e, per ognuna di esse, l'indagine si svolge su una parte o sulla totalità delle unità.

b) Campioni ragionati

Esistono diversi procedimenti di *scelta ragionata*, ma il più usato è quello *per quote*. Secondo tale metodo, il campione deve costituire una figura a scala ridotta del collettivo statistico. Le unità costituenti il campione sono scelte facendo riferimento ai cosiddetti *caratteri di controllo*, in base ai quali la composizione del campione deve rispecchiare la corrispondente *proporzione* o *quota* del collettivo, la quale deve essere rispettata dall'operatore che si occupa dell'indagine. Quest'ultimo ha, inoltre, un margine di discrezione elevato nella scelta delle unità.

2 Gli eventi aleatori

Un *evento aleatorio* è il risultato di un *esperimento casuale* o *prova*; il suo verificarsi è, appunto, incerto. Un esempio tipico di evento aleatorio è rappresentato dall'uscita di una data faccia (*evento*) nel lancio di un dado non truccato (*esperimento casuale*).

Se un evento aleatorio può essere suddiviso in eventi elementari si dice *composto*. Ad esempio, si consideri l'evento *E* «uscita di una faccia contrassegnata da un numero inferiore o uguale a 3» nel lancio di un dado; si tratta di un evento composto, in quanto esso si verifica se e solo se accade uno qualsiasi dei tre eventi elementari: «1», «2», «3».

Un evento elementare si rappresenta graficamente attraverso un punto, mentre un evento composto si rappresenta circoscrivendo con una linea chiusa (*diagramma di Venn*) i punti relativi agli eventi elementari costituenti.

Inoltre, un evento può essere:

- *possibile*, come l'uscita di una faccia contrassegnata da un numero dispari, nel lancio di un dado;
- *certo*, come l'uscita di una faccia contrassegnata da qualsiasi numero da 1 a 6, nel lancio di un dado;
- *impossibile*, come l'uscita di una faccia contrassegnata da un numero superiore a 6, nel lancio di un dado.

Due o più eventi si dicono *incompatibili* se il verificarsi dell'uno esclude il verificarsi dell'altro o degli altri; in caso contrario si dicono *compatibili*. Ad esempio, lanciando i dadi una sola volta, «l'evento 3» è incompatibile con «l'evento numero pari».

Due o più eventi incompatibili si dicono *necessari* se uno di essi deve verificarsi necessariamente.

Un evento $\bar{E}$ si dice *complementare* di un evento E , se esso corrisponde al *non verificarsi* di E .

a) Operazioni tra eventi

Unione (o somma)

L'*Unione* di due eventi E_1 ed E_2 è un evento costituito dagli eventi elementari dell'uno o dell'altro, prendendo una sola volta quelli comuni ad entrambi, e si indica con:

$$E_1 \cup E_2$$

Se gli eventi considerati non sono incompatibili, sicuramente esistono eventi elementari comuni.

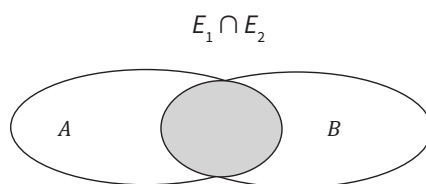
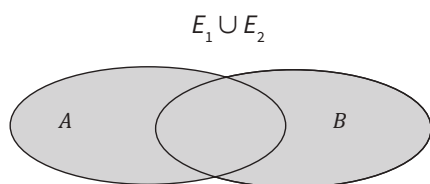
Intersezione (o prodotto)

L'*intersezione* di due eventi E_1 ed E_2 è un evento costituito dagli eventi elementari comuni all'uno e all'altro, e si indica con:

$$E_1 \cap E_2$$

L'intersezione è vuota se gli eventi considerati sono incompatibili.

Nelle due figure seguenti sono rappresentati, nell'area ombreggiata, rispettivamente, l'unione e l'intersezione di due eventi compatibili.



3 La probabilità

La *probabilità di un evento E* esprime il grado di fiducia circa il suo verificarsi; è un numero compreso tra 0 e 1 (estremi inclusi) e si indica così:

$$\Pr(E)$$

La probabilità dell'evento E è:

- $\Pr(E) = 0$ se l'evento è impossibile;
- $0 < \Pr(E) < 1$ se l'evento è incerto;
- $\Pr(E) = 1$ se l'evento è certo.

Esistono due orientamenti sul *concetto di probabilità*, il primo si rifà ad una *impostazione oggettivista* della probabilità, il secondo ad una sua *impostazione soggettivista*. Nella prima si fanno rientrare la *definizione classica* e la *definizione frequentista* o *statistica* della probabilità, nella seconda, invece, la *definizione personalistica*.

a) Impostazione oggettivista

- Definizione classica. *La probabilità di un evento aleatorio è il rapporto tra il numero dei casi favorevoli all'evento e il numero dei casi possibili, purché siano tutti ugualmente possibili* (Laplace).
- Definizione frequentista. *La frequenza relativa con cui si manifesta un evento aleatorio in n prove, tutte eguali ed eseguite nelle medesime condizioni, si avvicina sempre più alla sua probabilità al crescere del numero n* (legge dei grandi numeri o legge empirica del caso di Bernoulli).
- Definizione assiomatica. *La probabilità degli elementi di un insieme astratto di eventi si ottiene attribuendo numeri compresi tra zero e uno, e formulando, in conformità a detta attribuzione, delle regole per determinare la probabilità di altri eventi* (Kolmogorov).

b) Impostazione soggettivista

La probabilità di un evento è un valore che traduce, in termini numerici, un giudizio personale (*De Finetti*).

4 Probabilità composte, condizionate e totali

a) Teorema delle probabilità composte

Prima di dare l'enunciato del *teorema delle probabilità composte*, altrimenti detto *teorema del prodotto*, si deve fare una premessa: due o più eventi si dicono *indipendenti* se il verificarsi di uno di essi in una prova non influisce sulla probabilità di verificarsi degli eventi in altre prove; in caso contrario si dicono *dipendenti*. Si consideri un esperimento che consiste nell'estrarre una pallina da un'urna: se ad ogni estrazione, la pallina estratta è reimmessa nell'urna, la composizione della stessa non si altera, di conseguenza non si altera la probabilità degli eventi considerati (*estrazione con ripetizione*); se, invece, ad ogni estrazione, la pallina estratta non è reimmessa nell'urna, la composizione della stessa si modifica, per cui si dà luogo ad eventi dipendenti (*estrazione senza ripetizione*).

ENUNCIATO • *La probabilità di un evento composto da più eventi indipendenti è data dal prodotto delle probabilità degli eventi componenti.*

Siano E_1 ed E_2 i due eventi, per il *teorema delle probabilità composte*, si ha:

$$\Pr(E_1 \cap E_2) = \Pr(E_1) \Pr(E_2)$$

Ovviamente, il teorema si estende anche al caso in cui il numero degli eventi sia maggiore di 2.

ESEMPIO 1

Si consideri l'evento E «estrazione del 3 di denari da un mazzo di carte napoletane». Tale evento è composto da due eventi distinti: E_1 «estrazione di una carta di denari» ed E_2 «estrazione del 3».

La probabilità di estrarre una carta di denari è:

$$\Pr(E_1) = \frac{10}{40} = \frac{1}{4}$$

La probabilità di estrarre un 3 è:

$$\Pr(E_2) = \frac{4}{40} = \frac{1}{10}$$

Pertanto, la probabilità dell'evento considerato è:

$$\Pr(E) = \Pr(E_1 \cap E_2) = \frac{1}{4} \cdot \frac{1}{10} = \frac{1}{40}$$

b) Probabilità condizionate

La probabilità che un evento casuale E_2 si realizzi, nell'ipotesi che un altro evento E_1 si sia realizzato, si chiama probabilità condizionata dell'evento E_2 , rispetto all'evento E_1 , e si indica con il simbolo:

$$\Pr(E_2/E_1) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_1)}$$

dove $\Pr(E_1 \cap E_2)$ è la probabilità che si manifestino entrambi gli eventi E_1 ed E_2 .

c) Teorema delle probabilità totali

È anche detto *teorema della somma*.

ENUNCIATO • La probabilità del verificarsi, in una prova, di uno qualunque tra più eventi incompatibili è data dalla somma delle probabilità relative al verificarsi di ognuno degli eventi considerati.

Siano $E_1, E_2, \dots, E_s$ gli s eventi incompatibili considerati, per il teorema delle probabilità totali, si ha:

$$\Pr(E_1 \cup E_2 \cup \dots \cup E_s) = \Pr(E_1) + \Pr(E_2) + \dots + \Pr(E_s)$$

Ovviamente, se si tratta di eventi incompatibili e necessari, la loro somma è uguale a 1.

Se due eventi non sono incompatibili, la probabilità che si manifesti uno qualunque tra gli stessi è data dalla somma delle probabilità dei singoli eventi, diminuita della probabilità che i due eventi si manifestino contemporaneamente; se si indicano con E_1 ed E_2 i due eventi, si ha:

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$$

ESEMPIO 2

Si consideri la probabilità che, estraendo una carta da un mazzo di carte napoletane, si verifichi l'evento E_1 «uscita di una carta inferiore a 3» o l'evento E_2 «uscita di una carta superiore a 6». Per determinare la probabilità degli eventi considerati, si faccia riferimento al seguente schema.

Denari	1 2	3 4 5 6	7 8 9 10
Spade	1 2	3 4 5 6	7 8 9 10
Bastoni	1 2	3 4 5 6	7 8 9 10
Coppe	1 2	3 4 5 6	7 8 9 10

Gli eventi considerati sono incompatibili, per cui bisogna prima trovare la probabilità dei singoli eventi; i casi possibili sono 40 in entrambe le situazioni, essendo 40 il numero delle carte, mentre i casi favorevoli sono, rispettivamente, 8 (numero delle carte, di ognuno dei 4 semi, contrassegnate da un valore inferiore a 3) e 16 (numero delle carte, di ognuno dei 4 semi, contrassegnate da un valore superiore a 6):

$$\Pr(E_1) = \frac{8}{40} = \frac{1}{5} \quad \text{e} \quad \Pr(E_2) = \frac{16}{40} = \frac{2}{5}$$

La probabilità che si verifichi l'uno o l'altro evento è:

$$\Pr(E_1 \cup E_2) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$$

5 Il teorema di Bayes

Il *teorema di Bayes*, altrimenti detto delle *probabilità a posteriori*, ricalcola le probabilità delle cause che determinano un insieme di eventi.

Si indichino con A_i gli eventi causa, le cui *probabilità a priori*, ossia prima di verificare la prova, sono $\Pr(A_i)$, e con E l'evento effetto, con probabilità di verificarsi $\Pr(E)$. Per la formula della probabilità condizionata, si ha:

$$\Pr(A_i/E) = \frac{\Pr(A_i \cap E)}{\Pr(E)}$$

Poiché si ha:

$$E = \sum_i (E \cap A_i)$$

dal *teorema delle probabilità totali* per eventi incompatibili, si ha:

$$\Pr(E) = \sum_i \Pr(A_i \cap E)$$

ossia:

$$\Pr(E) = \sum_i \Pr(A_i) \Pr(E/A_i)$$

Pertanto, la formula del *teorema di Bayes* è la seguente:

$$\Pr(A_i/E) = \frac{\Pr(A_i) \Pr(E/A_i)}{\sum_i \Pr(A_i) \Pr(E/A_i)}$$

6 Le variabili aleatorie o casuali

Associando ad ognuno dei possibili eventi E_i , generati da un esperimento casuale, un numero reale x_i , si ottiene la *variabile casuale discreta* o *variabile aleatoria discreta* X , per la quale si ha:

$$\Pr(x_i) = p_i$$

$$\sum_i \Pr(x_i) = 1$$

Si chiama *legge* (o *funzione*) di *probabilità* di una variabile aleatoria, una relazione che stabilisce una corrispondenza tra i valori possibili di questa variabile e le loro probabilità. Essa viene indicata con la notazione: $p(X)$.

La *funzione di ripartizione della probabilità*, invece, indica la probabilità che la variabile casuale X assuma valori inferiori o uguali ad un dato valore x_j , e si indica in questo modo:

$$F(x_j) = \Pr(X \leq x_j) = \sum_{X \leq x_j} p_i$$

La funzione di ripartizione di una *variabile aleatoria* (o *casuale*) *continua* è anch'essa continua; per tale variabile la funzione di probabilità è detta *funzione di densità di probabilità* e la si indica ancora con $p(X)$. Non si può parlare, tuttavia, in questo caso di probabilità che in un esperimento casuale si abbia un dato valore della variabile casuale, ma solo di probabilità che si abbia un valore compreso in un dato intervallo $[x_i, x_j]$, cioè:

$$\Pr(x_i \leq X \leq x_j)$$

che non è altro che l'area delimitata dalla funzione di densità di probabilità in corrispondenza degli estremi x_i e x_j .

7 La distribuzione casuale binomiale

La *distribuzione casuale binomiale* è una distribuzione discreta, cui fanno riferimento tutti quei fenomeni che si presentano con *modalità dicotomiche* (del tipo *presente-assente*), le quali si verificano secondo lo schema di Bernoulli o delle *prove ripetute*.

La distribuzione casuale binomiale si distingue per le seguenti condizioni:

- *indipendenza delle prove*;
- *costanza della probabilità in ogni prova*;
- *la prova è ripetuta un prefissato numero di volte*;
- *il risultato della prova si classifica in una delle classi «successo» oppure «insuccesso»*.

Indicando con E il successo e con $\bar{E}$ l'insuccesso, ossia il non verificarsi dell'evento, si ha che:

$$\Pr(E) = p \quad \text{e} \quad \Pr(\bar{E}) = q$$

dove:

$$p + q = 1$$

Se ad E è associato il valore 1 e ad $\bar{E}$ il valore 0, al fenomeno «numero di successi in una prova» resta associata una variabile casuale X , che è la distribuzione casuale di Bernoulli, e si indica così:

$$\begin{cases} \text{variabile } x_i & 0 & 1 \\ \text{funzione } p_i & q & p \end{cases}$$

La media e la varianza della variabile casuale X sono, rispettivamente, uguali a:

$$\mu = p \quad \text{e} \quad \sigma^2 = pq$$

La distribuzione *binomiale* è detta così perché data dallo sviluppo del *binomio di Newton*, ossia $(p + q)^n$; essa, al variare di X , non è altro che la somma di n variabili casuali di Bernoulli. Per cui, al fenomeno numero di successi in n prove, è associata la distribuzione di probabilità:

$$\begin{cases} \text{variabile } x_i & 0 & 1 & 2 & \dots & n \\ \text{funzione } p_i & q^n & \binom{n}{1} p q^{n-1} & \binom{n}{2} p^2 q^{n-2} & \dots & p^n \end{cases}$$

La funzione di probabilità è:

$$p(X) = \binom{n}{X} p^X q^{n-X}$$

dove:

$$\binom{n}{X} = \frac{n(n-1)(n-2)\dots(n-X+1)}{X!}$$

La media e la varianza della variabile casuale binomiale sono, rispettivamente, uguali a:

$$\mu = np \quad \text{e} \quad \sigma^2 = npq$$

La curva rappresentativa della distribuzione in esame presenta un andamento che dipende dal valore assunto da p . A seconda che esso sia uguale, minore o maggiore di 0,5, la distribuzione presenta rispettivamente simmetria, asimmetria negativa o asimmetria positiva.

8 La distribuzione ipergeometrica

Lo schema di estrazione da cui scaturisce questa distribuzione è accomunabile a quello utilizzato per la distribuzione binomiale; una differenza sostanziale tra i due è che le prove sono *dipendenti* e quindi le condizioni per la realizzazione di questa distribuzione sono le seguenti:

- *dipendenza delle prove;*
- *la probabilità di un successo si altera ad ogni estrazione;*
- *il risultato della prova si classifica in una delle classi «successo» oppure «insuccesso»;*
- *la prova è ripetuta n volte;*
- *la prova è effettuata facendo riferimento ad una popolazione di N unità.*

Indicando con E e con $N - E$, rispettivamente, i successi e gli insuccessi in una prova, la distribuzione ipergeometrica ha per funzione di probabilità:

$$p(X) = \frac{\binom{E}{X} \binom{N-E}{n-X}}{\binom{N}{n}}$$

La media e la varianza della variabile casuale ipergeometrica sono, rispettivamente, uguali a:

$$\mu = np \quad \text{e} \quad \sigma^2 = npq \frac{N-n}{N-1}$$

La variabile casuale ipergeometrica risulta approssimata da una variabile casuale binomiale quando N è sufficientemente grande e la frazione di campionamento n/N è sufficientemente piccola. In tal caso, infatti, la probabilità di un successo, da prova a prova, cambia poco; detto in altri termini, l'estrazione senza reinserimento diventa praticamente uguale a quella con reinserimento.

9 La distribuzione di Poisson

È una distribuzione discreta di probabilità, cui fanno riferimento tutti quei fenomeni che si presentano con le modalità tipiche della distribuzione binomiale, ma con due condizioni in più:

- *numerosità n delle prove molto elevata;*
- *valore p , della probabilità di un successo, molto piccolo;*
- *il prodotto np è costante al crescere delle prove ed è uguale a $np = \lambda$.*

La funzione di probabilità della variabile casuale di Poisson è la seguente:

$$p(X) = \frac{e^{-\lambda} \lambda^x}{X!}$$

dove $e = 2,718\dots$ è la base dei logaritmi neperiani.

Tale distribuzione si ricava con un procedimento limite dalla distribuzione binomiale, infatti:

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np \rightarrow \lambda}} \binom{n}{X} p^X q^{n-X} = \frac{e^{-\lambda} \lambda^x}{X!}$$

La media e la varianza della variabile casuale di Poisson sono, rispettivamente, uguali a:

$$\mu = \lambda \quad \text{e} \quad \sigma^2 = \lambda$$

La curva rappresentativa della distribuzione di Poisson è asimmetrica positiva e, al crescere di λ , tende a diventare sempre più simmetrica.

La distribuzione in esame trova applicazione nel caso di *fenomeni rari*, ossia di fenomeni che hanno scarsa probabilità di verificarsi.

10 La distribuzione casuale normale o di Gauss

È una distribuzione continua di probabilità, che ha costituito il supporto ideale per ogni scienza, soprattutto in Fisica e in Astronomia. Gauss stesso la definì «*legge di probabilità degli errori accidentali*».

Essa presenta notevoli singolarità, che ne permettono un largo impiego nell'approssimazione di altre distribuzioni.

La variabile casuale normale varia da $-\infty$ a $+\infty$. La sua funzione di densità di probabilità è:

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

È di solito indicata con il simbolo $N(\mu, \sigma^2)$, che indica i suoi due parametri, media e varianza, rispettivamente.

La curva rappresentativa della funzione di densità di probabilità di una variabile casuale normale:

- è *simmetrica rispetto al valore di ascissa* $x = \mu$;
- *assume il valore massimo per* $x = \mu$;
- è *asintotica rispetto all'asse delle ascisse*;
- *presenta due flessi*.

L'area delimitata dalla curva normale è uguale a 1 e indica la totalità dei casi, mentre quella delimitata dalle perpendicolari condotte da due valori qualsiasi della variabile indica la probabilità che, in un'estrazione casuale, il valore della variabile sia compreso tra i due valori prefissati.

Per determinare le probabilità in questione si ricorre al calcolo integrale, che spesso risulta complicato; tuttavia, effettuando un mutamento di variabile, è possibile trasformare una qualsiasi distribuzione normale, del tipo $N(\mu, \sigma^2)$, in una *distribuzione normale standardizzata* $N(0, 1)$, avente, in altre parole, media 0 e varianza 1.

Tale trasformazione si effettua introducendo il concetto di *scarto standardizzato* s :

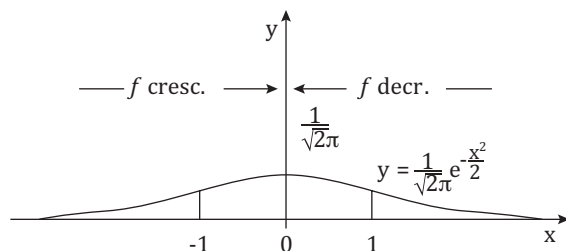
$$s = \frac{X - \mu}{\sigma}$$

Affinché l'area totale resti uguale a 1, si moltiplica la funzione di densità di probabilità per σ , ottenendo:

$$p(s) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}$$

Tale trasformazione si applica ad una qualunque serie di dati che presenti distribuzione normale.

Determinati gli scarti standardizzati per la serie di dati, la probabilità dei casi i cui valori sono compresi in un intervallo definito da due valori dati si desume da apposite tavole.



11 La distribuzione di Student

La funzione di densità di probabilità della *variabile casuale di Student* è:

$$p(g, t) = \frac{a(g)}{\left(1 + \frac{t^2}{g}\right)^{\frac{g+1}{2}}}$$

in cui g è il numero dei gradi di libertà. Il coefficiente $a(g)$ è fondamentale in quanto, grazie ad esso, l'area sottesa alla curva rappresentativa della distribuzione è uguale a 1.

Al crescere di g la distribuzione considerata tende ad una distribuzione normale standardizzata.

La media e la varianza della variabile casuale di Student sono, rispettivamente, uguali a:

$$\mu = 0 \quad \text{per } g > 1$$

$$\sigma^2 = \frac{g}{g-2} \quad \text{per } g > 2$$

La curva rappresentativa della distribuzione è:

- *simmetrica*;
- *unimodale*;
- *campanulare*.

12 La distribuzione χ^2

La funzione di densità di probabilità della *variabile casuale χ^2* è:

$$p(\chi^2) = a(g)(\chi^2)^{\frac{g}{2}-1} e^{-\frac{1}{2}\chi^2}$$

in cui g è il numero dei gradi di libertà. Anche in questo caso, il coefficiente $a(g)$ è fondamentale in quanto, grazie ad esso l'area sottesa alla curva rappresentativa della distribuzione è uguale a 1.

La media e la varianza della variabile casuale χ^2 sono, rispettivamente, uguali a:

$$\mu = g \quad \text{e} \quad \sigma^2 = 2g$$

Al crescere dei gradi di libertà, sia la variabile casuale χ^2 che la variabile casuale $\sqrt{2\chi^2}$ tendono ad una distribuzione normale.

La curva rappresentativa della distribuzione:

- non è simmetrica;
- è unimodale.

13 La distribuzione F di Fisher-Snedecor

La funzione di densità di probabilità della *variabile casuale* F di Fisher-Snedecor è:

$$p(F) = a(g_1, g_2) \frac{F^{\frac{g_1}{2}-1}}{\left(1 + \frac{g_1 F}{g_2}\right)^{\frac{g_1+g_2}{2}}}$$

in cui g_1 e g_2 sono i gradi di libertà. Il coefficiente $a(g_1, g_2)$ è fondamentale perché, grazie ad esso, l'area sottesa alla curva rappresentativa della distribuzione è uguale a 1.

Se i gradi di libertà g_1 e g_2 crescono, la variabile casuale F tende ad una variabile casuale normale.

La curva rappresentativa della distribuzione:

- non è simmetrica;
- è unimodale.

14 Teorema del limite centrale e legge dei grandi numeri

Il *teorema del limite centrale* afferma che *data una successione di n variabili casuali indipendenti* $X_1, X_2, \dots, X_n$ *aventi la stessa distribuzione di probabilità, con media μ e varianza σ^2 , la variabile casuale:*

$$S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$$

al crescere di n si distribuisce come una variabile casuale normale, ed ha media e varianza, rispettivamente, uguali a:

$$E(S_n) = n\mu \quad \text{Var}(S_n) = n\sigma^2$$

La *distribuzione binomiale* è somma di n variabili casuali di Bernoulli, tra loro indipendenti e con la stessa distribuzione, per cui, al tendere di n ad infinito, essa si approssima ad una normale avente media e varianza, rispettivamente, uguali a:

$$\mu = np \quad \sigma^2 = np(1-p)$$

La *legge dei grandi numeri* o *legge empirica del caso* enuncia la cosiddetta *definizione frequentista della probabilità*, secondo cui, effettuando un esperimento un gran numero di volte, si osserva che la frequenza relativa, data dal rapporto $f = \frac{m}{n}$ tra il numero m di volte in cui si verifica l'evento E e il numero n di esperimenti effettuati, tende ad avvicinarsi sempre più alla probabilità $\text{Pr}(E)$; in altri termini, *la frequenza relativa è una buona stima della probabilità di un evento quando $n \rightarrow \infty$.*

Nello schema della distribuzione binomiale, X è il numero di successi in n prove, per cui il rapporto $\frac{X}{n}$ è la proporzione di successi in n prove. Al variare di X , tale rapporto descrive la variabile casuale:

$$\gamma = \frac{X}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

che è la proporzione campionaria e/o la frequenza relativa; essa ha media e varianza, rispettivamente, uguali a:

$$\mu = \frac{1}{n}np = p \quad \sigma^2 = \frac{1}{n^2}npq = \frac{pq}{n}$$

Si ha, dunque:

$$\lim_{n \rightarrow \infty} \Pr\left\{\left|\frac{X}{n} - p\right| < \varepsilon\right\} = 1$$

Tale espressione indica che, al divergere di n , la probabilità che la proporzione campionaria si discosti, in valore assoluto, dalla probabilità p di successo in una prova di una quantità ε positiva e piccola a piacere è praticamente pari a 1.

15 La stima dei parametri

L'inferenza statistica si occupa, tra l'altro, della stima dei *parametri di una popolazione* (media, varianza etc. del collettivo), o semplicemente dei *parametri*, servendosi dei corrispondenti *parametri campionari* o *statistiche* (media, varianza etc. del campione).

L'insieme delle stime del parametro del collettivo e della corrispondente funzione di densità di probabilità (o della funzione di probabilità discreta) costituisce la distribuzione campionaria della statistica considerata, ed è una variabile casuale.

a) Stima puntuale e stima per intervallo

La *stima puntuale* di un parametro è quella fornita da un numero che coincide con la statistica ottenuta dal campione estratto.

La *stima per intervallo* indica se il valore del parametro incognito rientra, con una data probabilità, in un intervallo avente per estremi due numeri predefiniti. È ovvio che di stima per intervallo è possibile parlare solo se è nota la distribuzione della stima puntuale.

b) Proprietà di uno stimatore puntuale

Sia θ il parametro da stimare e $\hat{\theta}$ il corrispondente stimatore puntuale; quest'ultimo deve possedere determinate proprietà:

- deve essere *centrato* o *non distorto* o *corretto*, vale a dire che, al variare del campione, il valore medio della distribuzione delle diverse stime $\hat{\theta}$ deve essere uguale al valore del parametro della popolazione:

$$E(\hat{\theta}) = \theta$$

- deve essere *efficiente*, vale a dire che la distribuzione campionaria delle stime deve avere variabilità inferiore a quella di qualsiasi altra distribuzione campionaria di stime;

- deve essere *consistente*, vale a dire che, al tendere della numerosità n del campione alla numerosità N della popolazione, esso tende a concentrarsi completamente sul parametro da stimare; in altre parole, la probabilità che la stima si avvicini al valore del parametro tende a 1;
- deve essere *sufficiente*, vale a dire che deve utilizzare ogni tipo di informazione data dal campione.

Se uno stimatore non è centrato, l'espressione:

$$d = \theta - E(\hat{\theta})$$

fornisce la *distorsione* o *tendenziosità* dello stimatore.

c) Come si ottiene una stima per intervallo

Per ottenere una stima per intervallo di un dato parametro è necessario conoscere la sua distribuzione campionaria. A questo punto, supponendo che quest'ultima sia simmetrica, si fissa un intervallo denominato *intervallo di confidenza* o *intervallo fiduciario*, del tipo:

$$\hat{\theta} - \gamma \leq \theta \leq \hat{\theta} + \gamma$$

variabile al variare del campione e, dunque, della stima $\hat{\theta}$. Si fissa, inoltre, una probabilità β , denominata *livello di confidenza* o *livello di fiducia*, la quale indica che si ha fiducia nel fatto che il parametro θ cada entro i limiti dell'intervallo di confidenza; ossia:

$$\beta = \Pr(\hat{\theta} - \gamma \leq \theta \leq \hat{\theta} + \gamma)$$

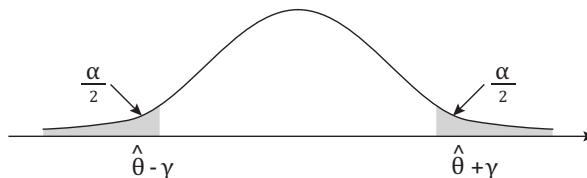
Così come esiste un universo di campioni, esiste un universo di intervalli, centrati sulle stime $\hat{\theta}$. Una percentuale di essi, pari proprio a β , contiene al suo interno il valore θ della popolazione.

In concreto, si fissa il valore complementare di $\beta = 1 - \alpha$, denominato *livello* o *soglia di significatività*, il quale misura la probabilità che, estraendo un campione casuale, da esso derivi un intervallo, sempre centrato sulla stima $\hat{\theta}$, che non contiene al suo interno il valore del parametro θ della popolazione.

In genere, si pone:

$$\Pr(\theta < \hat{\theta} - \gamma) = \frac{\alpha}{2} \quad \Pr(\theta > \hat{\theta} + \gamma) = \frac{\alpha}{2}$$

Nella figura seguente è indicato l'intervallo di confidenza del parametro θ .



d) Diversità nei metodi di stima

Esistono diversi metodi di stima puntuale dei parametri, i quali cercano di ottenere stimatori che soddisfino, se non tutte, almeno alcune delle proprietà auspicabili; in questa sede ne daremo solo un cenno.

e) Metodo della massima verosimiglianza

Il *metodo della massima verosimiglianza* consiste nel determinare il valore dello stimatore che rende massima la *funzione di verosimiglianza*:

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

dove $x_1, x_2, \dots, x_n$ è il campione osservato per stimare θ .

Massimizzare la funzione di verosimiglianza equivale a massimizzare il suo logaritmo; pertanto, il valore di $\hat{\theta}$ è la soluzione dell'equazione:

$$\frac{d \log L}{d \theta} = 0$$

Tale metodo può essere utilizzato qualunque sia la distribuzione della popolazione. Gli stimatori ottenuti sono consistenti, corretti ed efficienti.

f) Metodo dei minimi quadrati

Il *metodo dei minimi quadrati* sarà illustrato attraverso la stima dei parametri di un modello di regressione.

La stima dei parametri del modello con il metodo dei minimi quadrati, considerando il modello espresso in forma matriciale $Y = X\beta + e$, si ottiene cercando quel vettore β , costituito dagli elementi $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_k)$, per cui la quantità:

$$S = (Y - X\beta)'(Y - X\beta) = Y'Y + \beta'(X'X)\beta - 2\beta'X'Y$$

è minima.

Uguagliando a zero la derivata parziale rispetto a β si ha:

$$\frac{\partial S}{\partial \beta} = 2(X'X)\beta - 2X'Y = 0$$

da cui si ricava la stima dei minimi quadrati:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Gli stimatori $\hat{\beta}$ che si ottengono con questo metodo sono i più efficienti nella classe degli stimatori lineari e non distorti (*teorema di Gauss - Markov*). Tali stimatori sono anche consistenti.

g) Metodo dei momenti

Il *metodo dei momenti* è un metodo utilizzato per stimare qualsiasi parametro di una popolazione; esso si serve dei momenti campionari. Pertanto, se il momento dall'origine di ordine r di un campione di dimensione n è:

$$m_r = \frac{\sum_{i=1}^n x_i^r}{n}$$

per una stima della media della popolazione si pone $r = 1$, e la media della popolazione è fatta coincidere con la media del campione.

Per applicare tale metodo non è indispensabile conoscere la distribuzione della popolazione, è sufficiente conoscere le relazioni esistenti tra i parametri da stimare e certi momenti della popolazione.

Gli stimatori ottenuti sono consistenti ma poco efficienti.

h) Metodo di Bayes

Il *metodo di Bayes* si fonda sulla conoscenza della *probabilità a priori* p_i , che il parametro θ , in quanto considerato variabile casuale, assuma i valori $\theta_1, \theta_2, \dots, \theta_s$.

La *probabilità a posteriori* che il campione $x_1, x_2, \dots, x_n$ provenga dalla popolazione in cui $\theta = \theta_i$ è:

$$\pi_i = \frac{p_i L_i}{\sum_{i=1}^s p_i L_i} \quad i = 1, 2, \dots, s$$

dove $L_i(\theta_i; x_1, x_2, \dots, x_n)$ è la probabilità che il campione ha nell'universo dei campioni della popolazione in cui $\theta = \theta_i$. La stima del parametro θ col metodo di Bayes è quel valore θ_i (con $i = 1, 2, \dots, s$) che massimizza la *probabilità a posteriori* appena calcolata.

16 L'adeguatezza di un modello di regressione

In questo paragrafo sarà trattato un problema tipico dell'inferenza statistica: la *verifica di ipotesi*.

Le ipotesi sottostanti un modello di regressione si configurano alla stregua di congetture, la cui validità può essere confermata o messa in forse a seguito dell'evoluzione reale del fenomeno allo studio, così come si manifesta con le osservazioni empiriche; per questo occorre porsi in atteggiamento critico rispetto alle stesse.

L'inadeguatezza di un modello implica, potenzialmente, serie conseguenze. Numerose violazioni delle ipotesi possono condurre ad un modello instabile, nel senso che un campione differente potrebbe dare luogo ad un modello completamente diverso, comportando conclusioni opposte.

Consideriamo dapprima un modello di regressione semplice.

- L'ipotesi nulla da testare può essere del tipo $H_0: \beta = 0$; si può ipotizzare, in altre parole, che non esista alcuna relazione tra i due caratteri oggetto di osservazione. Il test da utilizzare si distribuisce, se l'ipotesi nulla è vera, come una variabile casuale t di *Student*, con $n - 2$ gradi di libertà; la regola di decisione, in presenza di un'ipotesi alternativa bidirezionale, consiste nel rifiutare l'ipotesi nulla se $\hat{t} < t_{n-2; \alpha/2}$ oppure $\hat{t} > t_{n-2; 1-\alpha/2}$.
- Volendo verificare l'ipotesi che il coefficiente di regressione β sia uguale ad un valore specificato β_0 (ossia volendo verificare l'ipotesi $H_0: \beta = \beta_0$), disponendo di una stima $\hat{\beta}$ fornita da un campione, il test da utilizzare si distribuisce come una variabile casuale t di *Student* con $n - 2$ gradi di libertà; la regola di decisione, in presenza di un'ipotesi alternativa bidirezionale, consiste nel rifiutare l'ipotesi se $\hat{t} < t_{n-2; \alpha/2}$ oppure $\hat{t} > t_{n-2; 1-\alpha/2}$.
Una volta costruito un modello di regressione multipla si può stabilire, attraverso l'uso di un test, se esiste o no una relazione significativa tra la variabile dipendente e l'insieme delle variabili esplicative.
- Talvolta si vuole verificare se esiste una relazione lineare tra la variabile y e le k variabili esplicative x ; pertanto, l'ipotesi nulla da testare è $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ e l'ipotesi alternativa è $H_1: \beta_j \neq 0$ (per almeno una j). Il rifiuto dell'ipotesi nulla implica che almeno una delle variabili $x_1, x_2, \dots, x_k$ contribuisca significativamente a *spiegare* la variabile dipendente. L'ipotesi nulla

può essere verificata utilizzando un test che si distribuisce come una variabile casuale F di *Fisher-Snedecor*, con k e $n - k - 1$ gradi di libertà. La regola di decisione consiste nel rigettare l'ipotesi nulla se $\hat{F} > F_{k; n-k-1; \alpha}$.

- Altre volte i test d'ipotesi riguardano i singoli coefficienti di regressione, nel senso che ci si chiede se il modello sia più valido con l'inclusione di un'ulteriore variabile esplicativa o con la sua esclusione. Le ipotesi per testare la significatività di ogni singolo coefficiente di regressione sono $H_0: \beta_j = 0$ e $H_1: \beta_j \neq 0$; l'accettazione dell'ipotesi nulla implica che la generica variabile x_j può essere cancellata dal modello, in quanto non è significativa nello *spiegare* la variabile dipendente. Il test per questa ipotesi si distribuisce, al variare del campione, come una variabile casuale t di *Student*; la regola di decisione consiste nel rigettare l'ipotesi nulla se $\hat{t} < t_{n-k-1; \alpha/2}$ oppure $\hat{t} > t_{n-k-1; 1-\alpha/2}$. Quello utilizzato è solo un test parziale, in quanto il generico coefficiente di regressione β_j dipende da tutte le altre variabili x_i (con $i \neq j$) incluse nel modello.

1 I numeri primi e la crittografia

Per definizione un numero *primo* è un numero *naturale* che non si può scomporre come prodotto di due numeri naturali più piccoli. L'importanza dei numeri primi deriva dal cosiddetto «Teorema fondamentale dell'aritmetica», il quale afferma che ogni numero naturale può essere rappresentato in un solo modo come prodotto di primi. Fin dai tempi di Euclide, cioè circa dal 300 a.C., gli antichi Greci sapevano che ci sono infiniti numeri primi e che essi sono distribuiti in maniera irregolare tra i numeri naturali. Da allora, molti matematici eminenti, come Fermat, Mersenne, Leibniz, Eulero, Gauss etc., hanno compilato e analizzato lunghi elenchi di numeri primi, nel tentativo di farsi un'idea della loro distribuzione e per cercare delle formule in grado di produrre, se non tutti, almeno «molti» numeri primi: in particolare, molti matematici si dedicarono a studiare i numeri del tipo $2^p - 1$, con p primo (ad esempio Fermat aveva dimostrato che, se non sono primi, i numeri di questo tipo hanno fattori tutti della forma $2kp + 1$, con k numero naturale).

Quello che spinge a cercare nuovi numeri primi sono alcune loro applicazioni pratiche: ad esempio, a partire dagli anni '80, grandi numeri primi sono stati usati per cifrare messaggi segreti con il metodo di crittografia a chiave pubblica detto RSA (dai nomi dei suoi ideatori Rivest, Shamir e Adleman). I metodi di crittografia a chiave pubblica prevedono che, chi deve ricevere l'informazione, renda pubblica la chiave (alcuni numeri) e il metodo per cifrare i messaggi. Egli poi, in base a dei «numeri segreti» in suo possesso è in grado di decifrare i messaggi che gli vengono inviati. Nel metodo RSA:

- i numeri segreti sono due numeri primi p, q e un numero N coprimo con $(p - 1)(q - 1)$;
- la chiave pubblica è costituita dal prodotto pq e da un numero M tale che $MN - 1$ sia multiplo di $(p - 1)(q - 1)$;
- il messaggio da trasmettere è un numero intero positivo x , minore di pq ;
- il messaggio cifrato è il numero y , ottenuto come resto della divisione di x^M per pq ;
- per decifrare il messaggio, è sufficiente calcolare il resto della divisione di y^N per pq .

Poiché il messaggio deve essere un numero positivo e minore di pq , in caso si tratti di un messaggio in lettere, bisogna innanzitutto trasformarlo in cifre con un metodo qualunque; poi, se il numero trovato fosse troppo grande, bisognerebbe spezzettarlo in numeri più piccoli da trasmettere separatamente. La spiegazione del metodo si basa sul seguente teorema, che a sua volta discende dal piccolo teorema di Fermat.

TEOREMA • Se $0 \leq x < pq$, allora $(x^M)^N$ è congruo a x modulo pq .

Nella pratica, bisogna scegliere p e q primi molto grandi, in modo che sia impossibile fattorizzare pq in tempo ragionevole. Il metodo, infatti, si basa sull'assunzione che sia estremamente difficile risalire a p e q conoscendo il loro prodotto. Se chi intercettasse un messaggio fosse in grado di fattorizzare pq , di conseguenza conoscerebbe anche $(p - 1)(q - 1)$, e per trovare N gli sarebbe sufficiente risolvere l'equazione $MN \equiv 1 \pmod{(p - 1)(q - 1)}$, con M e $(p - 1)(q - 1)$ primi tra loro; entrerebbe così facilmente in possesso di tutte le informazioni necessarie a decifrare il messaggio.

2 Sistema decimale e sistema binario

In un qualsiasi sistema di numerazione ogni cifra assume un diverso peso a seconda della posizione occupata nel numero. Numeriamo le cifre di un numero intero a partire dalla destra. Indicando con c_k la generica cifra, con $k = 1, 2, \dots, n$ (dove n è il numero di cifre che costituisce il numero assegnato), e con b la base del sistema di numerazione, si ha che un generico numero sarà esprimibile come:

$$\sum_{k=1}^n c_k \cdot b^{k-1} \quad (1)$$

La significatività di una cifra dipende dal suo peso. La cifra meno significativa è quella corrispondente alla potenza zero, quella più significativa è quella corrispondente alla potenza maggiore.

Nel **sistema decimale** una cifra può assumere i valori compresi fra 0 e 9. I pesi di ciascuna cifra di un numero, in ordine crescente, sono noti come unità, decine, centinaia etc. Sfruttando la relazione (1) si ha che un numero intero è esprimibile come:

$$\sum_{k=1}^n c_k \cdot 10^{k-1}$$

Ad esempio, il numero 235 può essere espresso come somma di:

$$(235)_{10} = 2 \cdot 10^2 + 3 \cdot 10^1 + 5 \cdot 10^0$$

La parentesi intorno al numero è contrassegnata con un pedice numerico che indica la base del sistema di numerazione.

Il **sistema numerico binario** è usato in informatica per la rappresentazione interna dei numeri, sia perché è più agevole realizzare fisicamente elementi con «due stati», anziché un numero di stati superiore, sia per la corrispondenza dei «due stati» con i valori logici di vero e falso. Nel sistema binario, infatti, una cifra (detta bit) può assumere solamente i valori 0 oppure 1. I pesi di ciascuna cifra di un numero binario seguono la stessa logica dei numeri decimali ma sono espressi da potenze di 2. Servendosi dell'equazione (1), un numero binario è esprimibile come:

$$\sum_{k=1}^n c_k \cdot 2^{k-1}$$

Ad esempio, il numero 1001 può essere espresso come:

$$(1001)_2 = 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$$

Se si sviluppano le potenze del due, i prodotti e le somme, si ottiene il numero in base 10:

$$(1001)_2 = (9)_{10}$$

Si dice anche che 1001 è la codifica nel sistema binario del numero 9 espresso in base decimale. Con n bit si possono codificare numeri decimali compresi tra 0 e $2^n - 1$.

Si definisce byte un numero composto da 8 bits. Con un byte si possono codificare i numeri decimali compresi 0 e 255 (infatti, poiché 1 byte è costituito da 8 bits, si ha: $2^8 - 1 = 255$).

3 Le operazioni nel sistema binario

Le regole per eseguire le *operazioni aritmetiche nel sistema di numerazione binario* sono le stesse di quelle del sistema decimale.

$0 \cdot 0 = 0$	$0 + 0 = 0$	
$0 \cdot 1 = 0$	$0 + 1 = 1$	
$1 \cdot 0 = 0$	$1 + 0 = 1$	
$1 \cdot 1 = 1$	$1 + 1 = 0$	con riporto di 1

Per convertire un numero binario in decimale si utilizza l'equazione (1) dove $b = 2$.

Per convertire un numero decimale in binario lo si divide ripetutamente per 2 finché non si ottiene come risultato della divisione il valore 0. Il risultato ottenuto dalla prima divisione avrà un *resto* che assume valore 0 oppure 1. Il valore del resto costituisce la cifra meno significativa del corrispondente numero binario. Il *resto* della seconda divisione fornirà la seconda cifra. Iterando il discorso si ottiene il numero convertito in bits.

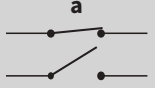
L'esempio mostra la conversione in binario del numero 30.

	resto
$30 : 2 = 15$	0
$15 : 2 = 7$	1
$7 : 2 = 3$	1
$3 : 2 = 1$	1
$1 : 2 = 0$	1
$(30)_{10} = (11110)_2$	

4 Gli operatori booleani

Gli *operatori logici booleani* possono essere rappresentati da semplici circuiti elettrici, costituiti da interruttori collegati fra loro in maniera opportuna.

Alle posizioni aperta e chiusa di un interruttore a corrispondono rispettivamente i valori 0 e 1. La condizione C del circuito sarà 1 oppure 0 a seconda che la corrente elettrica circoli o meno nel circuito.

a	C	
1	1	
0	0	

L'operatore **AND** è rappresentato da due interruttori disposti in serie:

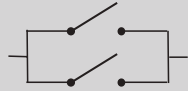
- $C = 0$ quando almeno uno dei due interruttori assume valore 0;
- $C = 1$ quando entrambi gli interruttori assumono valore 1.

a	b	C	
0	0	0	
0	1	0	
1	0	0	
1	1	1	

L'operatore **OR** è rappresentato da due interruttori disposti in parallelo:

- $C = 1$ quando almeno uno dei due interruttori assume valore 1;
- $C = 0$ quando entrambi gli interruttori assumono valore 0.

a	b	C
0	0	0
0	1	1
1	0	1
1	1	1



Combinando gli operatori elementari *NOT*, *AND*, *OR* si ricavano altri operatori logici come, ad esempio, *NOR*, *NAND*, *XOR*.

5 I contributi di Hilbert, Turing e Godel

Il programma ambizioso di *D. Hilbert* (1862–1943) prevedeva la risoluzione di tutti i problemi «aperti» della matematica: partendo da alcuni assiomi e procedendo attraverso passaggi logici non si può giungere a risultati contraddittori.

Nel 1937, a 25 anni, *A.M. Turing* (1912–54), logico matematico inglese, pubblicò un lavoro in cui è descritta una macchina concettuale che si prestava ad affrontare il problema di Hilbert. Questa macchina, nota come *macchina di Turing*, è l'essenza di tutti i calcolatori. Si tratta di un dispositivo astratto in cui sono descritti, in precisi termini matematici, tutti i processi logici atti a decidere la verità delle proposizioni attraverso una successione di operazioni ben definite. Utilizzando questo computer ideale, Turing dimostrò che alcuni problemi matematici non si possono definire mediante un numero finito di passaggi logici, convalidando il teorema dell'incompletezza (1931) di *K. Godel*, che dimostrava come la matematica non possa essere completa e totalmente consistente. Era chiaro che non poteva esistere un insieme di proposizioni logiche dalle quali possa essere derivata tutta la matematica. Successivamente Turing si occupò della progettazione di calcolatori digitali.

6 Il calcolatore elettronico

Il calcolatore tascabile è una macchina in grado di effettuare calcoli aritmetici. I numeri da elaborare e le operazioni vengono impostate tramite una tastiera. Le prime calcolatrici erano strumenti meccanici ed elettromeccanici in grado di eseguire le quattro operazioni. Le calcolatrici moderne sono elettroniche e permettono operazioni in automatico anche complesse (radice, elevamento a potenza, operazioni trigonometriche, calcoli di logaritmi etc.).

Un calcolatore numerico è un calcolatore automatico, elettronico e digitale.

Il calcolatore numerico è considerato:

- *automatico* perché, dopo averlo predisposto ed avviato, funziona da solo, senza altri interventi esterni;
- *elettronico* perché sfrutta la tecnologia elettronica (anche per il controllo delle eventuali parti meccaniche o elettromeccaniche);
- *digitale* perché opera su cifre (*digit*) opportunamente codificate.

La maggior parte dei calcolatori digitali codificano le informazioni da elaborare (numeri o stringhe di caratteri) in un numero binario. Ciascuna cifra di un numero binario è detta *bit* (acronimo di *binary digit*).

L'*architettura* di un calcolatore elettronico può essere semplificata in quattro blocchi:

1. una (o più) *unità centrale*, la CPU (Central Processing Unit), che rappresenta il «cuore» del sistema essendo responsabile dell'elaborazione dei dati;
2. una (o più) *unità di memoria centrale* (RAM, ROM etc.), dove vengono caricati i programmi di istruzione per la CPU. Anche i dati da elaborare e i risultati dell'elaborazione sono conservati nell'unità di memoria centrale;
3. una (o più) *unità d'ingresso* (input), che consente l'immissione nel calcolatore di dati e programmi. La tastiera, il mouse etc. sono esempi di unità d'ingresso;
4. una (o più) *unità di uscita* (output), che permette di visualizzare i risultati dell'elaborazione. Il display, il monitor, la stampante sono esempi di unità di uscita.

Si dice che il calcolatore è interfacciato con l'ambiente esterno attraverso le unità di ingresso ed uscita. Alcune interfacce sono sia di ingresso che di uscita (es. modem, porte seriali e parallele etc.).

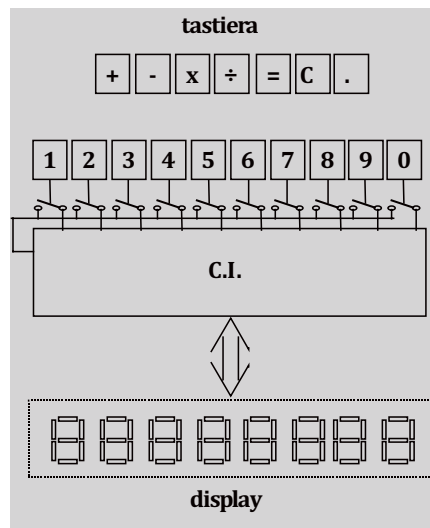
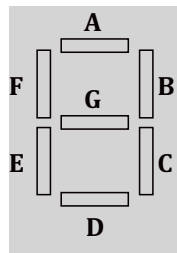
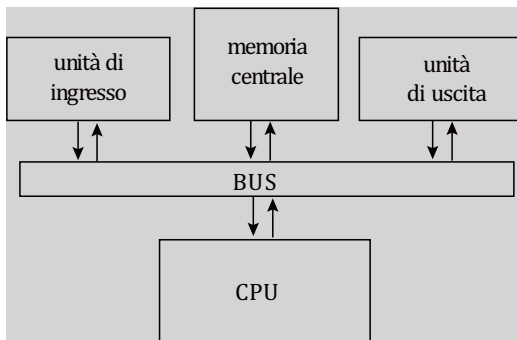
Le informazioni fra queste unità si trasmettono tramite un sistema detto *bus*.

Un *calcolatore tascabile* tradizionale è caratterizzato da una *tastiera* mediante la quale si impostano numeri e operazioni da eseguire. La tastiera rappresenta l'unità d'ingresso. I numeri vengono visualizzati su un dispositivo detto «*display*» (visore), il quale rappresenta l'unità di uscita. I display più moderni sono a *cristalli liquidi*, altri tipi di display sono a *LED* (Diodo ad emissione di luce, in inglese *light emission diode*), quelli più antichi sono, invece, display valvolari a gas. I display a cristalli liquidi sono a basso consumo di energia, perciò sono quelli preferiti. Alcune calcolatrici presentano anche una *stampante*, in grado di scrivere su un supporto cartaceo le operazioni eseguite ed i risultati. Le stampanti sono un altro esempio di unità d'uscita.

Un display *numerico* è capace di visualizzare solo numeri. Il numero viene visualizzato mediante l'accensione di segmenti luminosi disposti come in figura. È facile rendersi conto che per visualizzare il numero 8 è necessario che siano accesi tutti i segmenti (i segmenti sono 7 denominati A, B, C, D, E, F, G); per visualizzare il numero 3 dovranno accendersi i segmenti A, B, C, D, G.

La pressione su un pulsante della tastiera comporta la chiusura di un interruttore. Questa informazione viene trasferita al circuito elettronico interno alla calcolatrice. I circuiti elettronici di una calcolatrice sono essenzialmente *circuiti integrati* (C.I.), costituiti da materiali *semiconduttori* al *silicio*. Le tecnologie dei circuiti integrati hanno permesso di inglobare, in pochi centimetri quadrati, milioni di transistor, riducendo le dimensioni. È proprio questa tecnologia che ha permesso ai sistemi digitali di diventare sempre più sofisticati.

Quando il C.I. riceve l'informazione relativa alla chiusura di un interruttore sulla tastiera, provvede, automaticamente, ad accendere i segmenti relativi al numero corrispondente al tasto premuto.



ESEMPIO 1

Eseguiamo un'operazione di addizione fra due numeri utilizzando un calcolatore tascabile. L'operazione, ad esempio, è $7 + 3$. Si preme prima il tasto corrispondente al 7. Sul display verrà visualizzato il numero 7. Si preme il tasto «+». A questa operazione il *C.I.* provvede a memorizzare il numero 7 in un registro interno, detto *registro degli operandi*. Premendo il tasto corrispondente al numero 3, il 7 scompare dal display (ma comunque è presente nel registro degli operandi). Se, infine, viene premuto il tasto «=», il *C.I.* provvede automaticamente a sommare i due operandi, visualizzando il risultato 10 sul display e memorizzando lo stesso risultato in un registro detto *registro accumulatore*. I calcolatori tascabili moderni sono dei veri calcolatori numerici, capaci di eseguire programmi di calcolo anche complessi, di visualizzare caratteri alfanumerici e di rappresentare i diagrammi di funzioni. Inoltre, hanno la possibilità di interfacciarsi con l'ambiente mediante *porte seriali e parallele*. Tali porte, mediante cavi di collegamento ad altre unità, permettono l'utilizzo di diverse unità d'ingresso o uscita.

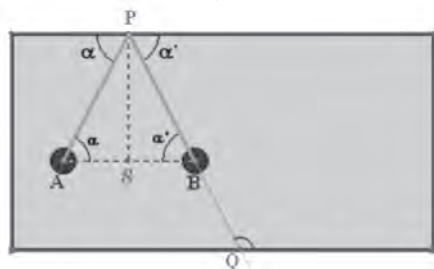
7 Un modello teorico della geometria del biliardo

Al gioco del biliardo, supposto rettangolare, si possono applicare alcune regole di geometria euclidea. Possiamo immaginare che la palla sia un punto materiale e che le sue traiettorie siano dei segmenti. Poiché le sponde del biliardo sono elastiche, in linea di principio valgono le leggi della riflessione della fisica ottica, proprio come uno specchio rifletterebbe un fascio di luce.

Se una biglia fosse un punto materiale e non avesse rotazione, scivolando senza attrito sul panno e sulle sponde, determinerebbe delle traiettorie in entrata e in uscita che rispettano le due seguenti leggi:

- la traiettoria in entrata e quella in uscita sono due segmenti che appartengono allo stesso piano, che nel nostro caso è il piano del biliardo;
- l'angolo acuto incidente, che la traiettoria in ingresso forma con il lato del biliardo, è uguale all'angolo acuto riflesso, che la traiettoria in uscita forma con lo stesso lato del biliardo.

Le due regole sono evidenziate nella figura seguente:



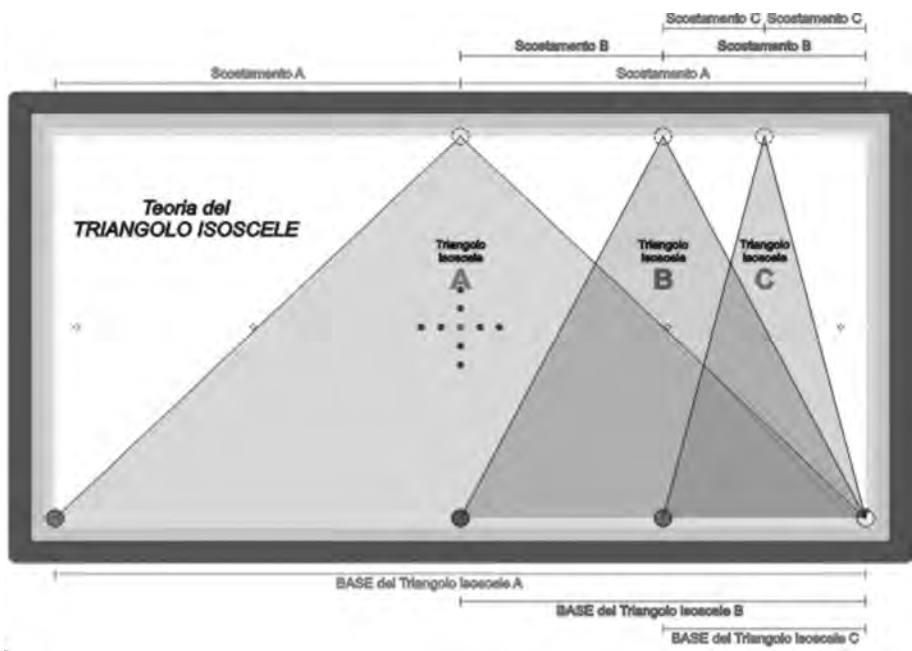
La palla posta nel punto A colpisce il punto P della sponda. Il segmento AP è la traiettoria in ingresso (incidente), che forma un angolo α con il lato del biliardo; il segmento PQ è la traiettoria in uscita (riflessa), che forma un angolo α' . I due angoli sono, in linea teorica, uguali. Se consideriamo il segmento parallelo al lato del biliardo uscente da A, esso incontrerà il segmento PQ nel punto B; il triangolo ABP, così ottenuto, risulta isoscele.

Per definizione l'altezza di un triangolo isoscele è anche la bisettrice dell'angolo tra i due lati uguali. L'altezza PS del triangolo APB è perpendicolare alla sponda del biliardo e passa per il punto P di mira. Lo spostamento AS del punto P rispetto al punto A di partenza è la metà della base.

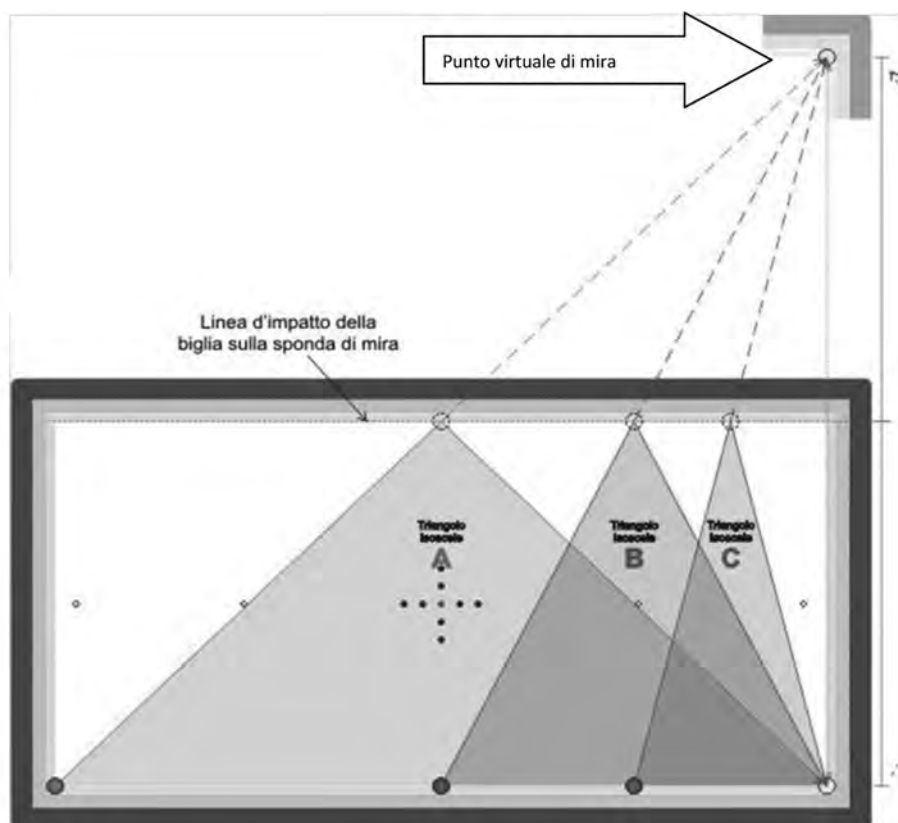
Se, ad esempio, partissimo da un vertice del biliardo, per raggiungere l'altro vertice adiacente, dovremmo indirizzare la biglia verso il punto medio del lato opposto. In questo caso i due vertici si identificano con i vertici della base, mentre il punto di mira con il vertice dell'altezza.



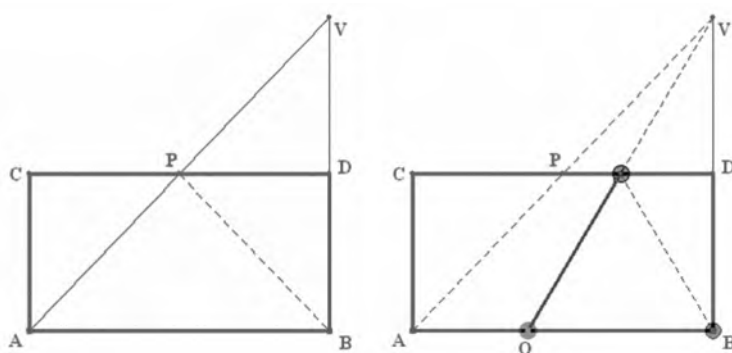
Nella figura sotto sono riportati alcuni triangoli isosceli, che determinano traiettorie per raggiungere sempre lo stesso punto di arrivo cambiando il punto di partenza.



È facile dimostrare che i prolungamenti di tutte le traiettorie in entrata (lati incidenti dei triangoli isosceli) si incontrano tutte in uno stesso punto, esterno al biliardo e detto *punto di mira virtuale*.

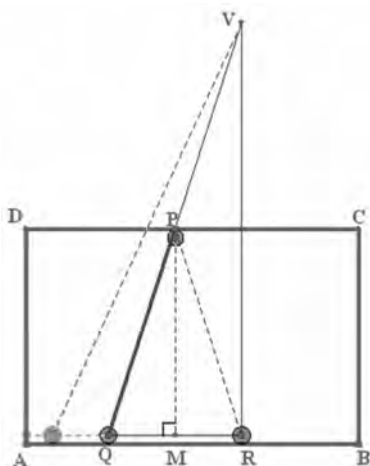


Possiamo verificare facilmente che il *punto di mira virtuale*, indicato con V, corrisponde al simmetrico del punto di arrivo B rispetto al lato CD: infatti, la traiettoria AV passante per il punto medio P di CD e la traiettoria BD si incontrano nel punto V. I triangoli BDP e VDP sono congruenti e, quindi, VD è congruente a BD; ne segue che V è il simmetrico di B rispetto al lato CD.



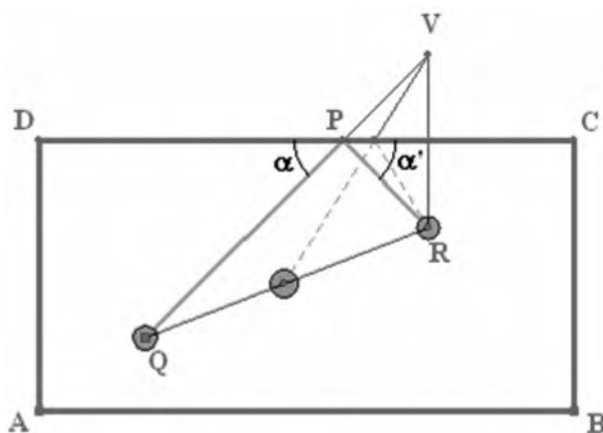
In conclusione, se vogliamo spedire nel vertice B (punto di arrivo) una biglia posta su un punto qualunque del lato AB del biliardo (punto di partenza), dobbiamo mirare nel punto virtuale V.

Nel caso in cui il punto di arrivo sia un punto qualunque della sponda, non necessariamente il vertice, il ragionamento è del tutto identico.



In figura la palla, posta in Q, deve raggiungere R. Si individua il punto medio M del segmento QR e si traccia la perpendicolare passante per M fino ad incontrare in P la sponda opposta DC (punto di impatto). Il triangolo QPR è isoscele e il punto di mira V si ottiene dall'intersezione fra il prolungamento di QP e la semiretta uscente per R parallela a BC.

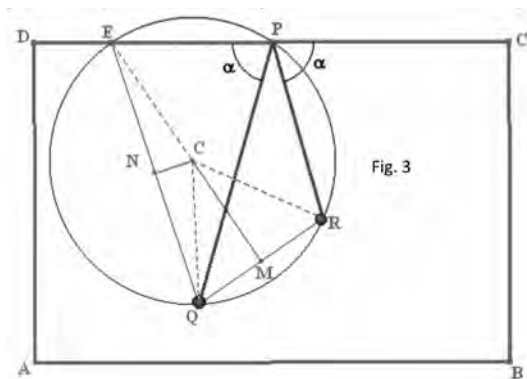
Oltre alle situazioni descritte, potrebbe accadere che il punto di partenza e di arrivo siano due punti qualunque interni al biliardo.



In figura, la biglia posta nel punto Q deve raggiungere il punto R dopo aver colpito la sponda nel punto P. L'individuazione per via grafica del punto di contatto P è più complessa.

Una volta individuato tale punto (si ricorda che l'angolo incidente α deve essere uguale all'angolo di riflessione α'), possiamo individuare il punto di mira V, che è dato dall'intersezione fra il prolungamento del segmento QP e la semiretta uscente da R e perpendicolare alla sponda DC. Ne segue che per inviare una biglia da un punto qualunque del segmento QR bisogna mirare nel punto virtuale V.

Una tecnica per l'individuazione del punto P, nel caso in cui si conosca soltanto il punto di partenza Q e il punto di arrivo R, è illustrata di seguito:



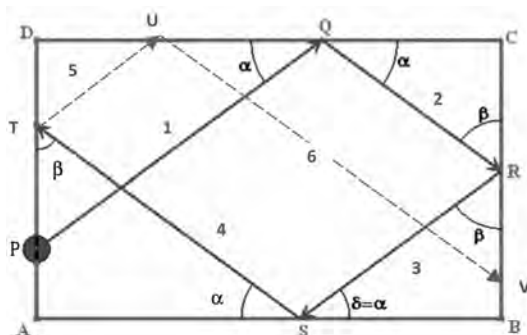
Si traccia l'asse del segmento QR fino a incontrare la sponda DC nel punto E, e dopo si disegna la circonferenza che passa per i punti Q, R, E.

Per ottenere tale circonferenza bisogna trovare il centro C ed il raggio: allora si traccia l'asse del segmento QE fino a incontrare in C il segmento EM. I triangoli ENC e QNC sono congruenti, così come lo sono i segmenti EC e QC. I triangoli QMC e RMC sono congruenti così come i segmenti QC e RC. Pertanto, i segmenti EC, QC e RC sono tutti della stessa lunghezza e sono quindi raggi della circonferenza in questione.

A questo punto, con il compasso puntato in C e con un'apertura uguale a uno dei tre segmenti, tracciamo la circonferenza: essa oltre a passare necessariamente per i tre punti E, R, C incontrerà il lato DC in un ulteriore punto P che è il punto che cercavamo (ovviamente si dovrebbe dimostrare che l'angolo incidente DPQ e l'angolo riflesso CPR, indicati con α , sono congruenti: questa dimostrazione la lasciamo ai lettori).

Per completare il discorso, ci sarebbe da chiedersi cosa accadrebbe alle traiettorie sia nel caso in cui la biglia rimbalzi sulla sponda consecutiva, sia nel caso in cui la biglia rimbalzi sulla frontale.

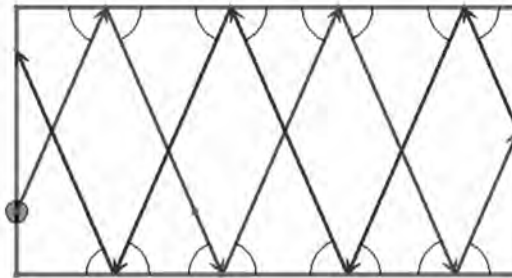
Nel primo caso, dopo il primo rimbalzo sulla sponda, la biglia continuerà il suo percorso incontrando le altre sponde fino a fermarsi. Se non ci fosse attrito continuerebbe il suo percorso all'infinito. In ogni punto di impatto fra la biglia e la sponda valgono le stesse leggi precedentemente esposte.



In figura, la biglia dalla posizione P è lanciata contro la sponda nel punto Q. A sua volta la biglia incontrerà le sponde consecutive nei punti R, S, T, U, V etc. Le traiettorie sono state numerate: quelle incidenti sono le dispari (1, 3, 5 etc.) e sono tutte parallele fra di loro; analogamente, le traiettorie pari (2, 4, 6 etc.) sono quelle riflesse e anch'esse sono parallele fra loro.

Ciò si spiega tenendo presente che, per la seconda legge, l'angolo incidente deve essere uguale all'angolo riflesso, e che in un triangolo rettangolo la somma dei due angoli acuti è 90 gradi. Quindi, se $\alpha + \beta = 90^\circ$, si ricava che $\beta = 90 - \alpha$. Nel triangolo RBS l'angolo di riflessione è β perché deve essere uguale a quello incidente della traiettoria 2. Con δ abbiamo indicato l'angolo d'incidenza della traiettoria 3. Poiché deve essere $\delta + \beta = 90^\circ$, si ha che $\delta = \alpha$. In definitiva, gli angoli che si formano alternativamente, lungo il percorso della biglia, sono due: α e $(90 - \alpha)$.

Le traiettorie successive, nel caso in cui la biglia rimbalzi sulla frontale, sono rappresentate nella figura seguente.



Anche in questo caso si ottengono traiettorie alternate, parallele fra loro, ma gli angoli che esse formano sulle due sponde opposte sono tutti uguali. In figura è mostrato il percorso di una biglia che, dopo il primo impatto, colpisce la sponda frontale, nonché il percorso di ritorno.

8 Le trasformazioni

Il concetto di *trasformazione* degli *spazi* è uno dei metodi più interessanti per l'indagine di un fenomeno. Spesso, infatti, per risolvere un problema è conveniente trasformare gli spazi in altri spazi rispetto ai quali il problema sia di più facile risoluzione. Le informazioni che si ottengono nello spazio convenientemente trasformato vengono poi correlate al problema originario.

Le trasformazioni negli spazi $\mathbb{R}^2$ ed $\mathbb{R}^3$ costituiscono il campo di applicazione geometrico. Sebbene le teorie delle trasformazioni hanno origine nella geometria, il concetto di trasformazione non è peculiare solo di questa branca ma interessa tutti i campi della matematica e della fisica.

È utile introdurre le trasformazioni nel senso più astratto, cioè applicandole ad insiemi (non necessariamente di soli punti e rette).

Si consideri un insieme $\mathbb{R}^k$ (con $k \in \mathbb{N}$) e sia X un sottoinsieme di $\mathbb{R}^k$ (se $k = 1, 2$ o 3 è possibile rappresentare graficamente l'insieme; comunque, con riferimento alla rappresentazione introdotta da Eulero, possiamo formalmente rappresentare un qualsiasi insieme).

Sia assegnata una legge di applicazione f in X (cioè una funzione) che associ ad ogni punto P di coordinate $P = (x_1, x_2, \dots, x_k)$ il punto Q dell'insieme $\mathbb{R}^n$ (con $n \in \mathbb{N}$):

$$\begin{aligned} Q &= f(P) \quad \text{oppure:} \quad Q = (f_1(P), f_2(P), f_3(P), \dots, f_n(P)) \\ \text{oppure:} \quad f(P) &= (f_1(P), f_2(P), f_3(P), \dots, f_n(P)) \end{aligned} \quad (1)$$

Queste particolari funzioni, o applicazioni da $\mathbb{R}^k \rightarrow \mathbb{R}^n$

$$f: X \subseteq \mathbb{R}^k \rightarrow \mathbb{R}^n \quad (2)$$

prendono il nome di trasformazioni di $\mathbb{R}^k$ in $\mathbb{R}^n$. Tale scrittura è la stessa se si tratta di trasformazioni vettoriali, infatti un vettore in uno spazio a k dimensioni sarà caratterizzato da k componenti.

All'applicazione f si estendono i concetti visti per le funzioni (suriettive, iniettive, biunivoche, inverse etc.).

Ogni trasformazione (2) caratterizzata dalla (1) individua n funzioni reali $f_1(P), f_2(P), f_3(P), \dots, f_n(P)$ delle k variabili $(x_1, x_2, \dots, x_k)$. Ciascuna f si chiama componente della trasformazione.

Caso particolare in cui $k = n$: trasformazioni da $\mathfrak{R}^n \rightarrow \mathfrak{R}^n$ (caso vettoriale)

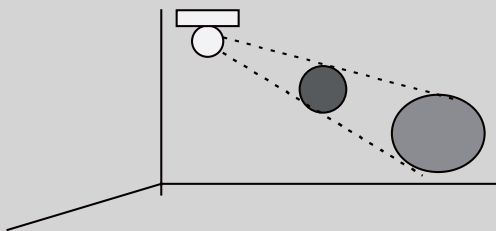
In questo caso lo spazio di partenza (dominio), dal punto di vista topologico, è lo spazio $\mathfrak{R}^n$ individuato da punti ad n componenti. Lo spazio di arrivo (codominio dell'applicazione) è uno spazio vettoriale di $\mathfrak{R}^n$.

Una tale trasformazione si chiama anche *campo vettoriale*. In particolare, una trasformazione di $\mathfrak{R}^2 \rightarrow \mathfrak{R}^2$ (leggi di erre 2 in erre 2) è una trasformazione piana.

ESEMPIO 2

1. Assegnata una massa m , è possibile associare allo spazio $\mathfrak{R}^3$ (spazio euclideo) circostante la massa, il campo vettoriale delle accelerazioni (ad esempio gravitazionali).
2. Assegnato un punto P che si muove secondo un'equazione oraria $s = s(t)$, è possibile associare ai punti occupati da P , nel tempo, i campi vettoriali delle posizioni, delle velocità o delle accelerazioni.
3. Assegnata una carica elettrica q , è possibile associare allo spazio $\mathfrak{R}^3$, circostante la massa, il campo vettoriale elettrico.
4. La funzione potenziale (elettrico, gravitazionale etc.), al contrario, associa ad un campo vettoriale un campo scalare (vedi campi conservativi).

L'aspetto di una trasformazione può essere di tipo geometrico, vettoriale o geometrico-vettoriale. Dal punto di vista didattico il problema delle trasformazioni lineari fra insiemi dello stesso ordine $\mathfrak{R}^k \rightarrow \mathfrak{R}^k$, con $k = 2$, può essere affrontato con un pantografo, cioè quello strumento per riprodurre disegni in una diversa scala, costituito da un parallelogramma articolato deformabile.



Un altro fenomeno, facilmente riproducibile, che costituisce una trasformazione lineare fra insiemi di $\mathfrak{R}^2$ è di tipo ottico-geometrico: la proiezione ottica di un disco (es. una moneta) su di una parete.

Se si accende una lampadina e la si assimila ad una sorgente puntiforme, e se la moneta è interposta fra la sorgente e la parete, allora la sua proiezione è un cerchio se, e solo se, la moneta è parallela alla parete; negli altri casi la proiezione ottica è un'ellisse.

Altri esempi di applicazione didattica relativi ad una trasformazione lineare di $\mathfrak{R}^2$ in $\mathfrak{R}^2$ sono le cartine geografiche, o in generale le cartine relative al rilevamento geologico di una entità fisica, viste in un dominio esteso della crosta terrestre e riportate, attraverso una opportuna scala di riduzione, in un altro insieme piano.

Se pensiamo ad un qualsiasi diagramma (es. diagramma polare, diagramma a bastoni o colonne, diagrammi a nastri, diagrammi di stato, superfici di stato, areogramma a torta, cartogramma, ideogramma, diagrammi di flusso etc.) ci rendiamo conto che quella rappresentazione geometrica è una trasformazione in generale di un evento numerico-vettoriale nel piano.

Dal punto di vista geometrico le assonometrie sono rappresentazioni nel piano di eventi che evolvono in uno spazio a tre dimensioni, ossia sono trasformazioni di $\mathfrak{R}^3 \rightarrow \mathfrak{R}^2$.

In molti casi un oggetto è la riproduzione in scala ridottissima o ingigantita del contorno di un altro oggetto macroscopico o microscopico. Ad esempio, i giocattoli oppure i modelli aerei e navali per studi aerodinamici e fluidodinamici sperimentali, in scala ridotta, sono trasformazioni di $\mathfrak{R}^3 \rightarrow \mathfrak{R}^3$.

L'ecografia, la radiografia, la tac, la risonanza magnetica, i diagrammi spirometrici, i diagrammi isostatici e i diagrammi degli stati tensionali di un solido deformato sono esempi in cui spesso si sfrutta il colore (variazione cromatica, sfumature del grigio ed un'unità di misura a volte anche qualitativa) per esprimere la variazione di un'entità fisico-chimica. In questi casi si riconosce che il sistema di trasformazione può assumere un carattere più generale, cioè del tipo $\mathfrak{R}^k \rightarrow \mathfrak{R}^n$.

Il punto di vista sperimentale-scientifico inquadra le trasformazioni fisiche in trasformazioni topologiche, o meglio in spazi topologici $\mathfrak{R}^n$; più «accurati» e riproducibili sono i risultati ottenuti, migliore è il modello analitico (matematizzazione) che studia l'evoluzione del fenomeno. Sotto questo punto di vista, il «calcolatore» (o elaboratore numerico) ha permesso lo studio simulato (si pensi alla realtà virtuale) in molti settori della ricerca scientifica e nei processi industriali.

9 Funzioni di due variabili e loro applicazioni

Una funzione di due variabili è una relazione che associa ad ogni coppia di numeri reali $(x, y) \in D \subseteq \mathbb{R}^2$, uno ed un solo valore reale di z . D è il *dominio* della funzione, x e y sono le *variabili indipendenti*, z la *variabile dipendente*.

Per indicare una funzione di due variabili si usa la scrittura $z = f(x, y)$.

Il dominio o campo di esistenza è l'insieme dei valori reali che si possono attribuire alle variabili indipendenti x e y affinché sia determinabile il corrispondente valore della variabile dipendente z .

Sono esempi di funzioni di due variabili:

$$z = 3xy + 2x^2y; \quad z = \frac{xy^2 - 3xy}{x + y - 4}; \quad z = \sqrt{x^2 + y^2 - 4}$$

Il dominio della prima funzione è costituito da tutte le coppie di numeri reali appartenenti ad $\mathbb{R}^2$, ossia da tutti i punti del piano; infatti, in questa funzione, le operazioni che occorre svolgere per trovare il valore di z , noti quelli di x e y , sono sempre possibili nel campo dei numeri reali. Considerato che i punti del piano si rappresentano mediante coppie di numeri reali, è indifferente parlare di queste ultime o di punti del piano.

Il dominio della seconda funzione è formato da tutte le coppie di numeri reali, escluse quelle che annullano il denominatore (ossia i punti che si trovano sulla retta di equazione $x + y - 4 = 0$).

La terza funzione è definita sui punti esterni della circonferenza di equazione $x^2 + y^2 - 4 = 0$ e sui punti della circonferenza stessa. Questo risultato si ottiene ponendo il radicando maggiore o uguale a zero e considerando il fatto che la radice quadrata di un numero è possibile solo quando il radicando è positivo o nullo.

Una funzione di due variabili può essere rappresentata in un sistema tridimensionale xyz e il suo dominio si trova nel piano xy .

In definitiva, per trovare il dominio di una funzione di due variabili si utilizzano le seguenti regole:

- se la funzione è razionale intera, il dominio è costituito da $\mathbb{R}^2$, cioè da tutte le coppie ordinate di numeri reali;
- se la funzione è razionale fratta, il dominio si determina ponendo il denominatore diverso da zero;
- se la funzione è irrazionale a indice pari, il dominio si calcola ponendo il radicando ≥ 0 ;
- se la funzione è logaritmica, si pone l'argomento > 0 .

Nei casi più complessi il dominio si determina mediante un sistema in cui si inseriscono le varie condizioni.

Ad esempio, per trovare il dominio della funzione $z = \sqrt{y - x^2 - x + 3} + \log(x + y - 5)$ si deve risolvere il sistema:

$$\begin{cases} y - x^2 - x + 3 \geq 0 \\ x + y - 5 > 0 \end{cases}$$

Il *grafico* di una funzione di due variabili è costituito da una superficie e, a differenza di quello di una funzione di una variabile (linea), è difficile da rappresentare non disponendo di un software *ad hoc*.

Una semplice rappresentazione grafica di una funzione di due variabili è possibile mediante le *linee o curve di livello*.

Le linee di livello si ottengono sezionando il grafico di una funzione di due variabili con dei piani paralleli al piano xy e proiettando le linee ottenute su quest'ultimo piano. Per determinarle si risolve il sistema:

$$\begin{cases} z = f(x, y) \\ z = k \end{cases}$$

In questo modo si trovano i punti di intersezione fra il grafico della funzione e un fascio di piani paralleli al piano xy . Per tracciare alcune di queste linee di livello basta attribuire numeri reali a piacere (generalmente numeri interi) a k .

La funzione in ogni punto di ciascuna linea di livello assume lo stesso valore.

Graficamente, le linee di livello possono essere *rette parallele*, *circonferenze concentriche*, *parabole*, *ellissi* etc.

Le rette parallele si hanno nel caso in cui la funzione è lineare (piano). In questa eventualità, attribuendo a k valori arbitrari, si ottengono rette con lo stesso coefficiente angolare e quindi parallele. La funzione:

$$z = x^2 + y^2 - 10$$

ha, ad esempio, per valori positivi di k , linee di livello che sono circonferenze concentriche. Le linee di livello della funzione:

$$z = 2x^2 + y - 8$$

sono parabole i cui grafici sono situati uno all'interno dell'altro. Dato che, sostanzialmente, le linee di livello si trovano sostituendo k al posto di z nella funzione, conoscendo la geometria analitica, è facile stabilire la natura delle linee di livello dall'espressione algebrica della funzione.

Le linee di livello vengono utilizzate in vari campi.

In *geografia* sono note le *isoipse*, linee che uniscono su una carta geografica o topografica tutti i punti aventi la stessa altitudine rispetto al livello del mare. Su una zona montana, ad esempio, ci sono la linea dei seicento metri, quella dei settecento metri, quella degli ottocento e così via. In meteorologia le *isoterme* sono le linee che uniscono, su una carta del tempo (carta sinottica), tutti i punti con la stessa temperatura; le *isobare* quelle che uniscono i punti con la stessa pressione.

In *analisi matematica* le linee di livello si utilizzano per determinare i *massimi* e i *minimi liberi* e *vincolati* di una funzione di due variabili. Nel caso di massimi e minimi liberi, se le linee di livello si restringono intorno ad un punto del dominio per valori crescenti di k , il punto rappresenta un *massimo*; se si restringono per valori decrescenti di k il punto è un *minimo*. Nel caso dei massimi e minimi vincolati, questi ultimi si trovano nei punti di tangenza fra le linee di livello della funzione e l'equazione del vincolo.

In *economia* si trovano le linee degli *isocosti*, rappresentate dalle curve i cui punti hanno costi uguali; quelle degli *isoquant*i, invece, sono rappresentate dalle curve che uniscono punti aventi quantità uguali.

Un'altra applicazione è nella *programmazione lineare*. In questo tipo di problemi, le linee di livello sono rette parallele e ci consentono di determinare il valore ottimo della funzione obiettivo.

10 La retta di bilancio

La teoria del comportamento del consumatore è molto semplice: si ipotizza che i consumatori scelgano la combinazione di beni migliore tra tutte quelle che possono essere acquistate; è ovvio, però, che esistano delle limitazioni all'acquisto.

La più semplice tra le limitazioni all'acquisto è, intuitivamente, il *vincolo di bilancio*: è ovvio che si può acquistare solo ciò che si può pagare.

Quindi, se indichiamo con (x_1, x_2) il paniere di consumo del consumatore (ovvero i due beni 1 e 2 che egli desidera) aventi, rispettivamente, un prezzo p_1 e p_2 , risulta chiaro che:

$$p_1x_1 + p_2x_2 \leq r$$

dove r è il reddito che è a disposizione per l'acquisto.

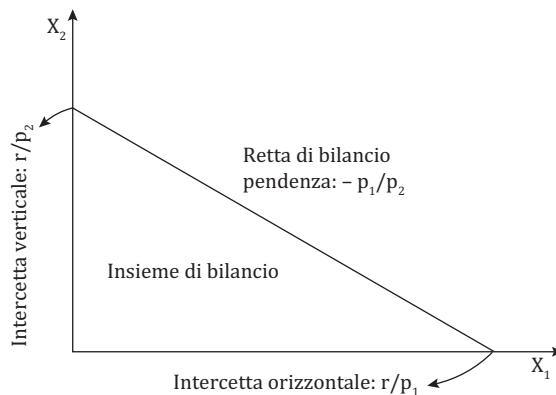
È bene notare che il fatto che la seguente trattazione si riferisca a soli due beni non ne limita affatto la generalità. È, infatti, possibile considerare uno dei due beni (in genere il bene 2) come bene composito, ovvero tutto ciò che il consumatore desidera oltre al bene 1.

Se consideriamo l'equazione:

$$p_1x_1 + p_2x_2 = r$$

si vede che, indicando x_1 sull'asse x e x_2 sull'asse y di un sistema di assi cartesiani ortogonali, si ottiene una retta avente le seguenti caratteristiche:

- intercetta verticale: $\frac{r}{p_2}$;
- intercetta orizzontale: $\frac{r}{p_1}$;
- pendenza: $-\frac{p_1}{p_2}$.



Questa retta è detta *retta di bilancio*.

Risulta chiaro che tutti i punti alla sinistra della retta e compresi nel primo quadrante ($x_1 \geq 0$ e $x_2 \geq 0$, ovviamente) sono quelli per cui:

$$p_1x_1 + p_2x_2 < r$$

e, insieme ai punti della retta di bilancio, costituiscono il cosiddetto *insieme di bilancio*, ovvero tutte le combinazioni di x_1 e x_2 che il consumatore può permettersi (considerando fissati i prezzi p_1, p_2 ed il reddito r). È chiaro che per comprendere come varia tale insieme è sufficiente analizzare il comportamento della retta. Ad esempio, un aumento del reddito (di un Δr) implica uno spostamento verso destra della retta senza modificarne la pendenza (che è indipendente da r). Infatti, si ha:

- intercetta verticale: $\frac{r + \Delta r}{p_2}$;
- intercetta orizzontale: $\frac{r + \Delta r}{p_1}$;
- pendenza: $-\frac{p_1}{p_2}$.

Dunque, l'aumento del reddito porta ad ampliare il numero di coppie (x_1, x_2) in maniera concorde alla retta di bilancio.

Diverso è il caso di un aumento dei prezzi dei beni desiderati dal consumatore.

Supponiamo che tali prezzi varino contemporaneamente.

Se l'aumento è il medesimo, si avrà:

$$ap_1x_1 + ap_2x_2 = r$$

che può anche essere scritta:

$$p_1x_1 + p_2x_2 = \frac{r}{a}$$

Ovvero un simile tipo di aumento equivale a una riduzione del reddito pari a $\frac{1}{a}$ volte. Se, oltre ad aumentare il prezzo, diminuisce anche il reddito, si ha uno spostamento verso sinistra della retta e, quindi, una diminuzione dell'insieme di bilancio.

Abbiamo visto come variano le intercette della retta di bilancio.

Si può sottolineare come la *pendenza* vari solo in dipendenza dei prezzi dei beni. Si può anche dare una diversa interpretazione a tale pendenza.

Ipotizziamo che il consumatore voglia aumentare il consumo del bene 1 di una quantità Δx_1 . Risulta chiaro che, se le altre condizioni permangono costanti, lo stesso consumatore dovrà anche ridurre il consumo del bene 2 di una quantità Δx_2 . Avremo:

$$p_1(x_1 + \Delta x_1) + p_2(x_2 + \Delta x_2) = r$$

Sottraendo a questa equazione quella della retta di bilancio prima delle variazioni, avremo:

$$p_1(x_1 + \Delta x_1) + p_2(x_2 + \Delta x_2) - (p_1x_1 + p_2x_2) = r - r \Leftrightarrow p_1\Delta x_1 + p_2\Delta x_2 = 0$$

da cui:

$$p_1\Delta x_1 = -p_2\Delta x_2$$

ovvero:

$$-\frac{p_1}{p_2} = \frac{\Delta x_2}{\Delta x_1}$$

L'ultima espressione ci permette di dire che la pendenza $\left(-\frac{p_1}{p_2}\right)$ della retta di bilancio rappresenta il *costo opportunità* del consumo del bene 1. Per aumentare il consumo del bene 1 occorre rinunciare a una parte del bene 2. Il costo economico vero e proprio di questo incremento è rappresentato dalla rinuncia all'opportunità di consumare il bene 2 ed è misurato dall'inclinazione della retta di bilancio.

11 Il tasso di cambio nominale

Il *tasso di cambio nominale*, indicato con e , è il prezzo della moneta nazionale rispetto alla moneta estera. Esso indica, ad esempio, quanti dollari, o yen giapponesi, o sterline inglesi etc. ci vogliono per comprare un euro. Se per comprare un euro ci vogliono più dollari oggi rispetto a ieri, si dice che l'euro si è apprezzato sul dollaro, mentre si dice che si deprezza in caso contrario. La scrittura:

$$\text{euro/dollaro} = 1,32$$

indica che per comprare un euro sono necessari 1,32 dollari. Viceversa, la scrittura:

$$\text{dollaro/euro} = 0,76$$

è l'inversa della precedente e indica che per comprare un dollaro sono necessari 76 centesimi di euro.

La relazione tra il tasso di cambio nominale tra la valuta A e la valuta B (Tc_{AB}) ed il corrispondente tasso di cambio tra la valuta B e la valuta A (Tc_{BA}) è dato, semplicemente, da:

$$Tc_{AB} = \frac{1}{Tc_{BA}}$$

12 L'ottimizzazione vincolata

a) Nozioni generali

Molto spesso in economia si affrontano problemi di ottimizzazione vincolata, che si risolvono mediante il calcolo differenziale.

Sia data una funzione $f(x, y)$ soggetta ad un vincolo $g(x, y)$ (il vincolo è un insieme di condizioni di natura economica o tecnica alla quale devono necessariamente sottostare x ed y); da esse si può formare una nuova funzione detta *lagrangiana*:

$$F(x, y, \lambda)$$

ottenuta uguagliando a 0 il vincolo, moltiplicandolo per λ (dove λ è detto *moltiplicatore di Lagrange*) e sommando il prodotto così ottenuto alla funzione originaria. Si calcolano le derivate parziali prime rispetto ad x, y, λ , si uguagliano a 0 e si trovano i *valori critici* (che sono le soluzioni del sistema ottenuto). Tali valori critici si sostituiscono nella $F(x, y, \lambda)$ (detta anche *funzione obiettivo*) permettendone l'ottimizzazione.

b) Programmazione lineare

In questo contesto ci occuperemo di problemi di programmazione lineare. L'obiettivo della programmazione lineare è quello di individuare l'*allocazione ottimale di risorse* in una situazione economica in cui ci sono prodotti o attività concorrenti.

La programmazione lineare permette di risolvere più facilmente problemi di ottimizzazione che presentano numerosi vincoli di disuguaglianza; se poi i vincoli, pur numerosi, sono limitati a due variabili, la soluzione grafica è particolarmente adatta.

c) Il duale

Nella programmazione lineare ad ogni problema (*primale*) di massimizzazione corrisponde un problema (*duale*) di minimizzazione (e viceversa).

Passare dal problema primale al duale può essere utile quando nel primale abbiamo *tre variabili* (e quindi non possiamo ricorrere alla soluzione grafica) e *due vincoli di disuguaglianza*; nel problema duale, in tal caso, troviamo due sole variabili (il che ci permette di servirci della soluzione grafica) e più di due vincoli di disuguaglianza.

Inoltre, risolvere un problema di minimizzazione in termini di massimizzazione può, a volte, risultare più facile.

Per passare dal problema primale al problema duale bisogna tener presente le seguenti regole:

- l'*ottimizzazione del primale* corrisponde alla minimizzazione del duale (e viceversa);
- i *segni di disuguaglianza* dei vincoli si invertono, tranne quelli delle variabili che rimangono positive anche nel problema duale;

- i *coefficienti dei vincoli del primale* formano una matrice dalla quale possiamo ottenere la matrice del duale dopo aver trasposto le righe in colonne;
- i *coefficienti della funzione obiettivo del problema primale* diventano le costanti dei vincoli del duale;
- le *costanti dei vincoli del primale* formano i coefficienti della funzione obiettivo del duale;
- le *variabili* x_i del primale si sostituiscono con le variabili z_i del duale.

È bene tener presente, inoltre, che:

- se esiste una soluzione ottimale ammissibile nel problema primale, questa è uguale alla soluzione ottimale ammissibile del duale (cd. *primo teorema del duale*);
- se nella soluzione ottimale ammissibile una variabile del primale ha valore diverso da zero, la variabile fittizia corrispondente nel duale deve avere valore ottimale uguale a zero;
- se, invece, nella soluzione ottimale ammissibile una variabile fittizia del primale ha valore diverso da 0, la corrispondente variabile del duale deve avere valore ottimale uguale a 0 (cd. *secondo teorema del duale*).

ESEMPIO 3

Ottimizzare la funzione:

$$f(x, y) = 7x^2 - 9y + 10x + 3$$

con il vincolo:

$$3x - 2y = 10$$

Si uguaglia il vincolo a 0 e lo si moltiplica per λ :

$$\lambda(3x - 2y - 10)$$

Tale risultato viene poi sommato alla $f(x, y)$, ottenendo così la funzione:

$$F(x, y, \lambda) = 7x^2 - 9y + 10x + 3 + \lambda(3x - 2y - 10)$$

Successivamente si uguagliano a 0 le derivate parziali prime:

$$\frac{\partial F}{\partial x} = 14x + 10 + 3\lambda = 0$$

$$\frac{\partial F}{\partial y} = -9 - 2\lambda = 0$$

$$\frac{\partial F}{\partial \lambda} = 3x - 2y - 10 = 0$$

Risolvendo il sistema formato dalle tre equazioni:

$$\begin{cases} 14x + 10 + 3\lambda = 0 \\ 2\lambda = -9 \\ x = \frac{2y + 10}{3} \end{cases} \Rightarrow \begin{cases} 14\left(\frac{2y + 10}{3}\right) + 10 + 3\left(-\frac{9}{2}\right) = 0 \\ \lambda = -\frac{9}{2} \\ x = \frac{2y + 10}{3} \end{cases}$$

Siamo così condotti alla soluzione dell'equazione:

$$\frac{28}{3}y + \frac{140}{3} + 10 - \frac{27}{2} = 0$$

$$56y + 280 + 60 - 81 = 0$$

$$56y + 259 = 0$$

$$y = -\frac{259}{56} = -\frac{37}{8}$$

$$x = \frac{2y + 10}{3} = \frac{2(y + 5)}{3} = \frac{2}{3}(y + 5) = \frac{2}{3}\left(-\frac{37}{8} + 5\right) = \frac{2}{3}\left(\frac{-37 + 40}{8}\right) = \frac{2}{3}\left(\frac{3}{8}\right) = \frac{1}{4}$$

Si hanno cos  le soluzioni:

$$\bar{x} = \frac{1}{4}, \quad \bar{y} = -\frac{37}{8}, \quad \bar{\lambda} = -\frac{9}{2}$$

L'operazione di calcolo delle derivate parziali di primo ordine e di eguagliarle a 0 trova la sua giustificazione teorica nel **teorema di Lagrange**, secondo cui una scelta ottima $(\bar{x}, \bar{y})$ deve soddisfare le tre condizioni del primo ordine:

$$\frac{\partial F}{\partial x} = 0; \quad \frac{\partial F}{\partial y} = 0; \quad \frac{\partial F}{\partial \lambda} = 0$$

Il valore della funzione $F(x, y, \lambda)$, calcolata sostituendo al posto delle incognite le soluzioni trovate (dette valori critici),  :

$$\begin{aligned} F\left(\frac{1}{4}, -\frac{37}{8}, -\frac{9}{2}\right) &= 7\left(\frac{1}{4}\right)^2 - 9\left(-\frac{37}{8}\right) + 10\left(\frac{1}{4}\right) + 3 - \frac{9}{2}\left(3 \cdot \frac{1}{4} + 2 \cdot \frac{37}{8} - 10\right) = \\ &= \frac{7}{16} + \frac{333}{8} + \frac{5}{2} + 3 - \frac{9}{2}\left(\frac{3}{4} + \frac{37}{4} - 10\right) = \frac{7}{16} + \frac{333}{8} + \frac{5}{2} + 3 - \frac{9}{2}\left(\frac{40}{4} - 10\right) = \\ &= \frac{7}{16} + \frac{333}{8} + \frac{5}{2} + 3 = \frac{7+666+40+48}{16} = \frac{761}{16} = 47,5625 \end{aligned}$$

La funzione sar  dunque ottimizzata per i valori critici:

$$\bar{x} = \frac{1}{4}, \quad \bar{y} = -\frac{37}{8}, \quad \bar{\lambda} = -\frac{9}{2}$$

13 L'interesse composto annuo

Si parla di *interesse composto annuo* quando gli interessi di un dato investimento vengono capitalizzati annualmente, ovvero quando sono aggiunti al capitale dell'anno precedente e reinvestiti.

Ad esempio, se si investissero 1.000 euro per 4 anni al tasso d'interesse del 7%, dopo il primo anno maturerebbero 70 euro d'interesse, che sarebbero aggiunti al capitale iniziale di 1.000 euro, formando un capitale di 1.070 euro. Nel secondo anno l'interesse del 7% sul nuovo capitale frutterebbe 74,90 euro.

Nel terzo anno, quindi, si avrebbero 1.144,90 euro (1.070 + l'interesse maturato di 74,90 = 1.144,90 euro), che, reinvestiti al 7%, frutterebbero 80,14 euro di interessi. All'inizio del quarto, la somma investita sarebbe 1.225,04 euro, che, con l'interesse del 7%, darebbero a fine anno 85,75 euro di interessi, per un montante totale di 1.310,79 euro.

Il ragionamento formulato si pu  sintetizzare con una formula matematica, che permette di calcolare il montante M ottenuto da un capitale iniziale C , investito per t anni al tasso d'interesse annuo i . La formula   la seguente:

$$M_t = C(1 + i)^t$$

Provando a reimpostare il calcolo proposto in precedenza con l'utilizzo di questa formula, e considerando che $C = 1.000$ euro, $t = 4$, $i = 7\% = 7/100$, si ha:

$$M_4 = 1.000 \cdot (1 + 7/100)^4 = 1.000 \cdot (107/100)^4 = 1.000 \cdot (1,07)^4 = 1.000 \cdot (1,31079) = 1.310,79$$

Vediamo come si arriva alla formula $M_t = C(1 + i)^t$.

Il montante, al termine del primo anno,   uguale al capitale iniziale pi  l'interesse maturato, ovvero: $M_1 = C + Ci = C(1 + i)$.

Il montante, al termine del secondo anno, è uguale al capitale presente all'inizio del secondo anno, che a sua volta coincide con il montante maturato al termine del primo anno (M_1) più l'interesse maturato sul capitale di inizio anno, vale a dire:

$$M_2 = M_1 + M_1 i = M_1(1 + i) = C(1 + i)(1 + i) = C(1 + i)^2$$

Il montante, al termine del terzo anno, è uguale al capitale presente all'inizio del terzo anno, che a sua volta coincide con il montante maturato al termine del secondo anno (M_2) più l'interesse maturato sul capitale di inizio anno, ossia:

$$M_3 = M_2 + M_2 i = M_2(1 + i) = C(1 + i)^2(1 + i) = C(1 + i)^3$$

Generalizzando, il montante dopo t anni si calcola con la formula $M_t = C(1 + i)^t$.

14 Lo spazio percorso da un punto

La velocità di un punto P è:

$$v = 0,2t^2 \text{ m/sec}$$

Calcolare lo spazio s percorso dal punto nell'intervallo di tempo $\Delta t = 10$ sec a partire dall'istante $t_0 = 0$. Qual è la velocità media del punto in questo intervallo di tempo?

Se un punto si muove lungo una curva γ ed il valore assoluto della sua velocità $v = f(t)$ è una funzione data del tempo t , lo spazio percorso da P , nell'intervallo di tempo $[t_1, t_2]$, è uguale a:

$$v = \frac{ds}{dt} \Rightarrow ds = v dt \Rightarrow \Delta s = \int_{t_1}^{t_2} f(t) dt$$

Si ha, nel nostro caso:

$$s = \int_0^{10} 0,2t^2 dt = 0,2 \left[\frac{t^3}{3} \right]_0^{10} = 66,7 \text{ m}$$

$$v_{med} = \frac{\Delta s}{\Delta t} = \frac{66,7}{10} = 6,67 \text{ m/sec}$$

15 Il lavoro di una forza

Quale lavoro deve essere compiuto per allungare una molla di 3 cm se, per una forza di 0,5 kg, la molla si allunga di 1 cm?

Se una forza variabile $F = f(x)$ agisce su un punto P che si sposta dalla posizione x_1 verso x_2 , il lavoro di questa forza è uguale a:

$$L = \int_{x_1}^{x_2} f(x) dx$$

In virtù della *legge di Hooke*, la forza F che allunga la molla di x metri è:

$$F = kx$$

dove k è il coefficiente di proporzionalità. Poiché $x = 0,01$ m ed $F = 4,9$ N, si ha che $k = 490$ N/m e, di conseguenza, $F = 490x$.

Quindi, il lavoro da compiere per allungare la molla di 3 cm è:

$$L = \int_0^{0,03} 490x dx = 490 \left[\frac{x^2}{2} \right]_0^{0,03} = 0,22 \text{ N/m}$$

16 L'equazione oraria del moto rettilineo uniforme

Assegnata una funzione $f(x)$, ricerchiamo una funzione primitiva. È noto che, per una funzione reale, se esiste una primitiva ne esistono infinite che differiscono a meno di una costante. Se $g(x)$ è una primitiva di $f(x)$, nel senso che l'integrale indefinito:

$$g(x) = \int f(x) dx$$

allora tutte le funzioni del tipo $g(x) + k$ (con $k \in \mathbb{R}$) sono primitive.

Il problema della ricerca dell'unica primitiva è associato alle condizioni iniziali. Ad esempio, nei problemi cinematici in cui il moto è rettilineo uniforme, indicando con s lo spazio e con $s(t)$ la funzione oraria, ed essendo nota la velocità $v(t) = s'(t) = \text{cost}$, per determinare un'unica funzione $s(t)$ bisogna non solo conoscere il valore della costante ($v = \text{cost}$), ma anche la posizione ad un prefissato istante (in molti casi si ha che, per $t = 0$, $s(0) = 0$).

Ad esempio, considerate un moto rettilineo uniforme (sottintendiamo le unità di misura) con $s'(t) = 2$, e tale che, all'istante $t = 3$, si abbia $s(3) = 1$. Per determinarne l'equazione oraria, bisognerebbe risolvere il seguente integrale:

$$\int s'(t) dt = \int 2 dt \Rightarrow s(t) = 2t + k$$

Fra tutte le funzioni $s(t)$ vi è un'unica primitiva che soddisfa la condizione iniziale $s(3) = 1$. Per determinare tale funzione (legge oraria), basta risolvere l'equazione:

$$2t + k = 1 \quad \text{per } t = 3$$

Si ottiene così un'equazione algebrica di primo grado nell'incognita k :

$$2 \cdot 3 + k = 1 \quad \Rightarrow \quad k = -5$$

In definitiva, la legge oraria è:

$$s(t) = 2t - 5$$

17 Equazioni differenziali e funzione armonica

Consideriamo il sistema costituito da un corpo rigido di massa m e una molla elastica il cui coefficiente di elasticità è K (dimensionalmente è una Forza/spostamento).

Supponiamo che il sistema sia costituito da una serie fra la massa e la molla e sia vincolato in modo che la massa possa spostarsi o muoversi lungo una sola direzione x . Supponiamo, inoltre, che i vincoli siano lisci e privi di attrito.

Orientiamo l'asse x in accordo alla suddetta direzione, e poniamo come origine la posizione del baricentro della massa m corrispondente alla condizione in cui la molla è scarica.

La condizione di molla scarica è una condizione di equilibrio statico del sistema. Se si sposta la massa m (cioè si disturba il sistema in equilibrio), avverrà la reazione della molla ($R = Kx$) rivolta sempre verso l'origine O della retta orientata x .

Se la massa m viene spostata dalla sua condizione di riposo e quindi lasciata libera, essa ritorna verso la condizione di riposo animata da un moto lineare (non uniforme) $x(t)$, che dipende da m , K e dalla posizione in cui è stata spostata la massa (x_0). Il problema consiste nell'individuare l'equazione del moto $x = x(t)$.

Utilizziamo il modello del corpo libero, cioè isoliamo la massa m ed applichiamo nel suo baricentro tutte le reazioni vincolari e le forze che agiscono su di essa.

La forza peso P e la reazione vincolare N si annullano e non contribuiscono al moto, in quanto i vincoli sono privi di attrito. Sul corpo agirà la reazione della molla R .

Tale reazione è un vettore che ha come direzione x , verso quello diretto verso l'origine O e modulo $|R| = K \cdot \ell$ (dove ℓ è la distanza da O).

Non è difficile rendersi conto che la funzione:

$$R = -K \cdot x(t)$$

rappresenta proprio il vettore R . Infatti, quando x è positivo (molla compressa) la reazione è opposta alla direzione di x ; quando x è negativa (molla dilatata) la reazione ha direzione «positiva» (la funzione $R = -Kx$ assume valori positivi).

Se $x(t)$ è la funzione posizione nel tempo (legge oraria), la sua derivata $x'(t)$ è la funzione velocità e la derivata seconda $x''(t)$ è la funzione accelerazione.

Applicando la legge fondamentale della dinamica di Newton, $F = ma$, si ha che:

$$R = mx''(t)$$

Sostituendo ad R , si ottiene:

$$-Kx(t) = mx''(t)$$

Ordinando:

$$mx''(t) + Kx(t) = 0$$

Quest'equazione è un'equazione differenziale, ordinaria, lineare, del secondo ordine, a coefficienti costanti, omogenea.

Risolviamo tale equazione, cioè troviamo l'integrale generale $x(t)$ che fisicamente esprime il moto di m .

Consideriamo l'equazione algebrica associata:

$$m \cdot \alpha^2 + K = 0$$

Tale equazione ammette due soluzioni complesse e coniugate:

$$\alpha = \pm i \sqrt{(K/m)}$$

È possibile dimostrare che l'integrale generale è:

$$x(t) = A \cos(\sqrt{K/m} \cdot t)$$

Verifichiamo che tale funzione soddisfi l'equazione differenziale. Bisogna calcolare le derivate di $x(t)$ e sostituirle nell'equazione.

$$x'(t) = -A \sqrt{(K/m)} \sin(\sqrt{(K/m)} \cdot t)$$

$$x''(t) = -A (K/m) \cos(\sqrt{(K/m)} \cdot t)$$

Sostituendo l'ultimo risultato così ottenuto, si ha:

$$m \cdot \{-A(K/m)\cos(\sqrt{(K/m)} \cdot t)\} + K \cdot \{A\cos(\sqrt{(K/m)} \cdot t)\} = 0$$

da cui:

$$-A \cdot K\cos(\sqrt{(K/m)} \cdot t) + K \cdot A\cos(\sqrt{(K/m)} \cdot t) = 0$$

poiché si tratta di una identità ($\forall t \in R$ il primo membro è uguale al secondo membro), la funzione $x(t)$ è una soluzione dell'equazione differenziale $mx''(t) + Kx(t) = 0$, detta *equazione dei moti armonici*.

La funzione $y = A\cos(\omega x)$, dove $\omega = \sqrt{K/m}$ e A sono chiamate, rispettivamente, *pulsazione* ed *ampiezza*, è detta *funzione armonica*.

Le *condizioni ai limiti* o *condizioni iniziali* di tale problema sono due:

- all'istante $t_0 = 0$ $x(0) = A$;
- all'istante $t_0 = 0$ $x'(0) = 0$;

che esprimono, rispettivamente, la posizione e la velocità della massa nel momento in cui è lasciata libera (istante iniziale).

I principali software per imparare e sperimentare la matematica

Sono numerosi i software attualmente applicati allo studio della matematica. Tra i più conosciuti, soprattutto in ambito didattico, ci sono *Derive* e *Cabri*.

1 Derive

Derive è un software di calcolo simbolico (Computer Algebra System - CAS). Alternative a tale software sono i programmi: *Maxima*, *Sage*, *X Cas* etc. Esso esegue calcoli sia numerici che simbolici: ad esempio, è in grado di risolvere equazioni, di semplificare espressioni letterali, tracciare grafici di curve e di superfici, di risolvere problemi di calcolo vettoriale e matriciale etc. *Derive* può essere usato a vari livelli. Il più elementare, ma meno efficace didatticamente, consiste nell'utilizzarlo come puro esecutore di calcoli. Un impiego *ad hoc* di questo programma è come mezzo di previsione teorica dei risultati. Può anche essere un utile supporto, soprattutto in alcune branche della matematica, quali ad esempio l'Analisi, per la comprensione o il rafforzamento di concetti decisamente complessi.

2 Lo schermo di Derive

Lo schermo di *Derive* comprende (dall'alto verso il basso):



1. **la barra del titolo**, la quale riporta: l'icona; il nome del programma; il nome del file associato al foglio di lavoro (se non è già stato salvato con un nome diverso, gli viene associato un titolo del tipo Algebra N, con N numero progressivo da 1 in poi); infine, a destra, i caratteristici pulsanti delle finestre Windows;

2. **la barra dei menu**, la quale è composta da diverse voci. Facendo clic su ciascuna di esse si apre un menu a tendina, il quale contiene vari comandi che Derive può eseguire. Quando un menu viene aperto, appare una finestra che contiene a sua volta altri comandi. Si può osservare che:
 - una sola lettera di ciascun comando è sottolineata; ciò vuol dire che si può accedere al comando premendo il tasto corrispondente alla lettera sottolineata;
 - alcuni comandi sono preceduti da un'icona analoga a quella riportata sulla barra dei comandi: questo significa che si ottengono risultati identici sia cliccando sul comando nel menu a tendina, sia cliccando sul comando della barra dei comandi;
 - alcuni comandi sono seguiti da tre puntini «...»: la loro scelta apre una finestra che fa sparire la precedente;
 - alcuni comandi sono seguiti da una freccia: in questo caso si apre un'ulteriore finestra senza che scompaia la precedente;
 - alcuni comandi sono seguiti da «Ctrl + lettera»: questo significa che si ottiene lo stesso risultato premendo la combinazione dei tasti indicata;
 - alcuni comandi sono preceduti dal simbolo di spunta: questo simbolo indica la scelta attiva, cioè che il comando che lo segue è selezionato. L'effetto dei comandi può essere disattivato con un clic sul comando stesso;
3. **la barra dei comandi o strumenti**, in cui sono presenti alcuni pulsanti che corrispondono ai comandi di utilizzo più frequenti. Per ricevere informazioni schematiche su un pulsante, basta spostarsi sopra di esso con il mouse. Si può notare che alcuni pulsanti appaiono «spenti» (si dice che sono in *dim light*), il che significa che in quello stato non sono utilizzabili. È semplice capire questo perché sono comandi che si riferiscono ad operazioni che agiscono su qualcosa di definito, per esempio risolvono equazioni. Poiché all'avvio del programma non è stato definito nulla, non è possibile utilizzare questi pulsanti. Quando verranno inseriti i dati necessari, assumeranno l'aspetto normale e potranno essere selezionati;
4. **una finestra algebrica**, che corrisponde alla zona in cui vengono visualizzate le espressioni immesse e i calcoli che si svolgono su di esse;
5. **la barra di stato**, che visualizza alcune informazioni sul documento e sulle opzioni scelte. Essa è suddivisa in tre parti: nella prima vengono riportati eventuali messaggi di errore oppure delle informazioni relative ai pulsanti delle barre; nella seconda dei messaggi relativi alle diverse istruzioni; nell'ultima eventuali tempi di calcolo delle espressioni;
6. **la barra di immissione**, la quale è suddivisa in due parti: la prima contiene pulsanti di esecuzione; la seconda è formata da una riga per l'inserimento delle espressioni, le quali possono essere scritte servendosi, eventualmente, dei pulsanti delle lettere e dei simboli presenti sotto di essa;
7. **le barre dei simboli greci e matematici**.

3 Alcune funzioni di Derive

L'immagine seguente mostra una delle innumerevoli funzioni [1] di questo programma, ovvero il calcolo del M.C.D. tra due numeri (questa funzione in Derive è indicata con GCD).

^[1] Derive incorpora al suo interno molte funzioni predefinite, che forniscono automaticamente, assegnando i valori in ingresso e senza che l'utente debba svolgere alcun calcolo, il risultato. Queste funzioni sono divise in categorie. Per ottenere il risultato basta scrivere nella barra di immissione la funzione con i valori in ingresso, seguita dall'uguale. Ad esempio, la funzione $GCD(a,b,c,\dots)$ restituisce il massimo comune divisore dei numeri $a, b, c, \dots$. Quindi, scrivendo « $GCD(6,15)=$ » e premendo invio, Derive restituisce il numero 3.



Di seguito è proposto un elenco di alcune tra le più utilizzate funzioni di Derive:

1. **PRIME(m)** ritorna true (vero) se m è un numero primo, false (falso) se m non è un numero primo. Ad esempio, PRIME(22564845703) restituisce falso perché 22564845703 è uguale a $106219 \cdot 212437$;
2. **NEXT_PRIME(m)** restituisce il primo «numero primo» maggiore del numero m . Ad esempio, NEXT_PRIME(1000) restituisce 1009 perché è il primo numero primo più grande di 1000;
3. **PREVIOUS_PRIME(m)** restituisce il primo numero primo minore del numero m , se esiste, altrimenti restituisce «?». Ad esempio, PREVIOUS_PRIME(1000) restituisce 997;
4. **NTH_PRIME(n)** restituisce il numero primo che occupa la posizione n nell'insieme ordinato degli infiniti numeri primi;
5. **FLOOR(a, b)** fornisce il quoziente della divisione intera tra a e b ;
6. **MOD(m, n)** calcola il resto della divisione fra i numeri interi m ed n ;
7. **FACTOR(n)** restituisce la scomposizione del numero n in fattori primi. Ad esempio, FACTOR(10) restituisce 2×5 ;
8. **DIVISORS(n)** restituisce il vettore ordinato di tutti i divisori positivi di n . Ad esempio, per determinare i divisori di 28, è sufficiente digitare l'espressione DIVISORS(28) = e si otterrà $\text{DIVISORS}(28) = [1, 2, 4, 7, 14, 28]$;
9. **GCD($a, b, c, \dots$)** restituisce il massimo comune divisore dei numeri $a, b, c, \dots$;
10. **LCM($a, b, c, \dots$)** restituisce il minimo comune multiplo dei numeri $a, b, c, \dots$. Ad esempio, LCM(6, 15, 14) restituisce 210;
11. **ABS(n)** calcola il valore assoluto del numero reale n ;
12. **APPROX(e, c)** approssima l'espressione e (se e è numerica), oppure ne approssima le parti numeriche; in ogni caso visualizza c cifre;
13. **QUOTIENT[A(x), B(x)]** restituisce il quoziente della divisione del polinomio $A(x)$ per il polinomio $B(x)$;
14. **REMAINDER[A(x), B(x)]** restituisce il resto della divisione del polinomio $A(x)$ per il polinomio $B(x)$;
15. **NSOLVE(e, v)** risolve numericamente un'espressione (equazione o disequazione che sia);
16. **SOLVE(e, v)** risolve l'equazione $e(v) = 0$ rispetto alla variabile v ;
17. **SOLUTIONS(e, v)** risolve l'equazione $e(v) = 0$ rispetto alla variabile v , scrivendo le soluzioni in un vettore;
18. **SUBST(e, v, n)** sostituisce nell'espressione $e(v)$ tutte le occorrenze della variabile v con l'espressione n ; inoltre, semplifica e restituisce il risultato;
19. **SUM[e(x), x , inf, sup]** calcola la somma degli elementi dell'espressione $e(x)$ al variare di x (assume tutti i valori interi da inf a sup);

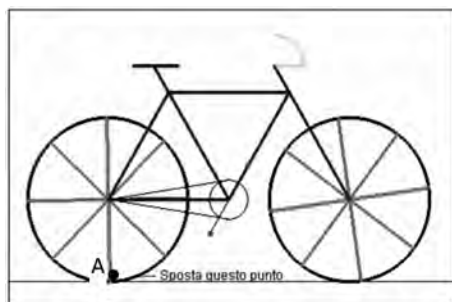
20. **IF(test, espressione 1, espressione 2)** è una funzione che consente di esprimere un'alternativa dove:
- test è la condizione che Derive valuta stabilendo la sua verità o la sua falsità;
 - espressione 1 è l'istruzione che viene eseguita se test è vera;
 - espressione 2 è l'istruzione che viene eseguita se test è falsa;
21. **ELEMENT(A,r,c)** seleziona l'elemento che, nella matrice A , occupa la riga r e la colonna c ;
22. **DET(A)** calcola il determinante della matrice quadrata A ;
23. **LN(x)** calcola il logaritmo naturale di x ;
24. **LOG(a,b)** calcola il logaritmo in base b di a ;
25. **SIN(x)** calcola il seno di x radianti o gradi, a seconda dell'unità di misura attiva;
26. **COS(x)** calcola il coseno di x radianti o gradi, a seconda dell'unità di misura attiva;
27. **TAN(x)** calcola la tangente di x radianti o gradi, a seconda dell'unità di misura attiva;
28. **COT(x)** calcola la cotangente di x radianti o gradi, a seconda dell'unità di misura attiva;
29. **SEC(x)** calcola la secante di x radianti o gradi, a seconda dell'unità di misura attiva;
30. **CSC(x)** calcola la cosecante di x radianti o gradi, a seconda dell'unità di misura attiva;
31. **ASIN(x)** calcola il minimo arco, in valore assoluto, il cui seno è x , nell'unità di misura attiva (radiante per default);
32. **ACOS(x)** calcola il minimo arco, in valore assoluto, il cui coseno è x , nell'unità di misura attiva (radiante per default);
33. **ATAN(x)** calcola il minimo arco, in valore assoluto, la cui tangente è x , nell'unità di misura attiva (radiante per default);
34. **ACOT(x)** calcola il minimo arco, in valore assoluto, la cui cotangente è x , nell'unità di misura attiva (radiante per default);
35. **LIM(f,v,p,d)** calcola il limite di un'espressione (successione, funzione), per v che tende a p e nell'eventuale direzione indicata da d . Se $d > 0$ si intende da destra, se $d < 0$ da sinistra;
36. **DIF(f,v,n)** calcola la derivata della funzione f , rispetto alla variabile v . Se il parametro n non è specificato, si calcola la derivata prima, altrimenti quella di ordine n ;
37. **INT(f,v,c)** calcola l'integrale indefinito della funzione f rispetto alla variabile v ; se è specificato c , scrive una famiglia di primitive, altrimenti scrive solo quella di parametro 0.

4 Cabri

Il *Cabri Geometre* è un software interattivo che consente di fare geometria in modo «dinamico» (Dynamic Geometry Software [2] - DGS): considera i punti, le rette, i cerchi come oggetti fondamentali e consente di creare sullo schermo oggetti complessi, specificando le relazioni che intercorrono fra essi. Le figure tracciate possono essere spostate, oppure deformate, mantenendo inalterate le relazioni e le proprietà iniziali. Ad esempio, disegnare con Cabri due punti A e B qualsiasi e, per mezzo del comando appropriato, rappresentare la retta r passante per questi due punti, equivale a imporre la relazione «La retta r passa per A e B qualunque sia la posizione nel piano dei punti A e B ». Quindi, se muoviamo con il mouse uno dei due punti, automaticamente anche la retta r si muoverà, ma passerà sempre per i due punti; non è possibile, invece, muovere la retta r .

[2] *Cabri* non è l'unico software di geometria dinamica: altri esempi sono *OpenEuclide*, *Geogebra* etc.

Con Cabri si possono formulare delle ipotesi, verificarle visivamente e numericamente, oppure è possibile chiedere la verità di una congettura geometrica. In figura è rappresentata una bicicletta, disegnata con Cabri utilizzando gli oggetti *Punto*, *Circonferenza* e *Segmento*. Se si spostasse con il mouse il punto A, la bicicletta si allontanerebbe come nella realtà: le due ruote, i pedali e il telaio si muoverebbero in sintonia. Questo perché nel disegnarla sono state imposte le opportune relazioni fra gli oggetti che la compongono.



5 Avvio di Cabri

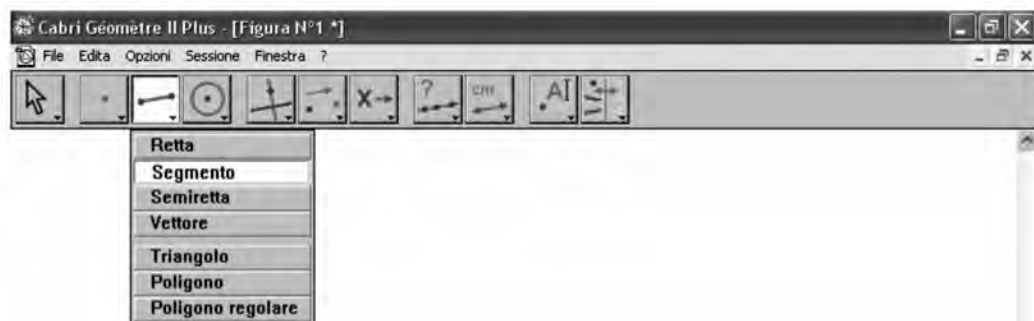
All'avvio, Cabri presenta un foglio bianco, la barra del menu e la barra dei pulsanti.



La **barra del menu** (File – Edita – Opzioni – Sessione Finestra) serve per i comandi non-geometrici, cioè per quelli di utilità (ad esempio, salvare, stampare etc.).

La **barra dei pulsanti** serve per le costruzioni geometriche. Tenendo schiacciato ciascun pulsante appare un sotto-menu con tutte le opzioni possibili. Ecco i sottomenu nelle seguenti immagini:









Il significato dei sotto-menu è più o meno intuitivo.

6 Elenco di siti web da cui trarre materiale inerente ai principali software utilizzati in matematica

Sui software applicati allo studio della matematica moltissimo ci sarebbe ancora da dire, ma sarebbe arduo esaurire l'argomento nello spazio di poche pagine. Al di là dei cenni forniti nei paragrafi precedenti, preferiamo indicarvi un elenco di *link* che rimandano a siti dai quali potrete trarre materiale inerente al tema in questione, oltre a consultare schede di approfondimento degli argomenti trattati nel volume.

<http://ww2.unime.it/weblab/matematica.php>

Da questo link potrete accedere ad un elenco di risorse online, che consentono di disegnare una superficie tridimensionale nello spazio (x,y,z) , di risolvere un sistema lineare di m equazioni in n incognite, di visualizzare graficamente la somma di due vettori in un piano etc.

<http://www.atuttascuola.it/matematica/index.htm>

In questo sito troverete compiti, ipertesti, presentazioni multimediali e cenni alle funzioni di Cabri e Derive.

http://www.matematicamente.it/software_matematico/

Il portale offre contributi a diversi aspetti della matematica, fra cui l'analisi. Vi sono molti argomenti interessanti e di ottima qualità, soprattutto in merito alla descrizione di alcuni dei più noti software didattici (Derive, GeoGebra, Excel, Mathcad etc.).

<http://www.extrabyte.info>

Si tratta di un weblog dedicato ad analisi matematica, algebra lineare, analisi funzionale, fisica teorica, programmazione.

http://w3.uniroma1.it/dcnaps/memoli/documenti/Spss_6_per_Windows.pdf

Da questo link è possibile scaricare una dispensa che spiega le principali funzioni di SPSS, uno dei più conosciuti e utilizzati software di elaborazione statistica dei dati.

<http://mathworld.wolfram.com/>

Math World è un'enciclopedia interattiva in inglese, continuamente aggiornata, sulla matematica. Il livello di utenza è ampio: il sito, infatti, è destinato a studenti, insegnanti e ricercatori ed offre argomenti di ogni genere (algebra, geometria, probabilità, statistica ed altro), trattati in maniera semplice e chiara.

<http://www.chem.polimi.it/homes/dmanca/CDPDIC/>

Da questo sito è possibile scaricare un tutorial di MATLAB.

<http://www.sosmath.com/>

Sito in inglese, che permette di reperire materiale inerente alla matematica. Gli argomenti trattati vanno dall'algebra alle equazioni differenziali.

<http://forrest.psych.unc.edu/research/>

È un sito da cui potete trarre utili informazioni relative a ViSta, programma che consente di effettuare calcoli e grafici statistici.

<http://math.ec.unipi.it/algebra/matric/pagine/introd.htm>

Da questo sito potete accedere ad un corso di algebra lineare, con tanto di definizioni ed esercizi.

<http://www.batmath.it/>

È un sito per amanti della matematica e della fisica. Si pone l'obiettivo di stimolare la curiosità e la fantasia, indispensabili per riuscire nello studio di queste discipline. Il materiale proposto è raggruppato in sezioni, elencate nel menu di navigazione. Sono presenti nozioni inerenti ad alcuni software, come Cabri, Geogebra etc.

<http://www.math.it/>

È un sito di applicazioni didattiche e di divulgazione della matematica. Esso affronta argomenti quali la storia della matematica, lo studio di funzioni, la geometria euclidea; sono presenti, inoltre, autoverifiche strutturate con cui esercitarsi.

<http://lugaversa.altervista.org/galleria.php>

Da questo sito è possibile scaricare una guida a MATLAB.

<http://dinamico.unibg.it/ctd/matgen/>

Sul sito sono pubblicati elementi di analisi matematica. Esso è curato da professori dell'Università di Bergamo e contiene, inoltre, alcune esercitazioni con relative soluzioni.

<http://www.istat.it/it/strumenti/metodi-e-software>

Da questo link è possibile accedere a software di tipo generalizzato per la produzione statistica.

<http://www.vialattea.net/esperti/math.html>

Il sito contiene numerose domande, con relative risposte, riguardanti le principali branche della matematica.

<http://www.chihapauradellamatematica.org/>

Il sito è finalizzato a promuovere la conoscenza dei concetti matematici, attraverso lezioni integrali di analisi, calcolo combinatorio, probabilità e geometria.

<http://www.letterioguglielmo.it/>

Da questo sito è possibile scaricare dispense, software e javascript per l'apprendimento della matematica.

<http://www.mclink.it/personal/MC5834/analisi.htm>

Il sito presenta un corso di Analisi I. Sono riportate, inoltre, le biografie succinte di alcuni grandi matematici.

<http://www.ciaburro.it/indmat/doc.htm>

Dal sito è possibile scaricare un manuale di MATLAB.

<http://www.ripmat.it/>

Il sito contiene appunti di matematica, algebra e analisi, consultabili online.

<http://www.antonioeteresa.net/matematica/index.html>

Il sito contiene appunti di analisi e interessanti nozioni relative all'applicazione di Excel allo studio della matematica.

<http://www.ciaramella.net/modules.php?name=Mathscript>

Il sito contiene grafici di funzioni, integrali, derivate ed elementi di geometria analitica realizzati con Mathscript.

<http://www.torinoscienza.it/>

Si tratta di un'enciclopedia interattiva sulla scienza, da cui reperire molte nozioni di interesse. Sono presenti anche articoli dedicati alle applicazioni della matematica, della fisica e della chimica.

<http://areeweb.polito.it/didattica/polymath/>

Il sito intende offrire l'occasione di un incontro e di un confronto per una «nuova matematica» nella scuola. Vuole contribuire a diffondere un'immagine di questa materia meno fredda e antipatica di quella che è diffusa fra gli studenti e in generale nell'opinione pubblica.

<http://www.lorenzoroi.net/>

Il sito contiene lezioni e dispense su argomenti di matematica e di fisica.

http://users.libero.it/prof.lazzarini/geometria_sulla_sfera/geo.htm

In questo sito sono elencati interessanti articoli di geometria nello spazio euclideo e non.

<http://progettomatematica.dm.unibo.it/>

Il sito si pone l'obiettivo di presentare, nella maniera più discorsiva possibile, i principali concetti di matematica.

<http://math.ec.unipi.it/>

Si tratta di risorse didattiche online per la matematica applicata alle scienze economiche e sociali. Il sito contiene interessanti note ed esercitazioni concernenti software quali Maple e MATLAB.

<http://alpha01.dm.unito.it/personalpages/cerruti/mathblog.html>

Il sito contiene pagine dedicate ai numeri primi e a tutti gli argomenti correlati, curate da Umberto Cerruti del Dipartimento di Matematica dell'Università di Torino.

<http://www.ing.unitn.it/~sabatini/comp/compitiargo.html>

Il sito contiene prove d'esame su equazioni differenziali, integrali, limiti di funzioni di variabili reali, serie, studi di funzione ed altro.

<http://users.dimi.uniud.it/~gianluca.gorni/>

Il sito contiene esercizi di Analisi I e II, con relative soluzioni.

<http://www.descrittiva.it/>

Il sito è dedicato alla geometria e a tematiche connesse agli elementi geometrici. Numerose immagini illustrano problemi di parallelismo, ortogonalità, proiezioni, nonché intersezioni tra piani e rette.

<http://www.dia.uniroma3.it/autom/FdAcomm/>

Da questo sito è possibile trarre informazioni su alcuni software didattici e scaricare un breve manuale di MATLAB.

II

Scienze

Sommario Parte II

Introduzione | Metodologie didattiche per lo studio delle Scienze

Sezione I | Scienze chimiche

Sezione II | Scienze fisiche

Sezione III | Scienze biologiche

Sezione IV | Astronomia e Scienze della Terra

Sezione V | Educazione alla salute

Introduzione

Metodologie didattiche per lo studio delle Scienze

L'insegnamento scolastico delle **Scienze** è rimasto molto legato ad una visione di scienza compatta e progressiva di saperi lineari cumulativi, benché aggiornata.

Non va certamente abbandonato l'*aspetto informativo e trasmissivo della conoscenza scientifica*, ma l'insegnamento delle Scienze in genere deve confrontarsi con una nuova immagine della disciplina, meno uniforme e meno convergente, più critica e più dialettica, ma anche più storica, *attenta alle diversità interne e alle contestualizzazioni esterne*.

La moderna conoscenza scientifica del mondo si è costruita nel tempo, attraverso un metodo di indagine fondato sull'**osservazione dei fatti e sulla loro interpretazione**, *con spiegazioni e modelli sempre suscettibili di revisione e di riformulazione*.

La constatazione di questa evidenza fa concludere che l'osservazione dei fatti e lo spirito di ricerca si pongono quale presupposto per un efficace insegnamento delle scienze.

Per fare emergere queste attitudini è necessario un **coinvolgimento diretto degli alunni**, i quali vanno incoraggiati, senza un ordine temporale rigido e senza forzare alcuna fase, a *porre domande sui fenomeni e sulle cose, a progettare esperimenti ed esplorazioni seguendo ipotesi di lavoro, a costruire i loro modelli interpretativi*.

Sovente, quando a scuola si parla di **laboratori**, si pensa a qualcosa di separato dalla normale attività educativa scolastica, qualcosa di aggiuntivo alla scuola comunemente intesa: da una parte le lezioni e le spiegazioni in classe, dall'altra, qualche volta, i laboratori, appunto, opzionali e facoltativi, dove anche gli studenti possono «sperimentare».

La verità è che non è mai stata superata la difficoltà di far considerare, al mondo della scuola, come equivalenti sul piano didattico ed educativo le *lezioni frontali* e il *laboratorio*.

1 Il metodo tradizionale: lezione frontale

Necessitando, come e forse più delle altre discipline, di trasmissioni di nozioni, si pone la necessità, per le scienze, di affiancare alle metodologie di laboratorio un'attività più tradizionale, senza la quale non ci si può addentrare nell'approfondimento della materia.

La **lezione frontale** costituisce una modalità di presentazione, o di esposizione didattica, fondamentalmente supportata dalla comunicazione orale. Essa appartiene ai metodi denominati *espositivi*, che possiamo distinguere in tre tipi differenti:

- **metodo espositivo puro**: il professore spiega, gli studenti ascoltano. Si basa su una concezione sostanzialmente ricettiva dell'apprendimento, anche se presuppone l'ascolto attivo da parte dell'allievo;
- **metodo espositivo interrogativo**: durante l'esposizione, o alla fine di essa, il professore formula domande agli studenti. Lo scopo dell'interrogazione è il *feedback*, che consiste nel verificare se il messaggio è stato compreso correttamente e, in caso negativo, nel modificarlo e riformularlo;
- **metodo espositivo partecipativo**: durante la lezione gli studenti possono porre domande e intervenire secondo modalità negoziate. La loro partecipazione si completa con *esercizi applicativi* o altre *attività comuni*.

Spesso, in classe, i tre metodi sopra elencati vengono utilizzati senza una rigorosa distinzione.

L'esposizione verbale, se ben organizzata, torna **didatticamente utile** in determinate circostanze, in quanto:

- *in breve tempo si presentano numerosi contenuti, dati e informazioni;*
- *pone le basi e organizza il campo per lo studio individuale o di gruppo.*

La sua **efficacia** dipende dalle competenze personali del docente, ossia da quanto egli sia in grado di adottare strategie comunicative, quali:

- *costruire interventi affascinanti ai fini del coinvolgimento espositivo;*
- *personalizzare l'esposizione, rapportandola alle caratteristiche del gruppo di allievi, senza abbassare la qualità dell'insegnamento;*
- *coinvolgere gli studenti con strategie partecipative;*
- *impiegare correttamente i sussidi e le tecnologie didattiche, allo scopo di integrare la comunicazione verbale con altri linguaggi particolarmente praticati dagli allievi.*

La lezione frontale si espone, invece, a molte **perplexità** quando:

- *non considera né il ritmo né la durata della capacità di attenzione degli studenti;*
- *non tiene conto degli interessi, delle curiosità, delle motivazioni degli allievi;*
- *la sua efficacia è limitata ai primi processi dell'apprendimento, relativi alla percezione e all'acquisizione delle conoscenze, e non considera i successivi processi di assimilazione, di accomodamento, di consolidamento etc.;*
- *riduce al minimo l'interazione tra insegnante e studente, tendendo ad annullare anche lo scambio di opinioni e di informazioni tra gli allievi stessi;*
- *costringe la valutazione al controllo delle capacità mnemoniche, ed in particolare alla memoria verbale e riproduttiva;*
- *nega agli allievi la possibilità di contrastare l'informazione ricevuta con proprie riflessioni o con giudizi critici;*
- *si presenta per lo più come ripetizione di ciò che è esposto nei libri di testo.*

2 Il metodo operativo: laboratorio

In più punti le *Indicazioni nazionali* per il curricolo sottolineano l'importanza di una metodologia didattica di tipo attivo, fondata sull'**esperienza di laboratorio**.

La **ricerca sperimentale, individuale e di gruppo**, rafforza nei ragazzi la fiducia nelle proprie capacità di pensiero, la disponibilità a dare e ricevere aiuto, l'attitudine ad imparare dagli errori propri e altrui, l'apertura ad opinioni diverse e l'abilità ad argomentare le proprie.

È importante disporre di tempi e modalità di lavoro che consentano, in modo non superficiale o affrettato, la produzione di idee originali da parte dei ragazzi, anche a costo di fare delle scelte sui livelli di approfondimento e di limitarsi alla trattazione di temi rilevanti. La valorizzazione del pensiero spontaneo dei ragazzi consentirà di costruire nel tempo le prime formalizzazioni in modo convincente per ciascun alunno. La gradualità e non dogmaticità dell'insegnamento favorirà negli allievi la fiducia nelle loro possibilità di capire sempre quello che si studia, con i propri mezzi e al proprio livello.

Data l'importanza del laboratorio, a scuola è necessario predisporre **ambienti di formazione** che permettano agli studenti di comportarsi come lo scienziato, seguendo le fasi del metodo sperimentale. Un *laboratorio didattico* dovrebbe essere una simulazione di un *laboratorio scientifico* in cui si mostrino tutte le fasi dell'esperimento, riconducendole a esempi emblematici; in tal modo, viene messo in luce il rapporto fra la realtà naturale e il mondo del laboratorio, collegando la comprensione all'azione.

Nell'eventualità in cui la scuola non disponga di aule utilizzabili per il laboratorio, è necessario che l'insegnante sia in grado di realizzare un *laboratorio in classe*, organizzando un ambiente

idoneo alla conduzione di attività pratiche e all'apporto del contributo personale da parte di ogni membro del gruppo.

Si deve, purtroppo, constatare come il ricorso alle attività laboratoriali sia ancora abbastanza limitato e spesso ridotto a poche esperienze svolte in genere dalla cattedra.

Il più delle volte, inoltre, nella pratica didattica, anche gli esercizi e gli esperimenti proposti si riducono ad attività per verificare il risultato previsto dalle leggi teoriche, più che come mezzi per la ricerca empirica.

3 Linguaggio comune e linguaggio scientifico

Un ulteriore elemento da tenere in considerazione nell'ambito dell'insegnamento scientifico è il problema della **differenza di linguaggio tra la conoscenza comune**, posseduta dai ragazzi e acquisita spontaneamente dal loro vivere quotidiano, **e la conoscenza scientifica**, connessa, invece, all'istruzione scolastica.

La difficoltà è dovuta al fatto che le parole del linguaggio comune presentano molti significati relativi allo stesso termine, tra i quali va cercato quello adatto, in base al contesto di riferimento; al contrario, le parole del linguaggio scientifico hanno un significato ben preciso e spesso in contrasto con quelli del linguaggio comune. Si comprende, quindi, la difficoltà dell'insegnamento delle discipline scientifiche a scuola e, al contempo, l'importanza della presenza di docenti preparati dal punto di vista disciplinare e metodologico, che siano in grado di fronteggiare il «problema linguistico» e di *trasmettere agli allievi la conoscenza scientifica, affrontando gli ostacoli legati alla complessità di questa rispetto a quella comune*.

Con lo **sviluppo dei linguaggi** e delle capacità di comunicazione, i ragazzi dovrebbero saper descrivere la loro attività di ricerca in testi di vario tipo (racconti orali, testi scritti, immagini, disegni, schemi, mappe, tabelle, grafici etc.), sintetizzando il problema affrontato, l'esperimento progettato (nonché la sua realizzazione e i suoi risultati), le difficoltà incontrate, le scelte adottate, le risposte individuate.

È opportuno potenziare, nel percorso di studio, l'impostazione metodologica, mettendo in evidenza i modi di ragionare, le strutture di pensiero e le informazioni trasversali, evitando così la frammentarietà nozionistica dei differenti contenuti. Gli allievi potranno così riconoscere in quello che studiano un'unitarietà della conoscenza.

Il percorso dovrà comunque mantenere un **costante riferimento alla realtà**, imperniando le attività didattiche sulla scelta di casi emblematici, quali l'osservazione diretta di un organismo o di un micro-ambiente, di un movimento, di una candela che brucia, di una fusione, dell'ombra prodotta dal Sole, delle proprietà dell'acqua etc.

Valorizzando le competenze acquisite dagli allievi, nell'ambito di una progettazione verticale complessiva, gli insegnanti potranno costruire una sequenza di esperienze che nel loro insieme consentano di sviluppare gli argomenti basilari di ogni settore scientifico.

4 Traguardi per lo sviluppo delle competenze e obiettivi di apprendimento

Nell'estensione delle **Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione** (*decreto ministeriale 16 novembre 2012, n. 254*), in relazione alla disciplina delle **Scienze**, vengono indicati i traguardi per lo sviluppo delle competenze al termine del percorso della scuola primaria e di quello della scuola secondaria, nonché gli obiettivi di apprendimento al termine della terza primaria, della quinta primaria e della terza secondaria di primo grado. Li riportiamo, limitando l'attenzione alla scuola secondaria di primo grado.

Traguardi per lo sviluppo delle competenze al termine della scuola secondaria di primo grado

L'alunno esplora e sperimenta, in laboratorio e all'aperto, lo svolgersi dei più comuni fenomeni; ne immagina e ne verifica le cause; ricerca soluzioni ai problemi, utilizzando le conoscenze acquisite.

Sviluppa semplici schematizzazioni e modellizzazioni di fatti e fenomeni, ricorrendo, quando è il caso, a misure appropriate e a semplici formalizzazioni.

Riconosce nel proprio organismo strutture e funzionamenti a livelli macroscopici e microscopici, ed è consapevole, inoltre, delle sue potenzialità e dei suoi limiti.

Ha una visione della complessità del sistema dei viventi e della loro evoluzione nel tempo; riconosce nella loro diversità i bisogni fondamentali di animali e piante, oltre che i modi di soddisfarli negli specifici contesti ambientali.

È consapevole del ruolo della comunità umana sulla Terra, del carattere finito delle risorse, nonché dell'ineguaglianza dell'accesso a esse, e adotta modi di vita ecologicamente responsabili.

Collega lo sviluppo delle scienze allo sviluppo della storia dell'uomo.

Ha curiosità e interesse verso i principali problemi legati all'uso della scienza nel campo dello sviluppo scientifico e tecnologico.

Obiettivi di apprendimento al termine della classe terza della scuola secondaria di primo grado

• Fisica e Chimica

- Utilizzare i concetti fisici fondamentali, quali pressione, volume, velocità, peso, peso specifico, forza, temperatura, calore, carica elettrica etc., in varie situazioni di esperienza; in alcuni casi, raccogliere dati su variabili rilevanti di differenti fenomeni, trovarne relazioni quantitative ed esprimerle con rappresentazioni formali di tipo diverso. Realizzare esperienze, quali, ad esempio, piano inclinato, galleggiamento, vasi comunicanti, riscaldamento dell'acqua, fusione del ghiaccio, costruzione di un circuito pila-interruttore-lampadina.
- Costruire e utilizzare correttamente il concetto di energia come quantità che si conserva; individuare la sua dipendenza da altre variabili; riconoscere l'inevitabile produzione di calore nelle catene energetiche reali. Realizzare esperienze, quali, ad esempio, mulino ad acqua, dinamo, elica rotante sul termosifone, riscaldamento dell'acqua con il frullatore.
- Padroneggiare concetti di trasformazione chimica; sperimentare reazioni (non pericolose) anche con prodotti chimici di uso domestico e interpretarle sulla base di modelli semplici di struttura della materia; osservare e descrivere lo svolgersi delle reazioni e i prodotti ottenuti. Realizzare esperienze, quali, ad esempio, soluzioni in acqua, combustione di una candela, bicarbonato di sodio + aceto.

• Astronomia e Scienze della Terra

- Osservare, modellizzare e interpretare i più evidenti fenomeni celesti, attraverso l'osservazione del cielo notturno e diurno, utilizzando anche planetari o simulazioni al computer. Ricostruire i movimenti della Terra da cui dipendono il dì e la notte, nonché l'alternarsi delle stagioni. Costruire modelli tridimensionali anche in connessione con l'evoluzione storica dell'astronomia.
- Spiegare, anche per mezzo di simulazioni, i meccanismi delle eclissi di Sole e di Luna. Realizzare esperienze, quali, ad esempio, la costruzione di una meridiana, la registrazione della traiettoria del Sole e della sua altezza a mezzogiorno durante l'arco dell'anno.

- Riconoscere, con ricerche sul campo ed esperienze concrete, i principali tipi di rocce ed i processi geologici da cui hanno avuto origine.
- Conoscere la struttura della Terra e i suoi movimenti interni (tettonica a placche); individuare i rischi sismici, vulcanici e idrogeologici della propria regione per pianificare eventuali attività di prevenzione. Realizzare esperienze, quali, ad esempio, la raccolta e i saggi di rocce diverse.
- *Biologia*
 - Riconoscere le somiglianze e le differenze, dal punto di vista biologico, tra le diverse specie di viventi.
 - Comprendere il senso delle grandi classificazioni, riconoscere nei fossili indizi per ricostruire nel tempo le trasformazioni dell'ambiente fisico, nonché la successione e l'evoluzione delle specie. Realizzare esperienze quali, ad esempio, l'osservazione, in coltivazioni e allevamenti, della variabilità in individui della stessa specie.
 - Sviluppare progressivamente la capacità di spiegare il «funzionamento» macroscopico dei viventi con un modello cellulare (collegando, per esempio, la respirazione con la respirazione cellulare, l'alimentazione con il metabolismo cellulare, la crescita e lo sviluppo con la duplicazione delle cellule, la crescita delle piante con la fotosintesi). Realizzare esperienze, quali, ad esempio, la dissezione di una pianta, la modellizzazione di una cellula, l'osservazione di cellule vegetali al microscopio, la coltivazione di muffe e microorganismi.
 - Conoscere le basi biologiche della trasmissione dei caratteri ereditari acquisendo le prime elementari nozioni di genetica.
 - Acquisire corrette informazioni sullo sviluppo puberale e sulla sessualità; sviluppare la cura e il controllo della propria salute, attraverso una corretta alimentazione; evitare consapevolmente i danni prodotti dal fumo e dalle droghe.
 - Assumere comportamenti e scelte personali ecologicamente sostenibili. Rispettare e preservare la biodiversità nei sistemi ambientali. Realizzare esperienze, quali, ad esempio, la costruzione di nidi per uccelli selvatici, l'adozione di uno stagno o di un bosco.

I Scienze chimiche

Sommario Sezione I

1 Il mondo della chimica	10 Le soluzioni
2 La costituzione della materia	11 La termodinamica chimica
3 La struttura dell'atomo	12 La cinetica chimica
4 Il sistema periodico degli elementi	13 L'equilibrio chimico
5 Il legame chimico	14 Acidi e basi
6 Composti chimici e loro nomenclatura	15 Gli equilibri ionici in soluzione acquosa
7 Stati di aggregazione e passaggi di stato	16 I fondamenti della chimica organica
8 La materia e le sue trasformazioni	17 Scienza e tecnologia dei materiali
9 I calcoli stechiometrici	

1 Il campo di studio della chimica

Per quanto ad un primo approccio possa apparire qualcosa di molto distante dalla realtà quotidiana, la chimica è ciò che governa tutto quello che avviene intorno e perfino dentro di noi. L'uomo, come tutti i viventi, è un complesso sistema chimico: la nascita, la crescita, l'invecchiamento sono il frutto di trasformazioni chimiche che avvengono all'interno del nostro organismo. Anche le più semplici azioni che svolgiamo nella quotidianità comportano un continuo succedersi di trasformazioni chimiche e sono, di conseguenza, governate dalle leggi di questa scienza.

La chimica è quindi una **disciplina sperimentale** che studia la **materia e le sue proprietà**, ovvero la composizione, la struttura, le trasformazioni che questa può subire e le energie in gioco in tali trasformazioni. Viene definita «sperimentale» una qualsiasi disciplina che parte da dati raccolti in osservazioni ed esperimenti opportunamente progettati per dedurre delle leggi generali che regolano i fenomeni osservati. Nel far questo, la chimica applica rigorosamente il **metodo scientifico**, che prevede:

- l'individuazione di grandezze fisiche caratteristiche del fenomeno (definite «osservabili sperimentali»);
- l'esecuzione di misure di queste grandezze in una serie di esperimenti;
- la deduzione di una relazione tra le «osservabili sperimentali», che costituisce la «legge» che regola il fenomeno;
- l'elaborazione di una «teoria» o di un «modello», attraverso il quale si possa dare un'interpretazione univoca e razionale ai dati raccolti.

Qualsiasi teoria è da ritenersi valida fintanto che sia in grado di spiegare tutte le osservazioni sperimentali. Se emerge anche un solo dato non interpretabile attraverso il modello o la teoria elaborata, questa va necessariamente corretta, ampliata o addirittura riformulata. Abbiamo detto che l'oggetto di studio della chimica è la materia, le sue proprietà e l'energia coinvolta nelle sue trasformazioni. Pertanto, prima di addentrarci nello studio della chimica, è indispensabile familiarizzare con due importanti entità fisiche: la **materia** e l'**energia**.

2 La materia

La **materia** costituisce tutto ciò che sia dotato di una **massa** e occupi uno **spazio**. L'osservazione del mondo esterno ci consente facilmente di distinguere per la materia tre diversi stati di aggregazione: solido, liquido e gassoso.

Allo **stato solido** la materia è dotata di un volume e di una forma propria. Ciò è dovuto al fatto che le particelle di cui è costituita sono stabilmente legate tra loro ed occupano posizioni fisse nello spazio.

Allo **stato liquido** la materia mostra, invece, un volume proprio ma è priva di forma: un liquido si adatta sempre alla forma del recipiente che lo contiene, poiché le particelle che lo costituiscono, pur rimanendo aderenti tra loro, non possiedono una rigorosa disposizione spaziale; ciò le rende libere di muoversi le une rispetto alle altre in una qualsiasi direzione dello spazio.

Allo **stato gassoso**, infine, la materia non possiede né forma, né volume proprio: infatti le particelle che costituiscono il gas non sono in contatto tra loro e tendono spontaneamente ad occupare tutto lo spazio a disposizione. Il rapido diffondere di un odore in un ambiente chiuso, così come la possibilità di eliminarlo semplicemente aprendo una finestra, sono chiare dimostrazioni di questo fenomeno.

Stato solido, liquido e gassoso sono quindi stati di rarefazione crescente della materia, poiché la distanza media tra le particelle aumenta passando dall'uno all'altro; ed anche di «disordine» crescente, poiché diminuisce gradualmente il grado di organizzazione interna delle particelle. A questa regola generale fa vistosamente eccezione l'acqua, che, passando dallo stato liquido allo stato solido, aumenta di volume: il ghiaccio, infatti, galleggia sull'acqua liquida. Andando avanti nello studio della chimica saremo in grado di capire il perché di questo fenomeno.

Quelle che avvengono nella materia possono essere distinte in **trasformazioni fisiche**, quando comportano unicamente il passaggio di una sostanza da uno stato di aggregazione ad un altro, e **trasformazioni chimiche**, quando determinano la trasformazione di alcune sostanze, i *reagenti*, in altre, i *prodotti di reazione*. Le reazioni chimiche possono eventualmente essere accompagnate da cambiamenti dello stato fisico nel passaggio dai reagenti ai prodotti, come nel caso dello sviluppo di gas o della «precipitazione» di prodotti solidi da reagenti in fase liquida.

Sperimentalmente è possibile verificare che la quantità di materia di un sistema che subisce una qualsiasi trasformazione si mantiene costante. Questa osservazione, ripetutamente confermata in vari esperimenti, ha portato **Lavoisier** ad enunciare, nel 1789, il **principio di conservazione della massa**: *«In una reazione chimica la somma delle masse delle sostanze reagenti risulta sempre uguale alla somma delle masse delle sostanze prodotte»*.

Questo principio, di estrema utilità in chimica e in fisica, è stato successivamente ridiscusso da **Einstein** nella sua *teoria della relatività*.

3 L'energia

L'**energia** rappresenta l'attitudine di un corpo o, più in generale, di un sistema materiale, a compiere lavoro. La materia stessa, per la sua struttura e costituzione, possiede una certa quantità di «**energia interna**» che, in opportune condizioni, può essere trasferita all'esterno; ad esempio, attraverso trasformazioni chimiche, convertita in energia di diversa natura ed utilizzata.

Questo è ciò che accade comunemente nelle combustioni: una combustione non è altro che la reazione di una sostanza, il combustibile, con l'ossigeno. Questa reazione libera una parte dell'energia interna dei reagenti trasferendola all'esterno come **calore**, una forma di energia diversa e utilizzabile dall'uomo.

L'energia posseduta dai sistemi materiali può essere di varia natura: possiamo distinguere energia meccanica, elettrica, termica, luminosa, magnetica, molecolare, atomica, nucleare etc. Tutte queste forme di energia sono convertibili l'una nell'altra attraverso opportune trasformazioni, e in un sistema isolato, cioè un sistema che non scambi con l'ambiente esterno energia o materia, la quantità totale di energia resta costante.

Questa osservazione, che non è mai stata contraddetta da alcuna osservazione sperimentale, rappresenta uno dei capisaldi della fisica e della chimica e costituisce il **principio di conservazione dell'energia**. Esistono varie formulazioni di questo principio. Una delle più note è la seguente: *«L'energia non può essere creata né distrutta ma unicamente trasformata da una forma all'altra; quindi l'energia totale di un sistema isolato resta costante»*.

In realtà, essendo l'Universo un sistema isolato, un modo alternativo ed estremamente semplice di enunciare il principio di conservazione dell'energia è:

«L'energia dell'Universo è costante».

Un'importante conseguenza del principio di conservazione è che qualsiasi trasformazione avvenga in un sistema isolato può soltanto cambiare la ripartizione dell'energia tra le diverse forme, lasciandone inalterato il contenuto di energia totale.

4 L'equivalenza tra materia ed energia

Materia ed energia sono state considerate per lungo tempo due entità fisiche nettamente distinte, ciascuna delle quali sottoposta ad un proprio principio di conservazione. **Einstein**, con la **teoria della relatività**, ha decretato un'unificazione di queste due entità dapprima disgiunte.

La conclusione fondamentale a cui Einstein giunse è che la massa di un corpo non è indipendente dall'energia che questo possiede. In particolare, se un corpo dotato di una **massa a riposo** [1] pari a m_0 si muove alla velocità v , la sua massa deve essere ridefinita come:

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Quindi imprimere ad un corpo un'energia cinetica pari ad E equivale ad incrementarne la massa di una quantità Δm pari a:

$$\Delta m = E / c^2$$

dove c è la velocità della luce nel vuoto.

Quanto detto per l'energia cinetica può essere esteso ad energia di qualsiasi natura: la relazione precedente stabilisce l'esistenza di un'equivalenza tra massa ed energia, che divengono due diverse manifestazioni di un unico ente fisico. La fisica classica prendeva in considerazione due diversi principi di conservazione per massa ed energia, la teoria della relatività fonde questi due aspetti. Nei sistemi macroscopici l'incremento della massa con l'energia non è rivelabile attraverso misure sperimentali e, dunque, per questi sistemi è sufficiente considerare come massa la massa a riposo m_0 . Nei sistemi microscopici come atomi, molecole, nuclei, elettroni, date le piccole dimensioni, è possibile invece impartire alle particelle energie molto alte provocando variazioni percentuali di massa significative e misurabili.

Il principio di equivalenza tra massa ed energia ha trovato diverse conferme sperimentali, tra cui la materializzazione di radiazioni ad alto contenuto energetico (raggi γ) e la trasformazione di materia in energia nelle reazioni di fissione o fusione nucleare.

5 Grandezze fisiche e sistemi di unità di misura

Abbiamo detto all'inizio di questo capitolo come per descrivere e caratterizzare la materia secondo il metodo scientifico sia indispensabile riferirsi a grandezze fisiche osservabili e misurabili, scelte opportunamente a seconda del fenomeno in studio. Misurare una grandezza fisica significa confrontarla con un campione scelto come riferimento e il risultato della misura, espresso nell'unità campione, rappresenta il numero di volte che il campione è contenuto nella grandezza misurata.

La definizione operativa di una **grandezza fisica** consiste, quindi, nello stabilire un procedimento di misura per la grandezza e nell'assegnarle un'unità di misura. Le grandezze fisiche sono numerosissime, ma non indipendenti una dall'altra. Ciò consente di sceglierne alcune, dette **fondamentali**, alle quali viene assegnata una propria unità di misura. Tutte le altre grandezze

[1] Per *massa a riposo*, o anche *massa at rest*, si intende la massa del corpo quando questo è fermo rispetto all'osservatore.

vengono considerate **derivate** dalle fondamentali, in base alla legge fisica che le correla a queste, e le corrispondenti unità di misura sono derivate dalle unità fondamentali.

Definire un **sistema di unità di misura** significa effettuare la scelta delle grandezze fondamentali e delle corrispondenti unità di misura. Va sottolineato che questa scelta è del tutto arbitraria e che una grandezza ritenuta fondamentale in un sistema di unità di misura può essere considerata derivata in un sistema diverso.

Grandezze che sicuramente incontreremo nello studio dei sistemi materiali sono:

- la **massa**, che misura la quantità di materia di cui è costituito un sistema;
- il **peso**, che rappresenta la forza con cui un corpo di massa m viene attratto dal campo gravitazionale terrestre ed è pari al prodotto della massa per l'accelerazione di gravità g :

$$\vec{P} = m \cdot \vec{g}$$

Il peso e la massa di un corpo sono, dunque, quantità proporzionali. La bilancia, strumento utilizzato generalmente per la misura della massa, realizza un confronto tra il peso del corpo C e il peso di una massa di riferimento. La condizione di equilibrio tra i due piatti della bilancia si realizza quando su di essi viene esercitata la stessa forza di attrazione gravitazionale, ossia quando:

$$m_c \cdot \vec{g} = m_{rif} \cdot \vec{g}$$

Ed essendo l'accelerazione di gravità uguale sui due piatti della bilancia, ciò corrisponde a:

$$m_c = m_{rif}$$

Nel Sistema Internazionale delle unità di misura (SI) la **massa** viene espressa in chilogrammi (kg), mentre l'unità di misura del **peso**, o delle forze in generale, è il newton (N).

Il **volume** è lo spazio tridimensionale occupato dalla materia; come abbiamo visto, questa può possedere o meno un volume proprio a seconda dello stato fisico in cui si trova: nel Sistema Internazionale il volume viene espresso in metri cubi (m^3).

La **densità** è una grandezza derivata dalle precedenti e rappresenta il rapporto tra la massa e il volume di un sistema; volumi uguali di sostanze diverse possono possedere masse molto diverse: in tal caso si dirà che queste sostanze hanno densità molto diverse; nel Sistema Internazionale la densità si esprime in kg/m^3 .

La **pressione**, una grandezza derivata che rappresenta la forza esercitata sull'unità di superficie, nel Sistema Internazionale si misura in pascal (newton/ m^2).

6 Il Sistema Internazionale delle unità di misura

La definizione di un sistema di unità di misura consiste nell'individuazione di alcune grandezze fisiche fondamentali dalle quali possono essere espresse tutte le altre (grandezze derivate).

Nel tempo sono stati definiti diversi sistemi di unità di misura adatti alle esigenze dei diversi settori di ricerca: ciò ha comportato l'introduzione di un gran numero di fattori di conversione per confrontare dati espressi in unità di misura diverse. Tuttavia, la necessità di un'ampia fruizione delle conoscenze scientifiche ha fatto sentire l'esigenza di un sistema unificato a livello internazionale. Per questo motivo nel 1960, nell'ambito della Conferenza Generale sui Pesi e le Misure, è stato definito un **Sistema Internazionale delle Unità di Misura**, attualmente in uso in Europa, negli Stati Uniti e sempre più diffusamente anche nei Paesi Anglosassoni, basato su sette unità fondamentali. La scelta dei campioni di riferimento per le unità fondamentali è del tutto arbitraria anche se risulta indispensabile che il campione goda dei requisiti di **accessibilità**, per

poter essere confrontato con le grandezze da misurare, e di **invariabilità**, poiché misure ripetute di una certa grandezza devono fornire sempre lo stesso risultato. Per questo motivo molti campioni sono stati definiti riferendosi a fenomeni naturali altamente riproducibili e misurabili con grande **precisione** [2] e **accuratezza** [3].

Il primo **campione di lunghezza** fu una barra di platino-iridio detta *metro campione*, conservata presso l'Ufficio Internazionale dei Pesi e delle Misure. Essendo preferibile riferirsi ad un campione di maggiore invariabilità e riproducibilità, attualmente nel Sistema Internazionale il **metro** è stato ridefinito come la distanza percorsa dalla luce nel vuoto in un intervallo di tempo di $1/299.792.458$ di secondo.

Il **campione di massa** invece continua ad essere un prototipo, ossia un cilindro di platino-iridio, anche questo conservato presso l'Ufficio Internazionale dei Pesi e delle Misure, a cui si è attribuita convenzionalmente massa pari ad un **chilogrammo**. Il **campione di tempo** è il **secondo**, definito come $9.192.631.770$ volte il periodo di una particolare transizione del cesio-133 (^{133}Cs). La tabella che segue elenca le grandezze fondamentali nel Sistema Internazionale, le relative unità di misura e i prefissi utilizzati per multipli e sottomultipli di queste unità.

UNITÀ DI MISURA DEL S.I.								
Grandezza	Unità	Simbolo	Multipli			Sottomultipli		
lunghezza	metro	m	tera	T	10^{12}	deci	d	10^{-1}
massa	chilogrammo	kg	giga	G	10^9	centi	c	10^{-2}
tempo	secondo	s	mega	M	10^6	milli	m	10^{-3}
corrente elettrica	ampere	A	kilo	k	10^3	micro	μ	10^{-6}
temperatura	kelvin	K	etto	h	10^2	nano	n	10^{-9}
luminosità	candela	cd	deca	da	10	pico	p	10^{-12}
quantità di sostanza	mole	mol				femto	f	10^{-15}
						atto	a	10^{-18}

Tab. 1

Dalle sette unità fondamentali si ottengono le unità per le grandezze fisiche derivate semplicemente tenendo conto delle leggi che le correlano alle grandezze fondamentali. Ad esempio, la velocità esprime lo spazio percorso nell'unità di tempo e nel Sistema Internazionale viene espressa in m s^{-1} ; l'accelerazione in m s^{-2} , la forza in kg m s^{-2} .

ALCUNE UNITÀ DERIVATE DEL S.I.			
Grandezza	Unità	Simbolo	Definizione
forza	newton	N	kg m s^{-2}
pressione	pascal	Pa	$\text{kg m}^{-1} \text{s}^{-2} (\text{N m}^{-2})$
energia	joule	J	$\text{kg m}^2 \text{s}^{-2}$
potenza	watt	W	$\text{kg m}^2 \text{s}^{-3} (\text{J s}^{-1})$
carica elettrica	coulomb	C	A s
differenza di potenziale	volt	V	$\text{kg m}^2 \text{s}^{-3} \text{A}^{-1}$
resistenza elettrica	ohm	Ω	$\text{kg m}^2 \text{s}^{-3} \text{A}^{-2}$
frequenza	hertz	Hz	s^{-1}

Tab. 2

[2] Una misura si dice *precisa* quando, eseguita un certo numero di volte, fornisce risultati molto prossimi tra loro.
 [3] Una misura si dice *accurata* quando è affetta da un errore piccolo e quindi fornisce un risultato molto vicino al valore vero.

1 La natura elettrica della materia

Le prime osservazioni sperimentali riguardanti la natura elettrica della materia risalgono al VI secolo a.C., quando il filosofo greco Talete osservò che, strofinando con un panno di lana una bacchetta di ambra (*élektron* in greco, da cui poi il termine *elettrone*), questa acquistava la capacità di attrarre a sé piccoli oggetti. Tale fenomeno, oggi denominato **elettrizzazione per strofinio**, permette di evidenziare come la materia, pur apparendo elettricamente neutra, sia intimamente costituita da cariche elettriche. Attraverso una serie di esperimenti è facile verificare che due oggetti della stessa sostanza, dopo essere stati elettrizzati, si respingono, mentre oggetti di sostanze diverse a seguito dell'elettrizzazione possono manifestare forze di tipo attrattivo o repulsivo.

Ciò ha portato a concludere che:

- in natura esistono due tipi di elettrizzazione o, per meglio dire, di **carica elettrica**: cariche **positive** (o vetrose, come quelle acquistate dal vetro per strofinio) e cariche **negative** (o resinose, come quelle acquistate dall'ambra);
- cariche positive e negative esercitano reciprocamente forze di attrazione o repulsione: cariche di segno opposto tendono ad attrarsi, cariche dello stesso segno tendono a respingersi;
- le forze di attrazione e repulsione elettrica sono azioni esercitate a distanza, senza necessità di contatto tra i corpi tra cui si instaurano.

Tutta la materia è, quindi, costituita da cariche elettriche. I corpi elettricamente neutri possiedono cariche positive e negative in ugual numero, mentre nei corpi elettricamente carichi, come quelli elettrizzati per strofinio, è presente un eccesso di un tipo di carica rispetto all'altro: un corpo carico positivamente possiede cariche positive in eccesso rispetto alle negative, viceversa un corpo carico negativamente possiede cariche negative in eccesso rispetto alle positive. L'elettrizzazione per strofinio, infatti, è il risultato del trasferimento di cariche elettriche da una sostanza ad un'altra.

2 L'atomo

L'idea che la materia fosse costituita da piccole particelle indivisibili dette **atomi** (dal greco *atomos*, «indivisibile») è molto antica e viene fatta risalire a Leucippo (V sec. a.C.) e al suo discepolo Democrito. Oggi sappiamo che l'**atomo** è la più piccola parte di ogni **elemento** che ne conservi le proprietà chimiche.

Tuttavia, contrariamente a quanto si credeva in passato, l'atomo non è affatto indivisibile ma è costituito da particelle subatomiche di tre tipi diversi:

- **protoni**, dotati di **carica elementare** [1] positiva;
- **elettroni**, dotati di carica elementare negativa;
- **neutroni**, privi di carica elettrica.

^[1] Una *carica elementare* è la più piccola frazione di carica elettrica che si possa misurare. Tutte le cariche elettriche presenti in natura sono multiple della carica elementare.

Protoni ed elettroni hanno cariche elettriche uguali in modulo e di segno opposto. La massa del protone, invece, è enormemente più grande della massa dell'elettrone (circa 1.836 volte). I neutroni, infine, sono particelle prive di carica elettrica e dotate di una massa molto simile a quella del protone. Queste particelle sono organizzate in una struttura praticamente «vuota», in cui protoni e neutroni sono concentrati in un piccolo nucleo centrale che contiene quasi tutta la massa dell'atomo, mentre gli elettroni, particelle di massa trascurabile rispetto alle precedenti, si muovono attorno al nucleo, generando una nube di carica elettrica negativa il cui raggio è circa 10.000 volte più grande del raggio nucleare. L'atomo è un sistema elettricamente neutro, in quanto il numero dei protoni è uguale al numero degli elettroni e, in tal modo, la carica totale è nulla.

3 Gli elementi e i loro isotopi

Protoni, neutroni ed elettroni sono le particelle elementari di cui è costituita tutta la materia, uguali in ogni elemento chimico. Ciò che distingue un elemento dall'altro è il **numero atomico Z**, cioè il numero dei protoni contenuti nel nucleo (coincidente con il numero degli elettroni): l'idrogeno ha numero atomico $Z = 1$ ed ha un nucleo contenente un solo protone; l'ossigeno ha numero atomico $Z = 8$ ed il suo nucleo contiene 8 protoni. Ogni elemento chimico viene rappresentato attraverso un simbolo costituito da una lettera maiuscola o da due lettere (la prima maiuscola e la seconda minuscola). La somma del numero dei protoni e del numero dei neutroni contenuti nel nucleo atomico è detto, invece, **numero di massa A**.

Due atomi aventi lo stesso numero atomico e diverso numero di massa vengono detti **isotopi**. Quasi tutti gli elementi possiedono due o più isotopi, presenti in natura con abbondanze diverse; generalmente gli isotopi dello stesso elemento vengono rappresentati tutti con lo stesso simbolo e distinti l'uno dall'altro indicandone il numero di massa in alto a sinistra: ^{12}C , ^{13}C e ^{14}C sono i tre isotopi naturali del carbonio e possiedono nuclei costituiti da sei protoni e sei, sette e otto neutroni rispettivamente. Solo l'idrogeno possiede simboli e nomi distinti per i suoi tre isotopi: l'idrogeno, ^1H , ha il nucleo costituito da un solo protone; il deuterio, ^2D , ha il nucleo costituito da un protone ed un neutrone; il trizio, ^3T , ha il nucleo costituito da un protone e due neutroni. Poiché le proprietà chimiche e fisiche degli atomi sono determinate essenzialmente dal numero dei protoni (e degli elettroni), gli isotopi hanno proprietà identiche e differiscono tra loro unicamente per la massa ed in alcuni casi per la stabilità del nucleo.

4 Gli ioni

Nella definizione del numero atomico abbiamo preso come riferimento il numero dei protoni del nucleo poiché il numero degli elettroni, numericamente coincidente con il numero dei protoni nell'atomo neutro, può subire delle variazioni: in opportune condizioni un atomo può perdere o acquistare elettroni. Si verifica, così, uno sbilanciamento tra la carica nucleare e la carica della nube elettronica: di conseguenza, la carica totale risulta diversa da zero. La specie generata in questo modo è detta **ione** e viene indicata semplicemente con il simbolo dell'elemento, accompagnato dal valore della carica totale in alto a destra. Ioni positivi o **cationi** si formano quando l'atomo perde uno o più elettroni e la loro carica coincide con il numero di elettroni persi (es.: Na^+ , Mg^{2+}); viceversa, ioni negativi o **anioni** si generano quando un atomo acquista elettroni (ad esempio: Cl^- , S^{2-}).

5 Le molecole

Quando due o più atomi si legano stabilmente tra loro si genera una nuova entità chimica, chiamata **molecola**. Molecole costituite da atomi uguali vengono definite **elementi**, come gli atomi isolati. Gli elementi in natura possono esistere allo stato monoatomico, quando l'atomo

isolato è sufficientemente stabile, o come molecole biatomiche o poliatomiche, più stabili degli atomi isolati. Molecole costituite da atomi di elementi diversi sono dette **composti**. In ogni caso, la molecola rappresenta la più piccola parte di una sostanza, sia questa un elemento o un composto, che ne mantiene le proprietà chimiche e fisiche. Le molecole vengono rappresentate attraverso diversi tipi di formule.

La **formula bruta** fornisce la composizione qualitativa e quantitativa della sostanza, indicando il tipo di atomi contenuti nella molecola e il loro numero, riportato come pedice a destra del simbolo corrispondente.

Ad esempio:

- l'acqua ha formula bruta H_2O , quindi contiene due atomi di idrogeno e un atomo di ossigeno;
- il metano ha formula bruta CH_4 e contiene un atomo di carbonio e quattro di idrogeno.

La formula bruta coincide con l'effettiva formula molecolare per i composti molecolari, cioè specie chimiche le cui molecole possono esistere come molecole isolate. Per i composti ionici, che possono esistere solo in strutture tridimensionali organizzate, dette *reticoli cristallini*, che raggruppano un gran numero di molecole, la formula bruta rappresenta invece una **formula minima**, che stabilisce in quale rapporto gli elementi sono presenti nel reticolo.

La **formula di struttura**, oltre a dare informazioni sulla composizione qualitativa e quantitativa della molecola, mostra in che modo gli atomi costituenti siano legati tra loro.

6 I pesi atomici e molecolari

Le masse degli elementi sono sempre state determinate per via sperimentale esprimendole come masse relative alla massa di un elemento scelto come riferimento. L'attuale scala dei pesi atomici riferisce pesi atomici all'isotopo 12 del carbonio, a cui si è attribuita convenzionalmente massa uguale a 12.

Viene definita **unità di massa atomica** (u.m.a.) la massa corrispondente ad 1/12 della massa dell'isotopo 12 del carbonio a cui si attribuisce per convenzione massa pari a 12. L'u.m.a. corrisponde a $1,66053886 \cdot 10^{-27}$ kg. Ciò che viene chiamato comunemente **peso atomico** di un elemento è in realtà una massa relativa ed esprime il rapporto esistente tra la massa assoluta dell'elemento, espressa in kg, e la massa dell'unità di massa atomica ($1,66053886 \cdot 10^{-27}$ kg); è quindi un numero puro, privo di dimensioni fisiche:

$$M_r = \frac{M_{\text{assoluta}}}{\text{u.m.a.}}$$

Poiché ogni elemento esiste in natura come una miscela di isotopi diversi, ciascuno dei quali dotato di una massa atomica relativa, come peso atomico dell'elemento viene utilizzata una media pesata attraverso le rispettive abbondanze naturali delle masse atomiche dei diversi isotopi. I **pesi molecolari** si ottengono direttamente dalla formula molecolare della sostanza sommando i pesi atomici degli elementi presenti, ciascuno moltiplicato per il numero di atomi contenuti nella molecola.

Per i composti ionici, la *formula bruta* rappresenta una *formula minima*, poiché la molecola isolata non esiste; in questo caso non si può parlare di peso molecolare quanto di **peso formula**.

7 Concetto di mole e numero di Avogadro

Poiché i pesi atomici sono «masse relative», cioè esprimono il rapporto esistente tra le masse degli elementi e quella dell'unità di massa atomica, **campioni di elementi diversi, contenenti**

un uguale numero di atomi, hanno masse (esprese in kg) che stanno tra loro nello stesso rapporto dei pesi atomici.

ESEMPIO

L'idrogeno ha $M_r = 1$ e l'ossigeno ha $M_r = 16$; 10 atomi di idrogeno pesano:

$$MassaH(kg) = 10 \cdot M_{assoluta} = 10 \cdot M_r \cdot u.m.a = 10 \cdot 1 \cdot 1,66053886 \cdot 10^{-27} \text{ kg}$$

Analogamente 10 atomi di ossigeno pesano:

$$MassaO(kg) = 10 \cdot M_{assoluta} = 10 \cdot M_r \cdot u.m.a = 10 \cdot 16 \cdot 1,66053886 \cdot 10^{-27} \text{ kg}$$

Il rapporto tra le masse espresse in chilogrammi è 1:16, come il rapporto della masse relative:

$$\frac{MassaH(kg)}{MassaO(kg)} = \frac{10 \cdot 1 \cdot 1,66053886 \cdot 10^{-27} \text{ kg}}{10 \cdot 16 \cdot 1,66053886 \cdot 10^{-27} \text{ kg}} = \frac{1}{16}$$

Viceversa, campioni aventi masse che stanno tra loro **nel medesimo rapporto dei pesi atomici** contengono **lo stesso numero di atomi**. Della stessa proprietà godono ovviamente i pesi molecolari, essendo semplicemente la somma di pesi atomici. Questa proprietà dei pesi atomici e molecolari ha permesso di introdurre un'unità di misura estremamente importante in chimica: la **mole**.

Viene definita **mole** di una qualsiasi sostanza una quantità in grammi di quella sostanza numericamente coincidente con il suo peso atomico o molecolare. In base alle considerazioni fatte precedentemente, una mole di qualsiasi sostanza contiene sempre lo stesso numero di particelle (atomi o molecole), detto **numero di Avogadro N_A** . Quest'ultimo è stato determinato con vari metodi ed è attualmente stimato pari a $6,022 \cdot 10^{23}$.

La mole è l'unità di misura della quantità di sostanza accettata dal Sistema Internazionale ed è di estrema importanza in chimica, poiché consente di conoscere con esattezza il numero di particelle contenute in un campione di massa conosciuta. Infatti, noto il peso di una mole, numericamente coincidente con il peso atomico o molecolare, per conoscere il numero di moli contenute in un campione di massa m sarà sufficiente dividere m per il peso di una mole:

$$n^\circ \text{ di moli} = \text{massa in grammi} / \text{massa in grammi di una mole}$$

In altri termini:

$$n^\circ \text{ di moli} = \text{massa in grammi} / P.A. \text{ o } P.M.$$

L'utilità della mole è, quindi, quella di poter mettere a reagire tra loro quantità di reagenti pesabili con una comune bilancia, cosa che non potremmo fare con la singola molecola, sapendo esattamente il numero di atomi o molecole che questi contengono e potendo fare previsioni sulla quantità delle specie ottenute dalla reazione.

1 Il modello di Rutherford

Tra le prime ipotesi sulla struttura dell'atomo va ricordata quella di **J.J. Thompson** (1898), che propose per l'atomo una struttura sferica in cui le cariche positive si trovassero uniformemente distribuite e all'interno della quale le cariche elettriche negative fossero dislocate in modo da realizzare la maggiore stabilità elettrostatica.

Il modello di Thompson, inizialmente apprezzato, venne abbandonato nel 1911 quando **E.R. Rutherford** propose il modello successivamente denominato «planetario» per l'evidente analogia con il Sistema Solare. In base a questo modello l'atomo è costituito da un nucleo centrale molto piccolo in cui si trovano concentrate le cariche positive, mentre attorno al nucleo gli elettroni si muovono in un volume quasi sferico, il cui raggio è 10.000 volte più grande rispetto al raggio nucleare. Questa ipotesi, benché supportata da osservazioni sperimentali (*scattering* delle particelle α), era in netto contrasto con la teoria elettromagnetica classica. Infatti, se gli elettroni fossero stati in quiete all'interno dell'atomo, sarebbero stati catturati dal nucleo; ipotizzandone un moto circolare attorno al nucleo, in base alla teoria elettromagnetica classica gli elettroni avrebbero dovuto emettere energia come radiazione elettromagnetica fino a ricadere sul nucleo stesso.

2 Il modello di Bohr

Si deve a **Niels Bohr** (1913) l'applicazione all'atomo di un importante concetto introdotto da Max Planck, quello di **quantizzazione** [1] dell'energia. Bohr accolse l'idea di Rutherford di un atomo avente struttura «planetaria», ipotizzando che nell'atomo esistessero, tra gli infiniti stati possibili, degli stati di energia costante, detti **stati stazionari**, nei quali l'elettrone potesse muoversi attorno al nucleo senza irradiare energia.

Questi stati corrispondono a ben definiti valori di energia, pari a multipli interi di un valore minimo, detto «**quanto di energia**». Bohr ricavò i valori di energia degli stati stazionari e i raggi delle corrispondenti orbite, imponendo la quantizzazione del momento angolare dell'elettrone, immaginato in moto circolare attorno al nucleo. Nell'ipotesi di Bohr, il momento angolare dell'elettrone mvr (dove m è la massa dell'elettrone, v la velocità e r il raggio dell'orbita) può assumere solo valori multipli interi di $h/2\pi$:

$$mvr = n (h/2\pi)$$

Il numero n che compare in questa espressione è detto **numero quantico principale** e può assumere solo valori interi positivi ($n = 1, 2, 3, \dots, \infty$). Dalla quantizzazione del momento angolare dell'elettrone discendono, come diretta conseguenza, la quantizzazione dell'energia e dei raggi delle orbite: al crescere del numero quantico principale n , crescono i raggi delle orbite percorse dall'elettrone e il valore delle energie ad esse associate: E_n . Gli elettroni negli atomi, quindi, possono occupare unicamente stati stazionari di energia univocamente definita dal numero quantico principale n ; ogni valore di n individua un **livello energetico**.

[1] Una grandezza fisica si dice *quantizzata* quando può assumere solo valori discreti, cioè valori discontinui, multipli interi di un valore minimo, detto «quanto» della grandezza stessa.

Nel modello di Bohr l'atomo risulta suddiviso in una serie di gusci concentrici di raggio crescente, corrispondenti ai livelli di energia crescente. Quando in un atomo un elettrone passa da un livello energetico ad un altro si dice che è avvenuta una **transizione elettronica**. Questo tipo di fenomeno è sempre accompagnato da uno scambio di energia con l'esterno che, il più delle volte, si concretizza nell'assorbimento o nell'emissione di una radiazione elettromagnetica ($\Delta E = E_2 - E_1 = h\nu_1$, dove h = costante di Planck). Per passare da uno stato di energia E_1 ad uno stato di energia superiore E_2 l'elettrone dovrà assorbire un fotone di energia, $E_2 - E_1 = h\nu_1$, ossia una radiazione elettromagnetica di frequenza ν_1 ; nel passaggio inverso, dallo stato di energia superiore E_2 allo stato di energia inferiore E_1 , l'elettrone dovrà emettere un fotone di energia, $E_2 - E_1 = h\nu_1$, cioè una radiazione di frequenza ν_1 . I fenomeni di emissione o di assorbimento di energia da parte degli elettroni negli atomi sono evidenziabili attraverso la registrazione di spettri rispettivamente di emissione o di assorbimento. Lo spettro di un elemento è l'insieme di tutte le radiazioni elettromagnetiche che l'elemento può emettere o assorbire in seguito alle transizioni elettroniche che possono verificarsi al suo interno; dall'insieme delle frequenze dello spettro si può risalire ai valori delle energie delle corrispondenti orbite stazionarie.

3 Le modifiche al modello di Bohr

Proprio lo studio degli spettri atomici ha richiesto l'introduzione di modifiche al modello di Bohr, necessarie ad interpretare la presenza negli spettri di alcuni atomi di frequenze non prevedibili in base a questo modello. In particolare, si osservava che alcune frequenze, le quali in base alla teoria di Bohr si sarebbero dovute presentare come righe singole (cioè corrispondenti ad un'unica frequenza emessa e quindi ad un'unica transizione elettronica), in realtà si presentavano come multipletti di righe molto vicine, dovuti all'emissione di radiazioni di frequenze diverse e vicinissime tra loro.

Una prima modifica al modello di Bohr si deve a Sommerfield, che introdusse un **numero quantico secondario** l . In pratica, se nel modello di Bohr si era assunto che l'elettrone si muovesse su orbite circolari, i cui raggi venivano quantizzati attraverso il solo numero quantico principale n , l'introduzione di un secondo numero quantico l consentiva di esprimere «quantitativamente» l'esistenza di una ellitticità nelle orbite percorse dagli elettroni. Attraverso i due numeri quantici n ed l era possibile quantizzare i due semiassi dell'orbita ellittica. I valori di n ed l sono intercorrelati tra loro:

- n può assumere tutti i valori interi positivi ($n = 1, 2, 3, 4, \dots, \infty$);
- l , invece, può assumere tutti i valori interi positivi dallo zero al valore $n-1$ ($l = 0, 1, 2, 3, \dots, n-1$).

Il numero quantico principale n individua nell'atomo una serie di livelli di energia crescente al crescere di n . Il numero quantico l individua, per ciascun valore di n , un certo numero di sottolivelli energetici: infatti l'energia dell'orbita dipende principalmente dal numero quantico n , ma subisce piccole variazioni al mutare di l . Pertanto, l'energia della singola orbita dovrà essere indicata più propriamente come $E_{n,l}$.

La transizione dallo stato $E_{1,0}$ allo stato $E_{2,1}$, che nel modello di Bohr avrebbe dovuto produrre nello spettro un'unica riga di frequenza, $E_2 - E_1 = h\nu_1$, in realtà produce un doppietto di righe aventi frequenze molto vicine tra loro, relative alle transizioni $E_{2,0} \rightarrow E_{1,0}$ e $E_{2,1} \rightarrow E_{1,0}$.

Un'ulteriore modifica a questo modello venne introdotta con la scoperta dell'**effetto Zeeman**, fenomeno per il quale negli spettri di atomi sottoposti ad un campo magnetico esterno alcune righe spettrali divengono dei multipletti. Ciò è frutto di una diversa interazione con il campo magnetico esterno delle orbite percorse dagli elettroni: orbite, che in assenza di campo magnetico esterno possiedono uguale energia e orientamenti diversi nello spazio, in presenza di un campo

esterno assumono energie diverse poiché la somma vettoriale del **campo magnetico orbitale** [2] e del campo esterno risulta diversa a seconda del reciproco orientamento.

Per tener conto dei diversi orientamenti delle orbite nello spazio è stato introdotto un terzo numero quantico, detto **numero quantico magnetico** m , i cui valori sono correlati a quelli del numero quantico secondario l ; m può assumere tutti i valori da $-l$ a $+l$, incluso lo zero. In altre parole:

$$m = -l, \dots, 0, \dots, +l$$

Va sottolineato che il numero quantico m non definisce ulteriori sottolivelli energetici poiché non influenza l'energia dell'orbitale, ma solo il suo orientamento nello spazio: orbite aventi gli stessi valori di n ed l e diversi valori di m possiedono la medesima energia e diversi orientamenti nello spazio. Una differenza di energia insorge unicamente in presenza di un campo magnetico esterno, nel quale l'energia totale dell'orbita cambia a seconda del suo orientamento rispetto al campo esterno.

4 La meccanica quantistica

Già con il modello di Bohr e con le sue successive modifiche si era fatta largo l'idea che i sistemi microscopici, costituiti, cioè, da particelle di massa molto piccola, come particelle subatomiche, atomi o molecole, non obbedissero alle leggi della fisica classica e che per loro dovesse essere elaborata un nuovo tipo di meccanica, detta **meccanica quantistica**. In essa, le grandezze fisiche che descrivono il moto delle particelle non possono assumere valori qualsiasi, variabili in modo continuo, come accade nei sistemi macroscopici, ma solo valori discreti, multipli interi di un valore minimo, detto «quanto».

Così, analogamente a quanto aveva assunto Bohr, l'energia totale dell'atomo è una grandezza quantizzata, che può assumere unicamente valori multipli interi del «**quanto di energia**». Oltre al concetto di quantizzazione per le particelle microscopiche, la meccanica quantistica introduce un altro importante principio, detto **principio di indeterminazione di Heisenberg**:

«Per una particella microscopica non è possibile conoscere contemporaneamente e precisamente posizione e quantità di moto».

Ciò significa che, mentre la meccanica classica consentiva di definire con esattezza posizione e velocità (ovvero la traiettoria nel tempo) di un corpo in base alle forze agenti su di esso, la meccanica quantistica non permette di fare lo stesso per una particella microscopica, per la quale posizione e velocità non possono essere conosciute contemporaneamente. Tutto ciò che possiamo conoscere è la probabilità che la particella occupi una certa posizione nello spazio. Ciò segna, quindi, il passaggio dal rigido determinismo della meccanica classica alla «**visione probabilistica**» della meccanica quantistica.

5 Il modello ondulatorio dell'atomo

L'applicazione della meccanica quantistica al sistema atomo si deve a **Schrödinger** (1926), che ha impostato e risolto l'equazione del moto dell'elettrone nell'atomo di idrogeno. Le soluzioni di questa equazione sono delle funzioni delle coordinate spaziali dell'elettrone che

[2] Qualsiasi circuito chiuso (o spira) percorso da corrente genera un campo magnetico. L'elettrone che percorre la sua orbita è assimilabile ad una spira percorsa da corrente e, quindi, genera un campo magnetico associato al suo moto orbitale.

associano ad ogni sua posizione attorno al nucleo un valore della probabilità che l'elettrone occupi, appunto, quella posizione. A queste funzioni è stato dato il nome di **orbitali atomici**, per sottolineare il fatto che non si tratta più di orbite ben definite che descrivono la traiettoria seguita dagli elettroni, bensì di funzioni di probabilità. La rappresentazione grafica degli orbitali atomici viene effettuata attraverso delle superfici che racchiudono al loro interno più o meno tutte le posizioni attorno al nucleo per le quali la probabilità di trovare l'elettrone sia diversa da zero. In tal modo, il volume racchiuso da queste superfici rappresenta la porzione di spazio in cui possiamo trovare l'elettrone con una probabilità prossima al 100%.

Le funzioni matematiche che descrivono gli orbitali atomici vengono indicate come $\psi_{n,l,m}(x,y,z)$ poiché sono funzioni delle coordinate spaziali dell'elettrone x, y, z e dipendono da tre parametri, detti **numeri quantici**, indicati con le lettere n, l, m . I valori che questi tre numeri quantici possono assumere sono gli stessi previsti nel modello di Bohr ma, mentre in questo modello erano stati introdotti semplicemente per dare una spiegazione alle osservazioni sperimentali, nel modello ondulatorio derivano direttamente dalla soluzione dell'equazione del moto dell'elettrone.

Il **numero quantico principale n** , che assume solo valori interi positivi, quantizza l'energia dell'elettrone ed indica la sua distanza dal nucleo: ciascun valore di n individua un livello energetico, detto anche «strato», il cui valore di energia cresce al crescere di n insieme alla dimensione dei corrispondenti orbitali:

$n = 1$	strato K
$n = 2$	strato L
$n = 3$	strato M

Il **numero quantico angolare l** , che assume solo i valori interi positivi dallo 0 ad $(n - 1)$, quantizza il momento angolare dell'elettrone e quindi definisce la forma dell'orbitale. Per ogni livello energetico n , il numero quantico l individua n sottolivelli di energia poco diversi tra loro, a ciascuno dei quali appartengono orbitali di forma diversa:

$l = 0$	orbitali di tipo s	(sottolivello s)
$l = 1$	orbitali di tipo p	(sottolivello p)
$l = 2$	orbitali di tipo d	(sottolivello d)

Le forme degli orbitali s e p sono riportate nella figura alla pagina seguente.

Il **numero quantico magnetico m** definisce l'orientamento delle funzioni orbitali nello spazio: per ogni valore di l , ossia per ogni tipo di orbitale, il numero quantico m assume $(2l + 1)$ valori, ognuno dei quali corrisponde ad un orbitale di tipo l diversamente orientato nello spazio. Dunque il numero di valori assunti da m , cioè $(2l + 1)$, indica quanti orbitali di tipo l ci sono nell'atomo.

Le funzioni orbitali aventi uguali valori di n ed l , e che differiscono solo per il valore di m , possiedono, in assenza di campi magnetici esterni, uguale energia e vengono dette «degeneri».

Tipo di orbitale	l	m	Numero di orbitali
s	0	0	1
p	1	-1,0,+1	3
d	2	-2,-1,0,+1,+2	5
f	3	-3,-2,-1,0,+1,+2,+3	7

Tab. 1

Come mostrato in figura, i tre orbitali p vengono distinti in p_x, p_y, p_z , a seconda dell'asse lungo il quale si sviluppano:

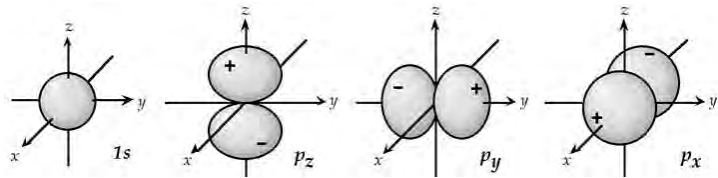


Fig. 1

I tre numeri quantici n, l ed m caratterizzano il moto orbitale dell'elettrone, cioè il suo moto attorno al nucleo. L'elettrone tuttavia è dotato anche di un moto di rotazione attorno al proprio asse, detto *moto di spinning*. Questo moto può avvenire in senso orario o antiorario, generando due campi magnetici di direzione opposta. Per definire lo *stato di spin* dell'elettrone è stato introdotto un quarto numero quantico, il **numero quantico di spin** m_s , che può assumere solo due valori: $+1/2$ e $-1/2$.

Quindi ogni elettrone nell'atomo è descritto da una propria funzione orbitale univocamente definita dai valori dei tre numeri quantici n, l ed m ed è contraddistinto da uno dei due possibili stati di spin.

Esiste un importante principio che regola il modo in cui gli elettroni possono «distribuirsi» nei diversi orbitali atomici; tale principio, denominato **principio di esclusione di Pauli**, afferma che:

«In un atomo non possono coesistere due elettroni descritti dagli stessi quattro numeri quantici».

Ciascuna delle funzioni orbitali, dunque, univocamente definita dalla terna di valori n, l, m , può descrivere al massimo due elettroni che differiscono tra loro per lo stato di spin. Questo principio viene generalmente espresso dicendo che ogni orbitale atomico può contenere al massimo due elettroni di spin opposto. Nella tabella che segue sono riassunti i valori attribuibili ai quattro numeri quantici: risulta evidente come in ogni livello n , in base al principio di esclusione di Pauli, si trovino $2n^2$ elettroni.

$n = 1$	$l = 0$	$m = 0$	$m_s = -1/2$
			$m_s = +1/2$
$n = 2$	$l = 0$	$m = 0$	$m_s = -1/2$
			$m_s = +1/2$
	$l = 1$	$m = -1$	$m_s = -1/2$
			$m_s = +1/2$
		$m = 0$	$m_s = -1/2$
			$m_s = +1/2$
		$m = +1$	$m_s = -1/2$
			$m_s = +1/2$

(Segue)

$n = 3$	$l = 0$	$m = 0$	$m_s = -1/2$
			$m_s = +1/2$
	$l = 1$	$m = -1$	$m_s = -1/2$
			$m_s = +1/2$
		$m = 0$	$m_s = -1/2$
			$m_s = +1/2$
		$m = +1$	$m_s = -1/2$
			$m_s = +1/2$
	$l = 2$	$m = -2$	$m_s = -1/2$
			$m_s = +1/2$
		$m = -1$	$m_s = -1/2$
			$m_s = +1/2$
		$m = 0$	$m_s = -1/2$
			$m_s = +1/2$
		$m = +1$	$m_s = -1/2$
			$m_s = +1/2$
		$m = +2$	$m_s = -1/2$
			$m_s = +1/2$

Tab. 2

Generalmente, un orbitale atomico viene indicato con un numero e una lettera. Il numero corrisponde al valore del numero quantico principale n e la lettera indica il tipo di orbitale: s se $l = 0$, p se $l = 1$, d se $l = 2$.

Graficamente si usa far corrispondere ad ogni orbitale atomico una casella, all'interno della quale l'elettrone viene indicato con una freccia verso l'alto o verso il basso per distinguere lo stato di spin.

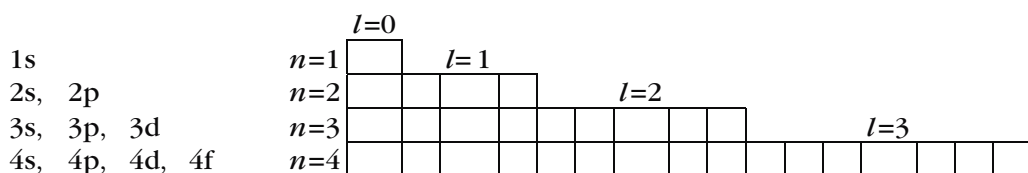


Fig. 2

La conoscenza del numero e del tipo di orbitali presenti nell'atomo e della loro sequenza energetica consente di stabilire le configurazioni elettroniche degli elementi.

In base a quanto detto sul significato dei numeri quantici possiamo affermare che, in linea generale, l'energia degli orbitali atomici ha un andamento crescente all'aumentare del numero quantico principale n . Nell'ambito dello stesso livello energetico, ovvero a parità di numero quantico principale, l'energia degli orbitali aumenta al crescere del numero quantico l . Ciò può determinare, in alcuni casi, delle inversioni nell'ordine di energia rispetto al numero quantico principale: gli orbitali $3d$, ad esempio, possiedono energia più alta dei $4s$; analogamente, i $4d$ hanno energia maggiore dei $5s$ etc.

Bisogna tener presente, infine, che l'ordine di energia degli orbitali può subire delle variazioni al mutare del numero atomico (cioè passando da un elemento ad un altro). Comunque, per la gran parte degli elementi risulta rispettata la sequenza energetica seguente:

1s, 2s, 2p, 3s, 3p, 4s, 3d, 4p, 5s, 4d, 5p, 6s, 4f, 5d, 6p, 7s, ...

6 La costruzione delle configurazioni elettroniche degli elementi

Come si è accennato precedentemente, il comportamento chimico degli elementi è determinato essenzialmente dalle caratteristiche della loro nube elettronica. L'interazione tra due atomi che si avvicinano è limitata allo strato esterno delle loro nubi elettroniche: gli elettroni dei livelli energetici sottostanti, insieme al nucleo, costituiscono una sorta di «guscio chiuso», non interagente con l'esterno. Per questo, la conoscenza della configurazione elettronica degli elementi consente di fare previsioni sulle loro proprietà chimiche e di suddividerli in gruppi dotati di proprietà affini. Da questo tipo di analisi è emerso, infatti, che elementi aventi nello strato esterno, cioè nell'ultimo livello occupato, lo stesso numero di elettroni mostrano proprietà chimiche molto simili.

Per costruire correttamente la configurazione elettronica di un elemento è indispensabile rispettare le seguenti regole fondamentali.

1. **L'energia totale dell'atomo deve essere la minima possibile.** Per *energia totale dell'atomo* si intende la somma delle energie di tutti gli elettroni presenti. Ciascun elettrone possiede l'energia dell'orbitale che occupa. Per questo motivo gli elettroni occupano preferibilmente gli orbitali di energia più bassa e solo quando questi sono completi vanno ad occupare orbitali di energia più alta. Lo stato in cui l'energia totale risulta la minima possibile viene indicato come **stato fondamentale**.
2. Il **principio di Hund**, detto anche **principio di massima molteplicità**. Questo principio impone che quando nell'atomo sono disponibili orbitali degeneri, cioè di uguale energia, gli elettroni ne occupino il massimo numero possibile, disponendosi singolarmente in ogni orbitale con spin paralleli (cioè dello stesso segno).
3. Il **principio di Pauli**. Abbiamo visto come, in base a questo principio, ogni orbitale atomico possa accogliere un massimo di due elettroni, dotati di spin opposto.

Di seguito sono mostrate le configurazioni elettroniche dei primi diciotto elementi, ordinate in modo da incolonnare tra loro elementi con lo stesso numero di elettroni sullo strato esterno.

H Z=1 1s ↑							He Z=2 1s ↑↓
Li Z=3 2p 2s ↑ 1s ↑↓	Be Z=4 2p 2s ↑↓ 1s ↑↓	B Z=5 2p ↑ 2s ↑↓ 1s ↑↓	C Z=6 2p ↑ ↑ 2s ↑↓ 1s ↑↓	N Z=7 2p ↑ ↑ ↑ 2s ↑↓ 1s ↑↓	O Z=8 2p ↑↓ ↑ ↑ 2s ↑↓ 1s ↑↓	F Z=9 2p ↑↓ ↑↓ ↑ 2s ↑↓ 1s ↑↓	Ne Z=10 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓
Na Z=11 3p 3s ↑ 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓	Mg Z=12 3p 3s ↑↓ 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓	Al Z=13 3p ↑ 3s ↑↓ 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓	Si Z=14 3p ↑ ↑ 3s ↑↓ 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓	P Z=15 3p ↑ ↑ ↑ 3s ↑↓ 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓	S Z=16 3p ↑↓ ↑ ↑ 3s ↑↓ 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓	Cl Z=17 3p ↑↓ ↑↓ ↑ 3s ↑↓ 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓	Ar Z=18 3p ↑↓ ↑↓ ↑↓ 3s ↑↓ 2p ↑↓ ↑↓ ↑↓ 2s ↑↓ 1s ↑↓

Fig. 3

Le configurazioni elettroniche possono essere scritte in forma molto più sintetica, indicando il tipo di orbitali occupati e facendo comparire su ognuno di essi, come apice, il numero di elettroni che ospitano.

Elemento	Configurazione elettronica
H	$1s^1$
He	$1s^2$
Li	$1s^2 2s^1$
Be	$1s^2 2s^2$
B	$1s^2 2s^2 2p^1$
C	$1s^2 2s^2 2p^2$
N	$1s^2 2s^2 2p^3$
O	$1s^2 2s^2 2p^4$
F	$1s^2 2s^2 2p^5$
Ne	$1s^2 2s^2 2p^6$

Tab. 3

Per ricavare rapidamente la configurazione elettronica di un elemento è sufficiente conoscere il suo numero atomico e disporre i suoi elettroni nei vari orbitali, rispettando il loro ordine di riempimento e tenendo a mente che gli orbitali di tipo *s* possono accogliere solo due elettroni, quelli di tipo *p* sei elettroni, quelli di tipo *d* dieci e quelli di tipo *f* quattordici.

Il sistema periodico degli elementi

1 Caratteristiche e struttura

Il **sistema periodico** degli elementi non è altro che una tabella in cui i diversi elementi chimici sono stati ordinati per numero atomico crescente, utilizzando il criterio di andare a capo ogni volta che si inizia a riempire un nuovo livello energetico. Ciò ha generato una struttura in cui sulla stessa riga vengono a trovarsi elementi aventi gli elettroni esterni sul medesimo livello energetico, e nella stessa colonna tutti gli elementi aventi un ugual numero di elettroni sul livello esterno.

Le righe orizzontali sono dette **periodi** e sono numerate in base al valore del numero quantico principale del livello energetico che si va riempiendo; le colonne sono dette **gruppi** ed hanno una numerazione corrispondente al numero di elettroni che si trovano sull'ultimo livello energetico. Così idrogeno ($H, Z = 1, 1s^1$), litio ($Li, Z = 3, 1s^2 2s^1$), sodio ($Na, Z = 11, 1s^2 2s^2 2p^6 3s^1$) etc., che hanno un solo elettrone sul livello esterno, fanno parte del I gruppo ed occupano la prima colonna a sinistra nella tavola periodica; berillio ($Be, Z = 4, 1s^2 2s^2$), magnesio ($Mg, Z = 12, 1s^2 2s^2 2p^6 3s^2$) etc., che hanno due elettroni sull'ultimo livello, fanno parte del II gruppo e costituiscono la seconda colonna da sinistra della tavola periodica; boro ($B, Z = 5, 1s^2 2s^2 2p^1$), alluminio ($Al, Z = 13, 1s^2 2s^2 2p^6 3s^2 3p^1$) etc., che hanno tre elettroni sull'ultimo livello, costituiscono il III gruppo, e così via.

Nella **tavola periodica** gli elementi risultano raggruppati in diversi blocchi: le prime due colonne da sinistra costituiscono il **blocco s**, che comprende gli elementi in cui l'ultimo elettrone si trova in un orbitale di tipo s. Il blocco s comprende solo due gruppi, il I e il II, poiché in ogni livello energetico vi è un solo orbitale s in cui prendono posto al massimo due elettroni. Le ultime sei colonne a destra costituiscono, invece, il **blocco p** e comprendono gli elementi in cui l'ultimo elettrone si trova in un orbitale di tipo p. I gruppi del blocco p sono sei, dal III al VII più il gruppo zero dei gas nobili, poiché in ogni livello energetico troviamo sei orbitali p che ospitano un massimo di sei elettroni. Al centro del sistema periodico troviamo il **blocco d**, detto anche blocco degli «elementi di transizione», che comprende dieci gruppi, essendo cinque gli orbitali di tipo d. Questo blocco mostra una particolarità rispetto ai precedenti, legata alle inversioni esistenti nell'ordine di energia degli orbitali rispetto al numero quantico principale: mentre nei blocchi s e p gli orbitali appartenenti al livello n -esimo vengono occupati nel periodo n -esimo, per il blocco d gli orbitali del livello n -esimo vengono occupati nel periodo $(n+1)$ -esimo; ossia la prima riga del blocco d, che corrisponde al riempimento degli orbitali 3d, si trova nel 4° periodo, poiché gli elettroni 3d possiedono energia più alta dei 4s e lo stesso accade per i periodi successivi. Quindi, il blocco d risulta traslato verso il basso di una riga rispetto alla numerazione dei periodi. Infine, inserito tra il blocco s e il blocco d, troviamo il **blocco f**, traslato di due righe in senso verticale rispetto al blocco s e al blocco p, poiché i primi orbitali f che si incontrano nell'atomo sono i 4f ed appartengono al quarto livello, ma vengono riempiti nel 6° periodo, dopo gli orbitali 6s. Il blocco f contiene i quattordici gruppi corrispondenti ai cosiddetti «elementi di transizione interna».

Questa struttura consente di mettere in evidenza l'esistenza di una «periodica» ripetitività nelle configurazioni elettroniche esterne degli elementi; nell'ambito di uno stesso gruppo troviamo elementi con numero atomico crescente spostandoci dall'alto verso il basso, ma dotati del medesimo numero di elettroni esterni, posti nello stesso tipo di orbitali.

[illegible]

È proprio questo strato esterno dell'atomo, detto **guscio di valenza**, a determinarne le proprietà chimiche e, dunque, possiamo aspettarci una discreta affinità di comportamento chimico per tutti gli elementi appartenenti ad uno stesso gruppo.

2 Proprietà periodiche e comportamento chimico

Spostandosi lungo un periodo, nell'ambito di questo sistema, si ha una variazione continua della configurazione elettronica esterna degli elementi, poiché ogni elemento possiede un elettrone in più rispetto a quello che lo precede. Ciò comporta anche una variazione continua di importanti proprietà, le quali determinano il comportamento chimico degli elementi. Tra queste vanno ricordate, in particolare, quelle descritte nei sottoparagrafi successivi.

a) Raggio atomico

Viene definito **raggio atomico covalente dell'atomo A** il valore della semidistanza tra i nuclei dei due atomi che costituiscono la molecola biatomica dell'elemento o di due atomi nel cristallo. Il raggio atomico ha un andamento tipico nel sistema periodico, mostrando un decremento continuo lungo i periodi ed aumentando progressivamente lungo i gruppi.

b) Energia di ionizzazione

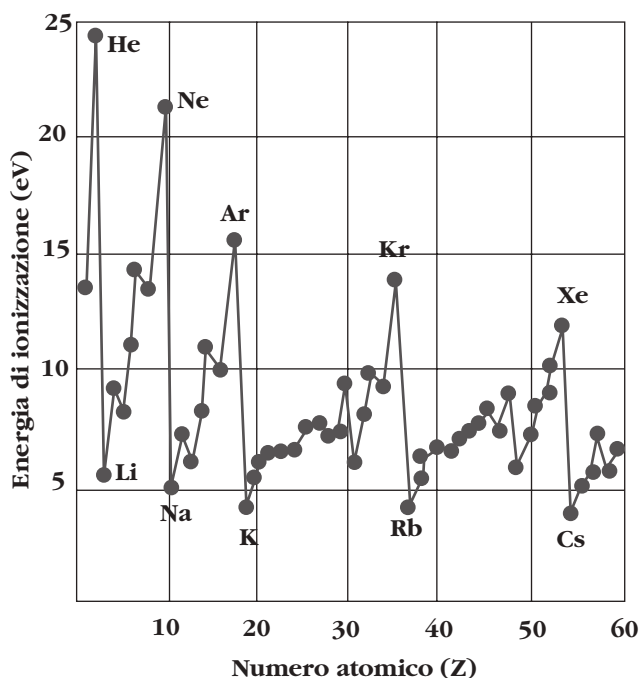


Fig. 2

Viene definita **energia di ionizzazione E_i** l'energia necessaria a strappare all'atomo isolato, a 0 K, il suo elettrone più esterno e a portarlo a distanza infinita, a stato di energia cinetica nulla. La ionizzazione produce uno *ione positivo*, o *catione*, meno stabile dell'atomo neutro e, quindi, è un processo che richiede apporto di energia dall'esterno: per questo l'energia di ionizzazione ha sempre valore positivo. Sottraendo all'atomo, successivamente, uno, due o tre elettroni, si formeranno ioni aventi carica +1, +2, +3, rispettivamente, e le energie spese per le successive ionizzazioni saranno dette energie di prima, seconda, terza ionizzazione. Parlando in modo generico di *energia di ionizzazione* ci si riferisce all'*energia di prima ionizzazione*.

Le energie di ionizzazione per gli elementi di uno stesso periodo mostrano un andamento all'incirca continuo e crescente passando dagli elementi del primo gruppo, per i quali si misurano le energie di ionizzazione più basse, a quelli dell'ottavo gruppo, che possiedono le energie di ionizzazione più alte. Nel primo periodo, di cui fanno parte solo l'idrogeno e l'elio, l'energia di ionizzazione ha valore massimo per l'He. Poiché l'entità dell'energia di ionizzazione può essere correlata al grado di destabilizzazione che la ionizzazione provoca rispetto allo stato di atomo neutro, questo

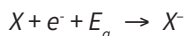
andamento ci permette di dedurre che le configurazioni elettroniche in cui il livello esterno è completo, come accade per l'He e per gli elementi del gruppo zero (i *gas nobili*), sono quelle che conferiscono all'atomo la maggiore stabilità.

c) Affinità elettronica

Viene definita **affinità per l'elettrone** la quantità di energia liberata o assorbita dall'atomo quando questo acquista un elettrone, trasformandosi nel corrispondente ione negativo. L'affinità per l'elettrone E_a , contrariamente all'energia di ionizzazione, può assumere valori sia positivi sia negativi. Se lo ione che si forma è più stabile dell'atomo neutro, il processo è spontaneo ed avviene con cessione di energia all'esterno: l'affinità per l'elettrone, in questo caso, ha valore negativo e il processo per un generico elemento X può essere scritto come:



Viceversa, se l'anione che si forma è meno stabile dell'atomo neutro, l'ingresso dell'elettrone nell'atomo non è spontaneo e può avvenire solo se forzato, fornendo energia dall'esterno: in questo caso l'affinità per l'elettrone ha valore positivo e il processo di formazione dell'anione può scriversi come:



Gli elementi che tendono ad acquistare spontaneamente un elettrone sono pochi e appartengono al gruppo degli alogeni, cioè al VII gruppo.

d) Carattere metallico

Il **carattere metallico** di un elemento è una misura della sua facilità a cedere elettroni: vengono definiti **metalli** gli elementi dotati di notevole tendenza a cedere elettroni; viceversa, vengono classificati come **non metalli** gli elementi con scarsa tendenza a cedere elettroni. È ovvio che una misura del carattere metallico di un elemento sia proprio la sua energia di ionizzazione: gli elementi dei primi gruppi del sistema periodico (I, II, III) hanno energie di ionizzazione basse, quindi una certa facilità a cedere elettroni, e possono essere classificati come «metalli»; al contrario, gli elementi degli ultimi gruppi (VII, VI) cedono difficilmente elettroni, come è evidente dai valori elevati delle energie di ionizzazione, e vengono classificati come «non metalli».

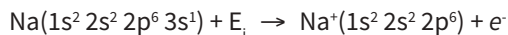
Come sempre accade in chimica, la separazione tra metalli e non metalli non è netta e le proprietà degli uni sfumano in quelle degli altri in modo graduale al variare della configurazione elettronica. A parità di configurazione elettronica, ossia nell'ambito di uno stesso gruppo, si osserva un aumento del carattere metallico al crescere del numero atomico e, quindi, delle dimensioni dell'atomo: ciò deriva, ovviamente, dal fatto che al crescere del raggio atomico gli elettroni esterni sono più lontani e, di conseguenza, più debolmente trattiene dal nucleo. Come criterio generale, possiamo dire che il carattere metallico degli elementi diminuisce lungo i periodi ed aumenta lungo i gruppi.

3 La regola dell'ottetto

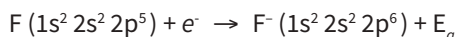
L'andamento dei valori delle energie di ionizzazione lungo un periodo ci ha permesso di concludere che la configurazione elettronica $s^2 p^6$, tipica dei gas nobili, conferisce all'atomo la maggiore stabilità possibile. Ciò comporta che:

*«Tutti gli elementi tendono a realizzare la configurazione ad otto elettroni esterni, detta comunemente **ottetto**, cedendo, acquistando o mettendo in comune elettroni con altri atomi, cioè formando legami chimici».*

Gli elementi del I gruppo mostrano energie di ionizzazione particolarmente basse proprio perché cedendo l'elettrone esterno, che si trova sull'orbitale ns , acquistano la configurazione elettronica del gas nobile che li precede nel sistema periodico, cioè $(n-1)s^2(n-1)p^6$:



Analogamente, gli elementi del VII gruppo, aventi configurazione $ns^2 np^5$, mostrano affinità per l'elettrone negativo, cioè acquistano spontaneamente un elettrone, poiché in tal modo assumono la configurazione $ns^2 np^6$ del gas nobile che li segue nel sistema periodico:



Gli elementi del I e del VII gruppo rappresentano due casi limite. Tutti gli altri elementi, per i quali il raggiungimento dell'ottetto richiederebbe l'acquisto o la cessione di più elettroni, ricorrono in genere alla formazione di legami covalenti con altri atomi; in tal modo la configurazione ad otto elettroni viene raggiunta mediante la «messa in comune» di elettroni. La regola dell'ottetto si applica bene agli elementi del secondo periodo: l'idrogeno fa chiaramente eccezione, poiché il suo livello elettronico esterno è completo già con due elettroni. Per gli elementi dei periodi successivi al secondo, che possiedono anche orbitali di tipo d potenzialmente utilizzabili nella formazione dei legami, si possono osservare deviazioni da questa regola.

4 La notazione di Lewis

Si è detto più volte che a determinare il comportamento chimico di ciascun elemento è la configurazione dello strato di valenza: tutto ciò che si trova al di sotto costituisce un guscio chiuso che non interagisce con l'esterno. Per questo le configurazioni elettroniche vengono spesso rappresentate in modo rapido, esplicitando la struttura del solo strato esterno ed indicando la parte sottostante attraverso il simbolo del gas nobile che precede l'elemento nel sistema periodico, racchiuso tra parentesi quadre. Ad esempio, per il potassio K, che ha numero atomico $Z = 19$, la configurazione elettronica può essere indicata come $1s^2 2s^2 2p^6 3s^2 3p^6 4s^1$ oppure come $[\text{Ar}] 4s^1$, dove $[\text{Ar}]$ indica la configurazione dell'Argon, cioè $1s^2 2s^2 2p^6 3s^2 3p^6$. Analogamente per l'atomo di zolfo S, che ha numero atomico $Z = 16$, la configurazione elettronica può essere scritta $1s^2 2s^2 2p^6 3s^2 3p^4$ oppure come $[\text{Ne}] 3s^2 3p^4$.

Le configurazioni elettroniche esterne degli elementi possono essere rappresentate in modo alternativo attraverso la cosiddetta **notazione di Lewis**, in cui gli elettroni del guscio di valenza vengono indicati come puntini attorno al simbolo dell'elemento. Ad esempio, con questo tipo di notazione il potassio viene rappresentato come $\text{K}\cdot$, il berillio come $\text{Be}\cdot\cdot$, il cloro come $\cdot\ddot{\text{Cl}}\cdot$ etc.

1 La specificità dei legami

Quanto detto finora riguardo la struttura atomica ci consente di comprendere come in natura gli unici elementi che possiedono stabilità sufficiente ad esistere allo stato monoatomico siano i gas nobili, essendo dotati di livelli elettronici esterni completi. Tutti gli altri elementi, invece, stabiliscono legami tra loro originando molecole la cui energia totale risulta inferiore rispetto alla somma delle energie degli atomi isolati, seguendo la naturale tendenza di qualsiasi sistema a minimizzare il proprio contenuto di energia.

Ogni legame chimico è caratterizzato da una **energia di legame**, che rappresenta l'energia necessaria a rompere il legame esistente tra due atomi portandoli a distanza infinita. L'energia di legame tra due atomi può variare a seconda della molecola nella quale sono inseriti.

Al crescere dell'energia di legame diminuisce la **distanza di legame**, intesa come la distanza esistente tra i nuclei dei due atomi legati. Questa distanza rappresenta in realtà un valore medio, poiché a temperature superiori allo zero assoluto (0 K), gli atomi o gli ioni legati tra loro sono in perenne oscillazione attorno alle proprie posizioni di equilibrio.

Abbiamo diversi tipi di legame chimico; in linea di principio, ogni volta che tra atomi o molecole esistono forze di entità sufficiente a generare aggregati rivelabili sperimentalmente si può affermare che tra di essi esiste un legame chimico. Una prima suddivisione può farsi tra:

- **legami intramolecolari**, che si instaurano, cioè, tra atomi uguali o diversi per generare una molecola (ad esempio: H_2O , CO_2 , H_2 , O_2);
- **legami intermolecolari**, che si instaurano tra molecole che costituiscono la materia nei suoi stati solido, liquido o gassoso.

Un secondo tipo di suddivisione può farsi in base al tipo di legame. Possiamo distinguere:

- **legami atomici**, che si instaurano a livello intramolecolare tra atomi che condividono elettroni, senza che si verifichi la cessione di uno o più elettroni da un atomo ad un altro;
- **legami di natura elettrostatica**, i quali possono instaurarsi sia a livello intramolecolare tra atomi che acquistano elettroni, formando ioni negativi, e atomi che cedono elettroni, formando ioni positivi, che a livello intermolecolare, tra le molecole dotate di polarità elettrica in seguito ad asimmetrie nella distribuzione della carica elettrica al loro interno;
- **legami metallici**, che esulano dai due modelli precedenti e si instaurano unicamente tra elementi a carattere metallico.

2 Il legame atomico

L'interpretazione del legame atomico può essere affrontata attraverso due diverse teorie: la *teoria del legame di valenza* (VB: *valence bond*) o la *teoria degli orbitali molecolari* (MO: *molecular orbital*).

La **teoria del legame di valenza** parte dall'assunto che la formazione del legame tra due atomi coinvolga unicamente lo strato elettronico esterno: due atomi mettono in comune uno o più elettroni disaccoppiati del guscio di valenza avvicinandosi ad una distanza inferiore alla somma dei rispettivi raggi atomici e determinando la sovrapposizione degli orbitali contenenti

gli elettroni disaccoppiati. La zona di ricopertura che si genera è ad alta densità di carica negativa ed è responsabile della formazione del legame. In tutto ciò ogni atomo mantiene inalterata la propria struttura interna e, quindi, la propria identità all'interno della molecola.

La **teoria degli orbitali molecolari**, al contrario, ammette che nella formazione di una molecola gli atomi perdano la propria struttura interna e che tutti gli elettroni si ridistribuiscono su dei nuovi orbitali estesi all'intera molecola, detti per questo orbitali molecolari. Questa teoria, in pratica, concepisce una struttura per la molecola del tutto analoga alla struttura dell'atomo: se nell'atomo gli elettroni sono descrivibili attraverso orbitali atomici monocentrici, cioè aventi come unico centro il nucleo atomico, nella molecola gli elettroni vengono descritti da orbitali molecolari policentrici, i cui centri sono i nuclei degli atomi che costituiscono la molecola stessa. Questi nuovi orbitali vengono ricavati, analogamente agli orbitali atomici, attraverso l'applicazione alla molecola delle leggi della meccanica quantistica.

a) Legame covalente omeopolare

Consideriamo il caso di due atomi di idrogeno che si leghino per formare la molecola H_2 . Affinché tra i due atomi possa formarsi un legame è necessario che questi si avvicinino: le forze che si manifestano in questa fase sono **forze coulombiane** [1] di tipo attrattivo tra il nucleo di ciascun atomo e la nube elettronica dell'altro e forze coulombiane di tipo repulsivo tra i due

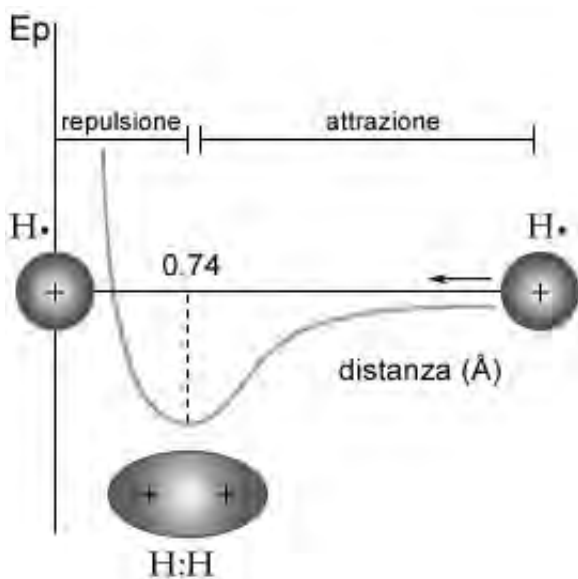


Fig. 1

nuclei e tra le due nubi elettroniche. Immaginiamo di avvicinare progressivamente due atomi di idrogeno partendo da distanza infinita e di calcolare l'energia potenziale del sistema riportandone i valori in un grafico: il risultato è mostrato in figura 1. Inizialmente si osserva una diminuzione di energia potenziale al diminuire della distanza, dovuta al prevalere delle forze di natura attrattiva; a 0,74 Å si registra un punto di energia potenziale minima, in cui si realizza il miglior compromesso possibile tra le forze di natura attrattiva e quelle di natura repulsiva; avvicinando ulteriormente i due atomi, si osserva una rapida crescita dell'energia potenziale, dovuta al prevalere di forze di natura repulsiva.

La distanza di 0,074 nm, per la quale si registra la minima energia potenziale, è l'effettivo valore della distanza di legame nella molecola di idrogeno ed è inferiore alla somma dei due raggi atomici, essendo il raggio dell'atomo di idrogeno pari a 0,053 nm. Evidentemente i due atomi si avvicinano fino a far compenetrare le loro nubi elettroniche; tra i due nuclei si genera una zona ad elevata densità di carica negativa, detta comunemente **zona di ricopertura**, che costituisce l'elemento legante tra i nuclei: infatti, è proprio l'attrazione tra la carica nucleare e la carica della zona di ricopertura a stabilizzare la

[1] Secondo la **Legge di Coulomb**, la forza che si esercita tra due cariche elettriche è direttamente proporzionale alle cariche ed inversamente proporzionale al quadrato della loro distanza. Questa forza è *attrattiva* tra cariche di segno opposto e *repulsiva* tra cariche dello stesso segno: $F = k \frac{q_1 \cdot q_2}{d^2}$.

molecola, e maggiore è l'estensione della zona di ricopertura, maggiore è la forza del legame. Quando, come in questo caso, la zona di ricopertura si trova lungo la congiungente dei nuclei, il legame è detto di **tipo σ** .

Legami di tipo σ possono instaurarsi tra orbitali di tipo s , come nella molecola di idrogeno, tra orbitali di tipo p , oppure tra un orbitale s ed un orbitale p . Esistono anche legami in cui le zone di ricopertura si trovano fuori della congiungente dei nuclei: in tal caso il legame che si instaura è detto di **tipo π** , ed è più debole rispetto al precedente poiché le zone di ricopertura sono più piccole e richiedono una certa deformazione degli orbitali atomici di partenza. I tipi di ricopertura che possono realizzarsi tra i diversi orbitali atomici sono mostrati in figura 3. Quando due atomi si legano, il primo legame che si instaura è sempre di tipo σ ; nel caso in cui questi dispongano di ulteriori elettroni disaccoppiati, possono formarsi legami di tipo π , fino ad un massimo di due.

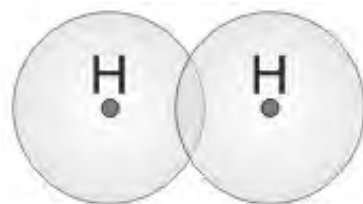


Fig. 2 - Molecola di H_2 ($H-H$)

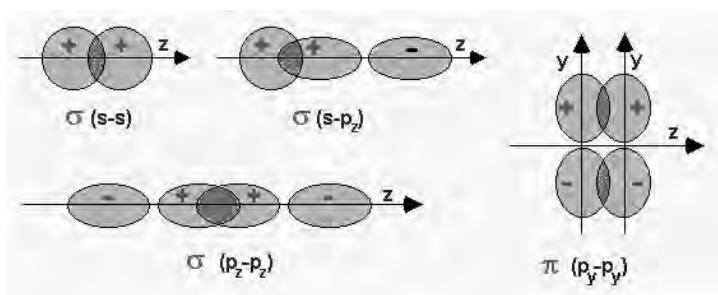


Fig. 3

Quindi due atomi possono essere legati attraverso un **legame semplice**, cioè un solo legame di tipo σ , o da un **legame doppio**, cioè da un legame σ e uno π , o da un **legame triplo**, cioè un legame σ e due legami π .

Il numero dei legami esistenti tra i due atomi è espresso dall'**ordine di legame**; ovviamente al crescere dell'ordine di legame cresce l'energia di legame e diminuisce la distanza di legame.

Esaminiamo a titolo di esempio la molecola di O_2 .

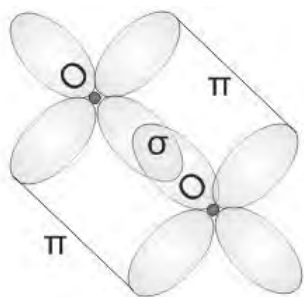


Fig. 4 - Molecola di O_2 ($O=O$)

L'atomo di O ha configurazione elettronica $1s^2 2s^2 2p^4$, quindi ciascuno dei due atomi di ossigeno possiede due elettroni disaccoppiati in orbitali di tipo p . Ricordiamo che gli orbitali di tipo p sono tutti uguali tra loro ed orientati nello spazio in modo che ognuno di essi abbia asse coincidente con uno dei tre assi cartesiani. Assumendo l'asse z come asse di legame tra i due atomi, si formerà un primo legame di tipo σ attraverso i due orbitali $2p_z$ in cui i due assi sono coincidenti, ed un secondo legame di tipo π al di fuori dell'asse di legame. La molecola di ossigeno è, quindi, caratterizzata dalla presenza di un legame doppio tra i due atomi di ossigeno. La rappresentazione in formula di struttura di questa molecola indicherà la presenza dei

due legami come due trattini identici tra i due atomi di ossigeno, anche se in realtà il legame σ e quello π differiscono tra loro sia per geometria sia per energia.

La molecola di azoto fornisce un esempio di molecola con triplo legame: ciascun atomo di azoto possiede tre elettroni disaccoppiati in orbitali di tipo p e, quindi, può formare un legame σ e due legami π situati fuori dalla congiungente dei nuclei su due piani ortogonali tra loro, che si intersecano lungo la congiungente.

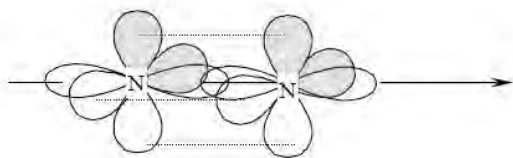


Fig. 5

In tutte le molecole descritte finora, gli atomi coinvolti nel legame sono identici: si tratta, dunque, di molecole biatomiche omonucleari nelle quali la carica elettrica è distribuita in modo perfettamente simmetrico tra i due atomi. Queste molecole, di conseguenza, non possiedono polarità elettrica e sono dette per questo *apolari*. Il legame covalente è definito, in questo caso, **puro** oppure **omeopolare**.

b) Legame covalente eteropolare

Consideriamo adesso il caso di una molecola costituita da atomi diversi, come, ad esempio, la molecola di HCl. L'atomo di idrogeno ha numero atomico $Z = 1$ e possiede un solo elettrone nell'orbitale $1s$; il cloro, invece, ha numero atomico $Z = 17$ e configurazione elettronica $1s^2 2s^2 2p^6 3s^2 3p^5$. Ognuno dei due atomi ha un elettrone disaccoppiato col quale poter formare un legame semplice; in questo caso la sovrapposizione avrà luogo tra un orbitale s ed un orbitale p . Tuttavia la profonda diversità dei due atomi determina una notevole differenza tra le azioni elettriche che questi esercitano sulla carica degli elettroni di legame, la quale si traduce in una pronunciata asimmetria nella distribuzione della carica stessa. La molecola HCl, infatti, mostra un'elevata polarità elettrica: l'atomo di cloro concentra su di sé la carica degli elettroni di legame in misura maggiore rispetto all'atomo di idrogeno e risulta carico negativamente, mentre l'idrogeno, depauperato di parte della sua carica ad opera del cloro, risulta carico positivamente.

Il legame che ne risulta è in realtà un *legame misto*, in parte di natura covalente, essendo determinato dalla messa in comune di elettroni, in parte di natura elettrostatica, poiché nella molecola si generano due poli elettrici di segno opposto che tenderanno ad attrarsi reciprocamente. In queste situazioni, che sono ovviamente le più frequenti, il legame che si instaura tra i due atomi è detto **covalente eteropolare**.

3 L'elettronegatività

Per **elettronegatività** di un atomo s'intende la capacità dell'atomo di addensare su di sé la carica degli elettroni di legame. L'elettronegatività è una proprietà che l'atomo esplica soltanto all'atto della formazione di un legame chimico, come azione competitiva con il partner di legame per l'attrazione della carica degli elettroni di legame. Ciò non toglie che anche all'elemento isolato possa essere associato un valore di elettronegatività, collocandolo in una scala di elettronegatività.

L'elettronegatività è in qualche misura correlabile sia all'energia di ionizzazione sia all'affinità per l'elettrone: infatti, un elemento con potenziale di ionizzazione elevato trattiene con forza i propri elettroni esterni ed, evidentemente, avrà poca tendenza a delocalizzare la carica del suo elettrone di legame sul partner di legame. Al contrario, elementi caratterizzati da basse energie di ionizzazione avranno una maggiore facilità a spostare la carica del proprio elettrone di legame sull'altro atomo. In modo analogo, possiamo immaginare che un elemento dotato di un'elevata affinità per l'elettrone abbia una spiccata tendenza ad accaparrarsi l'elettrone del partner di legame, poiché l'acquisto di un elettrone gli permette di liberare all'esterno l'energia E_p . Esistono diverse scale di elettronegatività: non a caso una di queste, la *Scala di Mulliken*, calcola l'elettronegatività degli elementi come media aritmetica tra i valori assoluti dell'energia di ionizzazione e dell'affinità per l'elettrone.

L'elettronegatività può essere correlata anche alla densità elettronica dell'elemento che è funzione dell'intensità del campo elettrico nucleare: infatti, gli elementi con densità elettroniche più alte (ossia, maggiore quantità di carica negativa per unità di volume) hanno campi elettrici nucleari elevati ed attraggono in maggior misura la carica degli elettroni di legame. La densità elettronica dipende sia dal numero atomico (cioè dal numero di elettroni della nube elettronica) sia dalle dimensioni dell'atomo ed è, quindi, una proprietà periodica degli elementi che aumenta lungo i periodi e diminuisce lungo i gruppi, con andamento inverso rispetto al raggio atomico.

Gli elementi di maggiore elettronegatività si trovano nella parte destra della tavola periodica, quelli di elettronegatività più bassa sono a sinistra; nell'ambito di uno stesso gruppo, l'elemento più elettronegativo è il primo, il più piccolo, e l'elettronegatività diminuisce gradualmente lungo il gruppo. L'elemento più elettronegativo del sistema periodico è dunque l'atomo di fluoro, il più piccolo del gruppo degli alogeni. La conoscenza dei valori di elettronegatività degli elementi ci permette di prevedere la natura del legame che si stabilisce tra due atomi: è ovvio che maggiore è la differenza di elettronegatività tra i due, più alto è il grado di polarità del legame che si forma.

4 Il legame dativo o di coordinazione

Si è visto finora che all'atto della formazione di un legame atomico ciascuno dei due atomi coinvolti mette in comune con l'altro un elettrone disaccoppiato. Il **legame dativo**, invece, è un tipo particolare di legame atomico in cui i due elettroni di legame provengono entrambi dallo stesso atomo, che mette a disposizione una coppia di elettroni presente in un suo orbitale completo, detta *lone pair* o **doppioetto elettronico**. Questo atomo, chiamato **datore**, condivide il suo doppietto elettronico con un altro atomo, definito **accettore**, il quale deve disporre di un orbitale vuoto di energia sufficiente oltre ad essere dotato di elettronegatività superiore al datore. Per evidenziare la particolarità di questo tipo di legame, spesso nelle formule di struttura viene indicato come una freccia che va dal datore all'accettore.

Un tipico accettore di legami dativi è l'atomo di ossigeno che, data la sua posizione nella tavola periodica, possiede elettronegatività abbastanza elevata e può liberare uno dei suoi orbitali accoppiando i due elettroni spaiati in uno degli orbitali *p*; in tal modo, l'altro orbitale, rimasto vuoto, può essere utilizzato per accogliere un doppietto elettronico.

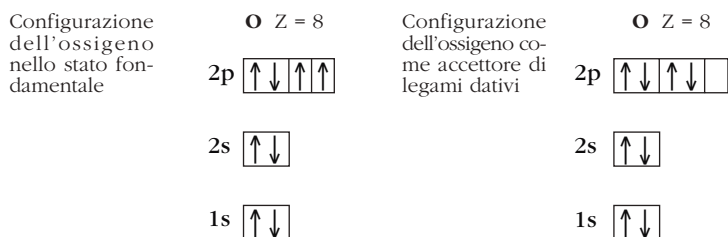
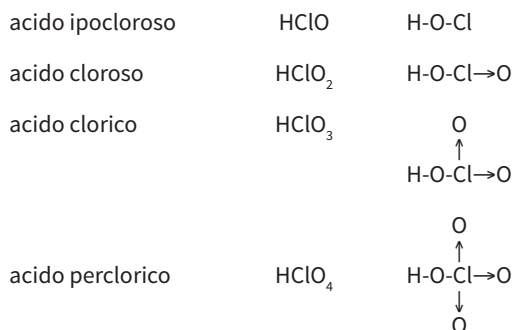


Fig. 6

La formazione di legami dativi consente agli atomi di stabilire un numero di legami superiore rispetto a quello prevedibile unicamente sulla base del numero di elettroni disaccoppiati. Consideriamo a titolo di esempio la serie degli acidi ossigenati del cloro: in base al numero di elettroni disaccoppiati, essendo la configurazione elettronica del cloro $1s^2 2s^2 2p^6 3s^2 3p^5$, dovremmo

aspettarci che tale atomo formi un solo legame covalente. In realtà, donando i suoi tre doppietti elettronici, il cloro può arrivare a formare quattro legami:



5 L'ibridizzazione degli orbitali

Abbiamo visto come ogni atomo tenda a completare la propria configurazione elettronica esterna raggiungendo l'ottetto e, quindi, una configurazione con quattro coppie di elettroni disposte attorno all'atomo. Qualsiasi coppia di elettroni, sia questa un doppietto non condiviso o una coppia di elettroni di legame, costituisce una zona ad alta densità di carica negativa che tende a respingere le altre. La repulsione elettrica porta queste zone a distanziarsi il più possibile tra loro, influenzando così la direzione dei legami nello spazio e, di conseguenza, la geometria molecolare.

La teoria dell'ibridizzazione degli orbitali è una delle trattazioni matematiche attraverso le quali si è riusciti a giustificare la geometria effettivamente osservata nelle molecole. In base a questa teoria, gli orbitali atomici impiegati nella formazione dei legami non sono gli orbitali dell'atomo isolato, ma orbitali modificati in modo da produrre zone di ricopertura più ampie, quindi legami più forti, e minimizzare le repulsioni elettriche tra i vari addensamenti di carica negativa presenti intorno al nucleo. Questi orbitali sono detti «ibridi» perché sono stati ottenuti matematicamente combinando tra loro gli orbitali dell'atomo isolato.

Nell'ibridizzazione degli orbitali risultano sempre osservate le regole seguenti:

- gli orbitali ibridi sono **ottenuti matematicamente** combinando tra loro orbitali atomici monoelettronici, cioè occupati da un solo elettrone. Il numero degli orbitali ibridi ottenuti è sempre uguale al numero di orbitali atomici mescolati;
- gli orbitali ibridi sono **tutti uguali tra loro** e, dunque, **isoenergetici**. Essi differiscono unicamente per l'orientazione nello spazio;
- gli orbitali ibridi mostrano **nubi elettroniche** sviluppate nella direzione del legame e possiedono due lobi, di cui uno molto più esteso dell'altro. Questi sono diretti nello spazio alla massima distanza possibile l'uno dall'altro.

Esistono diversi tipi di ibridizzazione, a seconda del genere e del numero di orbitali atomici mescolati. Il nome all'ibrido viene assegnato indicando gli orbitali atomici utilizzati nella combinazione e segnalando il numero di orbitali adoperati per ogni tipo con un apice.

a) Ibrido sp o lineare

È un tipo di ibridizzazione che deriva dal mescolamento di due orbitali atomici, uno di tipo s e uno di tipo p . Gli orbitali ibridi che ne risultano sono detti quindi sp e sono due, poiché provengono dalla combinazione di due orbitali atomici. Questi orbitali hanno assi giacenti sulla stessa retta e sono orientati in direzioni opposte, formando tra loro un angolo di 180° .

Le molecole in cui l'atomo centrale è ibridizzato sp sono lineari: ne sono esempi il cloruro di berillio BeCl_2 , l'anidride carbonica CO_2 e l'acetilene C_2H_2 .

Il Be ha configurazione $1s^2 2s^2$ e può ibridizzarsi sp promuovendo uno dei suoi elettroni $2s$ ad un orbitale $2p$ vuoto; ciascuno degli orbitali sp forma un legame σ con il cloro.

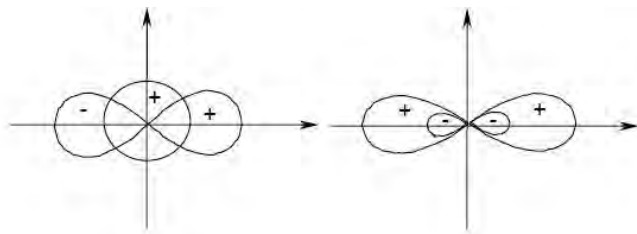


Fig. 7

Il carbonio C ha configurazione elettronica $1s^2 2s^2 2p^2$ e può anch'esso ibridizzarsi sp dopo aver promosso un elettrone $2s$ in un orbitale $2p$. Dall'ibridizzazione hanno origine due orbitali ibridi, che formano due legami σ , e rimangono due orbitali atomici p , che il carbonio può impiegare per formare due legami di tipo π . È questo il caso dell'anidride carbonica, in cui il carbonio centrale forma due legami doppi con i due ossigeni; oppure dell'acetilene, in cui i due carboni sono entrambi ibridizzati sp e formano un legame triplo tra loro e un legame semplice con i rispettivi idrogeni:



b) Ibrido sp^2 o planare

Gli orbitali ibridi sp^2 sono così denominati poiché derivano dal mescolamento di un orbitale s con due orbitali p . Essendo tre le funzioni atomiche di partenza, gli orbitali ibridi sono anch'essi tre ed hanno forma analoga agli ibridi sp . Per minimizzare la repulsione, questi orbitali giacciono tutti sullo stesso piano ed

hanno assi diretti verso i vertici di un triangolo equilatero, ossia a 120° l'uno dall'altro. Quindi le molecole i cui atomi centrali siano ibridizzati sp^2 sono planari.

Esempi di molecole con ibridizzazione sp^2 sono il cloruro di boro BCl_3 , l'acido carbonico H_2CO_3 o l'etene C_2H_4 . Sia il B che il C possono ibridizzarsi sp^2 in seguito alla promozione di un elettrone $2s$ ad un orbitale $2p$, ottenendo tre orbitali ibridi monoelettronici equivalenti. Al C rimane un orbitale p atomico con un elettrone dispari, che può essere impiegato per la formazione di un legame π :

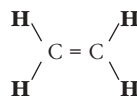
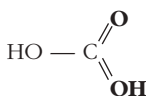
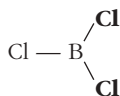


Fig. 9

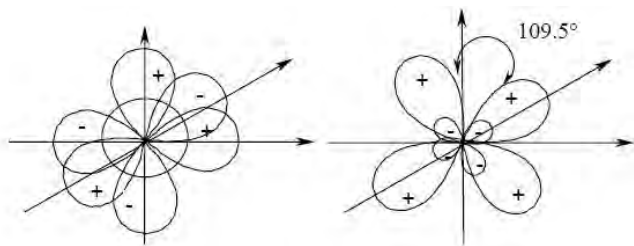
c) Ibrido sp^3 o tetraedrico

Fig. 10

Come indicato dal nome, l'ibrido sp^3 si ottiene dalla combinazione di un orbitale s con tre orbitali p ; gli orbitali ibridi generati sono quattro, tutti uguali, isoenergetici e diretti verso i vertici di un tetraedro, a $109,5^\circ$ l'uno dall'altro.

Questo tipo di ibridizzazione si osserva tipicamente nel carbonio,

ma anche in molecole come l'ammoniaca NH_3 o l'acqua H_2O , in cui gli atomi di N e O, pur non legando quattro atomi, si ibridizzano sp^3 ed impiegano alcuni degli orbitali ibridi per la formazione di legami σ , mentre gli altri ospitano i doppietti elettronici non condivisi. Poiché i *lone pair* occupano un volume maggiore attorno all'atomo rispetto ai doppietti di legame, proprio perché non condivisi con altri nuclei, nell' NH_3 e nell' H_2O si osserva un lieve restringimento dell'angolo di legame rispetto ai $109,5^\circ$ tipici dell'ibrido tetraedrico:

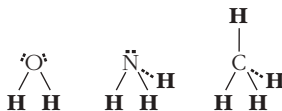


Fig. 11

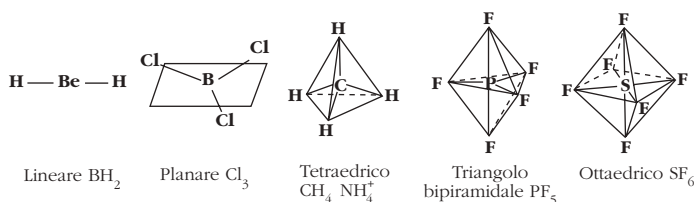
d) Ibridi sp^3d e sp^3d^2 

Fig. 12

Accenniamo brevemente ad altri due tipi di ibridizzazione che si riscontrano per atomi che possiedono anche orbitali di tipo d ; gli ibridi sp^3d e sp^3d^2 derivano dal mescolamento e dalla ridivisione di un orbitale s , tre orbitali p

ed uno o due orbitali d , rispettivamente. Gli orbitali ibridi ottenuti sono cinque o sei e generano geometrie molecolari bipiramidali con base triangolare o ottaedriche. Esempi di questo tipo di ibridizzazione sono PF_5 o SF_6 , mostrati in figura.

In linea generale, si può dire che per predire il tipo di ibridizzazione di un atomo in una molecola è sufficiente contare il numero delle coppie elettroniche presenti intorno al nucleo, includendo sia i legami σ che l'atomo forma (coincidente col numero di atomi che coordina) sia i doppietti non condivisi: il numero di orbitali ibridi è sempre uguale al totale delle coppie elettroniche presenti intorno al nucleo.

6 Il legame a elettroni delocalizzati

Ci sono alcune strutture molecolari che non possono essere efficacemente descritte da un'unica formula di struttura, ma mostrano proprietà compatibili con formule di struttura diverse. Un caso tipico è quello del benzene, molecola ciclica avente formula bruta C_6H_6 , la cui formula di struttura è stata discussa per lungo tempo. Ciascuno dei sei carboni del benzene forma tre legami σ , due con i carboni adiacenti nell'anello e uno con l'idrogeno. È plausibile

pensare, quindi, che i carboni siano ibridizzati sp^2 e ciascuno di essi abbia a disposizione un orbitale p atomico per formare un doppio legame. Ciò presupporrebbe una struttura esagonale irregolare con doppi legami, più corti, alternati a legami semplici, più lunghi, cosa che non si riscontra nella realtà. È noto, infatti, che il benzene ha la struttura di un esagono regolare con legami carbonio-carbonio tutti equivalenti e di lunghezza intermedia tra un legame semplice ed un legame doppio. Inoltre, immaginando di numerare i carboni dell'anello, così da poterli distinguere l'uno dall'altro, è chiaro che le strutture compatibili con questa ipotesi sono due, come rappresentato in figura, e possiedono la stessa probabilità di esistenza.

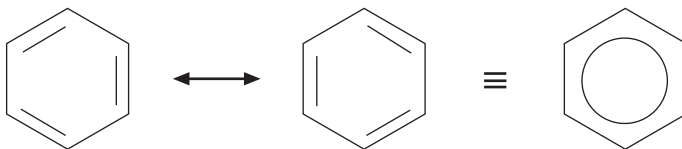


Fig. 13

La **teoria della risonanza** risolve questa apparente contraddizione, affermando che la struttura della molecola di benzene non è rappresentabile attraverso nessuna delle due strutture singolarmente, ma è un «ibrido di risonanza» delle due, cioè una forma intermedia tra le due strutture, dette *forme canoniche di risonanza*. Queste rappresentano solo delle situazioni limite, non realmente esistenti per la molecola. Ciò equivale a dire che i doppi legami non hanno una collocazione definita e gli elettroni π sono delocalizzati su tutto l'anello, costituendo una sorta di corrente elettrica che percorre la molecola. Per questo la molecola di benzene viene spesso rappresentata come un esagono regolare, al centro del quale si traccia un cerchio che simboleggia la circolazione degli elettroni π nella molecola. Generalmente, la risonanza tra due o più strutture, dette *strutture canoniche di risonanza*, viene indicata ponendo tra queste una doppia freccia, come in figura.

Il **legame ad elettroni delocalizzati** si incontra molto frequentemente nei composti organici e inorganici ogni qualvolta si possano scrivere per una stessa molecola più formule di struttura equivalenti; la delocalizzazione degli elettroni è un fenomeno limitato ai soli elettroni π e, quindi, le strutture canoniche di risonanza differiscono tra loro per la posizione dei doppi legami. Di seguito vengono mostrati alcuni esempi di risonanza in composti inorganici:

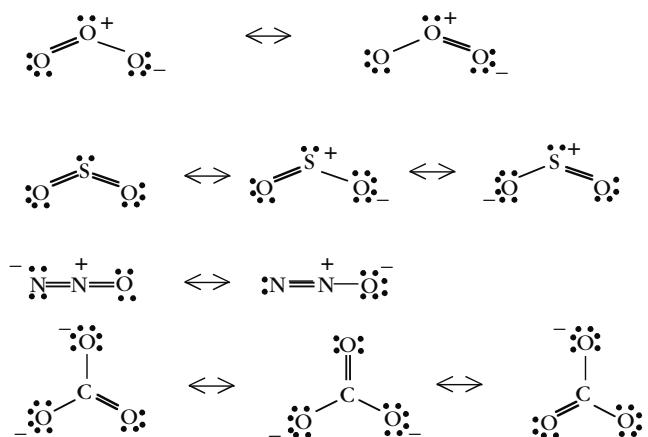
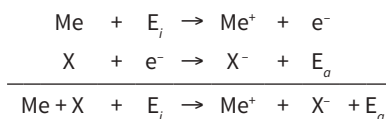


Fig. 14

7 Il legame ionico

Si è visto come il legame covalente che si instaura tra due atomi di diversa elettronegatività sia caratterizzato da una distribuzione asimmetrica della carica degli elettroni di legame tanto da generare un legame misto dovuto sia alla condivisione di elettroni tra i due atomi sia all'attrazione elettrostatica tra i due poli elettrici che si generano nella molecola. Quando la differenza di elettronegatività supera un certo limite, l'asimmetria nella distribuzione della carica di legame degenera in un vero e proprio trasferimento di un elettrone dall'atomo meno elettronegativo a quello più elettronegativo, con la formazione di due ioni di segno opposto che restano uniti in seguito all'attrazione elettrostatica.

Questo tipo di legame è detto **legame ionico** e si riscontra tipicamente tra un atomo a carattere spiccatamente metallico e uno a carattere non-metallico. La ionizzazione di un metallo richiede una quantità di energia pari alla sua energia di ionizzazione E_i ; l'acquisto di un elettrone da parte di un non-metallo, invece, libera una quantità di energia pari alla sua affinità per l'elettrone E_a . La formazione di un composto ionico tra un metallo Me e un non-metallo X può scriversi come:



Tuttavia, essendo i più alti valori di E_a inferiori ai più bassi valori di E_i , il processo scritto richiede comunque un apporto di energia dall'esterno e quindi risulta non spontaneo. Per questo, contrariamente a quanto accade per le molecole covalenti, per i composti ionici non si osserva mai la formazione della molecola isolata, costituita da un'unica coppia di ioni, ma solo di cristalli ionici in cui un gran numero di ioni positivi e negativi sono organizzati in un **reticolo cristallino**.

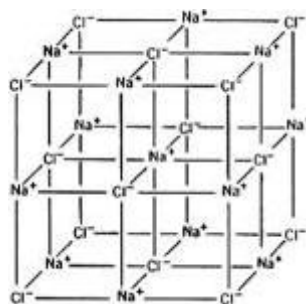


Fig. 15

All'interno del reticolo ogni ione esercita forze di attrazione coulombiana nei confronti di tutti gli ioni che lo circondano, il cui numero dipende dalla geometria del reticolo che si forma. In ogni caso, il numero delle interazioni attrattive è maggiore rispetto a quello della molecola isolata in cui l'attrazione coulombiana è limitata alla singola coppia ionica e ciò determina un guadagno di energia nettamente più alto. È questa energia, detta **energia reticolare**, a consentire la formazione del composto ionico.

8 I legami dipolari

Il *legame ionico* è il primo rappresentante di un'ampia famiglia di legami chimici di natura elettrostatica; tra questi, peraltro, è il più forte, poiché si instaura tra ioni dotati di carica unitaria o multipla dell'unità. Gli altri legami elettrostatici sono più deboli, poiché chiamano in causa frazioni della carica elementare, come quelle che originano i **dipoli [2] molecolari**. Va sottolineato che i dipoli molecolari possono essere **permanenti**, come quelli che si originano dai legami eteropolari, oppure **transitori**, o **indotti**, generati per **induzione elettrostatica**. L'insorgenza di un dipolo indotto in una molecola è dovuto alla temporanea distorsione della sua nube elettronica a causa della presenza di un campo elettrico esterno generato da uno ione, da una

[2] Un *dipolo* è un sistema costituito da due cariche elettriche di segno opposto, poste ad una distanza d .

molecola dipolare o da un dipolo indotto in una molecola vicina. I *legami dipolari* sono dovuti, quindi, ad interazioni *ione-dipolo*, *dipolo-dipolo*, *ione-dipolo indotto*, *dipolo-dipolo indotto*, *dipolo indotto-dipolo indotto*.

L'interazione ione-dipolo consiste, in pratica, nell'orientamento del dipolo nel campo elettrico dello ione: questo, infatti, volgerà verso lo ione la parte dotata di carica opposta a quella dello ione. Nel contempo, la presenza dello ione determina l'insorgere di un dipolo indotto nella molecola, che va a sommarsi al dipolo permanente. Quindi, all'interazione ione-dipolo si accompagna sempre quella ione-dipolo indotto. Il fenomeno descritto è alla base dell'idratazione degli ioni in soluzione acquosa.

Le interazioni tra dipoli si svolgono in modo analogo: la parte positiva di un dipolo attrae la parte negativa di altri dipoli, siano questi permanenti o indotti. Generalmente vengono indicate come **legami di Van der Waals** le interazioni *dipolo-dipolo* e *dipolo-dipolo indotto*, mentre le *interazioni dipolo indotto-dipolo indotto* vengono classificate come **forze di London**, responsabili delle forze coesive presenti nella materia. Queste forze sono originate dal fatto che in tutte le molecole si verificano fluttuazioni istantanee della carica, che originano dipoli istantanei; questi, a loro volta, inducono fluttuazioni e dipoli istantanei nelle molecole vicine. Nelle molecole apolari queste fluttuazioni mediate nel tempo hanno risultante nulla; tuttavia, l'esistenza di dipoli istantanei comporta l'esistenza di forze coesive tra le molecole, responsabili delle proprietà fisiche della materia. Sia le forze di Van der Waals sia quelle di London sono *forze a corto raggio*, poiché si sentono solo a breve distanza, decadendo come $1/r^6$.

9 Il legame idrogeno

Un discorso particolare spetta al legame dipolare in cui sia coinvolto l'atomo di idrogeno. Quando forma legami covalenti con elementi più elettronegativi, quest'ultimo, come tutti gli altri elementi, risulta impoverito di carica negativa e costituisce, quindi, la parte positiva di un dipolo elettrico. Tuttavia, il campo elettrico prodotto

dall'atomo di idrogeno, date le sue piccole dimensioni, è nettamente più alto rispetto a quello generato da tutti gli altri atomi: infatti l'idrogeno, privato del suo elettrone esterno, si riduce praticamente ad un protone. Ciò dà luogo a forti interazioni dipolari, indicate col nome di **legame idrogeno**.

La formazione di legami idrogeno si osserva in molecole come HF, H₂O, NH₃, in cui l'idrogeno può legare per attrazione elettrostatica un altro atomo elettronegativo di una molecola uguale o diversa. La forza di questo legame è tanto più elevata quanto maggiore è la differenza di elettronegatività tra l'idrogeno e l'atomo a cui è legato, ossia quanto maggiore è il suo grado di *protonazione*.

Legami idrogeno si incontrano non solo a livello *intermolecolare* (tra molecole diverse), come, ad esempio, tra le molecole di acqua allo stato solido, ma anche a livello *intramolecolare* (tra gruppi di atomi appartenenti alla stessa

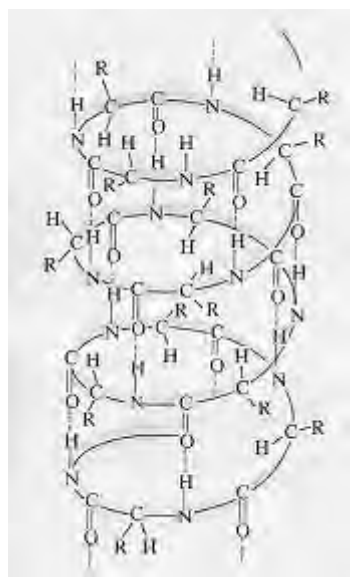


Fig. 17

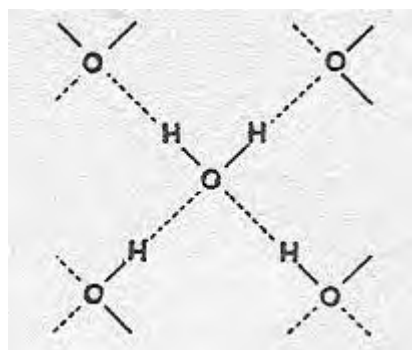


Fig. 16

molecola); ciò è estremamente frequente nelle macromolecole biologiche, ad esempio nelle proteine (Fig. 17).

10 Il legame metallico

Quello **metallico** è un tipo di legame del tutto particolare che si osserva esclusivamente tra atomi a carattere metallico (i 3/4 degli elementi naturali) ed esula dalle tipologie di legame osservate finora. Metalli come Li o Na, che possiedono un solo elettrone esterno su un orbitale di tipo s, mostrano scarsissima tendenza ad associarsi in molecole biatomiche, mentre formano reticoli cristallini compatti, in cui ogni atomo è circondato da 8 o 12 atomi. Ciascuno di essi mette



Fig. 18

a disposizione il proprio elettrone, che va a «delocalizzarsi» sull'intero cristallo: il risultato è un reticolo di ioni positivi immersi in un mare di elettroni.

Questa struttura spiega molte delle proprietà dei metalli. La conducibilità elettrica dei metalli è riconducibile all'estrema mobilità degli elettroni nel reticolo metallico, conferitagli dalla *delocalizzazione* su orbitali estesi fisicamente a tutto il cristallo.

La *malleabilità* e la *duttilità* (cioè il poter modificare la forma del metallo fino a costituirne dei fili) sono dovute al fatto che, nel reticolo metallico, gli ioni positivi sono tutti identici tra loro e, quindi, i legami non sono direzionali. Ciò consente di far scorrere i piani reticolari uno rispetto all'altro senza che il metallo perda la propria integrità. Un'azione analoga su un cristallo ionico ne provocherebbe immediatamente la frattura, poiché un minimo spostamento dei piani reticolari è sufficiente a portare affacciati gli uni agli altri ioni di carica uguale. Le forze repulsive che si originano determinano la frattura del cristallo; l'estrema fragilità dei cristalli ionici si deve proprio a questo fenomeno.

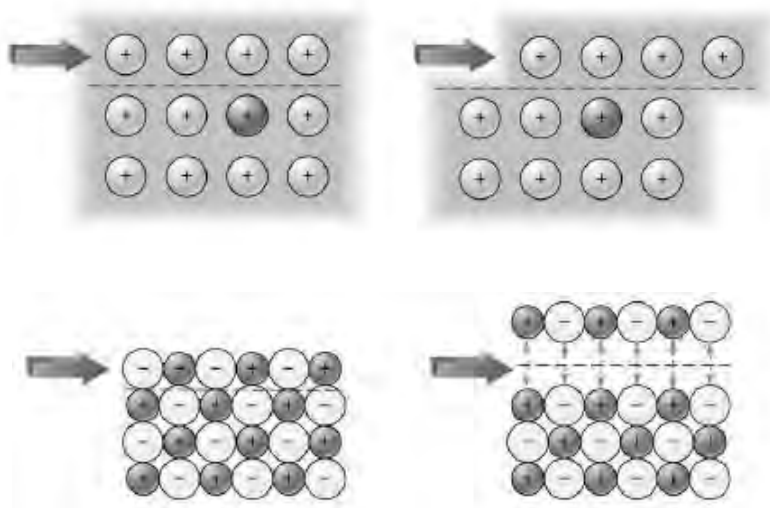


Fig. 19

Composti chimici e loro nomenclatura

1 Numero di ossidazione e valenza

Quando un atomo si lega ad altri atomi per formare una molecola modifica la sua struttura elettronica esterna cedendo, acquistando o mettendo in comune elettroni con gli altri. Gli elettroni interessati da questo processo sono detti **elettroni di valenza**. La **valenza** di un elemento in un composto *indica il numero di elettroni che questo impegna nel legame*. Abbiamo visto come il legame chimico possa essere di varie nature: nel caso dei legami ionici la valenza è espressa dalla *carica dello ione* (per questo talvolta si parla di *elettrovalenza*); nei composti covalenti, invece, ogni atomo ha una valenza coincidente col numero totale di elettroni messi in comune nei legami. Comunque, la valenza di un elemento è sempre un numero positivo.

In alcuni casi può risultare più utile parlare di **numero di ossidazione** anziché di valenza. Il numero di ossidazione di un elemento in un composto è la carica formale che questo assumerebbe supponendo di assegnare gli elettroni di ogni legame covalente all'atomo più elettronegativo.

Prendiamo a titolo di esempio la molecola d'acqua ($\text{H}-\text{O}-\text{H}$): tra H e O esiste un legame covalente eteropolare in cui ciascuno dei due atomi mette in comune con l'altro un elettrone. In termini di valenza diremo che ogni atomo di idrogeno ha valenza 1 mentre l'atomo di ossigeno, che forma due legami covalenti, ha valenza 2. In termini di numero di ossidazione, invece, essendo l'idrogeno meno elettronegativo dell'ossigeno, dovremmo assegnare l'elettrone di legame dell'idrogeno all'atomo di ossigeno (operazione puramente formale); gli idrogeni possiedono così numero di ossidazione +1 mentre l'ossigeno, avendo formalmente acquistato due elettroni, ha numero di ossidazione -2.

Nel caso degli ioni il numero di ossidazione coincide con la carica dello ione. Quindi numero di ossidazione e valenza hanno valore coincidente in modulo ma, mentre la valenza è rappresentata sempre da un numero positivo, il numero di ossidazione può assumere valori positivi o negativi.

Per calcolare il numero di ossidazione di un elemento in un composto il metodo migliore è costruirne la formula di struttura ed assegnare gli elettroni di ogni legame all'atomo più elettronegativo.

Il numero di ossidazione, però, può essere dedotto direttamente anche dalla formula bruta, tenendo presenti le regole seguenti:

- in una molecola, la somma dei numeri di ossidazione degli atomi che la costituiscono è sempre zero;
- in uno ione poliatomico, la somma dei numeri di ossidazione degli atomi che lo costituiscono è pari alla carica dello ione;
- ogni legame con un atomo più elettronegativo dà un contributo al numero di ossidazione dell'elemento pari a +1 (l'elettrone viene considerato ceduto);
- ogni legame con un atomo meno elettronegativo dà un contributo al numero di ossidazione dell'elemento pari a -1 (l'elettrone viene considerato acquistato);
- ogni legame di un atomo con sé stesso dà contributo nullo al numero di ossidazione;
- atomi dello stesso elemento possono comparire nella medesima molecola con numeri di ossidazione diversi. In tal caso si considera come numero di ossidazione il numero di ossidazione medio.

Ci sono elementi che assumono quasi sempre lo stesso numero di ossidazione: l'idrogeno ha sempre numero di ossidazione +1, tranne che nel legame con i metalli, in cui ha valore -1; l'ossigeno ha numero di ossidazione -2 in tutti i composti ad eccezione dei perossidi che contengono il gruppo -O-O-, in cui il numero di ossidazione è -1; i metalli del I gruppo hanno sempre numero di ossidazione +1; i metalli del II gruppo hanno sempre numero di ossidazione +2. Noti i valori dei numeri di ossidazione di tutti gli atomi di una molecola ad esclusione di uno, essendo sempre nulla la loro somma in una molecola neutra, quest'ultimo può essere ricavato per differenza.

ESEMPIOIn H_2SO_4 :

- ciascuno dei 2 atomi di H ha n.o. +1 [totale = $2 \cdot (+1) = +2$];
- ciascuno dei 4 atomi di O ha n.o. -2 [totale = $4 \cdot (-2) = -8$].

Indicando con x il numero di ossidazione dello zolfo, dovrà essere verificata l'uguaglianza:

$$x + 2 - 8 = 0$$

Da cui:

$$x = +6$$

2 Le categorie di composti inorganici

Si dicono **binari** i composti costituiti soltanto da due elementi; questi possono essere due elementi metallici, due elementi non-metallici, un metallo e un non-metallo, oppure un metallo o un non-metallo combinati con l'idrogeno o l'ossigeno. I nomi delle categorie di composti che ne derivano sono indicati nella tabella seguente:

	Metallo	Non-metallo	Idrogeno	Ossigeno
Metallo	leghe metalliche	sali ionici	idruri metallici	ossidi basici
				perossidi
Non-metallo	sali ionici	molecole inter-non-metalliche	idracidi	anidridi (ossidi acidi)
			idruri non-metallici	

Tab. 1

I **composti ternari** o **quaternari** sono composti contenenti tre o quattro elementi chimici. Composti contenenti un elemento metallico o non-metallico accompagnato da idrogeno ed ossigeno possono essere ottenuti, rispettivamente, per reazione di ossidi o anidridi con acqua.

Gli ossidi basici in acqua formano idrossidi, le anidridi o ossidi acidi originano gli ossiacidi. Come vedremo queste categorie di composti hanno comportamenti completamente diversi in soluzione acquosa, poiché i primi hanno carattere basico mentre i secondi hanno carattere acido.

La reazione tra un composto a carattere acido ed uno a carattere basico produce un sale con eliminazione di una o più molecole di acqua. Il sale derivato dalla reazione di un ossiacido con un idrossido può essere neutro, acido o basico. I sali acidi contengono ancora un residuo di atomi di idrogeno: ad esempio, NaHCO_3 , il bicarbonato di sodio, è un sale acido. I sali basici hanno tra i controioni del metallo uno o più gruppi OH: ad esempio, $\text{Fe}(\text{OH})\text{Cl}_2$, l'idrossidicloruro di ferro (III), è un sale basico.

3 Formule e nomenclatura

Qualsiasi composto chimico viene descritto attraverso una *formula bruta*, che ne definisce la composizione qualitativa e quantitativa, ed indicato con un *nome*. La IUPAC (International Union for Pure and Applied Chemistry) ha stabilito delle **regole di nomenclatura** affinché, attraverso il nome del composto, si possa risalire univocamente alla sua struttura. Le regole di nomenclatura IUPAC sono trattate qui di seguito.

a) Nomenclatura di composti binari

Si dicono **binari** i composti costituiti da due elementi chimici. Nella formula, l'elemento meno elettronegativo deve precedere quello più elettronegativo (ad esempio: H_2O , NaCl , H_2S , LiH ...). Il nome viene assegnato facendo seguire il suffisso **-uro** al nome del secondo elemento e lasciando inalterato il nome del primo. Se il secondo elemento è l'ossigeno, si usa il nome di **ossido** se il suo numero di ossidazione è -2 , **perossido** se è -1 . Se è presente più di un atomo dello stesso elemento, viene utilizzato il prefisso *mono-* (spesso omissso), *di-*, *tri-*, *tetra-* etc. oppure viene indicato tra parentesi il numero di ossidazione dell'elemento mediante un numero romano.

H_2O	ossido di idrogeno
H_2S	solfuro di idrogeno
NaCl	cloruro di sodio
LiH	idruro di litio
FeCl_2	cloruro di ferro (II)
FeCl_3	cloruro di ferro (III)
CO	monossido di carbonio; ossido di carbonio (II)
CO_2	diossido di carbonio; ossido di carbonio (IV)

b) Nomenclatura di ioni

Per i *cationi monoatomici* si usa semplicemente il nome dell'elemento seguito da un numero romano tra parentesi, che ne indica la valenza. Alcuni *cationi poliatomici* conservano nomi particolari, come lo ione **ossonio** H_3O^+ , lo ione **ammonio** NH_4^+ etc.

Cu^+	ione rame (I)
Cu^{2+}	ione rame (II)
Ca^{2+}	ione calcio (II)

Gli *anioni monoatomici* vengono indicati aggiungendo il suffisso **-uro** al nome dell'elemento. Per gli *anioni poliatomici* si attribuisce il suffisso **-ato** al nome dell'atomo centrale, di cui va indicato il numero di ossidazione tra parentesi, specificando il tipo e il numero degli atomi legati all'atomo centrale; ad esempio: PF_2O_2^- è lo ione difluorodiossifosfato (V).

Cl^-	ione cloruro	NO_3^-	ione nitrato (V)
SO_4^{2-}	ione solfato (VI)	NO_2^-	ione nitrato (III)
SO_3^{2-}	ione solfato (IV)	ClO^-	ione clorato (I)
SFO_3^-	ione fluorotriosolfato	ClO_3^-	ione clorato (V)
MnO_4^-	ione manganato (VII)	ClO_2^-	ione clorato (III)
MnO_4^{2-}	ione manganato (VI)	ClO_4^-	ione clorato (VII)

c) Nomenclatura di sali

I **sali** sono composti formati dall'associazione di uno o più **cationi** con uno o più **anioni**. Nella **formula**, i **cationi** sono posti sempre prima degli **anioni**. Se nel sale è contenuto più di un tipo di catione o di anione, questi vengono scritti in ordine di valenza o, a parità di valenza, in ordine alfabetico.

Il **nome**, invece, è composto dal nome dell'anione, a cui si fa seguire quello del catione, indicandone il numero, dove necessario, attraverso i prefissi *mono-*, *di-*, *tri-* etc. Per gli anioni idrogenati, si fa precedere al nome dell'anione l'espressione «monoidrogeno», «diidrogeno» etc., a seconda del numero di idrogeni che contiene. Per quanto concerne i sali basici, cioè contenenti ioni O^{2-} oppure OH^- , nella **formula bruta** questi anioni vengono scritti subito dopo il catione, mentre nel **nome** si inserisce tra l'anione e il catione la dicitura *ossido* o *idrossido*; ad esempio: Cu_2OCl_2 è il cloruro ossido di rame (II).

$BaSO_4$	solfato (VI) di bario	KHF_2	idrogenofluoruro di potassio
$BaSO_3$	solfato (IV) di bario	$NaTi(NO_3)_2$	bisnitrato di sodio e tallio
$NaHCO_3$	idrogenocarbonato di sodio	$KAl(SO_3)_2$	bissolfato (IV) di potassio e alluminio
NH_4HCO_3	idrogenocarbonato di ammonio	$Ca_5F(PO_4)_3$	fluoruro trisfosfato (V) di pentacalcio

d) Nomenclatura di acidi

Nella nomenclatura IUPAC, il termine **acido** è utilizzato solo per specie in soluzione acquosa, mentre le stesse specie allo stato puro dovrebbero essere indicate utilizzando le regole di nomenclatura viste per i sali. Tuttavia, essendo estremamente radicata l'abitudine all'uso dei nomi tradizionali, per le specie acide la IUPAC considera accettabili anche i nomi attribuiti dalla nomenclatura tradizionale.

Così gli **idracidi**, acidi binari costituiti dall'idrogeno e da un elemento non-metallico molto elettronegativo, vengono indicati facendo seguire il suffisso *-idrico* al nome del non metallo (HCl = acido cloridrico, HF = acido fluoridrico, HCN = acido cianidrico, H_2S = acido solfidrico). Per gli **ossiacidi**, ossia specie acide contenenti idrogeno, un non-metallo e l'ossigeno, il nome viene attribuito facendo seguire al nome del non-metallo il suffisso *-ico* se il non-metallo assume la sua valenza più alta, *-oso* se assume la sua valenza più bassa. Se il non-metallo può assumere più di due valenze si utilizzano prefissi associati ai suffissi; in ordine di valenza crescente, si hanno quindi gli acidi *ipo-...-oso*, *...-oso*, *...-ico*, *per-...-ico* (ad esempio: $HClO$ = ac. ipocloroso, $HClO_2$ = ac. cloroso, $HClO_3$ = ac. clorico, $HClO_4$ = ac. perclorico).

La nomenclatura tradizionale usa per gli acidi anche i prefissi *orto-*, *meta-*, *piro-*, che vengono utilizzati per distinguere acidi in cui l'atomo centrale assume sempre la stessa valenza e che si distinguono tra loro per il numero di molecole d'acqua che contengono. Il prefisso *orto-*, in genere, viene omissso e si utilizza per indicare l'ossiacido con il numero più alto di gruppi OH ; il prefisso *meta-* indica l'acido ottenuto dall'*orto-* per eliminazione di una molecola d'acqua; il prefisso *piro-* indica l'acido ottenuto da due molecole di *orto-* per eliminazione di una molecola d'acqua (ad esempio: acido *orto*-fosforico = H_3PO_4 , acido metafosforico = $H_3PO_4 - H_2O = HPO_3$, acido pirofosforico o difosforico = $2 H_3PO_4 - H_2O = H_4P_2O_7$). La IUPAC raccomanda unicamente di usare il prefisso *di-* in sostituzione di *piro-*.

Formula	Nomenclatura IUPAC	Nomenclatura tradizionale
H_2CO_3	carbonato (VI) di diidrogeno	acido carbonico
HCN	cianuro di idrogeno	acido cianidrico
HNO_3	nitrato (V) di idrogeno	acido nitrico
HNO_2	nitrato (III) di idrogeno	acido nitroso
H_3PO_4	fosfato(V) di triidrogeno	acido fosforico
H_3AsO_4	arseniato(V) di triidrogeno	acido arsenico
H_3AsO_3	arseniato(III) di triidrogeno	acido arsenioso
H_2SO_4	solfato (VI) di diidrogeno	acido solforico
H_2SO_3	solfato (IV) di diidrogeno	acido solforoso
H_2CrO_4	cromato (VI) di diidrogeno	acido cromico
$\text{H}_2\text{Cr}_2\text{O}_7$	dicromato (VI) di diidrogeno	acido dicromico
HClO	clorato (I) di idrogeno	acido ipocloroso
HClO_2	clorato (III) di idrogeno	acido cloroso
HClO_3	clorato (V) di idrogeno	acido clorico
HClO_4	clorato (VII) di idrogeno	acido perclorico
H_2MnO_3	manganato (IV) di diidrogeno	acido manganoso
H_2MnO_4	manganato (VI) di diidrogeno	acido manganico
HMnO_4	manganato (VII) di diidrogeno	acido permanganico
H_5IO_3	iodato (VII) di pentaidrogeno	acido ortoperiodico
HIO_4	iodato (VII) di idrogeno	acido periodico

Tab. 2

I **perossiacidi** sono acidi che contengono il gruppo perossido -O-O- e vengono chiamati semplicemente facendo precedere al nome dell'acido il prefisso *perossi-* (ad esempio: H_2SO_5 = ac. perossisolforico, $\text{H}_2\text{S}_2\text{O}_8$ = ac. perossidisolforico etc.). I **tioacidi** sono acidi derivati dai corrispondenti acidi ossigenati per sostituzione di uno o più atomi di ossigeno con atomi di zolfo e vengono chiamati facendo precedere al nome dell'acido i prefissi *tio-*, *ditio-*, *tritio-*, a seconda del numero di atomi di zolfo presenti (ad esempio: $\text{H}_2\text{S}_2\text{O}_3$ = ac. tiosolforico, $\text{H}_3\text{AsS}_2\text{O}_2$ = ac. ditioarsenico, H_2CS_3 = ac. tritio carbonico).

Stati di aggregazione e passaggi di stato

1 Lo stato solido

Dal punto di vista macroscopico un **solido** è definito come qualsiasi porzione di materia dotata di volume e forma propri. Dal punto di vista microscopico il solido si distingue dagli altri stati di aggregazione poiché possiede una rigida ed ordinata organizzazione interna delle sue particelle. Lo studio attraverso la diffrazione di raggi x ha messo in evidenza la presenza all'interno del solido di un **reticolo cristallino** risultante dalla ripetizione nelle tre dimensioni dello spazio di una **cella elementare**, cioè un'unità ripetitiva sempre uguale a se stessa. Le particelle che costituiscono il solido occupano definite posizioni, dette nodi reticolari, e generalmente non sono libere di spostarsi le une rispetto alle altre; tuttavia, ognuna di esse oscilla intorno alla propria posizione di equilibrio in misura tanto maggiore quanto più alta è la temperatura. Questo moto, infatti, è comunemente indicato col nome di agitazione termica delle particelle.

La **fusione del solido**, ovvero il suo passaggio allo stato liquido, si verifica quando l'energia delle oscillazioni dovute all'agitazione termica supera l'energia di legame delle particelle nel reticolo. Per questo la fusione di un solido richiede che se ne innalzi la temperatura.

Esistono sostanze che possiedono le proprietà macroscopiche dei solidi pur essendo prive della struttura interna ordinata tipica del solido: si tratta dei cosiddetti **solidi amorfi**, che possiedono una rete tridimensionale irregolare e disordinata; l'esempio più comune è il vetro.

A seconda della natura dei legami chimici che mantengono la struttura reticolare, distinguiamo solidi covalenti, ionici, molecolari e metallici.

Nei **solidi covalenti** i nodi reticolari sono occupati da atomi legati tra loro mediante legami covalenti; poiché l'energia del legame covalente è molto alta, i materiali covalenti sono estremamente duri ed hanno temperature di fusione molto alte (ne è un esempio il diamante).

Nei **solidi ionici**, invece, i nodi reticolari sono occupati da ioni mantenuti nelle loro posizioni di equilibrio da forze elettrostatiche attrattive e repulsive. Anche i legami ionici hanno energie elevate e quindi i solidi ionici mostrano temperature di fusione abbastanza alte; tuttavia, la natura rigorosamente direzionale del legame conferisce notevole fragilità a questi materiali che si fratturano per piccoli spostamenti di un piano reticolare rispetto all'altro.

Nei **solidi molecolari** i nodi reticolari sono occupati da molecole legate tra loro da interazioni deboli (legami dipolari, legami idrogeno), le cui energie sono nettamente inferiori rispetto a quelle di legami ionici e covalenti. Di conseguenza, questi solidi hanno temperature di fusione basse e scarsa durezza.

I **solidi metallici** sono costituiti da reticoli di ioni positivi in cui ogni atomo delocalizza i suoi elettroni di valenza in orbitali estesi a tutto il cristallo: gli elettroni possono migrare da un punto all'altro del reticolo e, quindi, i metalli sono buoni conduttori di corrente elettrica. Il legame non è direzionale e ciò consente ai piani reticolari di scivolare gli uni sugli altri; ciò conferisce ai metalli malleabilità e duttilità. Le energie reticolari variano molto a seconda del tipo di metallo, quindi le temperature di fusione assumono un'ampia gamma di valori.

2 Lo stato liquido

Tra le particelle che costituiscono il liquido esistono forze di interazione di entità insufficiente a mantenere un ordine rigoroso come nei cristalli, ma tali da non consentirne l'allontanamento. Lo studio tramite la diffrazione di raggi x ha evidenziato che nei liquidi esistono degli aggregati di particelle simili a piccole «isole cristalline» immerse in un mare di molecole in moto caotico. I cristalli liquidi rappresentano un caso limite di sostanze liquide in cui queste strutture cristalline embrionali sono sufficientemente definite da conferire al liquido delle particolari proprietà.

La natura e l'entità delle forze che legano le particelle di un liquido dipendono dalla sua natura chimica; anche per i liquidi possiamo effettuare la stessa distinzione vista per i solidi in liquidi covalenti, ionici, molecolari e metallici. Queste forze sono alla base delle forze coesive che tendono a mantenere unite le particelle del liquido originando alcune delle proprietà dei liquidi che ci accingiamo ad esaminare.

a) Tensione superficiale

Ogni particella nel corpo del liquido subisce forze di attrazione di uguale entità da parte delle altre particelle in tutte le direzioni dello spazio; la risultante di queste forze è nulla all'interno del liquido mentre per le particelle di superficie le forze attrattive hanno una risultante diretta verso il corpo del liquido, poiché le forze esercitate dalle particelle dell'aria sovrastante sono meno numerose.

Ciò determina la naturale tendenza della superficie dei liquidi a contrarsi, che trova la sua manifestazione più evidente nella sfericità delle gocce: ogni particella della superficie sferica subisce una forza attrattiva diretta verso il centro della sfera stessa.

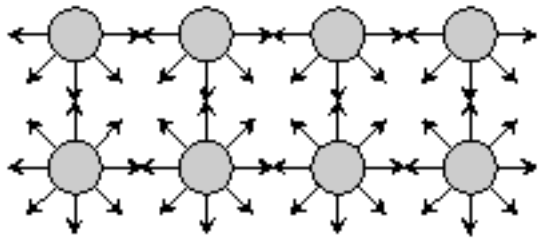


Fig. 1

b) Viscosità

Un'altra proprietà conferita ai liquidi dalle forze di interazione esistenti tra le particelle è la viscosità, che rappresenta la maggiore o minore facilità con cui uno strato di liquido scorre sull'altro. Maggiore è l'entità delle interazioni tra le molecole, maggiore è la viscosità. La viscosità cresce con le dimensioni molecolari, poiché aumenta l'entità delle forze di dispersione e di Van der Waals e varia a seconda del tipo di interazioni: molecole in grado di dare interazioni ad alta energia, come ad esempio i legami idrogeno, hanno viscosità maggiore delle altre a parità di dimensioni molecolari. Normalmente la viscosità diminuisce con la temperatura, poiché al crescere dell'agitazione termica delle particelle diventa più facile vincere le forze di interazione e dar luogo allo scorrimento.

c) Pressione di vapore

Ogni sostanza manifesta una certa volatilità, ossia una certa tendenza a passare alla fase gassosa. Tutte le sostanze di cui percepiamo l'odore sono caratterizzate da una discreta volatilità e le particelle che passano in fase gassosa stimolano i nostri recettori olfattivi, «producendo l'odore» della sostanza.

La volatilità delle sostanze viene quantificata attraverso una grandezza detta *pressione di vapore*. Consideriamo un recipiente chiuso parzialmente riempito con del liquido mantenuto alla temperatura T ; un certo numero delle particelle di superficie, dotate di agitazione termica

sufficiente a vincere l'attrazione delle altre particelle, abbandonano la fase liquida per portarsi nella fase vapore sovrastante. Contemporaneamente, alcune particelle della fase vapore ricadono sulla superficie e tornano in fase liquida. Quando il numero di particelle che passano in fase vapore eguaglia il numero delle particelle che tornano in fase liquida, il liquido è in equilibrio col suo vapore e la pressione del vapore in equilibrio con la fase liquida viene definita **pressione di vapore alla temperatura T considerata**.

Ogni sostanza possiede uno specifico valore della pressione di vapore ad ogni temperatura. La pressione di vapore cresce al crescere della temperatura poiché, aumentando l'agitazione termica delle particelle, è sempre maggiore la frazione di queste che riesce a passare in fase vapore e si mantiene costante a temperatura costante. Va sottolineato, però, che quello liquido-vapore è un equilibrio dinamico: ciò che resta costante nel tempo è il numero di particelle presenti in fase vapore; tuttavia, incessantemente, un certo numero di esse ricade in fase liquida ed un uguale numero abbandona la fase liquida per passare in fase vapore.

3 Lo stato gassoso

Abbiamo visto come quello gassoso rappresenti lo stato più rarefatto e disordinato della materia, tanto da non poter più riconoscere per quest'ultima una forma ed un volume propri. Tali caratteristiche sono dovute al fatto che le particelle che costituiscono il gas non sono in contatto tra loro e sono animate da un moto caotico e casuale che le porta ad occupare omogeneamente tutto lo spazio a loro disposizione. Lo stato di un gas viene definito attraverso tre parametri fondamentali: pressione, volume e temperatura.

La **pressione** di un gas esprime la forza esercitata dal gas sull'unità di superficie e nel Sistema Internazionale si misura in Pascal (Pa); un'unità ancora molto utilizzata è l'atmosfera (atm) o il millimetro di mercurio (mm Hg). I fattori di conversione tra queste unità sono: $1 \text{ atm} = 760,00 \text{ mm Hg} = 101325 \text{ Pa}$.

Se consideriamo una certa massa di gas racchiusa in un recipiente, la pressione esercitata sulle pareti può essere considerata, dal punto di vista microscopico, come il risultato dei continui urti delle particelle del gas animate da un moto casuale e dotate di un'energia cinetica crescente al crescere della temperatura.

Il **volume** è lo spazio occupato dal gas e nel Sistema Internazionale viene misurato in m^3 . In chimica generalmente vengono utilizzati il litro ($1 \text{ m}^3 = 10^3 \text{ l}$) ed i suoi sottomultipli.

La **temperatura**, come ci insegna la fisica, è una misura della tendenza del calore a sfuggire da un corpo: il trasferimento di calore tra due corpi può aver luogo solo nel momento in cui tra questi esista una differenza di temperatura. Si osserverà, allora, il passaggio di una certa quantità di calore dal corpo più caldo a quello più freddo fino al raggiungimento dell'equilibrio termico, ossia di un'uguale temperatura per entrambi. La temperatura viene misurata nel Sistema Internazionale in gradi Kelvin (K). La definizione di questa scala è stata realizzata in modo da far coincidere l'intervallo di temperatura corrispondente al grado centigrado ($^{\circ}\text{C}$) con quello corrispondente al grado Kelvin e così la scala Kelvin risulta soltanto traslata rispetto alla scala centigrada: le temperature di congelamento ed ebollizione dell'acqua in questa scala sono, rispettivamente, $273,15 \text{ K}$ e $373,15 \text{ K}$.

4 Il modello del gas ideale

La **teoria cinetica dei gas** fornisce un modello interpretativo del comportamento dei gas in termini puramente meccanici, senza fare alcun riferimento alla natura chimica delle molecole che lo costituiscono. Il modello del gas ideale, proposto dalla teoria cinetica, si adatta bene a gas reali in condizioni di pressione non troppo elevata e di temperatura sufficientemente alta.

In questo modello si assume che:

- le molecole del gas sono in moto perenne e caotico;
- nel loro moto le molecole interagiscono solo tramite urti elastici;
- il volume delle molecole è trascurabile rispetto al volume fisicamente occupato dal gas;
- tra le molecole di gas non esistono forze di interazione.

Ovviamente non sempre queste ipotesi sono verificate del tutto per il gas reale; tuttavia, a bassa pressione, il numero di particelle di gas per unità di volume è esiguo e ciò, da un lato, minimizza il numero di interazioni, dall'altro rende il volume proprio delle particelle trascurabile rispetto al volume del gas.

a) Legge di Maxwell-Boltzmann

Il moto delle particelle del gas è descritto efficacemente dalla legge di Maxwell-Boltzmann che fornisce la distribuzione statistica dell'energia cinetica delle molecole di un gas in condizioni stazionarie, cioè in un recipiente chiuso e a temperatura costante. Durante il loro moto caotico ed incessante le particelle del gas danno luogo ad urti elastici che trasferiscono energia cinetica dall'una all'altra, ridistribuendo continuamente l'energia totale. La funzione di Maxwell-Boltzmann fornisce, per ogni fissato valore di temperatura, la probabilità che le particelle di gas possiedano un certo valore di energia cinetica (e quindi di velocità di traslazione). Le distribuzioni dell'ossigeno a -100 , 20 e 600°C vengono riportate in figura, dove le curve mostrano come, al crescere della temperatura, la velocità corrispondente al

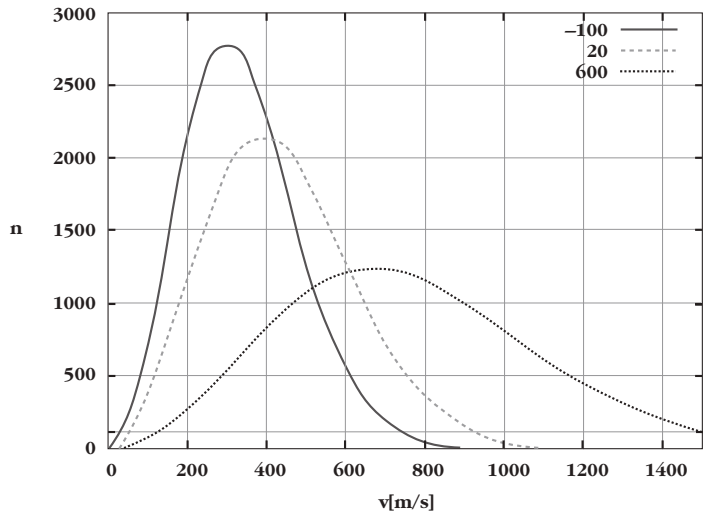


Fig. 2

massimo della curva, cioè la velocità più probabile, assume valori progressivamente più alti. Ciò significa che maggiore è la temperatura del gas, maggiore è la velocità di traslazione delle sue molecole; in tal modo abbiamo stabilito una correlazione tra una proprietà macroscopica del gas, la **temperatura**, e una proprietà microscopica, l'**energia cinetica** delle sue particelle. La meccanica statistica dimostra la validità della distribuzione di Maxwell-Boltzmann anche per gas reali, liquidi e solidi.

b) Leggi di Boyle, di Charles e di Gay-Lussac

Abbiamo visto come lo stato di un gas possa essere definito attraverso i valori delle tre variabili pressione (P), volume (V) e temperatura (T). Queste variabili non sono indipendenti tra loro e le leggi empiriche di Boyle, Charles e Gay-Lussac esprimono la relazione esistente tra due di queste variabili allorché la terza venga mantenuta costante.

La **legge di Boyle**, detta *isoterma*, esprime la relazione esistente tra pressione e volume a temperatura costante ed afferma, appunto, che:

«Il prodotto di pressione e volume è costante a temperatura costante».

$$PV = \text{cost} \quad \text{a} \quad T = \text{cost}$$

Quindi, se manteniamo costante la temperatura del gas, una diminuzione del volume comporta un aumento di pressione e, viceversa, un aumento del volume comporta una diminuzione di pressione. Ciò è abbastanza intuitivo, poiché, a temperatura costante, è costante la velocità media di traslazione delle particelle; chiaramente, minore è il volume a disposizione, maggiore il numero di urti delle particelle sulle pareti del recipiente e, dunque, la pressione. Viceversa, aumentando il volume a disposizione delle particelle, il numero di urti nell'unità di tempo diminuisce.

La **legge di Charles**, o **prima legge di Gay-Lussac**, o anche **prima legge di Volta e Gay-Lussac**, detta *isobara*, esprime una relazione di diretta proporzionalità tra volume e temperatura, espressa in gradi K, a pressione costante, e dunque:

«A pressione costante, volume e temperatura di un gas sono in rapporto costante tra loro».

$$\frac{V}{T} = \text{cost} \quad \text{a} \quad P = \text{cost}$$

Anche in questo caso l'interpretazione microscopica ci permette di giustificare tale relazione: affinché una massa di gas eserciti sulle pareti del suo recipiente sempre la stessa pressione, all'aumentare della temperatura, e quindi della velocità di traslazione delle molecole, deve corrispondere un proporzionale aumento dello spazio a loro disposizione.

La **seconda legge di Gay-Lussac**, o **seconda legge di Volta e Gay-Lussac**, o anche **legge di Amontons**, detta *isocora*, stabilisce una proporzionalità diretta tra pressione e temperatura, espressa in gradi K, a volume costante, e pertanto:

«A volume costante, pressione e temperatura di un gas sono in rapporto costante tra loro».

$$\frac{P}{T} = \text{cost} \quad \text{a} \quad V = \text{cost}$$

Anche questa legge trova un'immediata spiegazione alla luce della natura microscopica della pressione e della temperatura: mantenendo costante il volume a disposizione delle particelle di gas, al crescere della temperatura cresce la velocità di traslazione delle molecole ed il numero di urti sulle pareti cresce proporzionalmente.

c) Equazione di stato del gas perfetto

Le tre relazioni empiriche precedenti convergono in un'unica equazione, detta **equazione di stato del gas perfetto**, così chiamata perché è rigorosamente valida solo per il gas ideale.

Per n moli di gas ideale l'equazione di stato si esprime come:

$$PV = nRT$$

dove R rappresenta la costante universale dei gas. Il valore che R assume dipende dalle unità di misura in cui vengono espressi volume, temperatura e pressione. Generalmente si usa esprimere la pressione in atmosfere, la temperatura in gradi Kelvin e il volume in litri; per cui:

$$R = 0,0821 \text{ l} \cdot \text{atm} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$$

I gas reali obbediscono a questa legge in modo tanto più soddisfacente quanto minore è la pressione e quanto più alta è la temperatura: in queste condizioni, infatti, le interazioni tra le molecole risultano minimizzate e il volume proprio delle particelle è effettivamente trascurabile rispetto a quello fisicamente occupato dal gas. In base all'equazione di stato possiamo prevedere, per fissate condizioni di due delle tre variabili, quale valore assume la terza. Vengono definite **condizioni normali** per un gas la temperatura di 0°C e la pressione di 1 atm; in queste condizioni, il volume di una mole di qualsiasi gas a comportamento ideale è sempre lo stesso a prescindere dalla sua natura chimica e risulta pari a 22,414 l.

d) Legge di Avogadro

Amedeo Avogadro, studiando i gas, aveva enunciato un principio, noto oggi come legge di Avogadro, in base al quale:

«Volumi uguali di gas diversi, misurati nelle stesse condizioni di temperatura e pressione, contengono lo stesso numero di molecole».

Questo principio, rigorosamente valido solo per i gas a comportamento ideale, è implicito nell'equazione di stato. Infatti, se scriviamo l'equazione di stato per due gas, che distinguiamo con i pedici 1 e 2, si ha:

$$P_1 V_1 = n_1 R T_1 \quad \text{e} \quad P_2 V_2 = n_2 R T_2$$

se i gas sono in uguali condizioni di temperatura e pressione, ovvero se $P_1 = P_2$ e $T_1 = T_2$. Se sono uguali anche i volumi e quindi $V_1 = V_2$, affinché siano verificate le uguaglianze imposte dall'equazione di stato, devono necessariamente essere uguali anche il numero di moli $n_1 = n_2$.

La determinazione del numero di particelle per un volume noto di gas, in fissate condizioni di temperatura e pressione, può quindi essere estesa a qualsiasi altro gas a comportamento ideale. Inoltre, noto il volume occupato da 1 mole di gas, si può risalire al numero di molecole contenute in una mole, ossia al **numero di Avogadro** N_A (che, ricordiamo, ha un valore pari a $6,022046 \cdot 10^{23}$ molecole).

e) Legge di Dalton

Può essere enunciata come segue:

«La pressione totale di una miscela gassosa è data dalla somma delle pressioni parziali di ciascun componente».

$$P_{\text{tot}} = p_1 + p_2 + \dots + p_n$$

«La pressione parziale p_i del singolo componente può essere considerata come la pressione che questo eserciterebbe se occupasse da solo tutto il volume della miscela ed è pari al prodotto della sua frazione molare per la pressione totale».

$$p_i = x_i P_{\text{tot}}$$

Questa legge può essere spiegata alla luce dell'equazione di stato del gas perfetto che può essere applicata a qualsiasi gas, a prescindere dalla sua natura chimica; infatti ad una data temperatura e per un volume fissato, la pressione esercitata dal gas dipende unicamente dal numero di particelle presenti e non dalla loro natura chimica.

L'equazione di stato può essere utilizzata quindi per descrivere il comportamento di miscele gassose a comportamento ideale, ossia miscele di gas costituite da componenti di natura chimica

diversa. Se una miscela mostra comportamento ideale, evidentemente ad ognuno dei suoi n componenti può essere applicata l'equazione di stato e per ciascuno di essi si ha:

$$\begin{aligned} p_1 V &= n_1 RT \\ p_2 V &= n_2 RT \\ &\dots\dots\dots \\ p_n V &= n_n RT \end{aligned}$$

Sommando membro a membro queste espressioni otteniamo:

$$(p_1 + p_2 + \dots + p_n)V = (n_1 + n_2 + \dots + n_n)RT$$

La somma del numero di moli dei singoli componenti rappresenta il numero di moli totali che costituiscono la miscela. La pressione esercitata dalla miscela, indicata come pressione totale P_{tot} , è dovuta alla totalità delle particelle presenti ed è quindi la somma delle pressioni esercitate singolarmente da ciascuno dei componenti, dette pressioni parziali p_i . L'espressione precedente, dunque, non è altro che l'applicazione dell'equazione di stato dei gas all'intera miscela gassosa:

$$P_{\text{tot}} V = n_{\text{tot}} RT$$

Inoltre, facendo il rapporto membro a membro tra l'equazione di stato relativa al componente i -esimo e quella relativa all'intera miscela si ottiene:

$$\frac{p_i}{p_{\text{tot}}} = \frac{n_i}{n_{\text{tot}}} \text{ e quindi } p_i \frac{n_i}{n_{\text{tot}}} P_{\text{tot}} = \chi_i P_{\text{tot}}$$

dove il rapporto tra le moli del singolo componente e le moli totali n_i/n_{tot} è detto **frazione molare** ed è indicato generalmente con χ_i . Quindi la pressione parziale di ciascun componente è pari al prodotto della sua frazione molare per la pressione totale.

f) Legge di diffusione di Graham

Uno dei risultati a cui giunge la teoria cinetica dei gas è che esiste una proporzionalità diretta tra energia cinetica traslazionale media delle particelle e temperatura assoluta di un gas:

$$\frac{1}{2} m \bar{v}^2 = \frac{3}{2} kT$$

La legge di diffusione di Graham afferma che:

«Ad una determinata temperatura, due gas possiedono velocità di diffusione inversamente proporzionali ai loro pesi molecolari». In formule:

$$\frac{\bar{v}_2}{\bar{v}_1} = \frac{\sqrt{m_1}}{\sqrt{m_2}}$$

Quindi il gas più pesante si diffonde più lentamente di quello più leggero. Infatti, alla stessa temperatura, le particelle dei due gas possiedono la medesima energia cinetica media e pertanto:

$$\frac{1}{2} m_1 \cdot \bar{v}_1^2 = \frac{1}{2} m_2 \cdot \bar{v}_2^2 \text{ da cui: } \frac{m_1}{m_2} = \frac{\bar{v}_2^2}{\bar{v}_1^2} \text{ ossia: } \frac{\bar{v}_2}{\bar{v}_1} = \frac{\sqrt{m_1}}{\sqrt{m_2}}$$

g) Legge di Proust

Nel 1799, il chimico francese *Joseph-Louis Proust* notò che la composizione della *pirite*, un minerale molto abbondante in Spagna (Paese in cui si rifugiò durante la Rivoluzione francese), era sempre la stessa, indipendentemente dal luogo di provenienza; tale composizione era rispettata anche dalla pirite ottenuta in laboratorio. In particolare, determinò che la pirite era costituita da ferro e zolfo, presenti secondo un rapporto fisso e costante: per ogni grammo di ferro erano sempre presenti 0,57 grammi di zolfo. Lo studioso estese le sue ricerche anche ad altre sostanze e constatò che tale regolarità di comportamento nei rapporti di combinazione era comune anche agli altri minerali.

Queste considerazioni lo portarono a formulare la **legge delle proporzioni definite** (nota anche come **legge di Proust**), la quale può essere così enunciata:

«Quando due elementi si combinano per formare un composto, essi si uniscono secondo rapporti di peso fissi, che non possono essere variati a piacere».

La legge appare più chiara anche se si prende in esame una sostanza come l'*acqua*, la cui analisi era già stata effettuata dal chimico *Antoine-Laurent de Lavoisier*, il quale dimostrò che essa era costituita dall'85% di ossigeno e dal 15% di idrogeno. Ciò vuol dire che i due elementi si combinano rispettando un rapporto di 5,66:1 (i valori esatti sono, in realtà, 88,88% di ossigeno e 11,11% di idrogeno, ovvero un rapporto di 8:1). Questa proporzione, secondo Proust, non è modificabile, pertanto se volessimo far reagire quantità diverse di ossigeno e di idrogeno non otterremmo un'acqua diversa, più ricca dell'uno o dell'altro elemento, ma accadrebbe che quello dei due che risultasse in eccesso non entrerebbe nella combinazione. Ciò si verifica per tutti gli elementi e per qualunque combinazione, il che ci consente di affermare il principio sulla base del quale un elemento può combinarsi con un altro solo secondo «pesi» ben determinati.

Cerchiamo di comprendere meglio la *legge di Proust* utilizzando un semplice esempio. Immaginiamo di voler preparare in un laboratorio della pirite, un minerale composto, come abbiamo visto, da disolfuro di ferro (FeS_2). In tal caso, occorre far reagire ferro e zolfo nel rapporto di 1 g di ferro per 0,57 g di zolfo. Stando alla *legge di Lavoisier* (la quale afferma che «*in una reazione chimica il peso complessivo delle sostanze rimane inalterato*»), otterremo 1,57 g di disolfuro di ferro. Se vogliamo ottenere una quantità di pirite dieci volte maggiore, ossia 15,7 grammi, dobbiamo continuare a rispettare il rapporto di combinazione (1:0,57) tra ferro e zolfo, facendo reagire 10 g di ferro con 5,7 g di zolfo. Se, invece, mettiamo a reagire 10 g di ferro con 8 g di zolfo (una quantità superiore a quella richiesta dal *rapporto di combinazione*), otterremo sempre la medesima quantità di pirite (15,7 g), ma parte dello zolfo ($8 - 5,7 = 2,3$ g) non reagirà e rimarrà inalterato al termine del processo chimico.

5 L'equazione di Van der Waals

Il modello del gas ideale si adatta soddisfacentemente ai gas reali in condizioni di temperatura elevata, quando le particelle possiedono energia cinetica media sufficientemente alta da minimizzare gli effetti delle forze attrattive esistenti tra le particelle, e a basse pressioni, che riducono le interazioni tra le molecole e rendono trascurabile il loro volume proprio rispetto al volume fisicamente occupato dal gas. Al variare delle condizioni di temperatura e pressione si osserva per molti gas il graduale passaggio dal comportamento ideale al comportamento «reale», anche se in alcuni casi il gas si comporta da gas reale già a basse pressioni. La deviazione di un gas dal comportamento ideale può essere efficacemente quantificata attraverso il fattore di compressibilità del gas, ovvero il rapporto PV/RT , che vale esattamente 1 per una mole di gas ideale e devia dall'unità in misura tanto maggiore quanto più il gas in esame devia dal comportamento ideale.

Van der Waals ha riadattato l'equazione di stato del gas ideale perché riflettesse il comportamento del gas reale, introducendo due modifiche.

1. **Modifica del volume.** In un gas reale il volume fisicamente occupato dal gas non è l'effettivo volume a disposizione delle particelle, poiché queste possiedono un volume proprio che è inaccessibile alle altre. Il volume che compare nell'equazione di stato è, invece, il volume effettivamente a disposizione delle particelle e, quindi, per i gas reali il volume che compare nell'equazione di stato deve essere decurtato dal volume fisicamente occupato dalle particelle. Questo è calcolabile come il prodotto del numero di moli n per il volume di una mole di gas, detto **covolume b**:

$$V_{\text{ideale}} = V_{\text{reale}} - nb$$

2. **Modifica della pressione.** L'esistenza di forze di interazione non trascurabili tra le particelle di gas reale si traduce macroscopicamente in una diminuzione della pressione esercitata dalle particelle sulle pareti del recipiente. Le forze attrattive, infatti, tendono a far stazionare le particelle una vicina all'altra aumentando la durata media degli urti molecolari e diminuendo il numero di urti nell'unità di tempo. L'effetto è tanto più evidente quanto maggiore è il numero di molecole per unità di volume. Ciò significa che un gas ideale, privo di interazioni molecolari, esercita una pressione superiore rispetto ad un gas reale e il fattore addizionale introdotto da Van der Waals è proporzionale al numero di particelle per unità di volume:

$$P_{\text{ideale}} = P_{\text{reale}} + a \frac{n^2}{V^2}$$

Introducendo queste due modifiche nell'equazione di stato, si ottiene l'equazione di Van der Waals dei gas reali:

$$\left(P_{\text{reale}} + a \frac{n^2}{V^2} \right) (V - nb) = nRT$$

6 I passaggi di stato

Ogni sostanza può presentarsi in più stati di aggregazione a seconda delle condizioni di pressione e temperatura in cui si trova. Il passaggio da uno stato fisico ad un altro è detto **passaggio di stato**. L'acqua, ad esempio, è allo stato liquido in condizioni ambientali ($T=25^\circ\text{C}$ e $P=1\text{atm}$), ma è sufficiente abbassarne la temperatura al di sotto degli 0°C per osservarne il passaggio allo stato solido o elevarne la temperatura sopra i 100°C per osservarne l'ebollizione e il rapido passaggio allo stato vapore. Il diagramma che segue mostra i nomi assegnati ai possibili passaggi di stato.

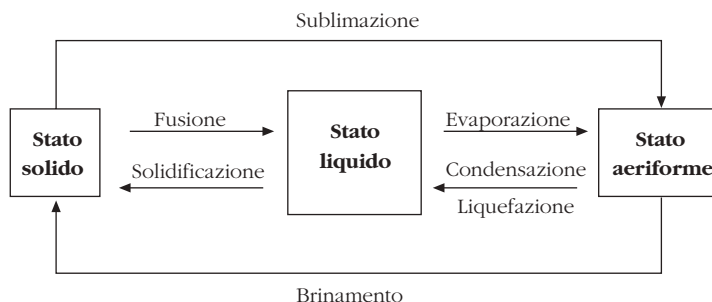


Fig. 3 - Schematizzazione dei passaggi di stato

Ogni passaggio di stato implica uno scambio di energia con l'esterno: i passaggi di stato che portano a stati di maggior disordine (tutti quelli verso destra) implicano l'assorbimento di energia dall'ambiente; viceversa, i passaggi verso stati di ordine maggiore (tutti quelli verso sinistra) comportano cessione di energia all'ambiente. Ciò, del resto, è del tutto intuitivo: fornendo energia ad un solido, ad esempio sotto forma di calore, aumenta l'agitazione termica delle particelle e, quindi, l'energia delle oscillazioni attorno alle loro posizioni di equilibrio nel reticolo cristallino. In questa fase la temperatura del solido aumenta di pari passo con l'agitazione termica, proporzionalmente al calore fornito. Quando l'energia delle oscillazioni supera l'energia di legame tra le particelle del solido ha inizio la fusione; il calore fornito durante la fusione non incrementa l'agitazione termica delle particelle, ma viene impiegato per rompere i legami presenti nel solido. Di conseguenza, fino alla completa fusione del solido la temperatura si mantiene costante al valore della temperatura di fusione.

Quando tutta la massa è passata allo stato liquido, continuando a fornire calore si osserverà nuovamente un incremento di temperatura proporzionale alla quantità di calore fornito; al crescere della temperatura, aumenta il numero di particelle di liquido dotate di energia sufficiente a passare in fase vapore e si assiste all'evaporazione lenta e continua di una quantità di liquido sempre maggiore.

Quando la temperatura raggiunge un valore tale che la pressione di vapore eguaglia il valore della pressione esterna, si assiste all'ebollizione, durante la quale l'evaporazione ha luogo non solo dalla superficie ma anche dal corpo del liquido; il vapore che si genera all'interno del liquido, essendo meno denso, risale formando le tipiche «bolle». Anche in questo caso, durante l'ebollizione, la temperatura si mantiene costante fintanto che tutto il liquido è evaporato e solo ad evaporazione ultimata, fornendo ulteriormente calore, si assiste all'incremento di temperatura della fase vapore. Riportando in un diagramma la temperatura di una sostanza in funzione della quantità di calore fornito, si ottiene un tipico diagramma caratterizzato da rampe e tratti costanti.

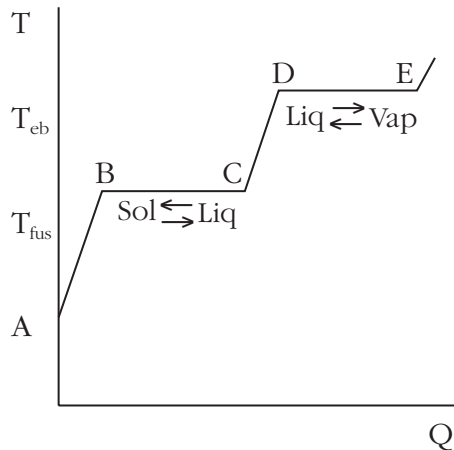


Fig. 4

Il tratto AB rappresenta la «rampa» di riscaldamento del solido, fino alla temperatura di fusione. Il tratto orizzontale BC corrisponde al **calore latente di fusione** del solido, così denominato poiché è una quantità di calore che non produce riscaldamento della sostanza ma ne permette il passaggio di stato fisico. Nel punto C tutta la sostanza è passata in fase liquida e l'ulteriore calore fornito determina, lungo il tratto CD, il riscaldamento della nuova fase liquida, fino alla temperatura di ebollizione. Il tratto orizzontale DE rappresenta il **calore latente di evaporazione**, che induce il passaggio di stato senza produrre riscaldamento. Infine, quando nel punto E tutta la massa è evaporata, inizia una nuova rampa di temperatura, la quale rappresenta il riscaldamento della nuova fase vapore.

7 I diagrammi di stato

Le temperature a cui avvengono i passaggi di stato di una certa sostanza mutano al variare della pressione esterna: ad esempio, l'acqua bolle a 100°C in condizioni ambientali normali, cioè

alla pressione di 1 atm. In alta montagna, dove il valore della pressione atmosferica è inferiore, la temperatura di ebollizione è più bassa o, come si dice colloquialmente, «l'acqua bolle prima».

I **diagrammi di stato** riportano i valori delle temperature alle quali avvengono i vari passaggi di stato di una sostanza a seconda della pressione esterna, definendo in tal modo i campi di stabilità delle diverse fasi. Va sottolineato che tutti i punti sulle curve, individuando le condizioni alle quali hanno luogo i passaggi di stato fisico, rappresentano le condizioni di coesistenza di due fasi diverse; tutti i punti contenuti tra una curva e l'altra rappresentano i campi di stabilità dei diversi stati fisici. Il punto in cui si incontrano le tre curve solido-liquido, liquido-vapore e solido-vapore è detto **punto triplo** ed è un valore, caratteristico per ogni sostanza, in cui si ha la coesistenza delle tre fasi. L'andamento di un tipico diagramma di stato è riportato in figura.

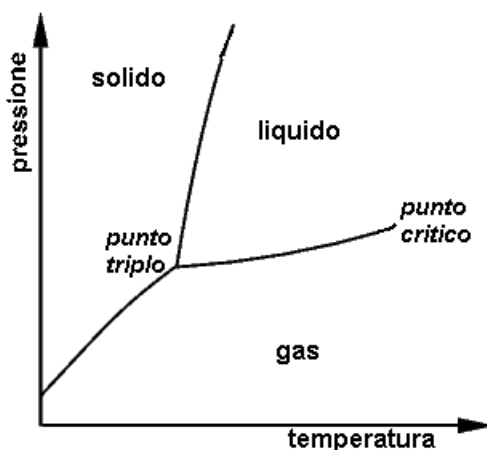


Fig. 5

La materia e le sue trasformazioni

1 Reazioni complete e di equilibrio

Abbiamo visto nei precedenti capitoli come la materia possa essere interessata da trasformazioni di tipo fisico, che comportano il passaggio di una sostanza da uno stato di aggregazione all'altro, senza che si verifichi un mutamento di composizione. Passiamo a parlare ora di trasformazioni di tipo chimico, ossia trasformazioni che modificano la composizione della materia trasformando delle sostanze di partenza (reagenti) in sostanze diverse (prodotti).

Una trasformazione chimica comporta la rottura di alcuni o tutti i legami chimici presenti nelle molecole dei reagenti e la riorganizzazione dei loro elementi costituenti in nuove specie chimiche, i **prodotti di reazione**, mediante la formazione di nuovi legami. Ogni trasformazione di tipo chimico viene indicata col nome di **reazione chimica** e può essere rappresentata attraverso una **equazione chimica** costituita da due membri: nel primo, a sinistra, sono indicate le specie reagenti, che entrano nella reazione; nel secondo, a destra, le specie prodotte. In una **reazione completa**, i due membri sono separati da una freccia unica, da sinistra verso destra ($\rightarrow$); in tal caso la reazione procede fino all'esaurirsi dei reagenti o del reagente presente in quantità minore. In una **reazione di equilibrio**, i due membri sono separati da una doppia freccia ($\rightleftharpoons$); questo simbolo sta ad indicare che la reazione scritta non si svolge unicamente dai reagenti verso i prodotti, ma anche in senso inverso. Ciò significa che appena formatesi alcune delle molecole dei prodotti si ricombinano a formare i reagenti. Al raggiungimento dell'equilibrio la reazione diretta ($\rightarrow$) e quella inversa ($\leftarrow$) assumono la stessa velocità e le concentrazioni dei reagenti e dei prodotti di reazione si mantengono costanti nel tempo.

La quantità di prodotti ottenibile da una reazione dipende dal bilancio energetico della reazione stessa: la rottura dei legami chimici nelle specie reagenti comporta una spesa di energia che deve essere sufficientemente ricompensata dall'energia liberata con la formazione dei nuovi legami nei prodotti di reazione. Se l'energia liberata dalla reazione è nettamente superiore a quella spesa, la reazione è spontanea e decorre in modo completo; se il guadagno di energia non è particolarmente elevato la reazione avviene facilmente sia dai reagenti ai prodotti sia in senso opposto: il risultato è una reazione di equilibrio e la quantità di prodotti ottenuta all'equilibrio rappresenta, quindi, la massima quantità ottenibile da quella reazione, nelle condizioni in cui è stata condotta. Infine, se il bilancio energetico della reazione è sfavorevole e l'energia spesa risulta superiore all'energia liberata, la trasformazione dei reagenti in prodotti non avviene spontaneamente, mentre sarà spontanea la reazione inversa.

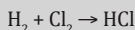
2 Equazioni chimiche e loro bilanciamento

L'equazione chimica che rappresenta una certa reazione viene scritta obbedendo alle regole seguenti:

- nell'equazione chimica si indicano **solo le specie che scompaiono e che si formano**, ignorando altre specie eventualmente presenti, le quali si conservano inalterate;
- se la reazione coinvolge specie dotate di cariche elettriche, la somma algebrica delle cariche deve essere uguale al primo e al secondo membro (**legge di conservazione della carica**);

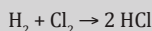
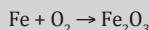
- tutti gli atomi che costituiscono le specie reagenti devono ritrovarsi nelle specie prodotte dalla reazione (**legge di conservazione della massa**);
- se la reazione ha luogo in **ambiente acquoso**, per le specie ioniche vengono indicati gli ioni costituenti. Se alcuni di questi non partecipano alla reazione vengono omessi;
- affinché siano soddisfatte le leggi di conservazione della carica e della massa, a ciascuna delle specie reagenti e prodotte si devono attribuire **coefficienti stechiometrici** che stabiliscono in quali proporzioni, in una data reazione, i reagenti e i prodotti debbano scomparire o formarsi. Questa operazione è indicata comunemente come **bilanciamento della reazione**.

L'attribuzione del *coefficiente stechiometrico* può essere eseguita facilmente considerando che, non potendosi verificare perdita di materia nel corso della trasformazione, tutti gli atomi di un elemento che compaiono nelle specie reagenti devono ritrovarsi in ugual numero, anche se diversamente combinati, nelle specie prodotte. Vediamo qualche esempio.

ESEMPIO 1

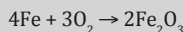
In questa reazione, idrogeno H_2 e cloro Cl_2 si combinano per formare acido cloridrico HCl. Poiché sia l'idrogeno che il cloro esistono come molecole biatomiche, la minima quantità di atomi di H (o di Cl) che possono entrare nella reazione è 2 e di conseguenza il numero minimo di molecole HCl prodotte è 2.

L'equazione bilanciata, quindi, è:

**ESEMPIO 2**

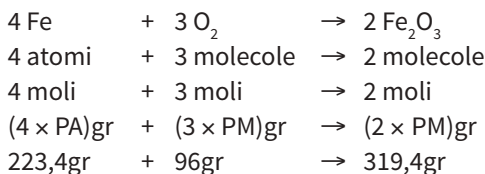
La molecola di ossigeno O_2 è biatomica, pertanto il numero di atomi di ossigeno che entrano nella reazione è necessariamente pari. L'ossido di ferro Fe_2O_3 , invece, contiene un numero dispari di atomi di ossigeno e, dunque, per poterne bilanciare il numero tra reagenti e prodotti si dovrà considerare almeno il loro minimo comune multiplo: se 3 molecole di ossigeno (6 atomi di O) entrano nella reazione, ne escono 2 molecole di ossido (6 atomi di O e 4 di Fe). Di conseguenza, il numero di atomi di Fe che devono essere posti a reagire è 4.

L'equazione bilanciata, quindi, è:



I coefficienti stechiometrici di una reazione possono essere letti ed utilizzati in vari modi: sicuramente essi esprimono il numero di atomi o molecole che si combinano tra loro per formare un certo numero di molecole di prodotti. Tuttavia, il rapporto di combinazione tra le molecole si mantiene tra le moli, poiché ogni mole contiene lo stesso numero di molecole, pari al numero di Avogadro. Quindi il coefficiente stechiometrico fornisce sia il rapporto molecolare sia il rapporto molare di combinazione dei reagenti. Infine, ricordando che il peso in grammi di una mole di sostanza coincide con il peso molecolare, il coefficiente stechiometrico fornisce informazioni anche sulle quantità in grammi di reagenti e prodotti che si combinano tra loro o che si generano dalla reazione.

Tornando all'esempio 2, noto il peso atomico del Fe (55,85), il peso molecolare dell' O_2 (32) e quello del Fe_2O_3 (159,7), possiamo sicuramente affermare, in base all'equazione precedente, che:



L'utilità della mole è, quindi, quella di far reagire quantità di sostanze pesabili con una comune bilancia, cosa non realizzabile per la singola molecola, sapendo esattamente il numero di atomi o molecole che queste contengono, e poter fare previsioni sulla quantità delle specie ottenute dalla reazione.

3 Le reazioni di ossidoriduzione

A seconda del tipo di reagenti e di prodotti di reazione è d'uso distinguere le reazioni chimiche in diverse classi. Una prima grande distinzione può essere fatta tra:

- **reazioni non di ossidoriduzione**, nelle quali si ha solo scambio di specie atomiche tra reagenti e prodotti senza che si verifichino variazioni dello stato di ossidazione. Nell'ambito di queste reazioni si possono ulteriormente distinguere reazioni di idrolisi, di precipitazione, di complessazione, di spostamento, reazioni acido-base, che affronteremo nel dettaglio parlando singolarmente dei diversi tipi di composti;
- **reazioni di ossidoriduzione**, nelle quali le specie coinvolte nella reazione assumono stati di ossidazione diversi passando dai reagenti ai prodotti di reazione.

Focalizziamo al momento l'attenzione su questo secondo tipo di reazioni e chiariamo, prima di tutto, il significato di alcuni termini che utilizzeremo.

- In una **reazione redox** le specie chimiche coinvolte cambiano il loro *stato di ossidazione* passando dai reagenti ai prodotti; queste reazioni comportano il trasferimento di elettroni da una specie chimica ad un'altra.
- La specie che cede elettroni si ossida e il suo *numero di ossidazione aumenta* passando dai reagenti ai prodotti (questa specie è detta **riducente**, poiché cedendo elettroni induce la riduzione di altre specie).
- La specie che acquista elettroni si riduce e il suo *numero di ossidazione diminuisce* passando dai reagenti ai prodotti (questa specie è detta **ossidante**, poiché acquistando elettroni induce l'ossidazione di altre specie).
- Data l'incapacità dell'elettrone di rimanere allo stato libero, un *processo di ossidazione* è necessariamente accompagnato da un *processo di riduzione*, ovvero una specie chimica non può ossidarsi e quindi perdere elettroni se non in presenza di un'altra specie chimica in grado di acquistarli.
- Come conseguenza di quanto detto prima, in una reazione redox il numero di elettroni ceduti dal riducente è pari al numero di elettroni acquistati dall'ossidante. In alcune reazioni ci possono essere più di una specie ossidante o più di una specie riducente; in ogni caso, il numero totale di elettroni ceduti dalle specie riducenti deve essere uguale al numero di elettroni acquistati dalle specie ossidanti.
- Vengono dette **reazioni di disproporzione** quei processi di ossidoriduzione in cui una stessa specie subisce contemporaneamente ossidazione e riduzione trasformandosi in due composti differenti, uno ad uno stato di ossidazione maggiore e l'altro ad uno stato di ossidazione minore rispetto al composto di partenza.

4 Il bilanciamento delle reazioni di ossidoriduzione

I metodi utilizzabili per il bilanciamento delle reazioni redox sono vari. I due più utilizzati sono il *metodo della variazione del numero di ossidazione* e il *metodo ionico-elettronico*.

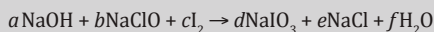
a) Metodo della variazione del numero di ossidazione

Il **metodo della variazione del numero di ossidazione** prevede la sequenza delle seguenti operazioni:

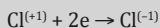
1. individuare la specie che si ossida e quella che si riduce;
2. stabilire il numero di elettroni n_1 acquistati da una molecola dell'ossidante e il numero di elettroni n_2 ceduti da una molecola del riducente, tenendo presente che il numero n_1 è il coefficiente del riducente e il numero n_2 è il coefficiente dell'ossidante;
3. una volta stabiliti i coefficienti di riducente e di ossidante, determinare, di conseguenza, quelli dei prodotti di ossidazione e di riduzione;
4. i coefficienti di tutte le altre specie presenti vengono assegnati in base al bilanciamento delle masse.

ESEMPIO 3

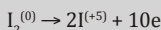
Bilanciamo la reazione:



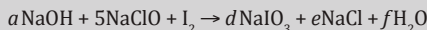
- Il cloro passa dallo stato di ossidazione +1 in ClO^- allo stato di ossidazione -1 in Cl^- e quindi subisce una riduzione acquistando 2 elettroni (n° di ossidazione iniziale + n° elettroni acquistati = n° di ossidazione finale: $+1 - 2e = -1$):



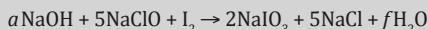
- Lo iodio passa dallo stato di ossidazione zero in I_2 allo stato di ossidazione +5 in IO_3^- e dunque subisce un'ossidazione cedendo 5 elettroni (n° di ossidazione iniziale = n° elettroni ceduti + n° di ossidazione finale):



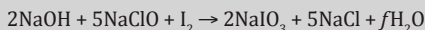
- Il coefficiente dell'ipoclorito (ClO^-) dovrebbe essere 10 e quello dello iodio 2, con un rapporto ipoclorito/iodio pari a 10:2. A questo punto si possono ridurre i coefficienti al rapporto minimo, ossia 5:1, attribuendo coefficiente 5 all'ipoclorito e coefficiente 1 allo iodio. Scrivendo i coefficienti nell'equazione chimica deduciamo immediatamente i coefficienti delle specie prodotte, cioè dello iodato e del cloruro:



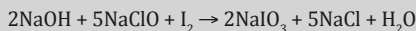
da cui:



- A questo punto i coefficienti assegnati non devono essere più modificati e gli altri coefficienti si devono attribuire in base al bilanciamento delle masse. Gli atomi di Na che troviamo a destra sono 7, a sinistra ce ne sono solo $(5+a)$; allora affinché $(5+a) = 7$, a dovrà essere pari a 2:



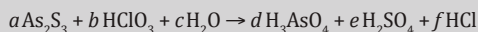
Essendo 2 gli atomi di idrogeno presenti a sinistra, il coefficiente f della molecola d'acqua deve essere unitario. Otteniamo così:



Il conteggio degli atomi di ossigeno (7 sia a sinistra sia a destra) fornisce un ulteriore controllo sulla correttezza dei coefficienti assegnati.

ESEMPIO 4

Bilanciamo la reazione:

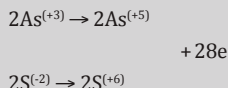


- Il cloro passa dallo stato di ossidazione +5 allo stato di ossidazione -1 e pertanto subisce una riduzione acquistando 6 elettroni:

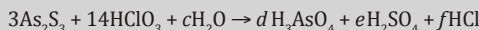


- L'arsenico passa dallo stato di ossidazione +3 allo stato di ossidazione +5 e quindi subisce un'ossidazione cedendo 2 elettroni; contemporaneamente anche lo zolfo passa dallo stato di ossidazione -2 allo stato di ossidazione +6 e, di conseguenza, subisce un'ossidazione cedendo 8 elettroni. La reazione di una molecola di solfuro

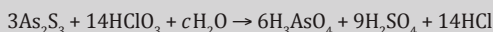
arsenioso As_2S_3 comporta l'ossidazione contemporanea di 2 atomi di arsenico e tre di zolfo, liberando nel complesso 28 elettroni ($2 \times 2 = 4$ liberati dall'arsenico e $3 \times 8 = 24$ liberati dallo zolfo):



- Il rapporto trovato tra i coefficienti di As_2S_3 e di HClO_3 è di 6:28, che semplificato diventa 3:14. Quindi all' As_2S_3 attribuiremo coefficiente 3, mentre all' HClO_3 attribuiremo coefficiente 14:



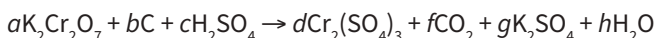
- I coefficienti dei prodotti di riduzione e di ossidazione si ricavano direttamente dal conteggio degli atomi di As, S e Cl che entrano nella reazione, i quali sono rispettivamente 6, 9 e 14. Nell'equazione avremo, dunque, $\text{d} = 6$, $\text{e} = 9$, $\text{f} = 14$:



- Resta da determinare il coefficiente dell'acqua che può essere facilmente ricavato indifferentemente dal conteggio degli atomi di ossigeno o di idrogeno: gli atomi di ossigeno che compaiono nei prodotti di reazione sono $6 \times 4 = 24$ nell'acido arsenico e $9 \times 4 = 36$ nell'acido solforico, per un totale di 60; nei reagenti $3 \times 14 = 42$ sono portati dall'acido clorico e i restanti 18 dovranno essere portati dall'acqua. Ne consegue che il coefficiente dell'acqua sarà $\text{c} = 18$.

b) Metodo ionico-elettronico

Il **metodo ionico elettronico**, sebbene apparentemente più complicato del precedente, risulta molto utile nel bilanciamento di reazioni redox aventi **stechiometria complessa**. Illustriamo la procedura da seguire direttamente con un esempio pratico, effettuando il bilanciamento della reazione:

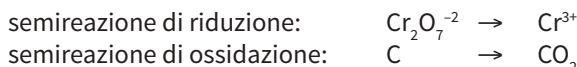


- *Individuare la specie o le specie che subiscono ossidazione e riduzione, attribuendo i numeri di ossidazione nella forma ossidata e nella forma ridotta.*

Nel nostro caso la specie che si ossida è il carbonio C, che passa dallo stato di ossidazione 0 allo stato di ossidazione +4 nell'anidride carbonica CO_2 . La specie che si riduce è il cromo che passa dallo stato di ossidazione +6 nel dicromato di potassio allo stato di ossidazione +3 nel solfato di cromo.

- *Scrivere separatamente le semireazioni di ossidazione (e di riduzione), indicando la specie che si è ossidata (o ridotta) e il relativo prodotto di ossidazione (o riduzione). Se le specie che partecipano alla reazione subiscono dissociazione in soluzione acquosa, nelle semireazioni devono comparire solo gli ioni che effettivamente si ossidano o si riducono; i controioni che non partecipano al processo redox vengono omissi.*

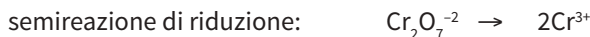
Il dicromato di potassio in soluzione acquosa è dissociato nei suoi ioni costituenti K^+ e $\text{Cr}_2\text{O}_7^{-2}$; analogamente il solfato di potassio, ottenuto come prodotto di reazione, è dissociato negli ioni costituenti Cr^{3+} e SO_4^{-2} . Di questi ioni quelli che effettivamente prendono parte alla reazione sono $\text{Cr}_2\text{O}_7^{-2}$ e Cr^{3+} , per cui la semireazione di riduzione dovrà essere scritta come:



- *Bilanciare le semireazioni scritte in base al principio di conservazione della massa: in ognuna delle semireazioni, il numero di atomi di ossidante o di riducente che compaiono nelle specie reagenti deve essere uguale al numero di atomi che troviamo nelle specie prodotte.*

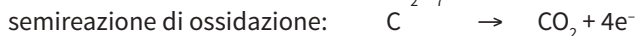
Nel nostro caso ogni ione dicromato che si riduce porta alla formazione di due ioni Cr^{3+} e quindi, nella semireazione di riduzione, il Cr^{3+} dovrà avere coefficiente 2. Le semireazioni bilan-

ciate rispetto alla massa degli elementi che subiscono i processi di ossidazione e di riduzione sono quindi:



- *Dopo aver effettuato il bilanciamento di massa, conteggiare il numero di elettroni ceduti dal riducente ed acquistati dall'ossidante. Gli elettroni ceduti compaiono come prodotti nella semireazione di ossidazione, gli elettroni acquistati, invece, compaiono tra i reagenti nella semireazione di riduzione.*

Per cui:



- *Bilanciare le semireazioni scritte in base al principio di conservazione della carica: in ogni semireazione, la carica complessiva delle specie reagenti deve risultare uguale alla carica complessiva delle specie prodotte. La carica totale di ciascun membro della semireazione deve essere calcolata come somma delle cariche degli ioni presenti (ognuno moltiplicato per il proprio coefficiente stechiometrico) e degli elettroni. Dove ci sia una disuguaglianza delle cariche, il bilanciamento dovrà essere effettuato aggiungendo ioni H^+ se la reazione avviene in ambiente acido, oppure ioni OH^- se la reazione è condotta in ambiente basico.*

Nel nostro esempio, nella reazione di riduzione il primo membro ha carica complessiva -8 (dovuta a due cariche negative del dicromato e sei elettroni), mentre il secondo membro ha carica +6 (dovuta ai due ioni Cr^{3+}). Poiché tra i reagenti della nostra reazione compare l'acido solforico, l'ambiente di reazione è acido e l'eccesso di carica negativa al primo membro viene bilanciato mediante l'aggiunta di ioni H^+ . Aggiungendo 14 ioni H^+ al primo membro, la carica totale diventa +6, come al secondo membro. Nella semireazione di ossidazione, invece, abbiamo un eccesso di carica negativa al secondo membro, dove la carica totale è -4 a fronte della carica 0 del primo membro. In questo caso dovremo aggiungere 4 ioni H^+ al secondo membro, riportando la carica totale a zero.

Per cui:



- *Procedere al bilanciamento della massa dell'idrogeno e dell'ossigeno, aggiungendo, dove necessario, delle molecole d'acqua.*

Nella semireazione di riduzione dovremo aggiungere 7 molecole di H_2O tra i prodotti, in quella di ossidazione 2 molecole di H_2O tra i reagenti.

Pertanto:



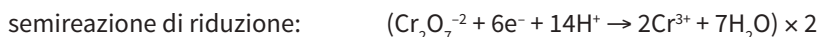
A questo punto le due semireazioni sono perfettamente bilanciate sia rispetto alla carica sia rispetto alla massa.

- *Poiché il numero di elettroni ceduti dal riducente deve essere pari al numero di elettroni acquistati dall'ossidante, ciascuna delle due semireazioni deve essere moltiplicata, se necessario, per un fattore tale da eguagliare il numero di elettroni persi ed acquistati.*

In pratica, è sufficiente calcolare il minimo comune multiplo di 6 (elettroni acquistati) e 4 (elettroni ceduti), ossia 12, e moltiplicare ciascuna delle semireazioni per un fattore tale che il

numero di elettroni scambiati risulti pari a 12. Nel nostro caso la semireazione di riduzione va moltiplicata per 2, quella di ossidazione per 3.

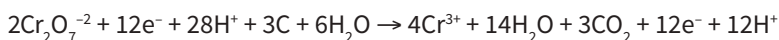
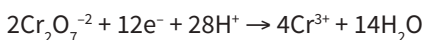
Pertanto:



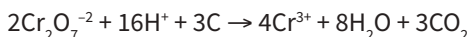
Eseguendo i prodotti otteniamo:



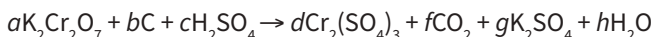
- *Sommare membro a membro le due semireazioni ottenute, semplificando i termini simili che compaiono nei due membri.*



Semplificando i termini simili, abbiamo:

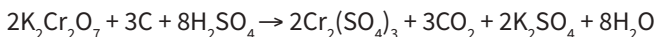


Riportando i coefficienti nell'equazione molecolare di partenza, si ottiene:

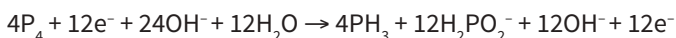
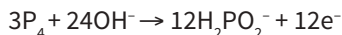
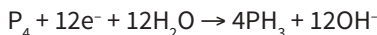
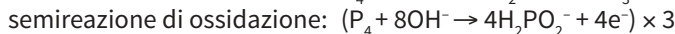
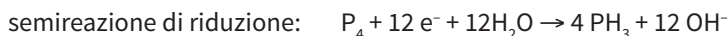
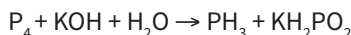


Avremo $a = 2$, $b = 3$, $c = 8$ (poiché ogni molecola di acido solforico fornisce due protoni), $d = 2$ (poiché ogni molecola di solfato di cromo contiene due ioni Cr^{3+}), $f = 3$, $g = 2$ (per bilanciare lo ione K^+), $h = 8$.

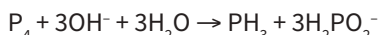
L'equazione bilanciata, dunque, è:



Come ultimo esempio, consideriamo la reazione di disproporzione del fosforo in ambiente alcalino:



Semplificando i termini simili e dividendo tutti i coefficienti per il loro massimo comun divisore, si ottiene:



5 Ruggine e ossidazione dello zinco

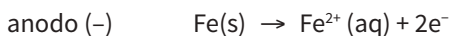
Dal punto di vista chimico, l'**ossidazione** è la perdita di elettroni con conseguente aumento, appunto, del numero di ossidazione: un elemento, cioè, perde elettroni (si ossida) a favore di un

altro vicino che li acquista (si *riduce*) in un processo combinato, detto di «ossido-riduzione» (o «*redox*»). Il termine fu inizialmente applicato alla reazione tra un metallo e l'ossigeno, per dare il corrispondente *ossido*. Il fenomeno può avvenire in diversi metalli (a esclusione di quelli «nobili»: oro, argento e platino) e in altre sostanze: anche il calcio, ad esempio, si «ossida», eppure non si «corrode» visibilmente.

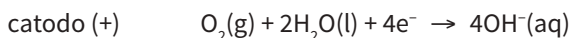
L'*ossidazione* è spesso seguita dalla *corrosione* e, in genere, ha effetti distruttivi: tipica è l'ossidazione del ferro (metallo enormemente diffuso), che dà luogo alla **ruggine**, una sostanza porosa e friabile, la quale, di solito, tende a staccarsi dalla superficie, lasciando le zone sottostanti esposte a ulteriore ossidazione e al procedere inesorabile della corrosione. In altre parole, è evidente, in questo fenomeno, l'alternanza del processo chimico e di quello fisico: prima si forma uno strato di ossido, che in seguito si fessura; quindi entra ancora ossigeno, che forma altro ossido e così via.

Per **corrosione** s'intende il *graduale deterioramento di un materiale, causato dal progressivo decadimento delle sue proprietà caratteristiche, per interazione chimico-fisica con l'ambiente circostante*, ovvero quando il materiale viene a contatto con aria, acqua (soprattutto del mare), acidi, gas inquinanti etc. Questo fenomeno, molto comune e di grandissima rilevanza, anche economica, può riguardare le sostanze più disparate, ma in genere, con il termine «corrosione» ci si riferisce a quella cui vanno soggetti i metalli e le loro leghe.

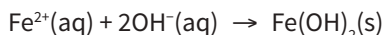
Abbiamo visto come l'«arrugginimento» sia un processo progressivo, il quale ha inizio perché nel metallo si crea un sistema che si comporta come una *cella elettrochimica*. Sulla superficie del manufatto si forma una «zona anodica» in cui il ferro si ossida:



Gli elettroni si trasferiscono, attraverso il metallo, verso un'altra zona a contatto con l'aria; quest'area si comporta da catodo ed è qui che avviene la *riduzione dell'ossigeno* in presenza di acqua:

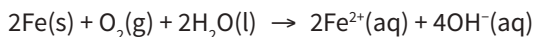


Gli ioni ferrosi Fe^{2+} vengono a contatto con gli ioni OH^{-} e generano un composto solido, l'*idrossido ferroso*, scarsamente solubile in acqua:



L'idrossido ferroso, in presenza di ossigeno e di acqua, si trasforma in *ruggine* ($\text{Fe}_2\text{O}_3 \cdot n\text{H}_2\text{O}$), un miscuglio di composti solidi, contenenti ioni ferrici Fe^{3+} derivanti dall'ossidazione degli ioni ferrosi Fe^{2+} .

Affinché avvenga la corrosione del ferro, occorre, quindi, che siano presenti contemporaneamente *ossigeno e acqua*:



La corrosione del ferro avviene tanto più rapidamente quanto maggiore è la *conducibilità elettrica* della soluzione acquosa; per questo, la più elevata presenza di elettroliti nell'acqua rende il processo corrosivo più veloce nei litorali marini, ricchi di salsedine.

C'è, da tempo, un notevole impegno da parte di chimici e scienziati nel ricercare metodi che impediscano o, almeno, rallentino il fenomeno della corrosione. Per proteggere i materiali ferrosi, si è soliti ricoprirli di uno strato protettivo; esso può essere costituito da «vernice» (a questo proposito si parla di veri e propri «cicli di pitturazione») contenente *sostanze antiossidanti* (o *autoprotettenti*, le quali mettono in atto un processo elettrochimico chiamato «**passivazione**»,

una specie di «ossidazione superficiale controllata») oppure da particolari metalli, come il *cromo*, lo *stagno* o lo *zinco*. In questo modo il ferro (o anche l'*acciaio*, una *lega* composta principalmente da ferro) viene isolato dall'aria e dall'umidità: pertanto non può ossidarsi.

Lo **zinco**, in particolare, viene adoperato nella fabbricazione di lamiere per tettoie, per grondaie e strutture del genere. Ma il suo utilizzo principale è legato proprio alla protezione del ferro dalla ruggine, ossia alla realizzazione del cosiddetto «*ferro zincato*» o «*galvanizzato*». Il processo di lavorazione consiste nel rivestire il ferro di una patina di zinco: le lamiere di ferro, in effetti, vengono prima accuratamente pulite con acido (per allontanare dalla loro superficie gli ossidi di ferro), quindi immerse in un bagno di zinco fuso. La deposizione dello strato protettivo di zinco può essere effettuata anche per via «elettrolitica», sebbene questa procedura sia senza dubbio più costosa. Il *ferro zincato* trova impiego nella costruzione di serbatoi per acqua, tubi, scatolame etc.

Per impedire o rallentare la corrosione, si può anche «legare» il ferro con altri elementi. L'*acciaio inossidabile*, ad esempio, è una lega che contiene *cromo* e *nicel*; questi metalli formano rivestimenti di ossidi che impediscono l'ossidazione del ferro. Gli oggetti in acciaio inossidabile marcati 18/10, molto comuni in commercio, contengono il 18% di *cromo* e il 10% di *nicel*.

1 Le reazioni complete

Vengono dette **complete** quelle reazioni in cui la formazione dei prodotti procede fintanto che nel sistema sono disponibili le specie reagenti. Consideriamo una reazione generica:



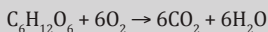
Possono verificarsi due casi:

1. i reagenti sono presenti in quantità esattamente stechiometrica, ossia il numero di moli di A e di B che sono state poste a reagire hanno un rapporto coincidente con a/b . In questo caso, al termine della reazione, le specie A e B sono completamente esaurite e sono presenti unicamente i prodotti di reazione;
2. i reagenti sono presenti inizialmente in quantità casuali e, quindi, la reazione procede fin quando uno dei due, il reagente in difetto rispetto all'altro, si esaurisce. Questo è indicato generalmente come **reagente limitante**.

Vediamo ora, in base a quanto detto, come procedere nel calcolo.

ESEMPIO 1

La combustione del glucosio nelle nostre cellule si svolge secondo la reazione:



Determinare quanti grammi di ossigeno vengono consumati dalla combustione di 10 gr di glucosio e quanti grammi di CO_2 vengono prodotti.

Soluzione

Dai coefficienti stechiometrici di reazione si deduce che ogni mole di glucosio ne richiede 6 di ossigeno. Dunque, determinato il numero di moli contenute in 10 gr di glucosio, il numero di moli di ossigeno consumate è sei volte il numero di moli di glucosio. Inoltre, essendo il rapporto stechiometrico tra O_2 e CO_2 6:6 (ovvero 1:1), il numero di moli di O_2 coincide con il numero di moli di CO_2 .

$$moli C_6H_{12}O_6 = \frac{10gr}{(6 \times 12 + 12 + 6 \times 16)} = 0,0555$$

$$moli O_2 = 6 \times moli C_6H_{12}O_6 = 0,33 = moli CO_2$$

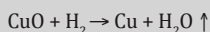
Per risalire dal numero di moli alle corrispondenti quantità in grammi, è sufficiente moltiplicare per il peso molecolare:

$$grammi O_2 = 0,33 \times 32 = 10,56 \text{ gr}$$

$$grammi CO_2 = 0,33 \times 44 = 14,52 \text{ gr}$$

ESEMPIO 2

L'idrogeno molecolare H_2 ad alta temperatura reagisce con l'ossido rameico CuO generando rame metallico Cu^0 e acqua, che si libera allo stato gassoso secondo la reazione:



Determinare:

1. quanti grammi di Cu metallico si generano a partire da 15 gr di ossido, considerando l'idrogeno in eccesso;
2. la minima pressione di idrogeno che si deve realizzare in un recipiente del volume di 30 L alla temperatura di 200 °C per riuscire a convertire totalmente 200 gr di ossido rameico in rame metallico.

Soluzione

1. Dalla stechiometria della reazione vediamo che 1 mole di ossido produce 1 mole di rame metallico; il numero di moli di ossido contenute nei 15 gr coincide, quindi, con il numero di moli di Cu ottenute:

$$\text{moli} \cdot \text{CuO} = \frac{\text{gr} \cdot \text{CuO}}{P.M.\text{CuO}} = \frac{15}{(63,55 + 16)} = 0,189 = \text{moli} \cdot \text{Cu}$$

Per risalire alla quantità di rame espressa in grammi, basta moltiplicare il numero di moli per il peso atomico del Cu:

$$\text{gr} \cdot \text{Cu} = \text{moli} \cdot \text{Cu} \times P.A.\text{Cu} = 0,189 \times 63,55 = 12,01 \text{ gr}$$

2. Per rispondere alla seconda domanda, occorre calcolare la quantità di idrogeno stechiometrica con i 200 gr di ossido; anche in questo caso l'equazione chimica ci fornisce un rapporto di reazione 1:1 e, pertanto, il numero di moli di H_2 e di ossido sono numericamente coincidenti:

$$\text{moli} \cdot \text{CuO} = \frac{200 \text{ gr}}{79,55 \text{ gr} \cdot \text{mol}^{-1}} = 2,51 \cdot \text{moli} \cdot \text{CuO} = \text{moli} \cdot \text{H}_2$$

La minima pressione di idrogeno da realizzare può essere calcolata in base all'equazione di stato del gas perfetto:

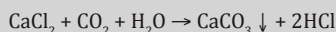
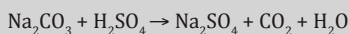
$$PV = nRT$$

da cui:

$$P = \frac{nRT}{V} = \frac{2,51 \text{ mol} \cdot 0,0821 \text{ L} \cdot \text{atm} \cdot \text{mol}^{-1} \cdot \text{K}^{-1} \cdot (200 + 273,15) \text{ K}}{30 \text{ L}} = 3,25 \text{ atm}$$

ESEMPIO 3

25 gr di Na_2CO_3 impuro vengono trattati con acido solforico in eccesso e l'anidride carbonica che si sviluppa dalla reazione viene fatta gorgogliare in una soluzione di cloruro di calcio dalla quale si osserva la precipitazione di 23 gr di CaCO_3 .



Determinare il grado di purezza del campione di carbonato di sodio.

Soluzione

Dalla stechiometria delle reazioni si evince che il numero di moli del carbonato di calcio ottenute sono numericamente coincidenti con quelle di anidride carbonica e queste con quelle di carbonato di sodio presenti nel campione iniziale:

$$\text{moli} \cdot \text{CaCO}_3 = \frac{23 \text{ gr}}{(40 + 12 + 3 \times 16) \text{ gr} \cdot \text{mol}^{-1}} = \text{moli} \cdot \text{CO}_2 = \text{moli} \cdot \text{Na}_2\text{CO}_3 = 0,23 \cdot \text{moli}$$

I grammi di Na_2CO_3 contenuti nel campione iniziale sono:

$$\text{gr} \cdot \text{Na}_2\text{CO}_3 = 0,23 \times (2 \times 23 + 12 + 3 \times 16) = 24,38 \text{ gr}$$

La percentuale di Na_2CO_3 presente nel campione è:

$$\% \text{Na}_2\text{CO}_3 = \frac{24,38}{25} \cdot 100 = 97,52\%$$

2 Le reazioni non complete: resa di reazione e resa percentuale

Alcune reazioni non decorrono in modo completo, ma sono caratterizzate da una **resa di reazione** inferiore al 100%. Ciò può verificarsi sia a causa dell'instaurarsi di un equilibrio tra reagenti e prodotti, in cui parte dei prodotti appena formati si ricombina per formare i reagenti; sia nel caso di reazioni che possono essere accompagnate da reazioni «collaterali» più o meno competitive con la reazione presa in esame.

La **resa percentuale** rappresenta la quantità di prodotto ottenuto rispetto alla quantità massima teoricamente ottenibile dalla reazione, nell'ipotesi che questa fosse completa o, analogamente, la quantità di reagente consumato dalla reazione rispetto alla quantità disponibile.

ESEMPIO 4

2 kg di solfuro ferroso vengono trattati per ottenere Fe^0 metallico. Se il rendimento del processo è del 75%, quanto ferro si ottiene?

Soluzione

In base alla stechiometria del composto, possiamo affermare che ogni mole di solfuro ferroso contiene 1 mole di zolfo e 1 mole di ferro:

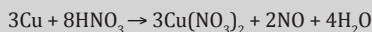
$$\text{moli} \cdot \text{FeS} = \frac{2000 \text{ gr}}{(55,85 + 32,06)} = 22,75 = \text{moli} \cdot \text{Fe}$$

Solo il 75% delle moli di Fe contenute nel solfuro sono trasformate in Fe^0 , cosicché la quantità di Fe ottenuta è:

$$\text{grammi} \cdot \text{Fe} = 22,75 \cdot \frac{75}{100} \cdot 55,85 = 952,94 \text{ gr}$$

ESEMPIO 5

Il rame elementare Cu^0 reagisce con l'acido nitrico per generare monossido di azoto secondo la reazione:



Calcolare la resa percentuale della reazione, sapendo che da 50 gr di rame si ottengono 9,9 gr di NO.

Soluzione

Le moli di NO ottenute e il numero di moli di rame effettivamente «reagite» sono nel rapporto di 2:3. Le moli di NO ottenute dalla reazione sono:

$$\text{moli NO} = \frac{9,9 \text{ gr}}{(14 + 16)} = 0,33$$

Il numero di moli di rame reagite si ricava dalla proporzione:

$$2:3 = \text{moli} \cdot \text{NO} : \text{moli} \cdot \text{Cu}$$

da cui:

$$\text{moli} \cdot \text{Cu} = \frac{3 \times 0,33}{2} = 0,495$$

$$\text{grammi} \cdot \text{Cu} = \text{moli} \cdot \text{Cu} \times P.A.\text{Cu} = 0,495 \times 63,55 = 31,45$$

Pertanto, la resa percentuale di reazione è:

$$\text{resa}\% = \frac{31,45}{50} \cdot 100 = 62,9\%$$

ESEMPIO 6

L'acido fosforoso H_3PO_3 si decompone ad alta temperatura generando fosfina e acido fosforico; sapendo che 24 gr di acido fosforoso hanno prodotto 2,3 gr di fosfina, calcolare la resa di reazione:

**Soluzione**

Le moli di acido fosforoso inizialmente presenti sono:

$$moli \cdot H_3PO_3 = \frac{24}{(3 + 30,97 + 3 \times 16)} = 0,293$$

Le moli di acido fosforoso decomposte sono ricavabili da quelle di fosfina prodotta e la stechiometria della reazione fornisce un rapporto 1:4 tra fosfina e acido fosforoso:

$$moli \cdot H_3PO_3 = 4 \times moli \cdot PH_3 = 4 \times \frac{2,3}{(3 + 30,97)} = 0,27$$

Dunque, la resa percentuale di reazione è:

$$resa\% = \frac{0,27}{0,293} \cdot 100 = 92,15\%$$

1 Definizione e composizione

Si definisce **soluzione** una dispersione omogenea di due o più componenti l'uno nell'altro in proporzioni variabili a piacere. L'omogeneità della soluzione è tale che campioni prelevati in punti diversi mostrano composizione identica tra loro e alla soluzione stessa. Generalmente viene indicato come **solvente** il componente che si trova nello stesso stato fisico della soluzione e come **soluto** l'altro componente; se tutti i componenti si trovano nello stesso stato fisico della soluzione, il solvente è il componente presente in quantità maggiore. La **composizione** delle soluzioni può essere espressa attraverso diverse **unità di concentrazione**, che possiamo distinguere in unità *massa/massa* (percentuale in peso, moli/kg solvente, frazione molare), unità *massa/volume* (moli/litro, equivalenti/litro), unità *volume/volume* (percentuali in volume).

La **percentuale in peso** esprime il numero in grammi di soluto contenuto in 100 gr di soluzione.

La **percentuale in volume** esprime il volume di un componente presente in 100 volumi di soluzione. Si tratta di un'unità poco utilizzata, poiché il volume è grandezza non sempre additiva.

La **frazione molare** è il rapporto tra il numero di moli del singolo componente e il numero di moli totali. Viene indicata con χ ed è per definizione una quantità minore di 1, mentre la somma delle frazioni molari di tutti i componenti della soluzione è unitaria.

La **molarità** esprime il numero di moli di soluto disciolte in un litro di soluzione e si indica con M. È una delle unità di concentrazione più utilizzate.

La **molalità** esprime il numero di moli di soluto disciolte in un chilogrammo di solvente e si indica con m. Per una soluzione diluita, la molalità e la molarità hanno valori praticamente coincidenti: infatti, nelle soluzioni diluite la densità della soluzione resta molto vicina a quella dell'acqua e, quindi, 1.000 gr di solvente corrispondono a circa 1 litro di soluzione.

La **normalità** esprime il numero di equivalenti di soluto contenuti in un litro di soluzione e si indica con N. È un'unità di interesse applicativo, poiché facilita molto i calcoli stechiometrici. Ricordiamo, infatti, che nelle reazioni le specie reagiscono sempre in un uguale numero di equivalenti.

2 Passaggio in soluzione e solubilità

La stragrande maggioranza di quelle che incontreremo sono **soluzioni acquose**, ossia soluzioni in cui il solvente è l'acqua. L'acqua è una molecola dotata di elevata polarità, essendo costituita da elementi di elettronegatività molto diversa; gli idrogeni rappresentano il polo positivo della molecola, l'ossigeno il polo negativo. Grazie a queste caratteristiche l'acqua ha ottime capacità solventi sia nei confronti di specie ioniche, sia di specie molecolari dotate di legami polari che in acqua possono andare incontro a ionizzazione, sia nei confronti di specie in grado di formare legami idrogeno.

Il passaggio in soluzione acquosa di **specie ioniche** è favorito dalla capacità della molecola d'acqua di attrarre con la stessa efficacia ioni sia negativi sia positivi: ogni ione che costituisce il reticolo cristallino viene «sfilato» dal reticolo dalle forze di attrazione esercitate dalle molecole d'acqua e, una volta passato in soluzione, viene completamente circondato da molecole del solvente che ne stabilizzano la carica. La struttura che ne deriva è detta «ione solvatato»,

generato sulla base di interazioni elettriche di tipo ione-dipolo; il processo di solvatazione, quindi, libera una certa quantità di energia, che ripaga dell'energia spesa per demolire la struttura cristallina. Una volta passata in soluzione, una specie ionica risulta completamente dissociata nei suoi ioni costituenti. Il passaggio in soluzione acquosa di **specie molecolari** si basa sull'instaurarsi di interazioni più deboli tra soluto e solvente, essenzialmente di tipo dipolo-dipolo se il soluto è una specie polare, oppure dipolo-dipolo indotto (legami di van der Waals) per le specie non polari; non va dimenticata l'ulteriore possibilità posseduta dalla molecola d'acqua di formare legami idrogeno con i soluti.

Alcune specie molecolari dotate di elevata polarità possono subire, in seguito al passaggio in soluzione acquosa, un **processo di ionizzazione** generando ioni positivi e negativi in seguito alla scissione di uno o più legami chimici fortemente polarizzati presenti nella molecola. Questa scissione si verifica poiché le molecole d'acqua esercitano forze di attrazione sugli atomi del legame polarizzato sottoponendolo a forti sollecitazioni che possono provocarne la rottura. L'entità delle interazioni tra il gruppo polare e le molecole d'acqua è tanto più grande quanto più grande è la polarità del legame, e il **grado di dissociazione** subito, inteso come rapporto tra il numero di molecole dissociate rispetto al totale delle molecole passate in soluzione, cresce al crescere della polarizzazione del legame che si rompe.

La quantità di sostanza che riesce a sciogliersi nell'acqua dipende, quindi, dalle capacità solventi dell'acqua nei confronti della sostanza stessa, ossia dalla sua possibilità di stabilire interazioni con il soluto che ne consentano una dispersione omogenea e stabile nel tempo. In linea generale possiamo dire che maggiore è la polarità di una molecola o la sua polarizzabilità, maggiore è la sua solubilità in acqua. Viene definita **solubilità** di una sostanza in un solvente la concentrazione di sostanza disciolta in quel solvente, ad una definita temperatura, in condizioni di saturazione. La **concentrazione di saturazione** rappresenta la concentrazione della soluzione in equilibrio con il corpo di fondo, ad una definita temperatura. Raggiunto il limite di saturazione, l'ulteriore aggiunta di sostanza alla soluzione ne determina la deposizione come corpo di fondo. Nella definizione di solubilità è presente un riferimento esplicito alla temperatura: ciò si deve al fatto che anche il passaggio in soluzione è un processo sensibile alla temperatura e, nella maggior parte dei casi, la solubilità di una sostanza aumenta al crescere della temperatura.

3 Le proprietà di soluzioni di non elettroliti

La presenza di un soluto altera le proprietà caratteristiche del solvente: la tensione di vapore, la temperatura di congelamento e quella di ebollizione risultano modificate. Queste proprietà sono il risultato delle forze di interazione presenti tra le molecole del solvente, che vengono necessariamente alterate dalla presenza del soluto. Si osserva che l'entità delle variazioni registrate risulta indipendente dal tipo di soluto disciolto, mentre dipende dalla sua concentrazione: queste proprietà vengono dette, quindi, **proprietà colligative**.

a) Variazione della pressione di vapore: legge di Raoult

La soluzione ottenuta sciogliendo in acqua un soluto non volatile (ovvero una sostanza con pressione di vapore inferiore a quella dell'acqua) mostra una diminuzione della pressione di vapore rispetto al solvente puro. Infatti il vapore in equilibrio con la soluzione ad una certa temperatura contiene sia il soluto sia il solvente e ciascuno di essi contribuisce alla pressione di vapore totale con la sua pressione parziale, come avviene per qualsiasi miscela gassosa:

$$P_{\text{totale}} = p_{\text{soluto}} + p_{\text{solvente}}$$

Raoult ha dimostrato che le pressioni parziali del soluto e del solvente nella fase vapore sono proporzionali alle rispettive pressioni di vapore allo stato puro (P^0_{soluto} e P^0_{solvente}) attraverso le rispettive frazioni molari:

$$P_{\text{totale}} = \chi_{\text{soluto}} P^0_{\text{soluto}} + \chi_{\text{solvente}} P^0_{\text{solvente}}$$

b) Innalzamento ebullioscopico

Diretta conseguenza della diminuzione della pressione di vapore della soluzione rispetto a quella del solvente puro è l'innalzamento della temperatura di ebollizione: l'acqua che porta disciolti soluti non volatili bolle ad una temperatura superiore ai 100°C, poiché è necessaria una temperatura maggiore affinché la pressione di vapore della soluzione riesca a raggiungere il valore della pressione atmosferica. Ovviamente al crescere della concentrazione del soluto l'abbassamento della pressione di vapore è maggiore e, di conseguenza, è maggiore l'innalzamento della temperatura di ebollizione. Si può dimostrare che la variazione della temperatura di ebollizione è proporzionale alla concentrazione molale del soluto attraverso una costante caratteristica del solvente, detta **costante crioscopica**:

$$\Delta T_{\text{eb}} = K_{\text{eb}} \cdot m$$

c) Abbassamento crioscopico

La presenza di un soluto provoca anche l'abbassamento della temperatura di congelamento del solvente. Tale fenomeno viene sfruttato nelle stagioni fredde, quando viene aggiunto all'acqua del radiatore delle auto il «liquido antigelo», un preparato liquido che abbassa la temperatura di congelamento dell'acqua fino ad un valore che, generalmente, non viene raggiunto dalla temperatura ambiente. Anche in questo caso si può dimostrare che l'entità dell'abbassamento è proporzionale alla concentrazione molale del soluto attraverso una costante crioscopia caratteristica del solvente:

$$\Delta T_{\text{cr}} = K_{\text{cr}} \cdot m$$

d) Pressione osmotica

Una proprietà di estrema importanza manifestata dalle soluzioni, soprattutto per le sue implicazioni nei sistemi biologici, è la pressione osmotica π . Questa proprietà si evidenzia quando una soluzione di concentrazione molare C_0 venga messa in contatto con il solvente puro attraverso una membrana semipermeabile, che permetta, cioè, selettivamente il passaggio di molecole di solvente e non di soluto (si ricordi che le membrane cellulari rappresentano le più sofisticate membrane selettivamente permeabili che si conoscano). Ponendo il solvente puro e la soluzione nelle due braccia di un tubo ad U, separate dalla membrana, si osserva il passaggio spontaneo di solvente dal braccio contenente il solvente puro a quello contenente la soluzione fino a che tra le due braccia si genera un dislivello tale da esercitare una pressione ΔP che coincide con la pressione osmotica della soluzione. Questa pressione è pari a:

$$\pi = C_0 RT$$

Anche la pressione osmotica, come l'abbassamento ebullioscopico, l'innalzamento crioscopico e la variazione della tensione di vapore del solvente sono proprietà dipendenti unicamente dal numero di particelle disciolte nel solvente e non dalla loro natura chimica. Queste proprietà delle soluzioni possono essere sfruttate per la **determinazione dei pesi molecolari** delle sostanze,

poiché ognuna delle relazioni precedenti contiene implicitamente il valore del peso molecolare della specie disciolta. Infatti:

$$\Delta T_{eb} = K_{eb} \cdot m = K_{eb} \cdot \frac{q}{PM} \quad (1)$$

$$\Delta T_{cr} = K_{cr} \cdot m = K_{cr} \cdot \frac{q}{PM} \quad (2)$$

$$\pi = C_0 RT = \frac{q}{PM} RT \quad (3)$$

dove q rappresenta la quantità di soluto, espressa in grammi, che si trova disciolta in 1 kg di solvente, nel caso delle relazioni (1) e (2), o in 1 litro di soluzione, nel caso della (3).

4 Le proprietà di soluzioni di elettroliti

Si definisce **elettrolita** una specie chimica che, in soluzione, si scinde parzialmente o totalmente in ioni. L'elettrolita si dice **forte** se la dissociazione è totale, **debole** se la dissociazione è parziale. Si parla di **dissociazione** per specie chimiche ioniche che, in soluzione acquosa, si dividono completamente negli ioni costituenti: le specie ioniche sono tutte elettroliti forti.

Per specie covalenti ad elevata polarità che generano ioni in soluzione acquosa a causa delle sollecitazioni esercitate dal solvente si parla più propriamente di **ionizzazione**. Il processo di ionizzazione può essere totale o parziale e viene definito **grado di dissociazione** α il rapporto tra il numero di moli dissociate e il numero di moli totali presenti in soluzione. Ovviamente il grado di dissociazione di un elettrolita forte è 1, mentre assume valore inferiore all'unità per tutti gli elettroliti deboli.

Per qualsiasi soluzione elettrolitica possiamo definire una concentrazione analitica C_{an} relativa all'elettrolita e più concentrazioni ioniche relative agli ioni generati dalla dissociazione: una soluzione a concentrazione analitica $[NaCl] = 0,1M$, ha concentrazioni ioniche $[Na^+] = 0,1M$ e $[Cl^-] = 0,1M$. Una soluzione a concentrazione analitica $[CaCl_2] = 0,1M$, ha concentrazioni ioniche $[Ca^{2+}] = 0,1M$ e $[Cl^-] = 0,2M$.

Ai fini delle proprietà colligative della soluzione ciò che conta è la concentrazione ionica totale, dal momento che queste proprietà non dipendono dal tipo ma dal numero di particelle disciolte.

Per un elettrolita forte, indicando con v il numero di ioni generati dalla dissociazione della singola molecola, la concentrazione totale è:

$$C_{tot} = v \cdot C_{an}$$

Per gli elettroliti deboli, invece, il numero di particelle presenti dopo la dissociazione è dato dalla somma del numero di moli rimaste indissociate e del numero di moli dissociate moltiplicato per v . Indicando con n_a il numero di moli introdotte in base alla concentrazione analitica e con n_{dis} il numero di moli che subisce dissociazione, $(n_a - n_{dis})$ rappresenta la frazione indissociata e $v \cdot n_{dis}$ è il numero di particelle prodotte dalla dissociazione.

Ricordando che il grado di dissociazione $\alpha = \frac{n_{dis}}{n_a}$, il numero totale di particelle dopo la dissociazione è:

$$n_{tot} = \left[(n_a - n_{dis}) + v n_{dis} \right] = n_a \left[1 - \frac{n_{dis}}{n_a} + v \frac{n_{dis}}{n_a} \right] = n_a [1 + \alpha(v-1)]$$

Dividendo tutto per il volume V della soluzione otteniamo la concentrazione totale:

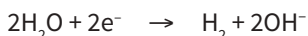
$$C_{\text{tot}} = C_a [1 + \alpha (v - 1)]$$

Quindi il valore delle grandezze colligative delle soluzioni di elettroliti può essere calcolato moltiplicando le concentrazioni molari o molali che compaiono nelle corrispondenti relazioni per v nel caso di elettroliti forti e per $[1 + \alpha (v - 1)]$ per gli elettroliti deboli. Una proprietà caratteristica delle soluzioni elettrolitiche è quella di poter condurre la **corrente elettrica** grazie alla presenza di specie dotate di carica (gli ioni positivi e negativi) libere di migrare sotto l'azione di un campo elettrico esterno.

5 L'elettrolisi dell'acqua

La presenza dell'**acqua** condiziona notevolmente l'andamento di un **processo elettrolitico** e ciò non perché essa interferisca nel libero movimento degli *ioni*, che anzi ne risulta facilitato, ma perché può intervenire nelle reazioni di scarica, facendole avvenire in maniera del tutto diversa.

Ad esempio, se invece di effettuare l'elettrolisi del cloruro di sodio fuso, operiamo con una soluzione acquosa di questo sale, impiegando un *catodo* di ferro e un *anodo* di titanio, vedremo che mentre all'anodo si forma cloro molecolare, al catodo non si ha deposito di sodio metallico, bensì sviluppo di idrogeno, che si libera sotto forma di bollicine gassose. Quest'idrogeno non può provenire altro che dalla decomposizione dell'acqua e, poiché in soluzione compaiono contemporaneamente ioni OH^- , possiamo spiegare ciò ammettendo che l'acqua venga dissociata e che l' H^+ formato si riduca a contatto con il catodo. Complessivamente, perciò, la reazione di scarica sarà:

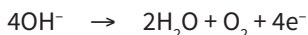


Al posto della riduzione dello ione Na^+ si ha, dunque, la riduzione dello ione H^+ proveniente dall'acqua. Il motivo per cui questo accade è legato al fatto che il potenziale di riduzione del sodio è molto più basso di quello dell'idrogeno: ciò vuol dire che la tendenza ad acquistare elettroni è nel sodio molto minore rispetto all'idrogeno. Possiamo anche affermare che l'idrogeno offre una minore resistenza alla riduzione di quella offerta dallo ione Na^+ e, pertanto, la sua scarica, a parità di altre condizioni, è favorita. Tale conclusione non può, purtroppo, essere generalizzata.

Il valore del potenziale di riduzione non sempre è sufficiente da solo a determinare la precedenza nella scarica degli ioni: occorre, infatti, tener conto anche del tipo di elettrodo sul quale essa avviene. Se operiamo l'elettrolisi del cloruro di sodio in soluzione acquosa, impiegando un catodo di mercurio anziché di ferro, tornerà a scaricarsi direttamente lo ione Na^+ e non si avrà più sviluppo di idrogeno. Il sodio metallico, in questo caso, rimane disciolto nel mercurio sotto forma di una lega che, come tutte le leghe di questo metallo, viene chiamata **amalgama**. L'influenza dell'elettrodo sulla scarica di uno ione viene, in linea di massima, espressa con la cosiddetta **sovratensione**, la quale può essere determinata e misurata nei vari casi. Per stabilire a chi spetta la precedenza nella scarica degli ioni su un determinato elettrodo è necessario, allora, sommare i valori del potenziale *redox* e della *sovratensione*: questa somma rappresenta il **potenziale di scarica** dello ione su quell'elettrodo.

Per quanto riguarda i *cationi* possiamo dire che la tendenza a ridursi è tanto maggiore quanto maggiore è il potenziale di scarica e che, di conseguenza, *potranno scaricarsi direttamente al catodo tutti quei cationi, il cui potenziale di scarica è maggiore di quello dell'idrogeno. Quelli, invece, il cui potenziale di scarica è minore rimarranno in soluzione e al loro posto si avrà la decomposizione dell'acqua con sviluppo d'idrogeno molecolare.*

Del tutto simile è la situazione all'*anodo*, dove però l'acqua non interviene più attraverso la riduzione dello ione idrogeno, ma tramite l'ossidazione dello ione OH^- , pure proveniente dalla sua dissociazione. L'ossidazione di questo ione avviene nel seguente modo:



E, partendo direttamente dall'acqua, la reazione completa sarà:

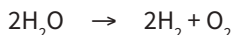


Il termine di paragone in questo caso sarà costituito, tenendo conto delle sovratensioni, dal potenziale di scarica dell' OH^- . *Tutti gli anioni il cui potenziale di scarica è minore di quello dell' OH^- hanno, rispetto all'elettrodo considerato, una maggiore tendenza a ossidarsi e quindi si scaricheranno direttamente a contatto con l'anodo. Quelli, invece, il cui potenziale di scarica è maggiore resteranno in soluzione e al loro posto si avrà la decomposizione dell'acqua, con conseguente ossidazione dell' OH^- e sviluppo di ossigeno molecolare.*

Se si sottopone ad elettrolisi la soluzione acquosa di un acido, si avrà al catodo, invariabilmente, sviluppo di idrogeno: l' H^+ , infatti, è l'unico catione presente e, pertanto, non possono esservi alternative. All'anodo, invece, possiamo avere o la scarica diretta di un anione oppure, più spesso, la decomposizione dell'acqua. Con l'acido solforico, ad esempio, le reazioni di scarica sono:

- al catodo, $4\text{H}^+ + 4\text{e}^- \rightarrow 2\text{H}_2$;
- all'anodo, $2\text{H}_2\text{O} \rightarrow 4\text{H}^+ + 4\text{e}^- + \text{O}_2$.

Come si vede, tutti gli ioni H^+ e gli elettroni che scompaiono al catodo vengono riprodotti all'anodo, per cui la reazione chimica complessiva, che avviene nella cella elettrolitica, è unicamente:



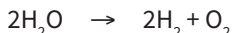
Essa, dunque, consiste esclusivamente nella scomposizione dell'acqua nei suoi elementi.

Identico risultato si ottiene se si parte dalla soluzione di una base forte, come NaOH o KOH . In tal caso, al catodo non si scaricherà il metallo ma reagirà l'acqua, mentre all'anodo si scaricherà lo ione OH^- .

Avremo cioè:

- al catodo, $4\text{H}_2\text{O} + 4\text{e}^- \rightarrow 2\text{H}_2 + 4\text{OH}^-$;
- all'anodo, $4\text{OH}^- \rightarrow 2\text{H}_2\text{O} + \text{O}_2 + 4\text{e}^-$.

Da cui deduciamo che la reazione chimica effettivamente verificatasi è, in definitiva:



Elettrolisi di questo tipo, nelle quali, almeno in apparenza, non partecipano gli ioni dell'elettrolita ma solo quelli dell'acqua, vengono chiamate, appunto, **elettrolisi dell'acqua**.

1 I concetti fondamentali

Il primo concetto con cui familiarizzare in termodinamica è quello di *sistema*, come entità distinta dall'ambiente. Un **sistema termodinamico** è una qualsiasi porzione di materia sulla quale si possano effettuare osservazioni o misure (un bicchiere in cui facciamo avvenire una reazione chimica può essere considerato un sistema termodinamico). L'**ambiente** comprende, invece, tutto ciò che è al di fuori del sistema, e quindi tutto l'Universo ad esclusione della porzione di materia che costituisce il sistema considerato. A seconda del tipo di relazione esistente con l'ambiente, un sistema termodinamico viene definito:

- **aperto**, se è in grado di scambiare materia ed energia con l'ambiente;
- **chiuso**, se è in grado di scambiare energia, ma non materia, con l'ambiente;
- **isolato**, se non scambia energia né materia con l'ambiente.

Ogni sistema termodinamico viene completamente descritto da un numero minimo di parametri, noti i quali è possibile costruirne una copia esatta; questi parametri sono detti **variabili di stato** e sono *temperatura*, *pressione* e *composizione*. Nell'ambito delle variabili di stato distinguiamo variabili **intensive**, ovvero indipendenti dalla massa, come *temperatura* e *pressione*, e variabili **estensive**, dipendenti dalla massa e additive rispetto a questa, come *volume* e *numero di moli*.

Lo **stato termodinamico** del sistema è definito dal valore delle sue *variabili di stato* e a ciascuno stato è associato un valore di ognuna delle *funzioni di stato* che lo descrivono. Viene definita **funzione di stato** una grandezza che assume valori dipendenti unicamente dallo stato in cui il sistema si trova, indipendentemente da come vi sia pervenuto. Il passaggio di un sistema da uno stato iniziale ad uno stato finale, distinto da quello di partenza, costituisce una **trasformazione termodinamica**. La variazione di una funzione di stato conseguente ad una trasformazione viene calcolata come differenza tra il valore assunto nello stato finale e quello che questa assumeva nello stato iniziale.

In base a quanto detto risulta chiaro che tutte le trasformazioni aventi in comune lo stato iniziale e lo stato finale sono caratterizzate dalle stesse variazioni delle funzioni di stato, indipendentemente dal modo in cui siano state prodotte. La **termodinamica** studia trasformazioni che avvengono tra stati di equilibrio, cioè trasformazioni che conducono il sistema da uno stato di equilibrio iniziale ad uno stato di equilibrio finale, e si occupa di stabilire quali scambi di energia tra sistema e ambiente avvengano nel corso di queste trasformazioni, nonché quali trasformazioni possa subire l'energia stessa convertendosi da una forma ad un'altra.

2 Entalpia e legge di Hess

Una **reazione chimica** è una trasformazione termodinamica che comporta un cambiamento della composizione del sistema; in questo caso il sistema termodinamico è costituito dalle **sostanze reagenti** che vengono trasformate in sostanze diverse, i **prodotti di reazione**. Abbiamo visto come una simile trasformazione comporti una spesa energetica, necessaria alla rottura dei legami chimici presenti nei reagenti, e la liberazione di energia, conseguente alla formazione di nuovi legami chimici nei prodotti. Complessivamente la reazione può determinare cessione di

energia dal sistema all'ambiente, se l'energia liberata dalla formazione dei prodotti è maggiore di quella spesa per la demolizione dei reagenti; nel caso contrario, la reazione richiede assorbimento di energia dall'ambiente.

Lo scambio di energia tra sistema e ambiente avviene spesso in forma di calore: se il sistema «si scalda» durante la reazione, sta cedendo energia all'ambiente; se viceversa abbiamo la sensazione che il sistema si raffreddi durante la reazione, evidentemente sta assorbendo energia dall'ambiente.

Per tutte le reazioni che avvengono in condizione di pressione costante, come le reazioni condotte in recipienti aperti, esposti alla pressione atmosferica, la quantità di calore scambiata può essere identificata con la variazione di una funzione di stato detta **entalpia** (dal greco *enthálpein*, «riscaldare»), indicata col simbolo H . La variazione di entalpia registrata nel corso di una reazione chimica viene definita **calore di reazione**.

Sono dette **endotermiche** le reazioni che avvengono mediante l'assorbimento di calore dall'ambiente e che sono caratterizzate da un aumento dell'entalpia ($\Delta H > 0$, dove $\Delta H = H_{\text{prodotti}} - H_{\text{reagenti}}$). Sono dette **esotermiche** le reazioni che avvengono mediante la cessione di calore dal sistema all'ambiente e che sono caratterizzate da una diminuzione dell'entalpia ($\Delta H < 0$). Il *calore di reazione* può essere considerato alla stregua di un reagente o di un prodotto della reazione, tanto che nelle reazioni esotermiche viene indicato tra i prodotti, in quelle endotermiche viene scritto tra i reagenti.

La **legge di Hess**, alla base dell'applicazione del concetto di *entalpia* alle reazioni chimiche, afferma:

«Il valore del ΔH di una reazione dipende unicamente dallo stato iniziale e dallo stato finale della reazione, mentre è del tutto indipendente dal cammino percorso».

Quindi, a parità di condizioni sperimentali, ovvero di temperatura e pressione, ad una stessa reazione è associato sempre lo stesso valore del ΔH di reazione, sia che questa avvenga in uno stadio unico, sia che venga condotta attraverso un certo numero di reazioni intermedie. Questa legge è una diretta conseguenza del fatto che l'entalpia è una funzione di stato del sistema.

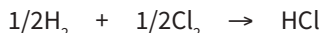
3 Stato standard ed entalpie di reazione

Abbiamo visto come tutte le funzioni di stato che descrivono un sistema termodinamico dipendano unicamente dallo stato del sistema e come questo sia definito, a sua volta, dal valore di un numero minimo di variabili di stato che abbiamo detto essere pressione, temperatura e composizione. Quindi, considerato il sistema costituito dai reagenti di una certa reazione, il loro contenuto entalpico dipende dalla pressione P e dalla temperatura T a cui questi si trovano, e lo stesso avverrà per i prodotti; di conseguenza, anche il valore del ΔH di reazione dipende dalla P e dalla T a cui la reazione è stata condotta. Per poter confrontare i valori dei ΔH di reazione e poterli riutilizzare nelle applicazioni pratiche è necessario riferirli a specifiche condizioni di pressione e di temperatura. A tal fine è stato definito per ogni specie chimica uno **stato standard, corrispondente alla specie pura nel suo stato stabile, alla temperatura di 25°C e alla pressione di 1 bar**.

Per le soluzioni, invece, lo stato standard è rappresentato dalla soluzione di **attività unitaria**. L'attività di un soluto in soluzione coincide con la concentrazione per soluzioni molto diluite; per soluzioni concentrate, invece, è correlata alla concentrazione dalla relazione $a = fC$, dove f è il coefficiente di attività e tiene conto delle interazioni esistenti tra le particelle di soluto nella soluzione.

La variazione di entalpia che si registra nel corso di una reazione, la quale partendo dai reagenti nei loro stati standard porti ai prodotti, anch'essi nei loro stati standard, è detta **entalpia**

standard di reazione e viene indicata con il simbolo ΔH^0 . I valori dei ΔH^0 di molte reazioni sono raccolti nella letteratura scientifica e possono essere utilizzati per calcolare i ΔH^0 non noti di altre reazioni. Tra i valori raccolti in letteratura, quelli di maggiore interesse nelle applicazioni pratiche sono i ΔH_f^0 , ossia le **entalpie molari standard di formazione**, le quali rappresentano il valore dei ΔH^0 delle reazioni che portano alla formazione di una mole di composto a partire dagli elementi costituenti, tutti considerati nei loro stati standard. Ad esempio, il ΔH^0 della reazione:



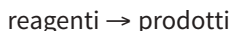
rappresenta l'entalpia molare standard di formazione dell'acido cloridrico. In base alla definizione stessa di entalpia molare di formazione, possiamo dire che le entalpie molari standard di formazione degli elementi sono nulle.

4 I cicli termochimici

In molti casi risulta indispensabile stabilire l'entità del riscaldamento che verrà prodotto da una reazione chimica (si pensi a reazioni condotte su scala industriale o alla scelta del combustibile più idoneo per un certo tipo di attività). Valutazioni di questo tipo possono essere effettuate in base alla conoscenza dei ΔH^0 di reazione. Una delle più importanti applicazioni pratiche della legge di Hess consiste, infatti, nella costruzione di **cicli termochimici** attraverso i quali calcolare i ΔH^0 di reazioni qualsiasi, a partire dai valori dei ΔH_f^0 dei reagenti e dei prodotti di reazione. Un'altra possibile enunciazione di questa legge, equivalente alla precedente, è allora:

«Il ΔH^0 di una reazione può sempre essere calcolato come differenza tra la sommatoria delle entalpie molari standard di formazione dei prodotti e quella delle entalpie molari standard di formazione dei reagenti, ognuna moltiplicata per il coefficiente stechiometrico con cui la specie compare nell'equazione chimica».

In effetti, essendo l'entalpia una funzione di stato, la variazione che questa subisce nel corso di una trasformazione dipende solo dallo stato iniziale e finale e non dal cammino percorso. Immaginiamo di voler calcolare il ΔH^0 di una generica reazione, che schematizziamo come:



Considerando la trasformazione che a partire dagli elementi conduce ai prodotti di reazione, la variazione di entalpia che questa comporta sarà la stessa tanto nel caso in cui il passaggio sia diretto (linea continua) quanto nel caso in cui si consideri come stadio intermedio quello della formazione dei reagenti (linea tratteggiata).

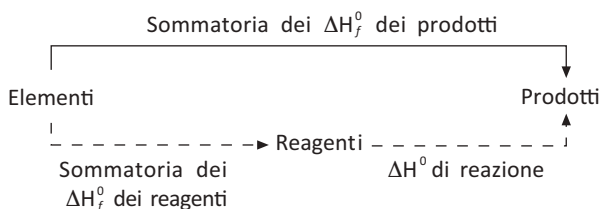


Fig. 1

Possiamo quindi dedurre che, per qualsiasi reazione, il ΔH^0 può essere calcolato come differenza tra le sommatorie delle entalpie molari standard di formazione dei prodotti e dei reagenti, come enunciato nella legge di Hess.

5 L'entropia: una misura del «grado di disordine»

È sicuramente un fatto acquisito nell'esperienza comune che tutti i fenomeni naturali avvengono spontaneamente in una specifica direzione e non in quella inversa: un corpo caldo si raffredda rapidamente se lasciato a temperatura ambiente, mentre non è mai stato osservato il suo riscaldamento spontaneo; se mettiamo una goccia di caffè nel latte osserviamo la sua spontanea diffusione in tutto il volume del liquido, fino ad ottenere una miscela di colore omogeneo, mentre non capiterà mai di osservare la goccia confinata in una specifica regione del liquido in cui l'abbiamo introdotta; se lanciamo una palla sul pavimento, dopo un certo numero di rimbalzi questa si ferma spontaneamente, avendo esaurito la sua energia cinetica, trasferita al pavimento sotto forma di calore sviluppato nell'urto. In modo analogo, alcune reazioni chimiche avvengono spontaneamente in una direzione: un combustibile, ad esempio la benzina, brucia producendo anidride carbonica e acqua; viceversa, anidride carbonica e acqua, sottoposte a riscaldamento, non restituiscono la benzina.

Tutto ciò accade poiché **qualsiasi sistema in natura tende ad evolvere verso uno stato di maggiore «disordine»** e l'energia stessa, che nel corso di qualsiasi trasformazione non viene creata né distrutta, ma solo convertita da una forma all'altra, tende spontaneamente a portarsi in una forma più dispersa e degradata. In termodinamica esiste una *funzione di stato* che misura il «grado di disordine» di un sistema: l'**entropia**, indicata con **S**. Utilizzando il concetto di entropia possiamo dire che:

«In natura sono spontanee tutte le trasformazioni per le quali l'entropia dell'Universo aumenta».

Va chiarito che per *entropia dell'Universo* si intende la somma dell'entropia del sistema e di quella dell'ambiente, le quali, nel loro insieme, costituiscono l'Universo.

Riprendendo in esame i fenomeni considerati in precedenza, possiamo effettivamente constatare la naturale tendenza dell'Universo a portarsi in uno stato di *maggiore disordine*: il raffreddamento spontaneo di un corpo è il risultato del trasferimento dell'agitazione termica delle sue particelle all'ambiente, ovvero della dispersione dell'energia; il fenomeno della diffusione di un liquido in un altro è il frutto del moto casuale delle particelle, che difficilmente determina una distribuzione non omogenea dei componenti (a meno che non intervengano altri fattori, come ad esempio l'immiscibilità); una palla che rimbalza trasferisce la propria energia cinetica (energia caratterizzata da un maggior grado di ordine, poiché tutte le particelle si muovono contemporaneamente in un'unica direzione) alle particelle del pavimento sotto forma di calore, una forma di energia più degradata (il moto dell'agitazione termica è disorganizzato e casuale); una reazione di combustione determina il passaggio del sistema da uno stato di ordine maggiore, costituito dall'unica molecola di combustibile in cui gli atomi sono legati tra loro e quindi ordinati, ad uno stato di maggiore disordine, ossia a un più alto numero di molecole, ognuna libera di muoversi indipendentemente dalle altre.

Dal punto di vista termodinamico, la variazione di entropia che un sistema subisce in una trasformazione in cui scambi con l'ambiente una quantità di calore Q è data dal rapporto tra la quantità di calore scambiata e la sua temperatura assoluta:

$$\Delta S = \frac{Q}{T}$$

L'entropia di un sistema allo zero assoluto (0 gradi Kelvin) è nulla, poiché a questa temperatura tutte le particelle sono immobili e il grado di disordine del sistema è nullo: ciò ci permette di conoscere il valore assoluto dell'entropia di un sistema a qualsiasi temperatura T , calcolando, in base alla relazione precedente, la variazione di entropia che si verifica portando il sistema dallo zero assoluto alla temperatura T .

Per valutare la *variazione di entropia* associata ad una reazione chimica è sufficiente calcolare la differenza tra l'entropia dei prodotti e quella dei reagenti; analogamente a quanto visto per l'*entalpia*, è opportuno riferire i valori allo stato standard, in modo da renderli confrontabili tra loro.

In linea generale, si può dire che una reazione chimica determina un **aumento di entropia**:

- quando la reazione comporta un aumento del numero di moli, cioè il numero di moli dei prodotti è maggiore del numero di moli dei reagenti;
- quando la reazione è accompagnata da passaggi di stato fisico e i prodotti di reazione si trovano in stati fisici più disordinati rispetto ai reagenti;
- nei processi di dissoluzione di specie solide nei liquidi.

6 Energia libera e spontaneità delle trasformazioni

In base a quanto detto finora, possiamo concludere che a qualsiasi reazione chimica possono essere associate variazioni delle funzioni termodinamiche che descrivono il sistema e che la spontaneità o meno della reazione dipende da due fattori:

- un **fattore entalpico**, legato al bilancio energetico tra l'energia richiesta per la demolizione dei reagenti e quella guadagnata nella formazione dei prodotti;
- un **fattore entropico**, che tiene conto della variazione del grado di disordine del sistema nel passaggio dai reagenti ai prodotti di reazione.

Ecco perché in chimica, per valutare la spontaneità di una reazione, viene impiegata la funzione di stato **energia libera**, indicata con **G**, la quale tiene conto contemporaneamente del fattore entalpico e del fattore entropico di una trasformazione. L'energia libera, o **funzione di Gibbs**, è definita come:

$$G = H - TS$$

La variazione di energia libera che accompagna una trasformazione condotta a pressione costante è:

$$\Delta G = \Delta H - T\Delta S$$

Le reazioni spontanee sono caratterizzate da una diminuzione dell'energia libera ($\Delta G < 0$) e vengono dette **esoergoniche**. Le trasformazioni che determinano un aumento dell'energia libera ($\Delta G > 0$) possono avvenire solo forzatamente, fornendo energia dall'esterno, e sono chiamate **endoergoniche**. Quando, invece, tra stato iniziale e stato finale della trasformazione non c'è variazione dell'energia libera ($\Delta G = 0$), il sistema si trova in equilibrio.

La tabella che segue riassume le condizioni in cui le reazioni avvengono spontaneamente. Le reazioni esotermiche, che comportano un aumento di entropia, sono spontanee a tutte le temperature; le reazioni endotermiche, invece, che comportano una diminuzione di entropia, non avvengono mai spontaneamente. Le reazioni favorite dal solo fattore entalpico avvengono spontaneamente solo per quei valori di temperatura per i quali questo diviene preponderante rispetto al fattore entropico; viceversa, le reazioni favorite dal solo fattore entropico avvengono spontaneamente solo per quei valori di temperatura per i quali questo diviene preponderante rispetto al fattore entalpico.

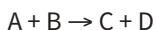
ΔH	ΔS	ΔG	T
< 0	> 0	< 0	Tutte le temperature
< 0	< 0	< 0	$T < \Delta H/\Delta S$
> 0	< 0	> 0	Tutte le temperature
> 0	> 0	< 0	$T > \Delta H/\Delta S$

Tab. 1

1 I meccanismi di reazione

Abbiamo visto come la termodinamica ci permetta di fare previsioni sulla spontaneità delle reazioni prendendo in esame unicamente lo stato di partenza e lo stato di arrivo della reazione. Indagare negli aspetti cinetici di una reazione chimica significa gettare uno sguardo al meccanismo con cui questa avviene. Va precisato, innanzitutto, che una reazione può aver luogo attraverso un unico atto reattivo (*reazione elementare*) o attraverso uno o più passaggi intermedi.

Consideriamo il caso semplice di una reazione elementare o del singolo atto reattivo di una reazione, che schematizziamo come:



Affinché le molecole A e B possano reagire è necessario che:

- si verifichi un urto;
- l'urto avvenga nel sito di reazione di entrambe le molecole (aspetto di notevole importanza in chimica organica, meno importante in chimica inorganica, in cui le molecole hanno minori dimensioni);
- l'energia cinetica traslazionale delle molecole che collidono sia sufficiente a far avvenire la reazione.

Si è osservato più volte come una reazione chimica comporti la rottura di legami esistenti nei reagenti e la formazione di nuovi legami nei prodotti; quando l'urto tra A e B avviene con energia sufficiente si genera un «aggregato» che costituisce uno **stato di transizione** tra reagenti e prodotti, caratterizzato da legami in via di rottura e legami in via di formazione. Questo stadio intermedio, detto **complesso attivato**, si forma a spese dell'energia cinetica delle particelle che collidono e possiede energia più elevata rispetto ai reagenti.

La differenza di energia tra le specie reagenti e lo stato di transizione (rappresentata in figure come ΔE_1) è detta **energia di attivazione** E_a e rappresenta la barriera energetica, totalmente ignorata dalla termodinamica, che il sistema deve superare perché la reazione possa aver luogo. Per la reazione inversa, l'energia di attivazione è data invece dalla somma di ΔE_1 e ΔE_2 , che corrispondono rispettivamente all'energia di attivazione e alla variazione di energia libera della reazione diretta. Il complesso attivato può generarsi ogni qualvolta la somma delle energie cinetiche traslazionali dei reagenti risulti maggiore o uguale all'energia di attivazione.

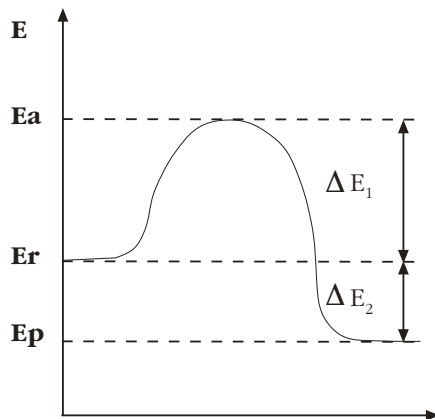


Fig. 1

Per riuscire ad aumentare la velocità di una reazione bisogna, da un lato, facilitare l'incontro delle specie reagenti, aumentandone la concentrazione o ponendole in un efficace mezzo di veicolazione; dall'altro, aumentare la percentuale di urti efficaci. A tal fine si può agire sulla temperatura, poiché sappiamo, in base alla distribuzione di Maxwell-Boltzmann, che al cresce-

re della temperatura aumenta la velocità media di traslazione delle molecole e, quindi, la loro energia cinetica.

2 La velocità di reazione

La conoscenza della velocità di reazione si acquisisce unicamente attraverso misure sperimentali e costituisce un'informazione di grande importanza non solo a fini pratici, ma anche per risalire al meccanismo di reazione. La misura della velocità di reazione può essere effettuata seguendo nel tempo la concentrazione di un reagente o di un prodotto. Nella trasformazione:



la velocità può essere espressa sia come diminuzione della [A] o della [B] nel tempo, sia come aumento della [C] o della [D]. In formule:

$$v = -\frac{\Delta[A]}{\Delta t} = -\frac{\Delta[B]}{\Delta t} \quad v = \frac{\Delta[C]}{\Delta t} = \frac{\Delta[D]}{\Delta t}$$

All'inizio della reazione, quando le concentrazioni di A e B sono elevate, gli urti sono più numerosi e la reazione avverrà sicuramente con velocità maggiore; al procedere della reazione è naturale aspettarsi una diminuzione della velocità fino al suo graduale arresto.

Sperimentalmente è possibile stabilire una dipendenza della velocità dalla concentrazione dei reagenti, e la relazione a cui si giunge, detta **equazione cinetica**, ha una forma del tipo:

$$v = k[A]^\alpha [B]^\beta$$

dove k è detta **costante di velocità** e assume un valore caratteristico per ogni reazione a una definita temperatura. Infatti, in base all'equazione di Arrhenius, la costante di velocità può essere espressa come:

$$k = A \cdot e^{-\frac{E_a}{RT}}$$

relazione che ne evidenzia la dipendenza dalla temperatura e dall'energia di attivazione della reazione considerata.

Ogni reazione è caratterizzata da un **ordine di reazione** globale, dato dalla somma degli esponenti con cui le concentrazioni delle specie reagenti compaiono nell'equazione cinetica. L'**ordine di reazione rispetto ad un reagente** è, invece, l'esponente con cui questo compare nell'equazione cinetica; ovviamente, maggiore è il suo valore, maggiore è l'influenza della sua concentrazione sulla velocità di reazione.

Questi esponenti coincidono poche volte con i coefficienti stechiometrici delle specie A e B, mentre nel caso più generale i valori di α e β sono diversi dai coefficienti stechiometrici e devono essere determinati sperimentalmente. Non è raro osservare, infatti, che uno o più reagenti non compaiono affatto nell'equazione cinetica e, quindi, possiedono ordine di reazione nullo, pur partecipando alla reazione: ciò significa semplicemente che anche aumentandone la concentrazione, la velocità di reazione non mostra significativi cambiamenti. Situazioni di questo tipo possono verificarsi quando la reazione in esame non è una reazione elementare, ma avviene attraverso stadi successivi, di cui uno solo, il più lento, determina la velocità di reazione; è ovvio che, in questo caso, la velocità può essere influenzata solo dalla concentrazione dei reagenti coinvolti nello stadio lento della reazione.

Spesso per avere un'idea efficace dell'ordine di grandezza della velocità di una reazione chimica viene utilizzato il **tempo di dimezzamento** $t_{1/2}$, definito come il tempo necessario affinché la concentrazione di un reagente si riduca alla metà del suo valore iniziale. Il valore del tempo di dimezzamento è inversamente proporzionale alla costante di velocità, e la relazione tra i due si ricava integrando l'equazione cinetica della reazione dal tempo zero al tempo $t_{1/2}$ a cui corrispondono le concentrazioni C_0 e $C_0/2$. La tabella che segue riporta le equazioni cinetiche e i corrispondenti tempi di dimezzamento per i diversi ordini di reazione.

Ordine di reazione	Equazione cinetica	Tempo di dimezzamento
zero	$v = -\frac{dc}{dt} = k$	$t_{1/2} = \frac{C_0}{2k}$
primo	$v = -\frac{dc}{dt} = kc$	$t_{1/2} = \frac{\ln 2}{k}$
secondo	$v = -\frac{dc}{dt} = kc^2$	$t_{1/2} = \frac{1}{C_0 k}$

Tab. 1

3 La catalisi

Da quanto detto finora si deduce che la velocità di una reazione dipende da vari fattori:

- la natura dei reagenti e dei prodotti, che incide sul valore dell'energia di attivazione;
- la temperatura, che influenza la velocità attraverso vari meccanismi, uno dei quali è l'aumento dell'energia cinetica media dei reagenti;
- la concentrazione dei reagenti, la cui influenza dipende dall'esponente con cui ciascuno compare nell'equazione cinetica;
- il mezzo di reazione, il quale può facilitare l'incontro tra i reagenti e, per le reazioni che avvengono in fase eterogenea (ad esempio tra liquidi e solidi, o tra solidi e gas), l'estensione dell'area di contatto.

Si osservi, tuttavia, che in alcuni casi la presenza nell'ambiente di reazione di una sostanza totalmente estranea al processo, che non viene modificata dalla reazione, può cambiarne notevolmente la velocità. Se la sostanza determina un aumento della velocità di reazione è detta **catalizzatore**, se ne determina un rallentamento è definita **inibitore**.

Si dice **catalizzatore** una specie chimica che, presente in piccole quantità nell'ambiente di reazione, è in grado di aumentarne la velocità, uscendo chimicamente inalterato dalla reazione stessa.

Per questo motivo il catalizzatore non compare mai nell'equazione chimica della reazione ed esplica la sua azione già a bassissime concentrazioni. I catalizzatori sono dotati generalmente di una buona selettività, intervenendo solo sulla specifica reazione, e svolgono la loro azione cambiando il meccanismo attraverso il quale la medesima si svolge, ossia facendole percorrere una via alternativa, caratterizzata da stati intermedi diversi, a cui corrisponde un'energia di attivazione minore. Nonostante ciò, lo stato iniziale e finale della reazione restano inalterati e, quindi, la resa della reazione, che dipende unicamente da parametri termodinamici, rimane immutata.

Ridisegnando il profilo di reazione in presenza di un catalizzatore, l'energia di attivazione risulta diminuita, e quindi la velocità di reazione diventa maggiore.

Se il catalizzatore si trova nello stesso stato fisico dei reagenti si parla di **catalisi omogenea**, mentre se si trova in uno stato fisico diverso si parla di **catalisi eterogenea**. In questo secondo caso, un aspetto di grande importanza è rappresentato dall'estensione e dalle caratteristiche

della superficie del catalizzatore, il quale spesso agisce assorbendo su di sé le specie reagenti e indebolendo i legami interessati dalla reazione.

Un tipo di catalisi di altissima specificità ed efficienza, che arriva ad aumentare la velocità di reazione fino a fattori di 10^9 , è la **catalisi enzimatica**, particolarmente importante nei sistemi biologici. Gli **enzimi** sono molecole proteiche dotate di strutture complesse, finalizzate alla catalisi di una specifica reazione; la specificità della catalisi enzimatica si basa spesso sulla complementarità della forma dell'enzima e di quella dei reagenti, che vengono accolti nel suo sito attivo e portati a contatto tra loro (di qui la metafora della *chiave-serratura*, spesso impiegata nella descrizione del meccanismo della catalisi enzimatica).

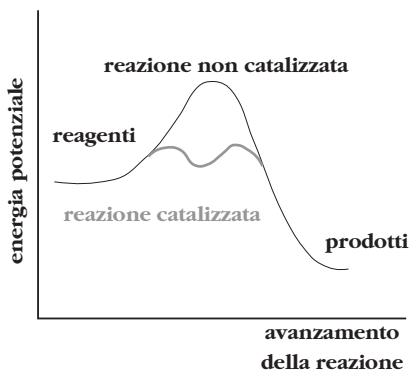


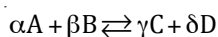
Fig. 2

13

L'equilibrio chimico

1 Gli stati fisici degli equilibri chimici

Parlando di trasformazioni chimiche abbiamo già accennato al fatto che una reazione possa svolgersi in modo completo, portando alla totale trasformazione dei reagenti in prodotti di reazione, o in modo parziale. Riferiamoci alla reazione generica:



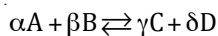
Immaginiamo di porre a reagire le specie A e B; inizialmente la reazione ha luogo soltanto verso destra, quindi le molecole A e B reagiscono generando i prodotti C e D. Man mano che cresce la concentrazione dei prodotti, comincia ad avvenire anche la reazione inversa e, quindi, i prodotti C e D si ricombinano tra loro rigenerando i reagenti A e B. La velocità della reazione inversa aumenta in misura proporzionale alla concentrazione delle specie prodotte C e D, fino ad eguagliare la velocità della reazione diretta. A questo punto la reazione è giunta all'equilibrio e le concentrazioni delle specie reagenti e prodotte si mantengono costanti nel tempo.

Da quanto detto dovrebbe apparire chiaro che l'**equilibrio chimico** è un equilibrio dinamico in cui, in modo continuo nel tempo, i reagenti si combinano tra loro per formare i prodotti e questi si ricombinano tra loro per rigenerare i reagenti. In tal modo, la concentrazione delle singole specie coinvolte nell'equilibrio si mantiene costante nel tempo.

Gli equilibri chimici possono stabilirsi in **fase omogenea**, quando tutti i reagenti e i prodotti di reazione si trovano nello stesso stato fisico; o in **fase eterogenea**, quando le specie coinvolte nell'equilibrio si trovano in stati fisici diversi. La trattazione che segue riguarda gli equilibri in fase omogenea; successivamente, estenderemo i risultati ottenuti agli equilibri in fase eterogenea.

2 La legge dell'azione di massa

Ogni *reazione di equilibrio* è governata da una **costante** caratteristica, la cui espressione deriva dalla **legge dell'azione di massa**. Per il generico equilibrio in fase omogenea:



questa può essere enunciata come segue:

«A temperatura costante, è costante anche il rapporto tra il prodotto delle concentrazioni delle specie prodotte e quello delle concentrazioni delle specie reagenti presenti all'equilibrio, ognuna elevata al rispettivo coefficiente stechiometrico».

$$K_c = \frac{[C]^\gamma \cdot [D]^\delta}{[A]^\alpha \cdot [B]^\beta} \quad (1)$$

Per gli equilibri in fase gassosa, le concentrazioni di equilibrio possono essere sostituite dalle pressioni parziali di equilibrio dei singoli componenti, e la *costante* assume la forma:

$$K_p = \frac{p_c^\gamma \cdot p_D^\delta}{p_A^\alpha \cdot p_B^\beta} \quad (2)$$

Un modo alternativo per esprimere la *costante di equilibrio* è nei termini delle frazioni molari di equilibrio dei componenti:

$$K_\chi = \frac{\chi_c^\gamma \cdot \chi_D^\delta}{\chi_A^\alpha \cdot \chi_B^\beta} \quad (3)$$

Queste costanti sono tutte correlate tra loro; assumendo per le specie gassose presenti all'equilibrio un comportamento ideale, ad ognuna può essere applicata l'equazione di stato dei gas perfetti, in base alla quale si ha:

$$PV = nRT \quad \text{ovvero} \quad P = \frac{n}{V}RT = cRT$$

dove c rappresenta la concentrazione molare. Sostituendo, nella relazione (2), ad ogni pressione parziale il valore ricavato dall'equazione di stato, si ottiene:

$$K_p = \frac{[C]^\gamma (RT)^\gamma [D]^\delta (RT)^\delta}{[A]^\alpha (RT)^\alpha [B]^\beta (RT)^\beta} = K_c (RT)^{(\gamma+\delta)-(\alpha+\beta)} = K_c (RT)^{\Delta n}$$

Una relazione analoga può essere ricavata tra la K_p e la K_χ ; ricordando che la pressione parziale di ciascun componente di una miscela gassosa è dato dal prodotto della sua frazione molare per la pressione totale, ossia $p_i = \chi_i P$, si ottiene:

$$K_p = \frac{\chi_c^\gamma P^\gamma \chi_D^\delta P^\delta}{\chi_A^\alpha P^\alpha \chi_B^\beta P^\beta} = K_\chi P^{(\gamma+\delta)-(\alpha+\beta)} = K_\chi P^{\Delta n}$$

L'esponente Δn che compare in queste relazioni rappresenta la variazione subita dal numero di moli gassose in seguito alla reazione. Le relazioni ottenute evidenziano come i valori delle costanti K_p , K_c e K_χ siano coincidenti solo per quelle reazioni in cui non si verifica una variazione del numero di moli gassose.

3 L'influenza della pressione sull'equilibrio gassoso

Il valore della costante di equilibrio K_p di una reazione dipende unicamente dalla temperatura, di cui a breve vedremo l'influenza; nonostante ciò, per le reazioni che comportano una variazione del numero di moli gassose, la composizione di equilibrio può risultare modificata a seconda del valore della pressione. Per comprendere questo è sufficiente riesaminare la relazione tra K_p e K_χ :

$$K_p = K_\chi P^{\Delta n}$$

Se non viene modificata la temperatura, il valore della costante K_p resta inalterato ed è dato dal prodotto di due fattori, il K_x , che rende conto della composizione dell'equilibrio ed assume valore tanto più grande quanto più l'equilibrio è spostato verso destra, e $P^{\Delta n}$. Dovendo rimanere costante il loro prodotto, l'aumentare di un termine comporta la diminuzione dell'altro.

L'influenza che la pressione ha sulla composizione di equilibrio dipende essenzialmente dal segno di Δn ed è inversa a seconda che il numero di moli gassose aumenti o diminuisca passando dallo stato iniziale allo stato finale.

Δn	K_p	P	K_x	Spostamento dell'equilibrio
> 0	$K_p = K_x P^{\Delta n}$	aumenta	diminuisce	←
> 0		diminuisce	aumenta	→
< 0	$K_p = \frac{K_x}{P^{\Delta n}}$	aumenta	aumenta	→
< 0		diminuisce	diminuisce	←

Tab. 1

Da quanto riportato in tabella risulta chiaro che tutte le reazioni che avvengono con un aumento del numero di moli gassose sono favorite dalla diminuzione della pressione; se, viceversa, il numero di moli diminuisce, la reazione è favorita dall'aumento della pressione.

4 La dipendenza della costante di equilibrio dalla temperatura

Il valore della costante di un equilibrio chimico può subire variazioni in seguito a variazioni di temperatura. Osserviamo, senza ricavarla teoricamente, l'equazione che esprime la dipendenza della costante di equilibrio dalla temperatura, nota come equazione di Van't Hoff. I valori K_{p_1} e K_{p_2} che la costante di equilibrio assume a due diverse temperature T_1 e T_2 sono tali che:

$$\ln \frac{K_{p_2}}{K_{p_1}} = \frac{\Delta H}{R} \left(\frac{T_2 - T_1}{T_1 \cdot T_2} \right)$$

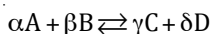
A seconda del segno del ΔH di reazione si può prevedere un diverso effetto della temperatura sulla costante di equilibrio. Per le reazioni endotermiche ($\Delta H > 0$), un aumento di temperatura determina un aumento della costante di equilibrio: infatti, essendo $\Delta H > 0$ e $(T_2 - T_1) > 0$, anche il

loro prodotto è positivo e quindi $\ln \frac{K_{p_2}}{K_{p_1}} > 0$; poiché il logaritmo assume valore positivo se il suo

argomento è maggiore di 1, ciò comporta che $K_{p_2} > K_{p_1}$. Viceversa, una diminuzione di temperatura determina una diminuzione della costante e, dunque, sfavorisce la reazione. Le reazioni esotermiche, avendo $\Delta H < 0$, si comportano in modo inverso e sono favorite dalla diminuzione di temperatura e sfavorite dal suo incremento.

5 Lo spostamento dell'equilibrio

Consideriamo il generico equilibrio in fase omogenea:



La sua costante K può essere espressa come:

$$K_c = \frac{[C]^\gamma \cdot [D]^\delta}{[A]^\alpha \cdot [B]^\beta}$$

Ponendo a reagire quantità qualsiasi dei reagenti A e B, il sistema evolverà in un tempo più o meno lungo, dipendente dalla velocità di reazione, fino al raggiungimento di concentrazioni di A, B, C e D tali da soddisfare il valore della costante di equilibrio.

In ogni istante t di questo intervallo di tempo, il sistema è caratterizzato da un quoziente di reazione Q avente forma analoga a quella della costante di equilibrio, ma che differisce da questa perché i valori delle concentrazioni che vi compaiono non sono i valori delle concentrazioni di equilibrio, bensì quelli che ciascuna specie possiede nell'istante t considerato:

$$Q = \frac{[C]^\gamma \cdot [D]^\delta}{[A]^\alpha \cdot [B]^\beta}$$

Nel corso della reazione il valore del quoziente di reazione cambia fino ad eguagliare, ad equilibrio raggiunto, il valore della costante. La costante termodinamica di equilibrio è il parametro che «governa» l'equilibrio chimico e viene soddisfatta sempre dal sistema, a prescindere dalle concentrazioni di partenza. Analogamente, se al sistema in equilibrio viene aggiunto o sottratto uno dei reagenti o dei prodotti, questo risponde alla perturbazione subita in modo da ripristinare dei valori di concentrazione (o di pressione) che soddisfino la costante di equilibrio.

Immaginiamo di aggiungere al sistema in equilibrio una concentrazione ΔA del reagente A. Il quoziente di reazione, dopo l'aggiunta, diventa:

$$Q = \frac{[C]^\gamma \cdot [D]^\delta}{[A + \Delta A]^\alpha \cdot [B]^\beta}$$

e il suo valore è inferiore a quello della costante di equilibrio, poiché la concentrazione ΔA aggiunta incrementa il denominatore del quoziente di reazione. Il sistema risponde a tale perturbazione «spostando l'equilibrio verso destra»; con questa dizione si intende che una certa quantità di A e di B reagiscono per formare C e D, incrementando così il numeratore del quoziente di reazione e diminuendo nello stesso tempo il denominatore, fino a riportarlo al valore della costante di equilibrio K .

Analogamente, se si sottrae al sistema in equilibrio una certa quantità di uno dei prodotti, il quoziente di reazione diminuisce rispetto al valore della costante di equilibrio e la reazione dovrà procedere verso destra per riportare il quoziente di reazione al valore della costante di equilibrio. Possiamo, quindi, concludere che:

- aggiungendo un reagente o sottraendo un prodotto l'equilibrio si sposta a destra;
- sottraendo un reagente o aggiungendo un prodotto l'equilibrio si sposta a sinistra.

Per questo motivo, le reazioni in cui uno o più prodotti si sviluppano allo stato gassoso, abbandonando l'ambiente di reazione, decorrono in modo completo.

6 Gli equilibri eterogenei

Gli **equilibri** che coinvolgono specie in diversi stati fisici vengono detti **eterogenei**. A questo genere di equilibri possono appartenere diverse tipologie di reazioni: *reazioni in fase gassosa* in cui siano coinvolti liquidi o solidi puri, oppure *reazioni in soluzione* che coinvolgono specie gassose o specie solide.

Anche questi equilibri possono essere trattati attraverso la *legge dell'azione di massa*, ma nell'espressione della *costante di equilibrio* non compaiono specie liquide o solide che partecipino a reazioni in fase gassosa né specie solide coinvolte in equilibri in soluzione. Per gli equilibri eterogenei che coinvolgono sia specie gassose sia specie in soluzione, la costante ha una *forma «mista»*, in cui compaiono le concentrazioni per le specie in soluzione e le pressioni (in atm) per i componenti gassosi.

1 Definizioni scientifiche degli acidi e delle basi

a) Definizione di Arrhenius

Nel 1887 Svante Arrhenius fornì una prima definizione di acidi e basi, in seguito a studi sulla dissociazione elettrolitica. In base alla sua teoria:

«Viene definito **acido** una specie chimica che in soluzione acquosa libera ioni idrogeno H^+ ».

Gli acidi possono distinguersi in **monoprotici**, se possono liberare un solo protone o ione H^+ ; e **poliprotici**, se possono liberare più di uno ione H^+ . In questo caso i protoni non vengono persi con un unico atto reattivo ma attraverso equilibri di dissociazione successivi, ognuno governato dalla propria costante di equilibrio.

Se la dissociazione del protone è *totale*, l'acido è detto **forte**; se la dissociazione è *parziale*, l'acido è detto **debole**. Per un acido poliprotico la dissociazione del primo protone può essere sia parziale sia totale, mentre le dissociazioni successive alla prima sono sempre parziali.

ESEMPIO 1

Acido forte	$HCl \rightarrow H^+ + Cl^-$
Acido debole	$CH_3COOH \rightleftharpoons CH_3COO^- + H^+$
Acido poliprotico	$H_2SO_4 \rightarrow HSO_4^- + H^+$
	$HSO_4^- \rightleftharpoons SO_4^{2-} + H^+$

A sua volta:

«Viene definita **base** una specie chimica che in soluzione acquosa libera uno o più ioni OH^- ».

Si distinguono basi **monoacide** e **poliacide**, a seconda che possano liberare uno o più ioni OH^- . La distinzione tra basi **forti** e **deboli** rimarca quanto detto per gli acidi; le *basi poliacide forti* hanno solo la prima dissociazione completa, mentre quelle successive sono parziali.

ESEMPIO 2

Base forte	$NaOH \rightarrow Na^+ + OH^-$	
Base debole poliacida	$Zn(OH)_2 \rightleftharpoons Zn(OH)^+ + OH^-$	K_1
	$Zn(OH)^+ \rightleftharpoons Zn^{2+} + OH^-$	K_2

b) Definizione di Brönsted-Lowry

La teoria di J.N. Brönsted e T.M. Lowry (1923) interpreta il comportamento di acidi e basi in modo totalmente diverso da quella di Arrhenius. Un punto di partenza fondamentale di questa teoria è la considerazione che lo ione H^+ , costituito praticamente da un protone nudo, genera intorno a sé un campo elettrico talmente elevato da non poter rimanere isolato in soluzione.

Di conseguenza, una specie chimica dotata di ioni H^+ può comportarsi da **acido** e, quindi, cedere il suo protone solo se in presenza di una specie chimica in grado di accettarlo. Questa specie «accettrice» di protoni è detta **base**. In quest'ottica:

«Viene definita **acido** una specie chimica in grado di cedere ioni H^+ se posta in presenza di una **base**, ossia di una specie chimica in grado di accettarli».

Nella teoria di Brönsted-Lowry non si concepiscono più acidi e basi come specie chimiche indipendenti e di comportamento opposto, né si riesce a darne una definizione separata. Il concetto di specie acida o di specie basica lascia il posto a quello di **sistema acido-base**, in cui il comportamento acido di una specie è necessariamente accompagnato dal comportamento basico di un'altra. La dissociazione di un acido in acqua rappresenta un sistema acido-base in cui l'acido cede il suo protone all'acqua, che, in questo caso, si comporta da base.

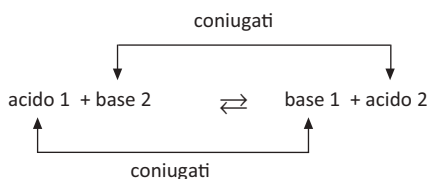
ESEMPIO 3

Acido forte $HCl + H_2O \rightarrow H_3O^+ + Cl^-$

Acido debole	CH_3COOH	+	H_2O	$\rightleftharpoons$	CH_3COO^-	+	H_3O^+	K_a
	acido		base		base coniugata		acido coniugato	

La cessione di un protone trasforma l'acido nella corrispondente **base coniugata**, ossia la forma «deprotonata» dell'acido che può, in opportune condizioni, comportarsi da base riacquistando il protone ceduto e riformando l'acido di partenza. Analogamente, in seguito all'acquisto di un protone, la base si trasforma nel suo **acido coniugato**, vale a dire la forma «protonata», che, a sua volta, in opportune condizioni, può comportarsi da acido cedendo il protone acquistato e rigenerando la base di partenza.

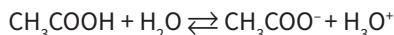
Un generico equilibrio acido-base può sempre essere scritto come:



Le **coppie coniugate** acido1-base1 e base2-acido2 sono coppie di specie chimiche che differiscono tra loro solo per un protone; nell'equilibrio precedente, la coppia1 si comporta da acido, la coppia2 si comporta da base.

Nella teoria di Brönsted-Lowry, il comportamento acido o basico di una specie chimica non è determinato dalla presenza del solvente (come l'acqua nella teoria di Arrhenius), quanto dalla presenza di una specie in grado di generare un secondo sistema coniugato acido-base. In quest'ottica, una stessa specie chimica può comportarsi da acido o da base, a seconda del partner di reazione.

La molecola d'acqua, ad esempio, si comporta da base in presenza di acido acetico, dando l'equilibrio:



mentre si comporta da acido in presenza di ammoniaca, con la quale si stabilisce l'equilibrio:

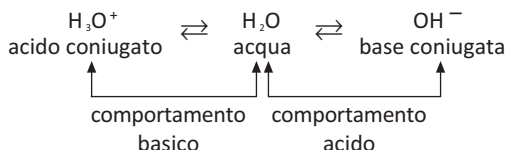


In tal modo, questa teoria rende ragione delle caratteristiche basiche che l'ammoniaca e i composti con comportamento analogo conferiscono all'acqua, cosa che non era immediatamente

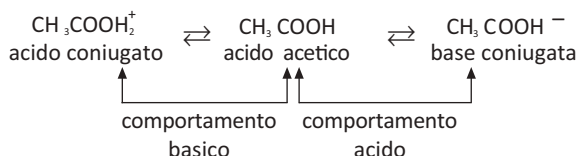
spiegabile in base alla teoria di Arrhenius. Lo stesso acido acetico, in presenza di un acido forte come l'acido perclorico HClO_4 , può comportarsi da base acquistando un protone e trasformandosi nell'acido coniugato $\text{CH}_3\text{COOH}_2^+$ (ione acidioacetato):



Quindi, il comportamento acido o basico assunto da una specie è relativo al partner di reazione:



L'acqua si comporta da base (acquistando un protone) con tutti gli acidi più forti di H_3O^+ , mentre si comporta da acido (cedendo un protone) con tutte le basi più forti di OH^- . Analogamente, per l'acido acetico possiamo osservare sia un comportamento acido sia un comportamento basico:



c) Definizione di Lewis

Una definizione ancora più generale di acidi e di basi è stata data da G.N. Lewis, che ne ha ampliato e profondamente mutato i concetti presi in considerazione fino a quel momento (la teoria di Lewis viene pubblicata nel 1923, ma comincia ad affermarsi solo quindici anni dopo).

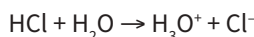
Viene definito **acido di Lewis** una specie chimica in grado di *accettare* uno o più *lone pairs*, mentre viene detta **base di Lewis** una specie chimica in grado di *fornire* uno o più *lone pairs*.

La categoria degli acidi di Arrhenius e di Brönsted-Lowry rientra in quella degli acidi e delle basi di Lewis come caso particolare, trattandosi di specie in grado di produrre un acido di Lewis, che è lo ione H^+ . Stando alla teoria di Lewis, le reazioni acido-base possono essere inquadrare nella categoria ben più ampia delle reazioni che portano alla formazione di *legami dattivi*: il datore rappresenta la base di Lewis, l'accettore è l'acido di Lewis.

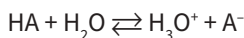
2 La forza degli acidi e delle basi

Rifacendoci alla definizione di acido e di base secondo Brönsted-Lowry, il confronto tra la forza di due acidi può essere effettuato sulla scorta della capacità di «protonare» la stessa base e, analogamente, il confronto tra la forza di due basi può effettuarsi sulla scorta della capacità di «deprotonare» lo stesso acido.

Esaminando il comportamento degli acidi in acqua, abbiamo definito **forti** gli acidi per i quali la reazione di dissociazione è totale, come ad esempio l'acido cloridrico:



Abbiamo, invece, indicato come **deboli** gli acidi che si deprotonano solo parzialmente, generando un equilibrio di dissociazione:



La costante di questo equilibrio è detta **costante acida** K_a e può essere scritta come:

$$K_a = \frac{[H_3O^+] \cdot [A^-]}{[HA]}$$

in cui non compare la $[H_2O]$ poiché, essendo l'acqua il solvente, la sua concentrazione può essere considerata costante. Maggiore è il valore della costante, più l'equilibrio di dissociazione è spostato a destra: la costante K_a di dissociazione degli acidi in acqua costituisce una misura della forza dell'acido (rispetto alla base H_2O) e può essere utilizzata per costruire, appunto, una scala di forza degli acidi.

Per gli acidi deboli, l'esistenza dell'equilibrio deriva dal fatto che, oltre alla reazione diretta ($\rightarrow$) di dissociazione dell'acido, avviene anche la reazione inversa ($\leftarrow$), assente per gli acidi forti, risultato dell'instaurarsi di una competizione tra l'acido HA, che fornisce protoni alla base H_2O , e l'acido H_3O^+ , che fornisce protoni alla base A^- . L'entità della reazione inversa è tanto più grande quanto maggiore è la competizione che si stabilisce tra le due basi (H_2O e A^-) e tra i due acidi (H_3O^+ e HA).

Possiamo, quindi, concludere che un acido HA si comporta da acido debole in acqua quando possiede acidità minore di H_3O^+ e la sua base coniugata A^- possiede elevata basicità rispetto ad H_2O , e che: *«Minore è l'acidità dell'acido HA, maggiore è la basicità della sua base coniugata A^- »*.

Il solvente acqua, agendo da base nei confronti degli acidi disciolti, ci permette di discriminare la forza degli acidi più deboli di H_3O^+ , mentre per gli acidi forti, che in acqua sono totalmente dissociati poiché donatori di protoni più forti di H_3O^+ , non risulta possibile costruire una scala di forza: questo fenomeno è detto *«effetto livellante del solvente»*.

Un discorso analogo può essere fatto per le basi; anche in questo caso distinguiamo *basi forti*, totalmente dissociate in acqua, e *basi deboli*, che generano equilibri di dissociazione:



con:

$$K_b = \frac{[NH_4^+] \cdot [OH^-]}{[NH_3]}$$

L'equilibrio che si instaura è frutto della competizione tra le due basi (NH_3 e OH^-) e tra i due acidi (H_2O e NH_4^+), e la costante di equilibrio corrispondente, detta costante basica K_b , costituisce una misura della forza della base. Anche in questo caso possiamo discriminare in acqua la forza delle sole basi più deboli di OH^- ; per tutte le altre basi la dissociazione è totale e, quindi, risultano livellate dal solvente.

Acidi poliprotici e basi poliacide danno luogo ad equilibri multipli di dissociazione in acqua, in cui i vari protoni dissociabili vengono liberati uno alla volta; questi equilibri, tutti concatenati tra loro, sono governati da costanti di equilibrio successive, il cui valore decresce di un fattore dell'ordine di 10^4 passando da una dissociazione alla successiva. Per questo motivo, in alcuni calcoli, si possono trascurare le dissociazioni successive alla prima.

ESEMPIO 4



3 Elettroliti anfoteri e autoprotolisi dell'acqua

Consideriamo il composto generico di struttura X-O-H, per il quale, in acqua, possiamo osservare tre tipi diversi di comportamento:

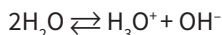
1. acido $\text{X-O-H} + \text{H}_2\text{O} \rightarrow \text{XO}^- + \text{H}_3\text{O}^+$
2. basico $\text{X-O-H} \rightarrow \text{X}^+ + \text{OH}^-$
3. anfotero $\text{X-O-H} + \text{H}_2\text{O} \begin{matrix} \nearrow \text{X}^+ + \text{OH}^- \\ \searrow \text{XO}^- + \text{H}_3\text{O}^+ \end{matrix}$

Il *comportamento acido* è determinato dalla scissione del legame O-H, il *comportamento basico* da quella del legame X-O; la scissione dell'uno o dell'altro dipende da quale dei due legami è più fortemente polarizzato e subisce, quindi, la maggiore sollecitazione alla rottura da parte dell'acqua.

L'entità della polarizzazione è, a sua volta, funzione dell'elettronegatività dell'atomo X e dell'atomo H: se l'atomo X è più elettronegativo di H, il legame che viene scisso è O-H e il composto ha comportamento acido; se, invece, l'atomo X è meno elettronegativo di H, il legame che viene scisso è X-O e il composto ha comportamento basico. Quando l'elettronegatività dei due atomi è simile, il composto assume **comportamento anfotero**.

«Un elettrolita anfotero o anfiprotico si comporta da acido debole con le basi e da base debole con gli acidi».

In questa definizione riconosciamo immediatamente l'acqua che, avendo l'atomo di ossigeno legato a due atomi uguali, rappresenta l'anfotero per eccellenza. L'acqua manifesta questo carattere anfiprotico anche con sé stessa, dando luogo al cosiddetto «**equilibrio di autoprotolisi**»:



la cui costante K_w , detta «prodotto ionico dell'acqua», è:

$$K_w = [\text{H}_3\text{O}^+][\text{OH}^-] = 10^{-14} \text{ (alla temperatura di } 25^\circ\text{C)}$$

Poiché il numero di ioni H_3O^+ e di ioni OH^- derivanti dall'equilibrio di autoprotolisi è lo stesso, le rispettive concentrazioni risultano uguali e quindi, sostituendo nell'espressione della costante la $[\text{H}_3\text{O}^+]$ con la $[\text{OH}^-]$ (o viceversa), otteniamo:

$$K_w = [\text{H}_3\text{O}^+][\text{OH}^-] = [\text{H}_3\text{O}^+]^2 = [\text{OH}^-]^2 = 10^{-14}$$

$$[\text{H}_3\text{O}^+] = [\text{OH}^-] = 10^{-7}$$

L'*autoprotolisi dell'acqua* è una *reazione endotermica*; di conseguenza, il valore della costante di equilibrio aumenta con la temperatura.

4 Il pH

La **scala del pH** rappresenta una misura del grado di acidità delle soluzioni acquose. Da quanto visto finora, in assenza di soluti a comportamento acido o basico, l'acqua contiene sempre una uguale concentrazione di ioni H_3O^+ e OH^- .

Si dice:

- **neutra**, una soluzione acquosa in cui la concentrazione degli ioni H_3O^+ e OH^- risulta uguale, il che si verifica in acqua pura e in tutte le soluzioni acquose in cui siano assenti specie a carattere acido o basico;
- **acida**, una soluzione acquosa in cui si ha una prevalenza di ioni H_3O^+ rispetto agli ioni OH^- , dovuta alla presenza di una specie acida in grado di protonare l'acqua;
- **basica**, una soluzione acquosa in cui si ha una prevalenza di ioni OH^- rispetto agli ioni H_3O^+ , dovuta alla presenza di una specie basica in grado di sottrarre protoni all'acqua o di dissociare ioni OH^- .

Poiché le concentrazioni degli ioni H_3O^+ e OH^- sono molto piccole, dell'ordine di 10^{-7} , invece della $[\text{H}_3\text{O}^+]$ si preferisce utilizzare il suo logaritmo negativo che, essendo negativo l'esponente di 10, è un numero positivo.

Viene definito **pH** il *logaritmo negativo della concentrazione degli ioni H_3O^+* e **pOH** il *logaritmo negativo della concentrazione degli ioni OH^-* :

$$\text{pH} = -\log[\text{H}_3\text{O}^+] \quad \text{pOH} = -\log[\text{OH}^-]$$

Ricordando il valore del prodotto ionico dell'acqua, $K_w = [\text{H}_3\text{O}^+][\text{OH}^-] = 10^{-14}$, e applicando la funzione $-\log$ ad entrambi i membri, si ottiene:

$$-\log K_w = -\log [\text{H}_3\text{O}^+][\text{OH}^-] = -\log 10^{-14}$$

$$\text{p}K_w = \text{pH} + \text{pOH} = 14$$

In **soluzioni neutre**: $[\text{H}_3\text{O}^+] = [\text{OH}^-] = 10^{-7} \rightarrow \text{pH} = \text{pOH} = 7$

In **soluzioni acide**: $[\text{H}_3\text{O}^+] > [\text{OH}^-]$

essendo $K_w = [\text{H}_3\text{O}^+][\text{OH}^-] = 10^{-14}$

$$[\text{OH}^-] = \frac{K_w}{[\text{H}_3\text{O}^+]}$$

Sostituendo nella disuguaglianza iniziale:

$$[\text{H}_3\text{O}^+] > \frac{K_w}{[\text{H}_3\text{O}^+]} \quad \text{da cui } [\text{H}_3\text{O}^+]^2 > 10^{-14}$$

$$[\text{H}_3\text{O}^+] > 10^{-7} \quad \rightarrow \quad \text{pH} < 7$$

In **soluzioni basiche**: $[\text{H}_3\text{O}^+] < [\text{OH}^-]$

ovvero, analogamente a prima:

$$[\text{H}_3\text{O}^+] < 10^{-7} \quad \rightarrow \quad \text{pH} > 7$$

5 Il calcolo del pH per soluzioni di acidi e basi

Il *calcolo del pH per soluzioni acquose di acidi e basi* deve essere condotto in modi diversi, a seconda della natura e della concentrazione degli acidi.

a) Acidi e basi forti diluitissime ($C_a < 10^{-6}M$)

Per *acidi e basi forti diluitissime*, nel conteggio degli ioni H_3O^+ e OH^- si deve tener conto anche degli ioni prodotti dall'autoprotolisi dell'acqua: in acqua pura $[H_3O^+] = [OH^-] = 10^{-7}$ e il contributo dell'acqua all'acidità o alla basicità delle soluzioni di acidi e basi a concentrazione ordinaria (da $0,1$ a $10^{-5}M$) è trascurabile. Dovendo calcolare il pH di una soluzione di HCl $10^{-8}M$, dobbiamo considerare da un lato la sua completa dissociazione, che produce in soluzione una $[H_3O^+] = [Cl^-] = 10^{-8}$, dall'altro l'equilibrio di autoprotolisi dell'acqua che, a causa dell'aggiunta in soluzione di ioni H_3O^+ , risulterà spostato a sinistra:

Concentrazione all'equilibrio	$2H_2O$	$\rightleftharpoons$	H_3O^+ $10^{-7} + 10^{-8} - x$	+	OH^- $10^{-7} - x$

essendo:

$$K_w = [H_3O^+][OH^-] = 10^{-14}$$

$$(10^{-7} + 10^{-8} - x)(10^{-7} - x) = 10^{-14}$$

Risolvendo l'equazione e ricavando la x abbiamo:

$$x = 2,051 \cdot 10^{-7}$$

da cui:

$$pH = -\log x = 6,68$$

Dunque un pH lievemente acido, come ci si attende per una soluzione così diluita. Se avessimo trascurato il contributo dell'acqua all'acidità totale, saremmo arrivati al paradosso di avere un pH basico (poiché $-\log [H_3O^+] = -\log [Cl^-] = -\log 10^{-8} = 8$) per una soluzione diluita di acido.

b) Acidi e basi forti diluite ($10^{-5} < C_a < 10^{-2}M$)

In questo caso il calcolo è molto semplice poiché, potendo considerare trascurabile il contributo dell'acqua, la concentrazione di ioni H_3O^+ (oppure OH^-) coincide con la concentrazione analitica dell'acido (o della base), da cui:

$$pH = -\log C_a$$

$$(o \text{ } pOH = -\log C_b)$$

c) Acidi e basi forti concentrate ($10^{-2} < C_a < 2M$)

Per *soluzioni a concentrazione superiore a $10^{-2}M$* , il calcolo rigoroso del pH richiederebbe che si tenesse conto dei «coefficienti di attività» f , ossia dei fattori correttivi della concentrazione dell'acido rispetto al valore della concentrazione analitica, che tengono conto delle interazioni esistenti tra le particelle di soluto.

A concentrazioni elevate, infatti, le interazioni non sono più trascurabili e le particelle si comportano come se la loro concentrazione in soluzione fosse minore rispetto a quella analitica (infatti $f < 1$):

$$pH = -\log f_a C_a$$

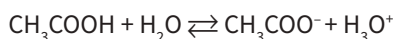
$$(o \text{ } pOH = -\log f_b C_b)$$

d) Acidi e basi deboli diluite ($10^{-4} < C_a < 10^{-1} \text{M}$)

Per soluzioni diluitissime di acidi deboli ($C_a < 10^{-4} \text{M}$), il calcolo del pH risulta matematicamente complicato, poiché per tener conto allo stesso tempo della dissociazione dell'acido e dell'auto-protolisi dell'acqua si arriva ad impostare un'equazione di 3° grado.

Per le soluzioni diluite, invece, si può trascurare il contributo dell'acqua e considerare soltanto la dissociazione dell'acido.

Calcoliamo, a titolo di esempio, il pH di una soluzione 10^{-2}M di acido acetico:



La costante di equilibrio è:

$$K_a = \frac{[\text{CH}_3\text{COO}^-][\text{H}_3\text{O}^+]}{[\text{CH}_3\text{COOH}]} \quad (1)$$

Potendo trascurare il contributo dell'acqua alla $[\text{H}_3\text{O}^+]$ all'equilibrio, questa risulterà uguale alla $[\text{CH}_3\text{COO}^-]$; inoltre la concentrazione dell'acido indissociato all'equilibrio, trascurando la quantità dissociata, che per gli acidi deboli è abbastanza piccola, può essere considerata coincidente alla concentrazione analitica iniziale C_a . Con queste approssimazioni, la precedente espressione diventa:

$$K_a = \frac{[\text{H}_3\text{O}^+]^2}{C_a} \quad (2)$$

da cui:

$$[\text{H}_3\text{O}^+] = \sqrt{K_a C_a} \quad \text{e} \quad \text{pH} = -\log \sqrt{K_a C_a}$$

Tale approssimazione può essere accettabile per soluzioni di acidi deboli, per i quali $K_a C_a \leq 10^{-5}$. Con ragionamento analogo si giunge a dimostrare che per soluzioni di basi deboli, per cui $K_b C_b \leq 10^{-5}$:

$$[\text{OH}^-] = \sqrt{K_b C_b} \quad \text{e} \quad \text{pOH} = -\log \sqrt{K_b C_b}$$

Laddove questa condizione non fosse verificata, nella relazione (2) alla concentrazione analitica dell'acido deve essere sottratta la concentrazione dissociata, coincidente con la $[\text{H}_3\text{O}^+]$ all'equilibrio; dunque, per calcolare il pH bisogna impostare e risolvere la seguente equazione di 2° grado:

$$K_a = \frac{[\text{H}_3\text{O}^+]^2}{C_a - [\text{H}_3\text{O}^+]}$$

da cui:

$$[\text{H}_3\text{O}^+]^2 + K_a [\text{H}_3\text{O}^+] - K_a C_a = 0$$

e) Acidi e basi deboli concentrate ($10^{-1} < C_a < 2 \text{M}$)

Come nel caso di acidi e basi forti, a concentrazioni elevate, le soluzioni perdono il loro comportamento ideale e le concentrazioni stesse devono essere corrette mediante fattori moltiplicativi, denominati **coefficienti di attività**.

Gli equilibri ionici in soluzione acquosa

1 L'idrolisi salina

Spesso la presenza di un sale in acqua può alterarne il valore del pH, poiché i suoi ioni possono reagire con l'acqua producendo ioni H_3O^+ oppure OH^- . Questo fenomeno, detto **idrolisi salina**, non ha nulla di diverso da quanto osservato precedentemente riguardo agli acidi e alle basi: anche le reazioni di idrolisi salina, infatti, sono reazioni acido-base tra gli ioni del sale e l'acqua, che hanno luogo quando gli ioni costituenti del sale sono dotati di proprietà acide o basiche.

a) Idrolisi basica

Consideriamo un sale contenente l'anione A^- , base coniugata dell'acido debole HA. Se disciolto in acqua, l'anione A^- esprime le sue proprietà basiche «deprotonando» l'acqua. Questa reazione, detta «**idrolisi basica**» poiché produce ioni OH^- , non è altro che una reazione acido-base tra la base A^- e l'acido H_2O :



La presenza in soluzione di un sale il cui anione sia la base coniugata di un acido debole impartisce alla soluzione un pH basico. La reazione di idrolisi è una reazione di equilibrio, la cui costante viene indicata spesso con K_i (oppure, nel caso dell'idrolisi basica, con K_b). In soluzione acquosa, a questa reazione si accompagna sempre l'equilibrio di autoprotolisi dell'acqua, regolato dalla costante K_w . Le due costanti sono sempre contemporaneamente soddisfatte e ciò, in termini matematici, significa che soddisfano un sistema di equazioni:

$$\left\{ \begin{array}{l} K_i = \frac{[\text{HA}][\text{OH}^-]}{[\text{A}^-]} \\ K_w = [\text{H}_3\text{O}^+][\text{OH}^-] \end{array} \right. \quad \text{ovvero} \quad \left\{ \begin{array}{l} K_i = \frac{[\text{HA}][\text{OH}^-]}{[\text{A}^-]} \\ [\text{OH}^-] = \frac{K_w}{[\text{H}_3\text{O}^+]} \end{array} \right.$$

Sostituendo la seconda equazione nella prima, otteniamo:

$$K_i = \frac{[\text{HA}]K_w}{[\text{A}^-][\text{H}_3\text{O}^+]} = \frac{K_w}{K_a}$$

dove K_a rappresenta la costante dell'acido debole di provenienza dell'anione A^- . La conclusione a cui siamo giunti è del tutto generale e può essere espressa come segue:

«Per qualsiasi coppia acido/base coniugata, il prodotto della costante K_a dell'acido e della costante K_b (o K_i) della base coniugata in acqua è pari al prodotto ionico dell'acqua K_w ». In formule:

$$K_w = K_a \cdot K_b$$

Consideriamo, a titolo di esempio, una soluzione 10^{-2} M di acetato di sodio CH_3COONa , un composto ionico che, in soluzione acquosa, si dissocia totalmente nei suoi ioni costituenti Na^+ e CH_3COO^- . Quest'ultimo, essendo l'anione coniugato di un acido debole, è dotato di proprietà basiche nei confronti dell'acqua e si comporta, quindi, come una base avente co-

$$\text{stante } K_i = \frac{K_w}{K_a} = K_b.$$

Il pH della soluzione di acetato si calcola esattamente come quello di una soluzione di base debole:



$$K_i = \frac{[\text{CH}_3\text{COOH}][\text{OH}^-]}{[\text{CH}_3\text{COO}^-]}$$

da cui:

$$K_i = \frac{[\text{OH}^-]^2}{[\text{CH}_3\text{COO}^-]} \quad \text{e} \quad \text{pOH} = -\log \sqrt{K_i [\text{CH}_3\text{COO}^-]}$$

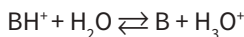
dove la concentrazione dello ione acetato all'equilibrio può essere tranquillamente approssimata con la concentrazione analitica del sale, poiché le costanti degli equilibri di idrolisi sono sempre molto piccole e la frazione idrolizzata è trascurabile rispetto alla quantità iniziale del sale. Dunque, in generale:

Il pH della soluzione di un anione che subisce idrolisi basica è dato da:

$$\text{pOH} = -\log \sqrt{K_i C_s}$$

b) Idrolisi acida

Consideriamo un sale contenente un catione BH^+ , acido coniugato di una base debole B. Analogamente a quanto visto prima, il catione BH^+ in acqua esprime la sue proprietà acide cedendo il protone all'acqua, ovvero dando «**idrolisi acida**»:



«La presenza in soluzione di un sale il cui catione sia l'acido coniugato di una base debole impartisce alla soluzione un pH acido».

L'equilibrio di idrolisi è regolato da una costante che indichiamo con K_i (o K_o , trattandosi di una idrolisi acida), esprimibile come:

$$K_i = \frac{[\text{B}][\text{H}_3\text{O}^+]}{[\text{BH}^+]} = \frac{K_w}{K_b}$$

A sua volta, il pH della soluzione può calcolarsi, analogamente a prima, come:

$$[\text{H}_3\text{O}^+] = \sqrt{K_i C_s}$$

da cui:

$$\text{pH} = -\log \sqrt{K_i C_s}$$

Quindi, in generale:

Il pH della soluzione di un catione che subisce idrolisi acida è dato da:

$$\text{pH} = -\log \sqrt{K_i C_s}$$

c) Idrolisi doppia

Se il sale disciolto in soluzione contiene sia un anione sia un catione in grado di dare idrolisi, si verifica una «**idrolisi doppia**», ovvero entrambi gli ioni reagiscono con l'acqua e il pH finale della soluzione dipende da quale dei due componenti subisce idrolisi in misura maggiore. Come si deduce dalla formula $K_w = K_a \cdot K_b$, quanto più forte è l'acido, tanto più debole è la sua base coniugata: dunque, l'idrolisi viene subito in misura maggiore dallo ione proveniente dal componente più debole. Se l'acido e la base di provenienza hanno forza comparabile, l'idrolisi dell'anione e del catione hanno all'incirca la stessa entità e la soluzione risulta neutra; se l'acido è più debole della base, prevale l'idrolisi dell'anione e la soluzione risulta basica; se la base è più debole dell'acido, prevale l'idrolisi del catione e la soluzione risulta acida.

Come tutti i problemi in cui sono coinvolti equilibri multipli, anche il calcolo del pH della soluzione di un sale che subisce idrolisi doppia è piuttosto complicato. Tuttavia, con alcune approssimazioni, si giunge alla formula semplificata, valida nella maggior parte dei casi:

$$\text{pH} = \frac{1}{2}(\text{p}K_a + \text{p}K_b)$$

2 I sistemi tampone

In alcune applicazioni pratiche e in molti processi biologici può risultare molto importante il fatto che nel corso di una reazione, la quale produce o consuma ioni H_3O^+ , il pH della soluzione si mantenga comunque costante; ciò viene realizzato attraverso i cosiddetti **sistemi tampone**.

«Un sistema tampone è l'insieme di sostanze che, nonostante l'aggiunta di ioni H_3O^+ o OH^- entro limitate quantità, è in grado di mantenere pressoché inalterato il valore del pH dell'ambiente in cui si trova».

a) Tipi di soluzioni tampone

Nella pratica di laboratorio possono essere preparate e utilizzate soluzioni dotate di questa capacità. Soluzioni che mostrano proprietà tampone sono, ad esempio, quelle costituite da un acido debole e da un suo sale (cioè un sale contenente la sua base coniugata) oppure da una base debole e da un suo sale in concentrazioni simili.

Esaminiamo, a titolo di esempio, il tampone costituito da acido acetico/acetato di sodio. In soluzione sono contemporaneamente presenti l'equilibrio di dissociazione dell'acido e l'equilibrio di idrolisi della base coniugata. Le $[\text{CH}_3\text{COOH}]$ e $[\text{CH}_3\text{COO}^-]$ che compaiono nei due equilibri sono ovviamente le stesse, trovandosi nella medesima soluzione, e hanno valore comparabile:



Poiché, in soluzione acquosa, $K_w = [\text{H}_3\text{O}^+][\text{OH}^-]$, le espressioni (1) e (2) si equivalgono; di conseguenza possiamo utilizzare uno qualsiasi dei due equilibri per descrivere la soluzione e derivarne il pH. Prendendo in considerazione la costante acida ed esplicitando il valore della $[\text{H}_3\text{O}^+]$, otteniamo:

$$[\text{H}_3\text{O}^+] = K_a \frac{[\text{CH}_3\text{COOH}]}{[\text{CH}_3\text{COO}^-]}$$

Da ciò, applicando la funzione $-\log$ ad entrambi i membri, ricaviamo:

$$\text{pH} = \text{p}K_a + \log \frac{[\text{CH}_3\text{COO}^-]}{[\text{CH}_3\text{COOH}]}$$

La contemporanea presenza in soluzione dell'acido debole e della sua base coniugata, in concentrazioni paragonabili tra loro, determina una notevole retrocessione sia dell'equilibrio di dissociazione dell'acido, sia di quello di idrolisi della base coniugata; questo autorizza a trascurare le concentrazioni di acido dissociato e di base idrolizzata rispetto alle concentrazioni iniziali, assumendo che all'equilibrio $[\text{CH}_3\text{COOH}] = C_a$ e $[\text{CH}_3\text{COO}^-] = C_s$, dove C_a e C_s rappresentano le concentrazioni analitiche dell'acido e del sale. Il pH della soluzione tampone è dato da:

$$\text{pH} = \text{p}K_a + \log \frac{C_s}{C_a}$$

nota come **formula di Henderson e Hasselbach**.

Soluzioni tampone possono essere preparate anche mescolando due sali provenienti da due dissociazioni successive di un **acido poliprotico**; ne sono esempi i tamponi $\text{HCO}_3^-/\text{CO}_3^{2-}$, $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}$, $\text{HPO}_4^{2-}/\text{PO}_4^{3-}$ e molti altri. Consideriamo il tampone $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}$, formato da monoidrogenofosfato e diidrogenofosfato. Anche in questo caso, per il calcolo del pH viene utilizzata la formula di Henderson e Hasselbach:

$$\text{H}_2\text{PO}_4^- + \text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{HPO}_4^{2-}$$

$$\text{pH} = \text{p}K_2 + \log \frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]}$$

b) Meccanismo di funzionamento delle soluzioni tampone

Una soluzione tampone possiede la capacità di mantenere inalterato il suo pH, in quanto l'aggiunta di una moderata quantità di acido (o di base) determina unicamente la trasformazione di una equivalente quantità di base coniugata A^- nell'acido HA (o viceversa), ripercuotendosi in modo trascurabile sul valore del pH.

Immaginiamo di aggiungere a un tampone di $[\text{CH}_3\text{COOH}] = [\text{CH}_3\text{COO}^-] = 0,1\text{M}$ ($K_a = 1,8 \cdot 10^{-5}$) una piccola quantità di acido cloridrico, ad esempio 0,002 moli (1/50 della concentrazione delle specie tamponanti). Ciò determina la trasformazione di 0,002 moli di CH_3COO^- in CH_3COOH . Il calcolo del pH iniziale e finale evidenzia come l'aggiunta sia ininfluente ai fini della variazione di pH.

pH iniziale:

$$\begin{aligned} (\text{tampone } [\text{CH}_3\text{COOH}] = & \quad \text{pH} = \text{p}(1,8 \cdot 10^{-5}) + \log \frac{0,1}{0,1} = \\ & [\text{CH}_3\text{COO}^-] = 0,1\text{M}) \quad \quad \quad = \text{p}(1,8 \cdot 10^{-5}) = 4,74 \end{aligned}$$

pH finale:

$$\begin{aligned}
 (\text{tampone} + 0,002 \text{ moli HCl}) \quad \text{pH} &= \text{p}(1,8 \cdot 10^{-5}) + \log \frac{C_s}{C_a} = \\
 &= 4,745 + \log \frac{0,1 - 0,002}{0,1 + 0,002} = 4,72
 \end{aligned}$$

Si osservi come l'aggiunta della stessa quantità di HCl in acqua avrebbe determinato una variazione di pH di 4,3 unità, che corrisponde ad un aumento di concentrazione degli ioni H_3O^+ di circa 20.000 volte.

$$\begin{aligned}
 \text{pH iniziale:} \quad [\text{H}_3\text{O}^+] &= [\text{OH}^-] = 10^{-7} \\
 (\text{acqua}) \quad \text{pH} &= 7
 \end{aligned}$$

pH finale:

$$(\text{tampone} + 0,002 \text{ moli HCl}) \quad \text{pH} = -\log[\text{H}_3\text{O}^+] = -\log(2 \cdot 10^{-3}) = 2,7$$

L'efficienza di un tampone può essere espressa, ovviamente, attraverso la sua *capacità tamponante*. La **capacità tamponante** di una soluzione è la variazione di pH che questa subisce, rapportata alla concentrazione di acido aggiunta.

Tale capacità è funzione della natura delle specie tamponanti e della loro quantità relativa, e non va confusa con la massima quantità di acido o di base che la soluzione può tamponare, la quale dipende invece dalla concentrazione delle specie tamponanti. Si può dimostrare matematicamente che ogni tampone realizza il **massimo della sua capacità** quando la concentrazione delle specie tamponanti è uguale. Tuttavia, il potere tamponante della soluzione può essere

considerato soddisfacente fintanto che $\frac{1}{10} \leq \frac{C_s}{C_a} \leq 10$, da cui si deduce che l'intervallo ottimale

di pH entro il quale lavora ciascun tampone risulta:

$$\text{pH} = \text{pK}_a \pm 1$$

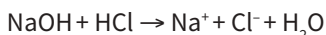
Oltre a quelle già prese in esame, altre soluzioni che manifestano proprietà tampone sono **soluzioni di amminoacidi**. L'amminoacido possiede due funzioni responsabili delle proprietà tampone: la funzione carbossilica e quella amminica, quindi risulta attivo in due campi di pH ($\text{pK}_{\text{acido}} \pm 1, \text{pK}_{\text{base}} \pm 1$).

Su un principio analogo si basano i **tamponi universali**: preparando miscele di acidi o di basi di forza diversa si possono realizzare sistemi tampone che lavorano in un campo di pH molto ampio. Un tampone universale efficiente dovrà contenere acidi i cui pK_a differiscano tra loro per non più di 2-2,5 unità, in modo che dove cessa di essere attivo un sistema tampone ne subentri immediatamente un altro.

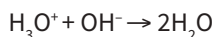
3 Le titolazioni acido-base

Effettuare una **titolazione** significa determinare la concentrazione di una specie presente in un campione materiale che si vuole analizzare, sfruttando una sua reazione specifica. Nel caso della **titolazione di acidi o basi**, si sfrutta la loro *reazione di neutralizzazione*: quando un acido e una base si trovano nella stessa soluzione, gli ioni H_3O^+ liberati dall'acido reagiscono con gli ioni OH^- liberati dalla base, generando acqua.

Ad esempio:



Oppure, omettendo gli ioni che non partecipano alla reazione:



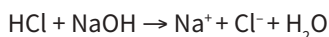
Questa **reazione**, detta, appunto, **di neutralizzazione**, è inversa rispetto all'equilibrio di autoprotolisi e quindi è fortemente spostata a destra (ricordiamo che $K_w = 10^{-14}$), tanto da poter essere ritenuta completa.

Nel corso di una titolazione di un acido con una base, alla soluzione dell'acido vengono aggiunte via via piccole quantità esattamente misurate della base. Mettendo in evidenza con un indicatore di pH o con altri mezzi il **punto di equivalenza** (o *punto equivalente*, oppure *punto finale*, o ancora *punto stechiometrico*) **della titolazione**, ossia il momento in cui la quantità di base aggiunta è esattamente equivalente a quella dell'acido presente all'inizio, si arriva a determinare la concentrazione dell'acido nel campione analizzato.

a) Titolazioni acido forte-base forte

L'andamento di una *titolazione acido-base* può essere seguito mediante la variazione del pH nel corso della titolazione: le **titolazioni acido forte-base forte** (o viceversa) sono le più semplici da questo punto di vista.

Consideriamo il caso della titolazione di HCl con NaOH: inizialmente il pH è determinato dalla concentrazione dell'acido cloridrico ($\text{pH} = -\log C_{\text{HCl}}$). Cominciando ad aggiungere soda, la *reazione di neutralizzazione*:



determina la scomparsa di una quantità di acido cloridrico uguale a quella di soda aggiunta e il pH è determinato dalla quantità di HCl residua.

Al punto di equivalenza, tutto l'acido forte ha reagito con la base forte e l'unica specie che determina il pH è l'acqua, col suo *equilibrio di autoprotolisi*. Questo equilibrio è presente anche prima e dopo il punto equivalente, ma non viene considerato nel calcolo del pH perché del tutto trascurabile nel caso di un eccesso di acido o di base. *Dopo il punto equivalente*, invece, il pH della soluzione dipende dalla quantità di soda aggiunta in eccesso rispetto all'acido. Il brusco salto di pH al punto di equivalenza della titolazione permette di impiegare un indicatore il cui pH di viraggio sia contenuto tra 4 e 10.

Inizio	$\text{pH} = -\log C_{\text{HCl}}$
Prima del punto di equivalenza	$\text{pH} = -\log \frac{C_{\text{HCl}} V_{\text{HCl}} - C_{\text{NaOH}} V_{\text{NaOH}}}{V_{\text{HCl}} + V_{\text{NaOH}}}$
Punto di equivalenza	$\text{pH} = -\log [\text{H}_3\text{O}^+] = -\log \frac{K_w}{[\text{OH}^-]} = 7$
Dopo il punto di equivalenza	$\text{pH} = -\log \frac{C_{\text{NaOH}} V_{\text{NaOH}} - C_{\text{HCl}} V_{\text{HCl}}}{V_{\text{HCl}} + V_{\text{NaOH}}}$

Tab. 1

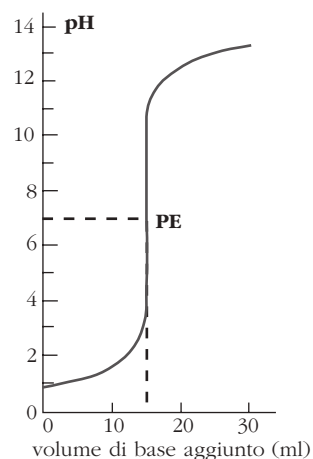
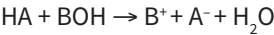


Fig. 1

b) Titolazioni acido debole-base forte

Il calcolo del pH nel corso della **titolazione di un acido debole HA con una base forte BOH** è un po' più complesso rispetto al precedente. Inizialmente la soluzione ha un pH deter-

minato soltanto dall'acido debole. Effettuando aggiunte successive di base forte, la *reazione di neutralizzazione*:



porta alla formazione della base coniugata dell'acido e in soluzione si genera il sistema tampone HA/A⁻.

Al punto di equivalenza, la quantità di base aggiunta è esattamente uguale all'acido presente all'inizio, che quindi è totalmente convertito nella sua base coniugata; il pH della soluzione, a questo punto, è determinato dalla *reazione di idrolisi* della base coniugata. *Dopo il punto di equivalenza*, invece, il pH è determinato dall'eccesso di base forte aggiunta, rispetto al quale la *reazione di idrolisi* diviene trascurabile.

Inizio	$\text{pH} = -\log \sqrt{K_a C_a}$
Prima del punto di equivalenza	$\text{pH} = \text{p}K_a + \log \frac{C_{\text{BOH}} V_{\text{BOH}}}{C_{\text{HA}} V_{\text{HA}} - C_{\text{BOH}} V_{\text{BOH}}}$
Punto di equivalenza	$\text{pOH} = -\log \sqrt{K_b \frac{C_{\text{BOH}} V_{\text{BOH}}}{V_{\text{BOH}} + V_{\text{HA}}}}$
Dopo il punto di equivalenza	$\text{pOH} = -\log \frac{C_{\text{BOH}} V_{\text{BOH}} - C_{\text{HA}} V_{\text{HA}}}{V_{\text{HA}} + V_{\text{BOH}}}$

Tab. 2

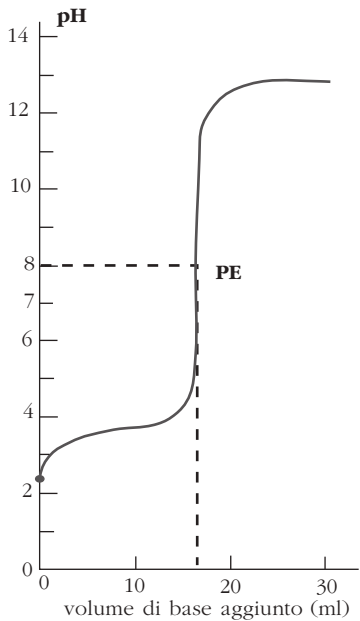


Fig. 2

c) Indicatori di pH

Abbiamo accennato alla possibilità di mettere in evidenza il raggiungimento del *punto di equivalenza* di una titolazione utilizzando degli **indicatori di pH**, ossia acidi o basi deboli, perlopiù di origine naturale, dotati di colorazioni diverse nella forma acida e nella forma basica. Negli schemi seguenti sono mostrate le strutture di alcuni tra gli indicatori di pH più comunemente utilizzati.

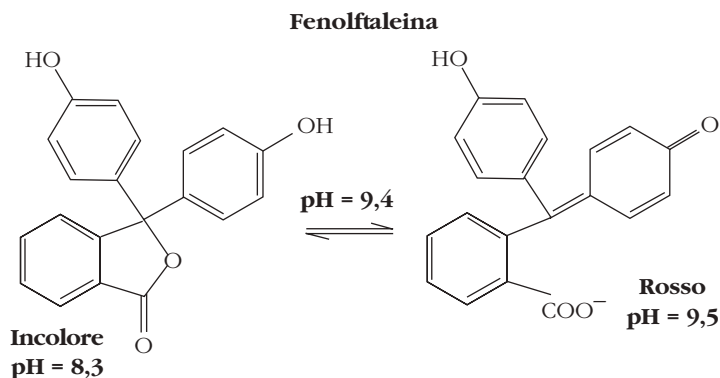


Fig. 3

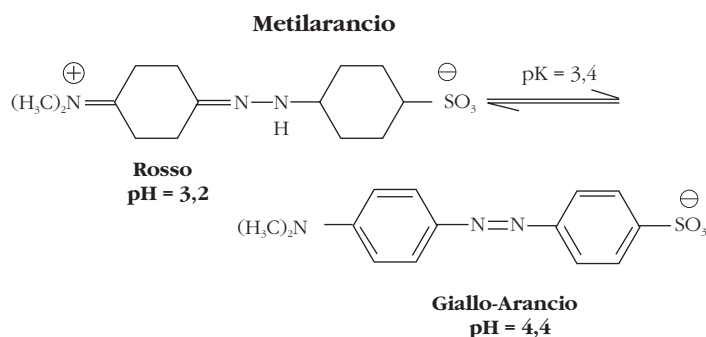


Fig. 4

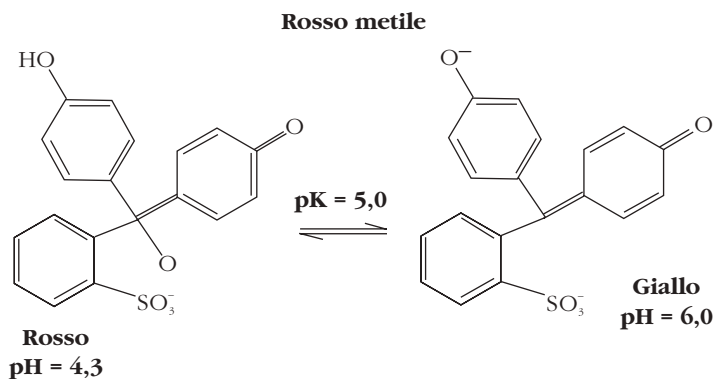
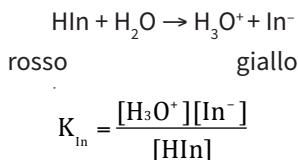


Fig. 5

Come per qualsiasi acido o base debole, la quantità relativa delle due forme coniugate HIn/In^- dipende dal pH dell'ambiente e il colore che si percepisce dipende dalla prevalenza di una forma sull'altra.

Trattandosi di sostanze acide o basiche, anche gli indicatori possiedono una propria costante di dissociazione:



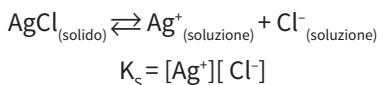
da cui:

$$[\text{H}_3\text{O}^+] = K_{\text{In}} \frac{[\text{HIn}]}{[\text{In}^-]} \quad \text{e} \quad \text{pH} = \text{p}K_{\text{In}} + \log \frac{[\text{In}^-]}{[\text{HIn}]}$$

Quando $[\text{In}^-] = 10[\text{HIn}]$, cioè a $\text{pH} = \text{p}K_{\text{In}} + 1$, il nostro occhio percepisce il colore giallo; quando $[\text{HIn}] = 10[\text{In}^-]$, cioè a $\text{pH} = \text{p}K_{\text{In}} - 1$, il colore che si percepisce è il rosso; quando, infine, $\text{pH} = \text{p}K_{\text{In}}$, il colore è intermedio tra quello della forma acida e quello della forma basica, e l'indicatore è al *viraggio*. Affinché l'indicatore possa essere impiegato per evidenziare il raggiungimento del punto di equivalenza di una titolazione, è necessario che il suo $\text{p}K$ sia il più vicino possibile al pH corrispondente al punto di equivalenza, in modo da segnalarne, appunto, il raggiungimento attraverso il suo viraggio.

4 Il prodotto di solubilità

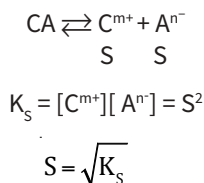
Il sistema costituito da un sale poco solubile in equilibrio, come corpo di fondo, con la sua soluzione satura è un esempio di *equilibrio eterogeneo*. Tra il sale solido e quello in soluzione esiste l'equilibrio:



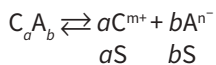
La costante di questo equilibrio è detta **prodotto di solubilità K_s** e in essa compaiono soltanto le concentrazioni degli ioni in soluzione. Il *prodotto di solubilità*, come tutte le *costanti di equilibrio*, è «costante» a temperatura costante. Inoltre, poiché il passaggio in soluzione è spesso un processo endotermico, l'aumento di temperatura accresce il valore del prodotto di solubilità e, quindi, della solubilità delle sostanze.

Conoscendo il prodotto di solubilità di una sostanza in un certo solvente, ad una fissata temperatura T , si può risalire alla **solubilità** della sostanza, ossia alla quantità di sostanza che può essere disciolta in quel solvente, alla temperatura T , in condizioni di saturazione.

La solubilità si può ricavare dal prodotto di solubilità e la relazione esistente tra le due grandezze dipende dalla stechiometria del sale. Se il rapporto tra il numero dei cationi e il numero degli anioni è 1:1, in seguito al passaggio in soluzione di ogni molecola vengono liberati un catione ed un anione. Indicando con S la solubilità, espressa come concentrazione molare di sostanza disciolta, si ha:



Per sali dotati di stechiometria più complessa, si ha:



$$K_s = [C^{m+}]^a [A^{n-}]^b = (aS)^a (bS)^b$$

$$S = \sqrt[a+b]{\frac{K_s}{a^a b^b}}$$

Dal valore del prodotto di solubilità si può stabilire, inoltre, se si possa verificare o meno precipitazione del composto poco solubile, a partire da una soluzione dei suoi ioni: la concentrazione degli ioni in soluzione, infatti, deve sempre soddisfare il prodotto di solubilità. Quando questa concentrazione eccede il valore del prodotto di solubilità, si verifica la precipitazione di una quantità di sale tale da riportare le concentrazioni entro i limiti stabiliti dal K_s .

1 I legami tra atomi di carbonio

Abbiamo potuto constatare come l'**atomo di carbonio** sia un elemento del tutto particolare. Rispetto a quanto prevedibile in base alla sua configurazione elettronica nello stato fondamentale (s^2p^2 , con due soli elettroni disaccoppiati), il carbonio sembra formare un numero di legami superiore a quello atteso. Infatti, promuovendo un elettrone s nell'orbitale p vuoto e portandosi in uno stato di energia leggermente superiore, il carbonio acquisisce la capacità di formare quattro *legami covalenti* con l'evidente guadagno di poter formare due legami in più e di raggiungere l'ottetto (formando solo due legami, il carbonio arriverebbe a sei elettroni).

Un'altra caratteristica estremamente importante di questo elemento è quella di potersi legare a sé stesso, generando lunghe *catene di atomi lineari, ramificate e cicliche*.

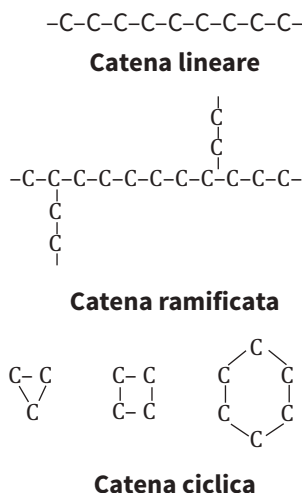


Fig. 1

Il carbonio può formare *legami semplici, doppi o tripli*, generando strutture di diversa geometria: il carbonio ibridizzato sp^3 forma quattro legami semplici, disposti ai vertici di un tetraedro; il carbonio ibridizzato sp^2 forma tre legami, di cui due semplici e uno doppio, disposti ai vertici di un triangolo equilatero; il carbonio ibridizzato sp , infine, forma due legami, entrambi doppi, oppure un legame semplice e uno triplo, che giacciono sulla stessa retta. Ciò dà vita ad un numero illimitato di strutture molecolari di forme, dimensioni, proprietà e funzioni estremamente variabili; varietà che ha permesso, tra l'altro, lo sviluppo dei sistemi viventi, la cui sorprendente complessità poteva essere realizzata unicamente da un elemento così versatile.

2 I gruppi funzionali

Tutte le specie chimiche che possiedono uno scheletro più o meno esteso di atomi di carbonio concatenati tra loro vengono dette **molecole organiche** e, nell'ambito di queste, una categoria particolare è costituita dalle **biomolecole**, le molecole, cioè, che costituiscono i sistemi viventi.

Come abbiamo visto, le molecole organiche possono essere rappresentate attraverso *diversi tipi di strutture*:

- la **formula bruta**, che descrive unicamente la composizione qualitativa e quantitativa della molecola, senza fornire alcuna indicazione su come gli atomi presenti possano essere concatenati tra loro;
- la **formula di struttura**, la quale, invece, indica sia la composizione sia la reale struttura molecolare, evidenziando come i singoli atomi siano legati tra loro;
- le **formule razionali** (ovvero un misto dei due tipi precedenti), in cui alcuni gruppi di struttura nota sono indicati attraverso la loro formula bruta, mentre altre parti caratteristiche della molecola in esame sono rappresentate evidenziandone la struttura.

L'uso di *formule di struttura* e di *formule razionali* si è reso necessario in **chimica organica** in quanto un fenomeno frequentemente osservato è quello dell'**isomeria strutturale**. Si dicono *isomeri strutturali* due o più composti che rispondono alla stessa formula bruta, ma possiedono formule di struttura diverse. L'unico modo per distinguere due isomeri strutturali è quello di scrivere le formule di struttura. Essendo formati da lunghi concatenamenti di atomi di carbonio, i composti organici sono facilmente raggruppabili in **serie omologhe** di composti, ovvero insiemi di molecole di struttura e proprietà simili, che rispondono alla medesima formula generale. I termini di una serie omologa di composti, in pratica, contengono gli stessi elementi, il cui rapporto è costante in ogni termine della serie, e differiscono tra loro unicamente per il numero di atomi di carbonio; in genere due termini consecutivi della serie omologa differiscono tra loro sempre della stessa unità strutturale (perlopiù un gruppo $-\text{CH}_2-$).

La presenza di gruppi di particolare struttura, che conferiscono alle molecole determinate proprietà chimiche, consente di effettuare una suddivisione dei composti organici in *classi funzionali*.

In tabella sono riportati i principali **gruppi funzionali** organici e i nomi della **classe di composti** corrispondente.

GRUPPO FUNZIONALE		Classe di composti
Struttura	Nome	
$-\text{C}=\text{C}-$	doppio legame	alcheni
$-\text{C}\equiv\text{C}-$	triplo legame	alchini
$-\text{OH}$	ossidrile	alcoli enoli fenoli
$-\text{O}-$	etereo	eteri
$\begin{array}{c} \text{H} \\ \\ -\text{C} \\ \\ \text{O} \end{array}$	aldeidico	aldeidi
$\begin{array}{c} \diagup \\ \text{C}=\text{O} \\ \diagdown \end{array}$	chetonico	chetoni
$\begin{array}{c} \text{OH} \\ \\ -\text{C} \\ \\ \text{O} \end{array}$	carbossile	acidi carbossilici
$\begin{array}{c} \text{O} \\ \\ -\text{C} \\ \\ \text{O} \end{array}$	estereo	esteri
$-\text{NH}_2$	amminico	ammine primarie
$-\text{NH}-$	amminico	ammine secondarie
$\begin{array}{c} \text{NH} \\ \\ -\text{C} \\ \\ \text{O} \end{array}$	ammidico	ammidi proteine

Tab. 1

3 Gli idrocarburi

Gli **idrocarburi** sono i composti organici più semplici, costituiti, come dice il nome, unicamente da carbonio e idrogeno. Nonostante ciò, rappresentano una classe molto ampia, dalla quale si possono far derivare (formalmente e, in alcuni casi, anche praticamente) tutti gli altri composti organici.

Nell'ambito degli idrocarburi, una prima grande suddivisione può essere fatta tra:

- **idrocarburi saturi**, in cui tutti gli atomi di carbonio sono ibridizzati sp^3 e formano quattro legami semplici;
- **idrocarburi insaturi**, contenenti anche carboni ibridizzati sp^2 ed sp , quindi doppi e tripli legami.

a) Alcani

Gli *idrocarburi saturi* sono detti **alcani**, o **paraffine**. La formula generale di questi composti è $C_n H_{2n+2}$ e il primo termine della serie omologa è il metano CH_4 , che contiene un solo atomo di carbonio. Come tutti i composti organici, anche gli alcani possono possedere *scheletri carboniosi lineari, ciclici o ramificati*.

• Alcani lineari

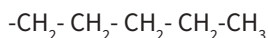
Idrocarburo		Radicale	
metano	CH_4	metile	$-CH_3$
etano	C_2H_6	etile	$-C_2H_5$
propano	C_3H_8	propile	$-C_3H_7$
butano	C_4H_{10}	butile	$-C_4H_9$
pentano	C_5H_{12}	pentile	$-C_5H_{11}$
esano	C_6H_{14}	esile	$-C_6H_{13}$
eptano	C_7H_{16}	eptile	$-C_7H_{15}$
ottano	C_8H_{18}	ottile	$-C_8H_{17}$
nonano	C_9H_{20}	nonile	$-C_9H_{19}$
decano	$C_{10}H_{22}$	decile	$-C_{10}H_{21}$
undecano	$C_{11}H_{24}$	undecile	$-C_{11}H_{23}$
dodecano	$C_{12}H_{26}$	dodecile	$-C_{12}H_{25}$
...	...	...	...
icosano	$C_{20}H_{42}$	icosile	$-C_{20}H_{41}$

Tab. 2

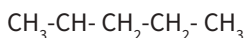
La nomenclatura degli **alcani lineari** è alla base di quella di tutti i composti organici: il nome dell'alcano lineare è formato da un prefisso che, ad esclusione dei primi quattro termini della serie, indica il numero di atomi di carbonio contenuti nella catena, e dal suffisso **-ano**, caratteristico della classe.

Quando ad uno dei carboni della catena di un alcano si toglie un atomo di idrogeno, questo dispone di un legame libero, attraverso il quale legarsi ad altre catene di carboni; si ottiene, dunque, un **radicale alchilico** il cui nome viene derivato da quello dell'alcano, sostituendo il suffisso **-ano** con il suffisso **-ile**. L'atomo dal quale è stato tolto l'idrogeno deve essere indicato prima del nome del radicale, attraverso un numero che ne individua la posizione nella catena.

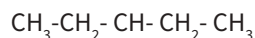
Di seguito sono mostrate le strutture dell'1-pentile, del 2-pentile e del 3-pentile, i tre radicali alchilici ottenibili dal pentano:



1-pentile



2-pentile



3-pentile

I radicali che si otterrebbero togliendo l'idrogeno al quarto e quinto carbonio sono analoghi all'1-pentile e al 2-pentile. Spesso per radicali come l'1-pentile, in cui l'idrogeno viene tolto

da un carbonio terminale della catena, l'indicazione 1- è omessa e il radicale viene chiamato *normal-pentile* o *n-pentile*.

- Alcani ciclici

Gli **alcani ciclici** contengono tutti atomi di carbonio ibridizzati sp^3 . Tuttavia, affinché la catena possa chiudersi su sé stessa, i due carboni terminali devono perdere due atomi di idrogeno e stabilire un legame C-C, motivo per il quale la loro formula generale è $C_n H_{2n+2}$. Il nome del *cicloalcano* si fa derivare da quello dell'alcano con lo stesso numero di carboni, antepo-ndonogli il prefisso **ciclo-**. Ovviamente il primo termine della serie dei cicloalcani è il *ciclopropano* (anello a 3 termini), seguito da *ciclobutano*, *ciclopentano*, *cicloesano* etc.

- Alcani ramificati

Gli **alcani ramificati** vengono considerati come derivati da *alcani lineari*, in cui alcuni atomi di idrogeno sono stati sostituiti con *radicali alchilici* (detti per questo *sostituenti*). Per denominare gli alcani a catena ramificata è necessario procedere come spiegato di seguito.

- Individuare la catena di atomi di carbonio più lunga, che darà il nome di base alla molecola, e attribuirle la numerazione, in modo da assegnare alla prima *ramificazione* che si incontra il numero più basso possibile. Se, numerando la catena a partire da un estremo o dall'altro, il primo punto di ramificazione assume sempre lo stesso numero, si deve preferire la numerazione per la quale la somma delle posizioni di tutti i sostituenti è minore. Nell'esempio proposto, la catena più lunga contiene sei carboni, dunque il nome di base della molecola sarà *esano*. La numerazione deve essere attribuita cominciando da sinistra e, in tal modo, le prime ramificazioni si trovano in posizione 2; altrimenti, se fossimo partiti da destra, la prima ramificazione sarebbe stata in posizione 3.

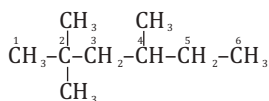


Fig. 2

- Attribuire il nome a tutte le ramificazioni presenti (contandone il numero di carboni e utilizzando il nome del radicale alchilico corrispondente), sistemarle in ordine alfabetico e anteporle al nome della catena principale, precedute dal numero corrispondente all'atomo di carbonio della catena a cui sono legate. Se sono presenti più ramificazioni dello stesso tipo, il nome della ramificazione porta un prefisso moltiplicativo (di-, tri-, tetra-), a seconda del loro numero, e le posizioni che queste occupano nella catena sono elencate tutte insieme, prima del nome della ramificazione, separate da virgole. Nel nostro esempio sono presenti tre radicali metilici, in posizione 2,2,4. Il nome del composto è, quindi, *2,2,4-trimetilesano*.

b) Alcheni e alchini

Alcheni e *alchini* appartengono alla classe degli *idrocarburi insaturi*.

Gli **alcheni** sono caratterizzati dalla presenza di doppi legami C=C e rispondono alla formula generale $C_n H_{2n}$; vengono denominati con gli stessi criteri utilizzati per gli alcani, utilizzando il suffisso **-ene**, caratteristico di questa classe di composti.

Gli **alchini**, invece, sono contraddistinti dalla presenza di tripli legami C≡C e la loro formula generale è $C_n H_{2n-2}$; vengono denominati utilizzando il suffisso **-ino**.

Sia per gli alcheni sia per gli alchini i termini più piccoli della serie contengono due atomi di carbonio, il numero minimo necessario a formare un legame multiplo.

Gli idrocarburi insaturi contenenti più di tre atomi di carbonio presentano *isomeria strutturale* legata alla posizione del legame multiplo lungo la catena: per questo, nel dare il nome al composto, si rende necessario specificarne la posizione. Il criterio per attribuire il nome ad un idrocarburo insaturo è quello di individuare la catena più lunga contenente il legame multiplo e numerarla in modo da attribuirgli la posizione più piccola possibile; questa viene indicata con il più piccolo dei numeri dei due carboni uniti dal legame multiplo. Gli *alcheni*, inoltre, presentano un secondo tipo di isomeria, dovuta al fatto che nel legame doppio C=C si perde la libera rotazione tipica dei legami semplici C-C e i sostituenti dei carboni occupano delle posizioni fisse rispetto al doppio legame.

Consideriamo, a titolo di esempio, il *butene*, l'alchene contenente quattro carboni. Il doppio legame può trovarsi sia tra il C1 e il C2, sia tra il C2 e il C3: nel primo caso abbiamo l'*1-butene*, nel secondo caso il *2-butene*. Quest'ultimo esiste, a sua volta, in forma di due isomeri differenti: il *cis-2-butene*, in cui i gruppi metilici si trovano dalla stessa parte del doppio legame, e il *trans-2-butene*, in cui i due metili si trovano da parti opposte.

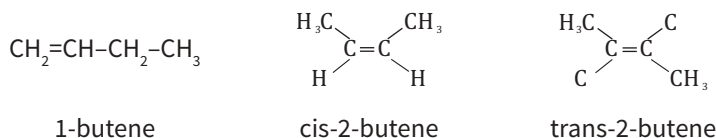


Fig. 3

Anche negli alchini la presenza del triplo legame impedisce la rotazione attorno al legame $\text{C}\equiv\text{C}$, tuttavia non si assiste ad isomeria *cis-trans* a causa della linearità della molecola.

Alchene		Radicale		Alchino		Radicale	
etilene o etene	C_2H_4	etilenile o etenile	$-\text{C}_2\text{H}_3$	acetilene o etino	C_2H_2	etinile	$-\text{C}_2\text{H}$
propilene o propene	C_3H_6	propilenile o propenile	$-\text{C}_3\text{H}_5$	propino	C_3H_4	propinile	$-\text{C}_3\text{H}_3$
butilene o butene	C_4H_8	butilenile o butenile	$-\text{C}_4\text{H}_7$	butino	C_4H_6	butinile	$-\text{C}_4\text{H}_5$
pentene	C_5H_{10}	pentenile	$-\text{C}_5\text{H}_9$	pentino	C_5H_8	pentinile	$-\text{C}_5\text{H}_7$
esene	C_6H_{12}	esenile	$-\text{C}_6\text{H}_{11}$	esino	C_6H_{10}	esinile	$-\text{C}_6\text{H}_9$

Tab. 3

c) Idrocarburi aromatici

Un'importante classe di **idrocarburi** è rappresentata dai **composti aromatici**, caratterizzati dalla delocalizzazione degli elettroni π . La molecola più conosciuta nell'ambito di questa classe è sicuramente il *benzene*, un idrocarburo di formula C_6H_6 , di cui abbiamo discusso proprio trattando del legame ad elettroni delocalizzati (Cap. 5, §6).

I composti aromatici, in realtà, sono numerosi e hanno strutture variegata su cui non ci soffermeremo. Ricordiamo soltanto che i derivati del benzene, ottenuti sostituendo uno o più idrogeni dell'anello benzenico con radicali di varia natura, sono ancora composti aromatici. Quando si introducono più sostituenti, a seconda delle posizioni relative che questi occupano, si possono distinguere *isomeri orto-*, *meta-* e *para-*.

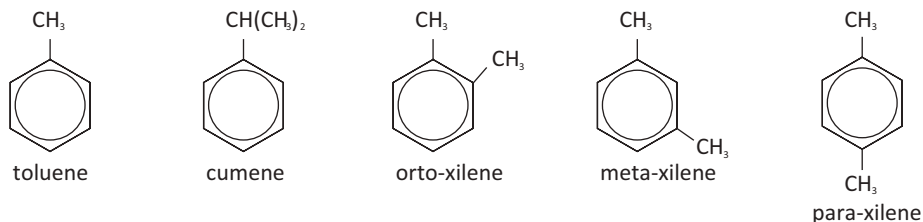


Fig. 4

Analogamente, composti derivanti dalla condensazione di più anelli benzenici sono a loro volta idrocarburi aromatici.

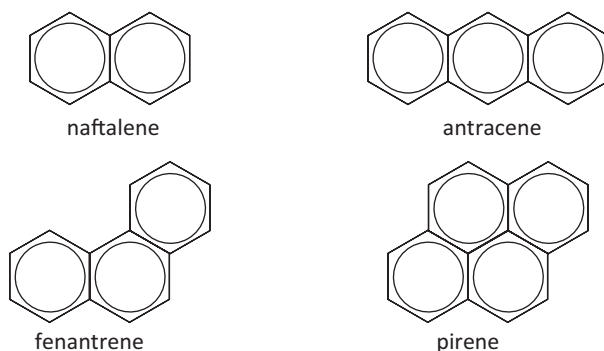


Fig. 5

I *radicali aromatici* vengono denominati in modo analogo agli altri *radicali idrocarburi*. Il radicale costituito da una parte *alifatica* (cioè una parte idrocarburi non aromatica) e una *aromatica* viene denominato come un *radicale alifatico* sostituito se il legame libero è sulla parte alifatica, come un *radicale aromatico* se il legame libero è sull'anello.

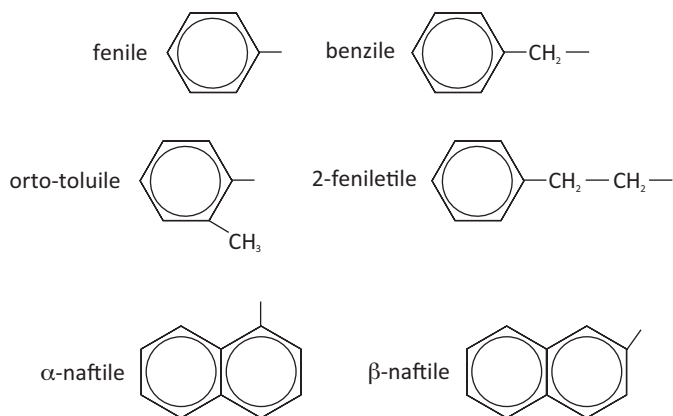


Fig. 6

4 Alcoli, fenoli ed eteri

La presenza del *gruppo idrossilico* -OH in una molecola organica dà vita a due grandi famiglie di composti:

- gli **alcoli**, quando il gruppo idrossilico è legato a una catena alifatica;
- i **fenoli**, quando il gruppo idrossilico è legato direttamente a un anello benzenico.

Gli *alcoli* sono denominati come derivati degli idrocarburi, sostituendo la -o finale con il suffisso **-olo**.

La posizione del gruppo idrossilico nella catena viene indicata attraverso il numero del carbonio che porta l'idrossile, al quale viene attribuito il valore più piccolo possibile. Un tipo di nomenclatura alternativa utilizza la parola «alcool», seguita dal nome del radicale che porta il gruppo -OH.

$\text{CH}_3\text{-OH}$	metan olo o alcool metilico	$\text{CH}_3\text{-CH}_2\text{-CH}_2\text{-OH}$	1-propan olo o alcool propilico
$\text{CH}_3\text{-CH}_2\text{-OH}$	etan olo o alcool etilico	$\begin{array}{c} \text{OH} \\ \\ \text{CH}_3\text{-CH-CH}_3 \end{array}$	2-propan olo o alcool isopropilico
	alcool benzilico	$\begin{array}{c} \text{CH}_2=\text{CH-CH}_2\text{-CH-CH}_3 \\ \\ \text{OH} \end{array}$	4-penten-2- olo

Fig. 7

Quando nella molecola sono presenti più gruppi alcolici, il nome porta l'indicazione della posizione dei gruppi, seguita dal suffisso *-diolo*, *-triolo* etc.

$\text{HO-CH}_2\text{-CH}_2\text{-OH}$		$\begin{array}{c} \text{OH} \\ \\ \text{HO-CH}_2\text{-CH}_2\text{-CH}_2\text{-OH} \end{array}$
1,2-etandiol	1,4-cicloesandiol	1,2,3-propantriolo o glicerolo

Fig. 8

Gli **eteri**, invece, sono composti in cui l'ossigeno è legato a due radicali organici R di natura alifatica o aromatica. Tali composti possono essere *simmetrici*, quando i due radicali R sono uguali; o *asimmetrici*, quando i due radicali sono diversi. Gli eteri vengono denominati semplicemente facendo seguire la parola «etere» ai nomi dei radicali R e R' legati all'ossigeno.

$\text{CH}_3\text{-O-CH}_3$	dimetiletere o etere metilico	
$\text{CH}_3\text{-CH}_2\text{-O-CH}_2\text{-CH}_3$	dietiletere o etere etilico	$\begin{array}{c} \text{CH}_3 \\ \\ \text{CH}_3\text{-O-CH} \\ \\ \text{CH}_3 \end{array}$
$\text{CH}_3\text{-O-CH}_2\text{-CH}_3$	etilmetiletere	2-propil-metiletere

Fig. 9

5 Aldeidi, chetoni e acidi carbossilici

Queste famiglie di composti sono tutte accomunate dalla presenza del **gruppo carbonilico** >C=O . Nelle **aldeidi**, il carbonile è legato a un idrogeno e a un radicale organico, ed è quindi un gruppo terminale della catena. Nei **chetoni**, il carbonile è legato a due radicali organici e, per-

tanto, si trova all'interno di una catena di carboni. Negli **acidi carbossilici**, infine, il carbonile è legato a un radicale organico e a un idrossile; di conseguenza, anche il gruppo carbossilico è sempre un gruppo terminale.



aldeide



chetone



acido carbossilico

Fig. 10

Nella nomenclatura IUPAC, individuata la catena principale della molecola, questi composti vengono indicati sostituendo alla -o finale del nome dell'alcano il suffisso **-ale** per le aldeidi, **-one** per i chetoni, **-ico** per gli acidi. Per i chetoni deve essere indicata la posizione del gruppo carbonilico, attribuendogli il numero più piccolo possibile. Il nome degli acidi è preceduto sempre, ovviamente dal termine «acido».

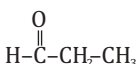
Aldeidi



metanale



etanale

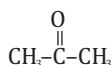


propanale

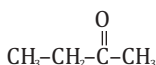


butanale

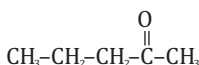
Chetoni



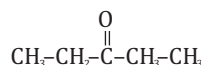
propanone



butanone



2-pentanone



3-pentanone

Acidi carbossilici

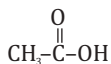
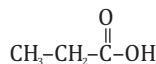
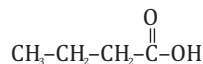

 acido
metanoico
o acido formico

 acido etanoico
o acido acetico

 acido propanoico
o acido propionico

 acido butanoico
o acido butirrico

Fig. 11

6 Gli esteri

Gli **esteri** sono composti derivati dagli *acidi carbossilici* per sostituzione dell'idrogeno acido con un radicale organico e possono essere preparati attraverso la *reazione di condensazione* (reazione di addizione di due molecole con eliminazione di una molecola d'acqua) di un acido carbossilico con un alcool.

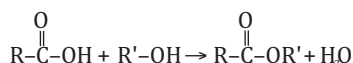


Fig. 12

Alla classe degli esteri appartengono composti di enorme diffusione e importanza come i **grassi** e gli **oli**. Chimicamente queste sostanze, indicate spesso come **trigliceridi**, sono esteri del glicerolo (o 1,2,3-propantriolo) con acidi dotati di lunghe catene alifatiche, detti *acidi grassi*.

La distinzione tra oli e grassi è basata sullo stato fisico in cui questi si trovano nelle condizioni ambientali: gli *oli* sono liquidi, mentre i *grassi* sono solidi. Ciò è riconducibile alla natura degli acidi grassi, che sono *insaturi* (quindi dotati di punto di fusione più basso) negli oli e *saturi* (dunque con punto di fusione più alto) nei grassi.

Come tutti i composti che derivano da reazioni di condensazione, anche gli esteri possono subire **idrolisi**, cioè la reazione inversa della condensazione, in cui le molecole si dividono, riacquistando la molecola d'acqua che avevano perso. L'idrolisi dei grassi in ambiente basico produce i **saponi**, reazione detta per questo *saponificazione*. I saponi sono sali sodici di acidi grassi a lunga catena, dotati, in particolare, di una «catena» *idrofoba* e di una «testa» *idrofila*, costituita dall'anione carbossilato. Questa duplice natura dei saponi rende ragione del loro impiego nella detergenza: disciolti in acqua ad una concentrazione sufficiente, si organizzano in strutture micellari (come mostrato in figura 14) in cui le code idrofobe sono disposte all'interno di una sfera, sulla cui superficie si trovano, invece, le teste idrofile. La micella è in grado di sciogliere al suo interno le particelle di grasso e di trasportarle nell'acqua.

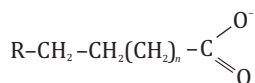


Fig. 13 - Molecola di un sapone

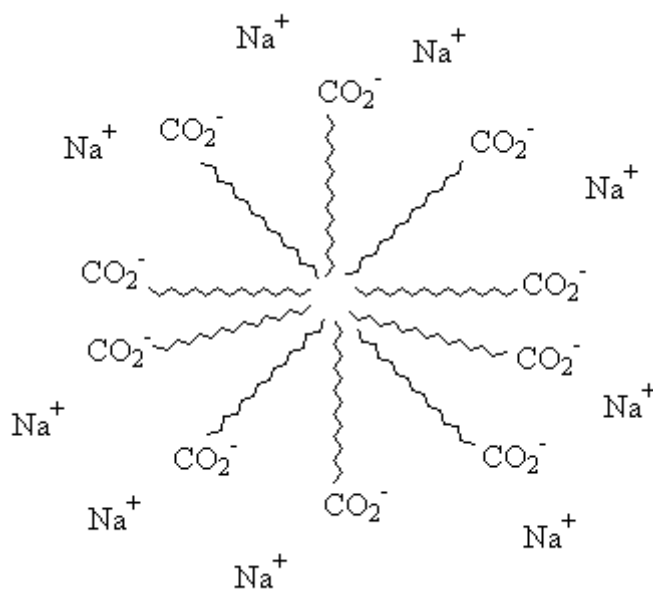


Fig. 14

7 I composti azotati

Fra i composti organici contenenti azoto vanno citate le **ammine**, sostanze formalmente derivate dall'ammoniaca per sostituzione di uno o più idrogeni con radicali organici. In base al numero di idrogeni sostituiti, si distinguono *ammine primarie* ($\text{R}-\text{NH}_2$), *secondarie* (R_2-NH)

e terziarie (R_3-N). Le **ammidi** sono composti azotati in cui l'azoto è legato ad un carbonio carbonilico.

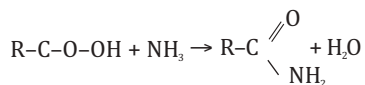
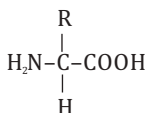


Fig. 15

L'ammide è il prodotto della condensazione di un acido con l'ammoniaca o con un'ammina; questa reazione è di estrema importanza in natura, poiché da essa dipende la formazione delle *catene polipeptidiche*, cioè delle *proteine*. Le proteine sono, infatti, il prodotto della condensazione di un gran numero di **amminoacidi**, molecole polifunzionali in quanto dotate contemporaneamente di una *funzione amminica* e di una *carbossilica*.

Gli amminoacidi che costituiscono le proteine naturali sono tutti α -*amminoacidi*, ossia molecole in cui la funzione amminica e quella carbossilica si trovano sullo stesso atomo di carbonio.


 Fig. 16 - Struttura generale di un α -amminoacido

Gli *amminoacidi naturali* sono in tutto 20 e differiscono tra loro per la natura del gruppo $-R$.

Le **proteine** vengono originate dalla condensazione del gruppo amminico di un amminoacido con il gruppo carbossilico di un altro amminoacido: nella proteina così generata, gli amminoacidi alle estremità della catena sono detti *N-terminale* e *C-terminale*, poiché uno possiede ancora il gruppo amminico libero e l'altro il gruppo carbossilico.

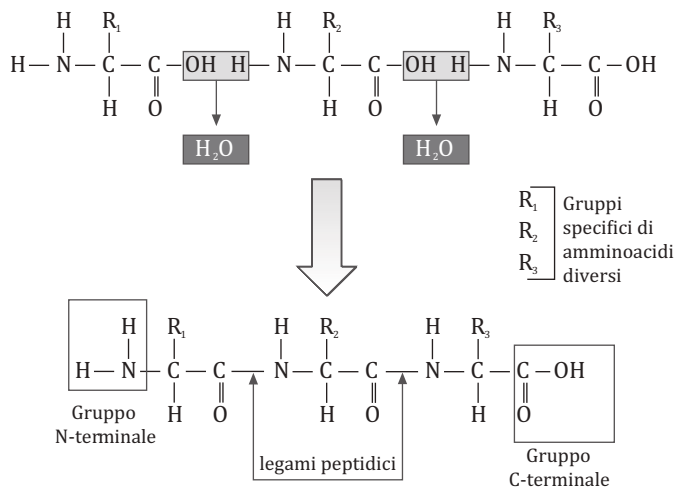


Fig. 17

1 I materiali e le loro proprietà

Tutti gli oggetti che ci circondano sono composti da uno o più **materiali**. Se è vero che con lo stesso materiale si possono ottenere svariati oggetti (col legno, ad esempio, si producono le matite, le sedie, le porte etc.), è anche vero che diversi oggetti possono essere composti da materiali differenti per caratteristiche meccaniche, tecnologiche, fisiche ed economiche.

La ricerca tecnologica crea di continuo **nuovi materiali** che ne vanno a sostituire altri perché migliori nelle prestazioni e nei costi: le vele delle barche, una volta di tela, oggi sono in materiale sintetico (kevlar e mylar); i fari delle autovetture, un tempo di vetro, oggi sono di plastica etc.

Gli oggetti costruiti dall'uomo sono composti da diversi materiali a seconda del loro uso specifico: vi immaginate cosa succederebbe se al posto dei vetri le vostre finestre fossero chiuse da pannelli di legno? E se il vostro martello fosse fatto di gommapiuma? E se le vostre scarpe fossero di carta?

Ciascun materiale ha delle **caratteristiche proprie**, delle qualità che lo rendono più o meno adatto ad un determinato scopo. **Queste proprietà si dividono in tre tipi:**

- *proprietà meccaniche;*
- *proprietà tecnologiche o lavorative;*
- *proprietà fisiche.*

Le **proprietà meccaniche** di un materiale ci rivelano quanto questo sia capace o meno di **resistere a sollecitazioni esterne**. Esse sono riportate nella seguente tabella.

PROPRIETÀ MECCANICHE	DESCRIZIONE	NOTE TECNICHE
Durezza	È la capacità del materiale di resistere alla scalfittura, ossia alla penetrazione di una punta	Devono essere «duri» i materiali sottoposti ad urti (come le piastrelle) o che formano attrezzi che devono scalfire oggetti (come un chiodo o la punta di un trapano)
Resistenza alle sollecitazioni	È l'attitudine del materiale a resistere alle forze che agiscono su di esso, le quali provocano sollecitazioni cosiddette di taglio, flessione, trazione, compressione, torsione (Fig. 1)	Tutte le parti «strutturali», ossia che costituiscono l'ossatura portante, di un oggetto devono poter resistere alle sollecitazioni (si pensi, ad esempio, al telaio di una bicicletta, ai pilastri di una casa, alle ruote di un'auto etc.)
Resilienza	Il suo opposto è la fragilità: è resiliente un materiale che non si rompe o fessura improvvisamente e bruscamente	Un tipico materiale fragile e non resiliente è il vetro. È capace di resistere alle sollecitazioni, tuttavia ha il grosso limite di rompersi improvvisamente e senza preavviso

(Segue)

PROPRIETÀ MECCANICHE	DESCRIZIONE	NOTE TECNICHE
Elasticità	È l'attitudine di un materiale ad assumere nuovamente la sua forma originaria una volta cessata l'azione della forza esterna. Il suo opposto è la plasticità, ossia la tendenza di un materiale a deformarsi e mantenere la forma acquisita se sottoposto a deformazioni	Gli pneumatici delle automobili devono potersi piegare per poi recuperare la forma originaria, così come la gomma per cancellare, o un pallone di calcio
Resistenza alla fatica	È la capacità di resistere a sollecitazioni ripetute	Se provate a piegare ripetutamente un filo di ferro vedrete che alla fine, dopo aver a lungo insistito, si romperà anche se la sollecitazione non è così forte

Tab. 1

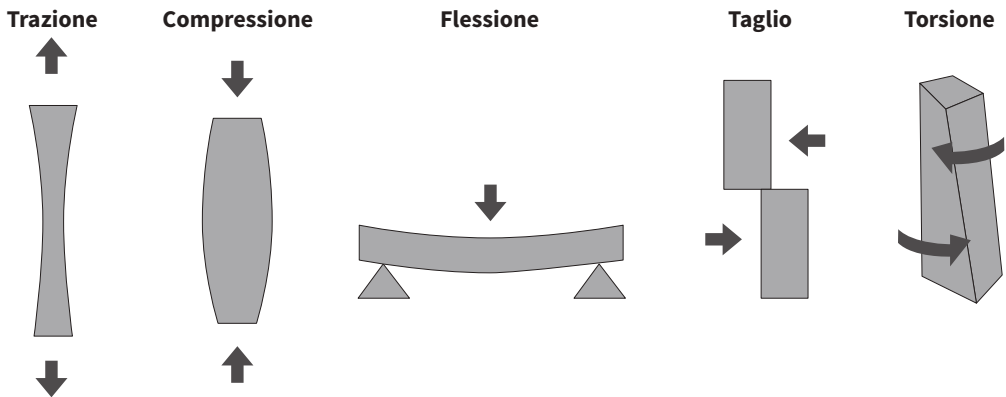


Fig. 1 - I diversi tipi di sollecitazioni a cui sono sottoposti i materiali

Le **proprietà tecnologiche** di un materiale ci dicono quanto sia propenso ad essere **lavorato e manipolato**. Sono riportate nella tabella seguente.

PROPRIETÀ TECNOLOGICHE	DESCRIZIONE	NOTE TECNICHE
Malleabilità	Attitudine di un materiale a farsi ridurre in fogli sottili	Alcuni metalli sono molto malleabili. Uno di essi è l'oro, tanto che già nel Medioevo i miniatori realizzavano le loro opere adoperando sottilissimi strati di foglia d'oro spesso meno di mezzo millimetro
Temperabilità	Alcuni materiali aumentano le proprie qualità meccaniche se vengono riscaldati e poi raffreddati di colpo	Succede ad esempio al vetro e all'acciaio. I vetri temperati permettono di ottenere chiusure trasparenti e allo stesso tempo difficili da rompere
Duttilità	Attitudine di un materiale ad essere ridotto in fili	Caratteristica utile ad ottenere fili, cavi metallici e fibre tessili
Fusibilità	Capacità di un materiale di diventare liquido se portato ad alte temperature	Tra i materiali fusibili ricordiamo i metalli ed il vetro: grazie a questa loro caratteristica, essi possono essere colati in stampi per poi acquisirne la forma una volta solidificati

Tab. 2

Le **proprietà fisiche** di un materiale riguardano il suo **comportamento dinanzi a calore, elettricità e forza di gravità**. Sono riportate qui di seguito.

PROPRIETÀ FISICHE	DESCRIZIONE	NOTE TECNICHE
Peso specifico	Peso dell'unità di volume del materiale	Un materiale può essere più o meno «leggero»: se è molto leggero sarà particolarmente adatto a galleggiare o comunque servirà per far diminuire il peso di un oggetto
Conduttività elettrica	Propensione del materiale a trasmettere corrente elettrica	Alcuni materiali sono usati come «isolanti», ossia proteggono dall'elettricità perché quest'ultima non è capace di passarli attraverso. Altri materiali invece, come il rame ed i metalli in genere, lasciano passare l'energia elettrica con gran facilità
Conduttività termica	Propensione a trasmettere, cioè a far passare, il calore	Alcuni materiali, come il sughero o la lana di vetro, oppongono forte resistenza al passaggio del caldo e del freddo: verranno dunque adoperati come isolanti. Altri invece, come l'acciaio o la ghisa, non isolano dal caldo/freddo e vengono adoperati, ad esempio, nei radiatori
Temperatura di fusione	Temperatura alla quale il materiale da solido diventa liquido, ossia fonde	I materiali si «sciolgono» a diverse temperature. Quando arrivano a queste temperature mutano, dunque, le loro caratteristiche
Inalterabilità	Capacità di non deteriorarsi sotto l'azione di agenti esterni, quali pioggia, inquinamento etc.	I materiali, specie quelli da costruzione, devono poter mantenere inalterate le loro caratteristiche altrimenti rischiano, a lungo andare, di non essere più idonei all'uso che se ne vorrebbe fare

Tab. 3

2 Dalle risorse ai prodotti

Non tutti i materiali che costituiscono gli oggetti che ci circondano si trovano in natura così come sono. Perfino il legno, a ben pensarci, in natura si trova sotto forma di tronco d'albero, mentre nelle nostre case è utilizzato soprattutto in pannelli piani, ben levigati e tagliati.

Ma nel caso del legno, se non altro, è intuitivo capire dove si deve andare a cercare la materia prima: nei boschi. Non lo è affatto per altri materiali, come ad esempio l'alluminio, che in natura non esiste. Lo si ricava da un minerale chiamato bauxite, contenuto in particolari rocce metallifere. Restando al caso dell'alluminio, le rocce metallifere costituiscono la **risorsa naturale**, ossia il primo passo verso la realizzazione del prodotto finito. Sono risorse naturali gli alberi per il legno, le cave di pietra per il marmo, ancora gli alberi per la carta. L'alluminio, una volta estratto dalla bauxite, costituisce la **materia prima** che subisce una prima lavorazione grossolana: viene trasformata, ad esempio, in lamina per poterne fare poi delle lattine. La lamina di alluminio è un **semilavorato**, qualcosa che non è più semplicemente materia prima, ma che non è nemmeno prodotto finito. A che serve una lamina di alluminio? Così com'è, a niente. Ecco dunque che la lamina, il semilavorato, viene ulteriormente trasformato, fino a dargli la forma della lattina, che costituisce il **prodotto finito**.

Dunque le materie prime devono subire diverse lavorazioni prima di divenire prodotto finito. Queste lavorazioni prendono il nome di **operazioni unitarie**.

Tra le più importanti vi sono le operazioni di *laminazione*, *estrusione* e *stampaggio*; ma ve ne sono tante altre comunque indispensabili a creare gli oggetti che ci circondano: *saldatura*, *tornitura*, *calandratura*, *taglio al laser* etc.

La **laminazione** è il processo per cui un materiale viene ridotto in lamine (cioè fogli sottili) e si attua nel laminatoio, un macchinario composto da due pesanti rulli, tra i quali passa il pezzo da lavorare.

L'**estrusione** è il processo mediante il quale si ottengono lunghe strisce di uguale sezione: sono estrusi i serramenti in alluminio delle nostre abitazioni o le «pignatte» in terracotta che costituiscono i pavimenti. Per capire il concetto di estrusione, pensate al tubetto di maionese che avete nel frigo: schiacciando il tubetto la maionese passa attraverso il buco a forma di stellina ed esce sotto forma di striscia, anch'essa con sezione a stellina. Questo è un classico esempio di processo di estrusione: un materiale fluido viene pressato contro una matrice opportunamente sagomata, in modo che dalla parte opposta fuoriesca il materiale con una sezione uguale a quella della matrice.

Lo **stampaggio** è un processo simile a quello che usavamo sulla spiaggia da piccoli per creare le formine di sabbia: c'è uno stampo all'interno del quale una pressa spinge del materiale, più o meno liquido, che una volta solidificato prende la forma dello stampo.

Per quanto riguarda gli altri tipi di lavorazione, in rapida carrellata ricordiamo:

- la **saldatura**, ossia il procedimento mediante il quale vengono unite in modo definitivo due strutture di metallo;
- la **tornitura**, ovvero l'operazione che permette di ottenere forme con simmetria di rotazione (i vasi, le gambe delle sedie in legno etc.);
- la **calandratura**, cioè il processo attraverso cui le lamine metalliche assumono forme curve e cilindriche;
- il **taglio a laser a controllo numerico**, vale a dire il processo mediante il quale una lamiera viene tagliata con estrema precisione (secondo un disegno realizzato al computer) da un raggio laser.

3 Le materie seconde

Si è parlato di *materie prime* e di *prodotti finiti*; ma sia durante la produzione degli oggetti, sia alla fine del loro ciclo di utilizzo, si creano degli **scarti**, ossia dei residui inutilizzati.

Ad esempio, mentre si produce un tavolo di legno, si creano molti trucioli. Questi non vengono gettati bensì **riutilizzati**: mescolati a speciali colle e pressati con forza generano pannelli cosiddetti *truciolati*, meno pregiati di quelli di vero legno (*legno massello*) ma comunque utilissimi.

Il caso dei trucioli è un tipico esempio di *scarto di produzione*, ma è anche vero che i prodotti una volta utilizzati lasciano comunque spesso degli *scarti inutilizzabili*: i **rifiuti**.

Se fino a qualche anno fa non c'era una diffusa sensibilità al problema dei rifiuti, oggi molte città hanno intrapreso la strada della **raccolta differenziata**: non si butta più tutto nello stesso cestino della spazzatura, ma si raccolgono separatamente alluminio, vetro e carta perché si sa già

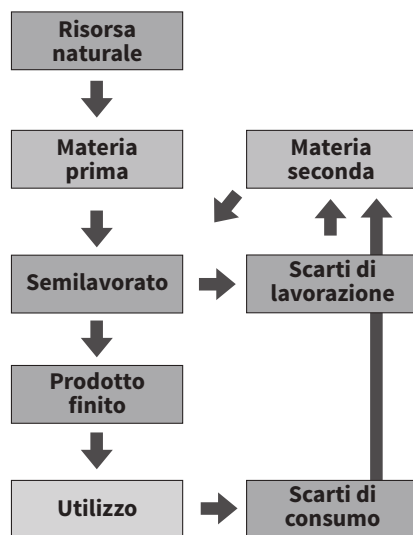


Fig. 2 - Dalla risorsa naturale al prodotto finito

che questi scarti potranno essere recuperati e riutilizzati, avranno cioè una nuova vita. Molti dei nostri quaderni sono prodotti con carta riciclata, così come le bottigliette di plastica (Pet) o di vetro.

4 Il legno

Il **legno** è un materiale che si trova copioso in natura poiché costituisce il tronco degli alberi.

Le radici e le fronde dell'albero hanno scarsi utilizzi pratici, dal tronco invece si ricava legname adoperabile nel campo dell'edilizia, dell'arredamento, come combustibile etc.

Il tronco (Fig. 3) è costituito da (Fig. 4):

- una **parte interna**, che è quella da cui viene ricavato il legname d'uso. A sua volta è costituita dal **durame** e dall'**alburno**;
- una **parte periferica**, la meno utilizzata dall'uomo, costituita dal **cambio** (che crea il legno nuovo), dal **libro** (una sorta di «autostrada» che porta la linfa proveniente dalla chioma) e dalla **corteccia** (che funge da protezione dell'albero).

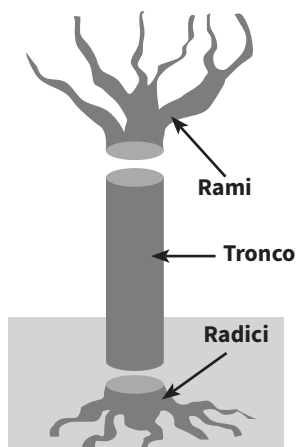


Fig. 3 - Struttura esterna dell'albero

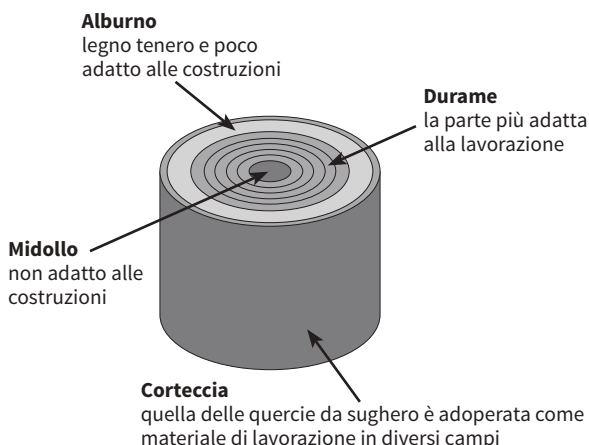


Fig. 4 - Struttura interna del tronco

a) Proprietà

Il legno contiene molta acqua ed è capace di assorbirne. Quando la assorbe cambia dimensioni e tende a *deformarsi*, ovvero a **imbarcarsi**. Per questo, ad esempio, è sconsigliata la posa di pavimenti di parquet in bagno, dove l'umidità è molta. Quando poi perde acqua, ossia viene essiccato, il legno *diminuisce di volume*, si **ritira**.

In pratica, la peculiarità del legno di *assorbire acqua* (**igroscopicità**) lo rende un materiale «vivo», che tende (seppur minimamente) a cambiare dimensioni. Per questo, quando si posa un parquet, si lasciano sempre 1-2 cm di distacco dalle pareti, per permettere che le assi possano dilatarsi leggermente.

Il legno resiste ottimamente alla *flessione*: per tale motivo in passato (ma ancor oggi) è stato molto adoperato per realizzare le travi dei pavimenti e quelle di copertura dei tetti.

Al contrario di quanto si creda, è sì *pericoloso negli incendi*, ma non quanto altri materiali (ad esempio l'acciaio) che crollano senza preavviso. Le travi in legno iniziano a bruciarsi esteriormente, ma la crosta carbonizzata rallenta notevolmente il prosieguo della dissoluzione del resto della struttura, sicché la sua funzione di sostegno non è immediatamente compromessa.

Alcuni legni sono più duri, specie quelli di origine tropicale, e dunque più adatti per realizzare pavimenti. Molto meno duri sono i legni delle conifere, di pini ed abeti. Il frassino ed il faggio sono estremamente elastici, adatti quindi per realizzare archi o sci.

Il legno **si taglia facilmente, si può curvare** bagnandolo e scaldandolo, e talvolta (pino, abete, pioppo) è **fendibile**, ossia si taglia con estrema facilità lungo la direzione delle fibre con l'ausilio di accette o cunei. Una proprietà simile la ritroveremo in certe rocce, come ad esempio l'ardesia, la pietra nera ligure che ancor oggi viene spaccata in sottili lastre, con l'utilizzo di piccoli cunei metallici, dai cosiddetti «spacchini».

Il legno è **molto adoperato nell'arredamento**. A seconda dell'essenza (cioè, del tipo di albero), potrà avere **colore più chiaro o più scuro**, più giallognolo o più rossastro. Inoltre, è forse il materiale d'arredo più «odoroso», caratteristica che lo rende inconfondibile e assai amato.

Oltre al colore, a distinguere i vari tipi di legno è la **venatura**, quella che in gergo tecnico viene definita *fiammatura del legno*. Un legno come l'ulivo, ad esempio, ha una fiammatura bellissima, ricca di nodi e curve molto piacevoli; altri, come il frassino, hanno invece una venatura più regolare ed ordinata.

b) Legname

Una volta **abbattuti**, gli alberi vengono **sramati** (si lascia solo il tronco e si taglia il resto) e trasportati in segheria. Qui i tronchi subiscono un processo di **lisciviazione**, cioè vengono ripuliti di tutte le sostanze interne che alla lunga potrebbero farli imputridire. La lisciviazione avviene mettendo i tronchi a bagno nell'acqua o sottoponendoli artificialmente a getti di vapore acqueo (un po' come quando portiamo dei vestiti in lavanderia: possono essere lavati con l'acqua o a secco, cioè col vapore).

La **segagione** è il processo mediante il quale i tronchi vengono tagliati, in modo da creare assi lunghe e sottili, o in maniera da formare travi e travetti. Naturalmente dallo stesso tronco potranno ricavarsi molte *assi* ma poche *travi*, il più delle volte anche una sola.

Una volta tagliate, le assi e le travi vengono lasciate a stagionare... sì, proprio come i prosciutti! La **stagionatura** è un processo che serve ad eliminare l'acqua contenuta nel legno, che altrimenti tenderebbe (col tempo) a deformarsi (e precisamente a *imbarcarsi* e a *svergolarsi*). Se la stagionatura è naturale, il **legname** (è questo il nome che assumono i pezzi di legno tagliati, specialmente da impiegare come materiale da costruzione e da lavoro) viene lasciato per alcuni anni accatastato, lasciando che l'umidità in esso contenuta evapori da sola. Altrimenti esiste anche un tipo di stagionatura artificialmente accelerata, nella quale sul legno viene soffiata dell'aria calda che ne velocizza la deumidificazione.

c) Il legno nell'arredamento: massello, pannelli impiallacciati e nobilitati

Nell'**arredamento** delle nostre case il legno viene utilizzato per realizzare sedie, tavoli, armadi, mensole, finestre, porte e pavimenti (il cosiddetto *parquet*).

Un «pezzo» di legno pieno, interamente proveniente da uno stesso tronco, è detto **massello**. Un legno massello è **del tutto autentico**, è esteticamente molto bello, però costa anche molto.

Per questo, al posto del legno pieno, spesso si adoperano pannelli che sono rivestiti solo da una sottile «fettina» di legno pregiato: l'aspetto così sarà uguale (più o meno) a quello del legno massello, ma la sostanza no. Dentro questo involucro, vale a dire all'interno della sottile lamina di legno pregiato (detta **piallaccio**), si nasconde in realtà un pannello molto meno costoso, ottenuto assemblando insieme i trucioli di scarto delle segherie o incollando diversi strati di legno «povero» (*compensato, multistrato, paniforte*).

Questi pannelli, oltre a poter essere rivestiti con piallacci di essenze pregiate, possono anche essere verniciati con smalti (**laccatura**) o ricoperti con sottili fogli di materiali plastici (laminati plastici).

MATERIALE	CARATTERISTICHE E LAVORAZIONE
Compensati e multistrati	Si ottengono incollando strati di legno disposti con le fibre ora in un senso, ora nell'altro, in modo alternativo, così da ammortizzare le tensioni interne della struttura ed evitare quegli imbarcamenti e svergolature tipici del legno massello. I fogli di compensato e i multistrato sono, dunque, molto stabili, non si dilatano né si deformano e non sono soggetti a ritiro
Tamburati	Sono pannelli molto leggeri perché costituiti da una struttura di listelli, riempita da cartone pressato a nido d'ape. Come nel caso dei paniforti, i pannelli sono rivestiti con impiallaccature che li nobilitano, ossia li rendono esteticamente più piacevoli
Paniforti	Hanno uno spessore che varia da 8 mm a 26 mm e si ottengono mediante incollaggio di listelli a sezione quadrata, che successivamente sono impiallacciati, ossia rivestiti con fogli di legno pregiato
Truciolari	Sono i meno pregiati tra i pannelli di legno: si ottengono incollando i trucioli che rappresentano lo scarto di precedenti lavorazioni. I truciolari sono molto economici ma anche molto pesanti. Inoltre, le viti non fanno buona presa su di essi, dunque appaiono poco adatti all'assemblaggio

Tab. 4

d) Attrezzature da falegname

Quello del falegname è uno dei lavori più antichi del mondo, e fino a qualche decennio fa si basava sull'utilizzo di tecniche e strumenti millenari. Solo di recente nella bottega di questo artigiano sono comparsi, accanto agli attrezzi «classici», altri tecnologicamente più evoluti e moderni. In particolare sono presenti diversi **strumenti alimentati ad energia elettrica**, che velocizzano e rendono meno faticoso il lavoro. Inoltre, i vari pezzi di legno non vengono più assemblati adoperando incastri, ma sono ormai d'uso comune **viti e chiodi**.

Gli **attrezzi classici** si dividono in:

- attrezzi per **tagliare** (sega, segaccio, saracco);
- attrezzi per **sgrossare** (raspa, lima, pialla, scalpello, cartavetro);
- attrezzi per **forare** (trapano a mano, succhiello);
- attrezzi per **usi diversi** (martello, cacciavite, pinze, tenaglie etc.).

Nella bottega di un moderno falegname, o di una piccola azienda di falegnameria, troveremo anche **macchinari più complessi**, che potremmo suddividere in:

- macchinari per **spianare** (piallatrice a filo e a spessore);
- macchinari per **tagliare** (sega a nastro e a disco);
- macchinari per **modellare** (fresatrici);
- macchinari per **ottenere incavi e forare** (mortasatrice e trapano elettrico);
- macchinari per **tornire** (tornio);
- macchinari per **le finiture** (cartavetratrice).

5 La carta

Esistono svariati tipi di **carta** e ciascuno ha le sue specifiche caratteristiche; tuttavia alcune proprietà li accomunano. Tra queste vi è l'**igroscopicità**, ossia l'attitudine ad assorbire acqua, che abbiamo già visto essere tipica anche del legno.

La carta **resiste bene alle sollecitazioni di trazione, molto meno a quelle di taglio**. È anche **resistente alla compressione**, ma solo quando assume ben precise forme: è il caso del cartone ondulato e dei pannelli a nido d'ape.

La carta è **malleabile**, è **stampabile** e può avere diverse **finiture superficiali**.

a) Produzione artigianale

La carta è un'invenzione cinese che risale all'inizio del II secolo d.C. Nel metodo di **produzione tradizionale ed artigianale**, la carta si otteneva da **tessuti naturali** (soprattutto cotone) ricchi di fibre vegetali, raccolti di casa in casa dagli straccivendoli. I tessuti venivano maciullati da pesanti pestelli e mescolati all'acqua, dando origine ad un impasto molle con il quale venivano riempiti dei tini. A questo punto, un operaio immergeva nel tino un setaccio (forma) e, quando lo faceva riemergere, sulla trama di quest'ultimo rimaneva un po' dell'impasto molle, che ormai, per diventare carta, doveva solo essere asciugato. Per farlo, si ribaltava il setaccio e si creava una pila di fogli, che venivano poi pressati per fare uscire l'acqua in eccesso. Il resto dell'*asciugatura* si effettuava appendendo i fogli all'aria aperta, come fossero panni stesi.

In Italia la carta arriva solo nel 1276, a *Fabriano*. Nelle sue celebri cartiere si apportano anche importanti modifiche alla tecnica di produzione: dall'utilizzo dell'acqua corrente dei fiumi per la produzione dell'impasto fibroso, all'introduzione della filigrana, ossia di quel marchio di fabbrica, visibile in controluce, che viene impresso sulla carta durante la sua fabbricazione.

b) Produzione industriale

Industrialmente la carta viene prodotta con metodi diversi da quelli artigianali, benché il concetto di fondo non cambi: la carta nasce sempre da un *impasto di acqua e fibre naturali che viene lasciato asciugare* su un setaccio.

Oggi, tuttavia, la pasta non si ottiene più dalla frantumazione di tessuti di cotone, ma è per lo più costituita da **cellulosa estratta dal legname e opportunamente trattata**.

In breve, il **processo produttivo della carta industriale** può essere così riassunto (Fig. 5):

- gli alberi vengono tagliati e portati all'industria. Di solito questo spostamento avviene facendo trasportare i tronchi dalla corrente di un fiume;
- nella fabbrica, il legname viene tagliato in pezzi più piccoli e scortecciato;
- i pezzi di legno subiscono sminuzzatura mediante mole, dando origine alla **pastalegno**;
- la pastalegno viene ulteriormente trattata chimicamente, mediante idrato di sodio e solfito;
- le fibre vengono ammorbidite e spezzetate dalla liscivia in un apposito contenitore;
- la pasta viene sbiancata utilizzando cloro o acqua ossigenata;
- la pasta così ottenuta è già *cellulosa*; viene ulteriormente trattata, aggiungendo le cosiddette «cariche»: *cariche minerali* (gesso, caolino, talco), che le danno opacità e levigatezza superficiale; *collanti*, che la rendono meno igroscopica ed evitano che assorba troppo inchiostro; *coloranti*, per avere diverse tonalità.

A questo punto l'intruglio di cellulosa è molto umido e semisolido. La pasta viene dunque posta su un **nastro trasportatore continuo** che, vibrando, la riduce a un foglio abbastanza sottile. Un cilindro le imprime quindi la **filigrana**, ovvero il marchio di fabbrica visibile in controluce. La pasta viene poi pressata attraverso alcuni **rulli asciugatori**, che le danno lo spessore desiderato, eliminando contemporaneamente gli eccessi d'acqua.

L'ultima umidità contenuta nella carta viene eliminata facendola passare attraverso dei rulli caldi di ghisa, detti **essiccatori**.

La carta è infine *lisciata* e *calandrata*, cioè viene fatta passare tra grandi rulli, i quali le donano lo spessore finale e la rendono più liscia superficialmente.

Il prodotto finito viene avvolto in enormi rulli e immagazzinato.

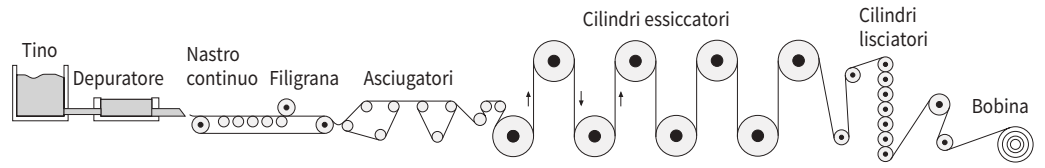


Fig. 5 - Catena industriale di produzione della carta

c) Tipologie

Non esiste un solo tipo di carta: esistono carte per scrivere, da stampa, da pacco, da giornale etc.; inoltre, una carta può avere una finitura superficiale patinata, uso mano, pigmentata, collata, paraffinata etc.

I **vari tipi di carta** possono distinguersi anche in base alle **materie prime utilizzate** e al **processo di produzione**.

A seconda dell’*impasto iniziale da cui si ottiene*, distinguiamo:

MATERIALE	CARATTERISTICHE E LAVORAZIONE
Carta con legno	L’impasto che le dà origine è costituito da un’alta percentuale di pasta-legno (legno sminuzzato) ed è tra le carte considerate di scarsa qualità
Carta senza legno	Costituita esclusivamente con pasta chimica, è più pregiata della precedente
Carta di pura cellulosa	Nasce da un impasto di sole cellulose tecniche
Carta kraft (da pacco)	È composta da una cellulosa tecnica particolarmente resistente, detta kraft. È la carta dei sacchetti del pane
Carta riddata	Carta riciclata. Non è mai perfettamente bianca ed è usata per stampati di qualità media

Tab. 5

A seconda del *metodo di produzione e della finitura superficiale*, distinguiamo:

MATERIALE	CARATTERISTICHE E LAVORAZIONE
Carta patinata	È una carta sulla cui superficie viene applicata una patina, al fine di avere aspetto e finitura migliori, nonché un’eccellente stampabilità. Dopo la <i>patinatura</i> assai spesso si passa alla <i>calandatura</i> , che permette di rendere la carta ancora più liscia e lucida
Carta accoppiata	È formata da diversi fogli uniti insieme. Di solito quelli che rimangono non visibili sono di scarsa qualità
Carta a mano	È la carta che viene fabbricata con il procedimento manuale classico, senza l’impiego di macchine. È molto pregiata e lussuosa ed ha la caratteristica di avere i bordi imperfetti, sfrangiati (detti «intonsi» o «a filo d’acqua»). Può essere <i>filigranata</i> , cioè mostrare in controluce un disegno caratteristico

(Segue)

MATERIALE	CARATTERISTICHE E LAVORAZIONE
Carta a manomacchina o uso mano	Carta pregiata, con una finitura superficiale che ricorda quella irregolare della carta a mano. Come quest'ultima, talora ha i bordi sfrangiati e può essere <i>filigranata</i> o <i>vergata</i>
Carta bibbia o india	Carta senza legno molto sottile, che viene adoperata quando si deve produrre un libro con moltissime pagine
Carta calandrata	Carta che è stata lisciata facendola passare tra due grandi rulli metallici; può essere più o meno lucida
Carta filigranata	Carta che reca una filigrana, ossia un segno distintivo leggibile in controluce. La filigrana è tipica delle carte a mano
Carta lucida	Carta superficialmente liscia e lucida, di solito è stata patinata
Carta mat	Carta liscia, ma non lucida
Carta naturale	Carta non patinata
Carta <i>pelure</i> e velina	Carta leggera di bassissima <i>grammatura</i> (*)
Carta concettata	Carta che, nell'impasto, contiene alcune fibre di colore diverso dal fondo
Carta vergata	Carta che presenta una <i>vergatura</i> , cioè una trama leggera di linee chiare. Tutte le carte a mano sono vergate, poiché la vergatura è proprio l'impronta che i fili (<i>verghelle</i>) della forma lasciano sull'impasto. Un caso particolare di vergatura è la filigrana
Carta voluminosa o <i>bouffant</i> o spessorata	Tipo di carta di elevato spessore, che serve a dare più volume a libretti di poche pagine
(*) La parola « <i>grammatura</i> » è un lemma tecnico utilizzato nel campo della tipografia e, in generale, nel settore della produzione e del consumo della carta da legno e dei suoi derivati. Essa rappresenta, in modo semplicistico, la «consistenza» del foglio di carta, ma anche di fibre tessili animali o vegetali, e non va confusa con lo «spessore» del foglio (fibra), anche se, di fatto, una carta di grammatura superiore avrà tendenzialmente uno spessore superiore. Tecnicamente, la grammatura viene definita come «il peso di un foglio di carta di un metro quadro di superficie».	

Tab. 6

A seconda dell'uso, distinguiamo infine:

MATERIALE	CARATTERISTICHE
Carta da pacco o da imballo	Serve a contenere della merce
Carta da disegno	Serve a disegnare, pertanto deve resistere alle cancellature
Carta da lettere	È usata per la corrispondenza
Carta per scrivere	In generale, tutte le carte su cui bisogna scrivere con matite, penne a sfera, pennarelli, stampanti
Carta da stampa	Carta adatta alla stampa professionale
Carta di sicurezza	Tipo di carta che contiene reattivi chimici particolari. Quando si tenta di cancellare quanto stampato, i reattivi generano una macchia indelebile. È utilizzata per le carte valori e le carte per documenti
Carta per manifesti	È usata per le affissioni. Da un lato deve essere liscia per favorire la stampa, dall'altro ruvida per la colla. Inoltre, deve resistere bene alla luce
Carta giornale	È usata per i quotidiani, ha una grammatura compresa tra 52 e 56 g/m ² , è calandrata e contiene molta pastalegno

Tab. 7

d) Supporti scrittori alternativi alla carta

La carta non è il solo supporto per la scrittura: in passato sono stati diversi i materiali usati per conservare il pensiero dell'uomo in forma scritta. Dalla *pietra* all'*argilla*, dal *papiro* alla *pergamena* e alla *cera*, nel corso della storia la scrittura è apparsa su diversi materiali, assumendo ogni volta caratteristiche diverse.

Il primo supporto alternativo alla carta è costituito dall'**argilla** della scrittura cuneiforme mesopotamica, ottenuta, appunto, premendo dei cunei sulla superficie dell'argilla fresca. Tale scrittura era certo più pratica di quella direttamente impressa sulla **pietra** con lo scalpello, modalità, questa, che troviamo nei geroglifici egiziani qualche secolo più tardi. Nella terra dei faraoni, gli ideogrammi erano dunque scolpiti nella pietra; ma è anche vero che in una civiltà così evoluta e fiorente negli scambi culturali ed economici non poteva esistere solo un tipo di scrittura così rigido come quello su pietra. Per gli usi quotidiani si adoperava, infatti, il papiro, sul quale i grafismi venivano tracciati col pennello o col calamo, ovvero lo stelo di una graminacea opportunamente tagliato.

• Papiro

La tecnica di preparazione dei fogli di **papiro**, nota già nel 3000 a.C., consisteva nel tagliare il midollo della pianta (il *Cyperus Papyrus*) a strisce sottili, che venivano successivamente sovrapposte in direzione ora verticale, ora orizzontale, e poste ad asciugare. I singoli fogli di papiro venivano poi uniti e avvolti intorno ad un bastone a formare il «rotolo», andando così a costituire un *volumen* che veniva riposto in un astuccio cilindrico. Potevano essere vergati (scritti) da un solo lato.

Dal IV secolo a.C., l'uso del papiro si diffuse anche in Grecia. Il suo impiego iniziò a scemare solo a partire dal II sec. a.C., ossia con la diffusione di un nuovo supporto scrittorio originario dell'Asia Minore: la *pergamena*.

• Tavolette di cera

Le prime **tavolette cerate** risalgono al IX sec. a.C. Consistevano in semplici assi di legno leggermente scavate, in modo da poter ospitare una colata sottile di cera, incisa con uno stilo di avorio, ferro o legno. Quest'ultimo era appuntito da un lato e largo, a spatola, dall'altro, per permettere di rasare e cancellare la tavoletta.

Le tavolette di cera furono adoperate fin dai tempi dei Fenici e potevano essere unite a formare documenti complessi, riuniti in *codex*, precursori del libro di oggi. I *codex* venivano rilegati con cordicelle e, nel caso si fosse trattato di documenti giuridici, sigillati.

• Pergamena

Come già accennato, l'uso delle tavolette e del papiro fu soppiantato dall'introduzione, nel II secolo a.C., della **pergamena**, ad opera di Eumene II, re di Pergamo, che dovette rispondere alla carenza di papiri successiva al divieto di esportazione di tale supporto scrittorio dall'Egitto.

La pergamena si otteneva dalla pelle di un animale, solitamente capra o agnello. Nel Medioevo si ricorreva anche a cervi, montoni e conigli, secondo la disponibilità. La pelle veniva dapprima posta a macerare nella calce, quindi raschiata, poi messa a seccare in forte trazione. La trazione serviva a rompere le fibre e a rendere opaca la pergamena.

Quest'ultima veniva infine trattata superficialmente, perché la pelle è sempre un po' grassa e dunque repelle i pigmenti. A tale scopo si adoperava una mistura di polvere di pomice, ottenuta da calcinazione (guscio d'uovo oppure osso di seppia) e di resina (sandracca dell'Africa settentrionale).

La pergamena restò sempre un bene di lusso, spesso impreziosito dall'immersione in bagni di porpora. A differenza del papiro, poteva essere scritta su entrambi i lati: lato carne e lato pelo. Se nel vitello (carta-pecora) i due erano quasi indistinguibili, ciò non avveniva nelle altre pelli. Nella miniatura, per le dorature, era preferibile lavorare sul lato pelo, che tuttavia risultava difficoltoso per la calligrafia.

La pergamena poteva, altresì, sopportare cancellature mediante lavaggio o raschiamento. Questa sua proprietà, unitamente al costo elevato, introdusse l'uso di recuperare supporti già scritti, cancellando per intero il loro contenuto.

La preziosità della pergamena fece sì che anche dopo la diffusione della carta (XIII secolo) il suo utilizzo fosse ancora considerato il più appropriato per la riproduzione dei testi sacri.

6 Le fibre tessili

Le **fibre tessili** sono filamenti piuttosto corti e sottili, che vengono uniti insieme con opportuni procedimenti in modo da formare lunghi **filati**. Questi sono poi **tessuti** insieme, così da realizzare oggetti di abbigliamento ed arredo: giacche, pantaloni, maglie, gonne, ma anche tende, lenzuola, coperte, tappeti etc.

Le fibre si distinguono, a seconda della loro origine, in *fibre naturali* e *fibre chimiche*.

Le **fibre naturali** si trovano, appunto, in natura e si possono ricavare dagli animali (fibre animali), dalle piante (fibre vegetali) o dai minerali (amianto).

Le **fibre chimiche** si ottengono *sinteticamente* nei laboratori scientifici.

A metà tra le due tipologie sta il **rayon**, una fibra che potremmo definire **artificiale**, in quanto ottenuta mediante un'artificiosa lavorazione di una sostanza naturale: la *cellulosa*.

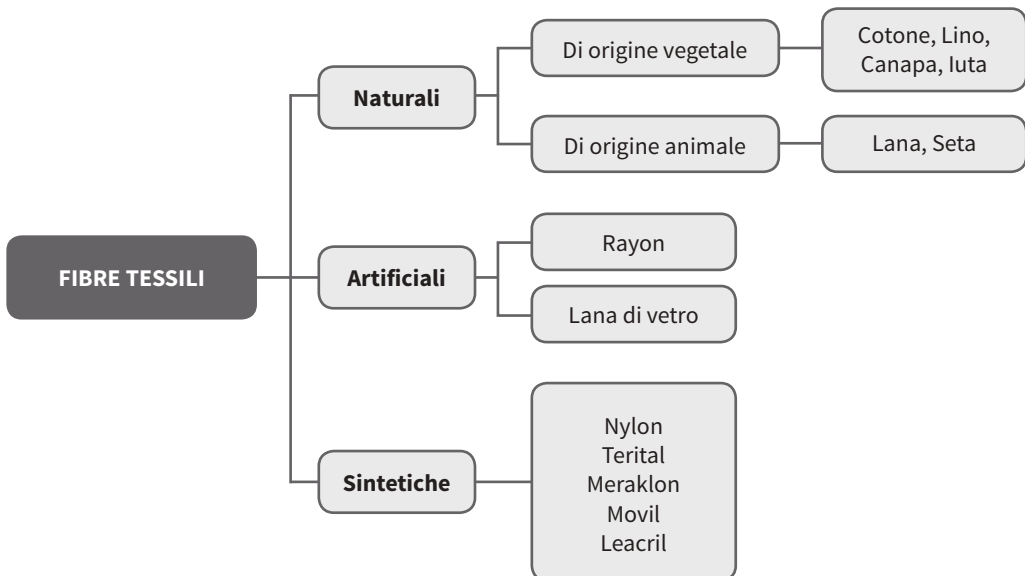


Fig. 6 - Classificazione delle fibre

a) Cotone

La fibra del **cotone** si ottiene da una pianta tipica dei climi tropicali. Il frutto è una capsula fatta da filamenti che ne avvolgono i semi. Questi filamenti, che costituiscono la bambagia, sono lunghi 2-3 cm e costituiscono la fibra del cotone.

Il cotone è una fibra alquanto economica, è molto igienica e si adatta bene alla pelle lasciandola «respirare». Viene adoperato tanto nell'abbigliamento quanto nella biancheria per la casa.

Esso è raccolto nei campi con speciali macchinari, dopodiché si lascia **essiccare**. Dopo l'essiccazione, le fibre vengono separate dai semi con un processo detto di **sgranatura**.

b) Lino

Il **lino** è una pianta erbacea coltivata tanto nelle regioni a clima umido-temperato quanto in quelle a clima freddo. Dal fusto del lino si ricavano lunghe fibre, che altro non sono se non i vasi che conducono la linfa dalle radici al fogliame.

Le fibre si ottengono facendo **macerare** i fusti in vasche d'acqua stagnante. Dopo questo trattamento si passa all'**essiccazione** e alla **separazione** della parte fibrosa dal resto della pianta.

Le fibre così ottenute sono quindi pettinate e **raccolte in matasse** pronte per la filatura e la tessitura.

c) Lana

La **lana** costituisce il vello di molti animali. Consiste di lunghe fibre arricciate (10-20 cm) ed è raccolta mediante la **tosatura** degli animali, che avviene una o due volte ogni anno.

La lana va poi lavata, sgrassata e infine **pettinata**. La pettinatura genera filamenti lunghi, adatti a tessuti leggeri. Le fibre più corte, invece, subiscono un particolare trattamento, che prende il nome di **cardatura** e che fornisce filati più pesanti e morbidi.

Esistono diversi tipi di lana, di provenienza diversa.

d) Seta

La *bombice del moro* è una farfalla che, prima della metamorfosi, quando è ancora allo stato di larva, si nutre di foglie di gelso. A un certo punto il baco inizia a secernere un lungo filamento che lo andrà ad avvolgere completamente, come un guscio protettivo, e lo custodirà durante la metamorfosi: il **bozzolo**.

Il filamento che il baco secerne è costituito da una lunga fibra, denominata **seta**.

I bozzoli vengono bolliti in modo che la crisalide muoia e non rompa il lungo, prezioso filo di seta uscendo dal bozzolo stesso. Una volta recuperato il capo del filo, si riesce dunque a svolgere l'intero bozzolo, che di solito dà origine ad un filamento lungo anche 1 chilometro.

La seta è una fibra assai pregiata, lucente, leggera e resistente. Molto ricercata, ha la caratteristica di mantenere la piega e di costituire un buon isolante termico.

e) Rayon

Il **rayon** è una fibra artificiale, ossia che non esiste in natura ma che viene prodotta a partire dalla cellulosa degli alberi. Questa viene fatta macerare fino a diventare una poltiglia densa, chiamata **viscosa**, la quale viene schiacciata contro un disco di platino microforato (filiera) con un processo di estrusione da cui hanno origine i fili della fibra. Questi poi si seccano (all'aria o per immersione in acido solforico) e vengono infine avvolti su bobine metalliche.

f) Fibre sintetiche

Le **fibre sintetiche** sono prodotti chimici, ovvero nascono in laboratorio in seguito a processi di elaborazione del petrolio e del carbone. Da questi procedimenti si ottengono i prodotti intermedi, che sono **polimeri**, ossia **plastiche**, le quali, opportunamente disciolte, vengono pressate nella filiera dando vita, analogamente a quanto detto per il rayon, al filo sintetico.

Sono fibre sintetiche il **nylon**, il **leacril**, il **movil**, il **terital**. Tutte sono accomunate da alcune caratteristiche, che le distinguono nettamente dalle fibre naturali:

- non vengono attaccate dalle tarme;
- non ammuffiscono;
- sono leggere;
- non si restringono in lavatrice;
- sono impermeabili.

Proprio quest'ultima proprietà è anche il limite delle fibre sintetiche: essendo impermeabili non permettono alla pelle di traspirare in modo naturale.

I filati, che siano di lana, di lino o di altro tipo, vengono successivamente lavorati, mediante la tessitura, per ottenere i tessuti come noi li conosciamo. Questi sono strutture di fili perpendicolari: l'**ordito** è l'insieme dei fili disposti nel senso della lunghezza, la **trama** è invece costituita dai fili ad essi perpendicolari. A seconda di come si intrecciano ordito e trama, si possono avere diverse strutture, dette **armature**: due delle principali sono l'**armatura a tela** e quella **a raso**.

7 Il vetro

Alcune delle caratteristiche del **vetro** lo rendono tra i materiali esteticamente più belli e funzionalmente più utili per le necessità della vita domestica.

Si ottiene dalla fusione della **silice**, una sabbia costituita da piccoli cristalli di quarzo (immaginatevi alla vista qualcosa di molto simile allo zucchero), che ad elevate temperature si sciolgono, per poi solidificarsi nuovamente una volta raffreddati.

Quando si produce il vetro, alla silice vengono aggiunti **fondenti** (carbonati di sodio e di potassio) che ne abbassano la temperatura di fusione, e **stabilizzanti** (calce) che gli conferiscono maggior resistenza. Alcuni particolari ossidi metallici sono adoperati per donare al vetro una certa colorazione.

a) Proprietà

Il vetro è un **buon isolante termico ed elettrico**, è solido a temperatura ambiente, ha scarsa resilienza (ossia è **fragile**, il che non significa che non resista agli urti e alle sollecitazioni, ma solo che si rompe senza prima deformarsi o creparsi) ed è **malleabile** (cioè può essere facilmente ridotto in fili e lastre).

Inoltre, è altamente **trasparente**, dunque ideale per essere usato come contenitore laddove si voglia controllare costantemente livello e quantità del contenuto.

È caratterizzato, infine, dalla **temperabilità**, ossia dalla capacità di migliorare le sue proprietà meccaniche con un processo, di riscaldamento e poi di raffreddamento improvviso, che prende il nome di *tempera*.

b) Lavorazione

Gli oggetti in vetro si dividono in due grandi famiglie:

- **oggetti con incavi**;
- **lastre di vetro**.

Lastre di vetro sono, ad esempio, quelle delle nostre finestre, o i piani di lavoro di certi splendidi tavoli. *Oggetti con incavi* sono, invece, tutti i recipienti: bicchieri, bottiglie, vasi etc.

Volendo, gli oggetti in vetro incavato potrebbero essere realizzati anche con materiali differenti (ad esempio la ceramica), ma il vetro è spesso preferibile per alcune ragioni:

- si può vedere cosa contiene;
- ha spessori inferiori;
- è relativamente economico;
- è neutro, ossia non cede particelle alle sostanze contenute.

Gli oggetti incavati possono essere realizzati industrialmente mediante **stampaggio con soffiatura** o artigianalmente per mano di abilissimi artisti che adoperano la tecnica della **soffiatura libera**. Gli oggetti ottenuti artigianalmente sono spesso di qualità estetiche eccelse e hanno costi molto elevati, ma giustificati dalla grande perizia dell'autore che li ha realizzati come vere opere d'arte.

c) Produzione degli oggetti in vetro con incavi

Gli **oggetti artigianali in vetro** vengono realizzati con il seguente metodo:

- l'artigiano infila nel forno un lungo tubo di ferro cavo e arrotola sulla sua estremità un po' di vetro fuso;
- il vetro viene fatto rotolare avanti e indietro su un piano per dargli forma cilindrica;
- il vetraio soffia all'estremità della canna per gonfiare il vetro come un palloncino, nel frattempo lo lavora con vari attrezzi metallici per dargli la forma desiderata;
- alla fine separa il manufatto dalla canna, adoperando un'apposita forbice.

Industrialmente, invece, i vetri cavi si ottengono, come detto, mediante *stampaggio*: il vetro incandescente viene posizionato tra uno stampo ed un controstampo, che lo pressano in modo da formarlo a loro immagine. Nel caso delle bottiglie, una goccia di vetro fuso viene infilata in uno stampo, dopodiché una cannucchia getta aria compressa all'interno della goccia, creando così la cavità interna alla bottiglia.

d) Produzione delle lastre di vetro

Le **lastre di vetro** sono usatissime nell'edilizia e nell'industria automobilistica.

La trasparenza del vetro e la sua **coibenza termica** (cioè la sua capacità di isolare) lo rendono un materiale perfetto per le nostre finestre, anche perché, a differenza dei nuovi materiali plastici trasparenti (policarbonato), il vetro non si riga facilmente, non ingiallisce se esposto per lungo tempo al sole e non si impolvera subito.

Oltre alla lastra in vetro classica, esistono *due tipi particolari di lastre*: quelle di *sicurezza* e quelle in *vetro temperato*.

Le **lastre di sicurezza** sono utilizzate al fine di rendere meno pericolosi i vetri in caso di rottura. Sappiamo bene, infatti, quanto sia facile tagliarsi con un pezzo di vetro e come possa essere pericolosa la presenza di una finestra, ad esempio, in un asilo.

I vetri di sicurezza sono costituiti da due lastre accoppiate, tra le quali viene stesa una pellicola di nylon sottile e perfettamente trasparente che, in caso di rottura, trattiene i frammenti, evitando che staccandosi possano ferire qualcuno.

I **vetri temperati**, invece, sono materiali ad alta resistenza, ottenuti mediante un processo detto di **tempera**. La resistenza agli urti di un vetro temperato è molto superiore ad uno di uguale spessore ma non temperato. Inoltre, quando si rompe un vetro temperato si producono tanti minuscoli frammenti che possono provocare piccole scalfitture sulla pelle e mai gravi tagli.

Quanto alla **produzione** delle lastre di vetro, un tempo i maestri vetrai le creavano «srotolando» un tubo di vetro ottenuto mozzando le due estremità di un cilindro.

Oggi la tecnica industriale usata è quella del **floating glass**, secondo la quale:

- il vetro fuso viene accumulato in un grande bacino;
- di qui trasborda su un letto di stagno fuso, su cui, in pratica, galleggia e si distribuisce uniformemente, trasformandosi in lastra;
- la lastra così ottenuta viene fatta lentamente raffreddare in forni di **ricottura**;
- infine, il vetro viene tagliato nei formati commerciali.

TIPO DI VETRO	CARATTERISTICHE	IMPIEGHI
Vetro comune	Trasparente e poco resistente alle sollecitazioni	Finestre, bottiglie, bicchieri
Pyrex	Resiste al calore	Recipienti usati in cucina e nei laboratori chimici
Vetro fotocromatico	Si scurisce, se esposto alla luce, grazie alla presenza di sali di argento fotosensibili	Lenti per occhiali e macchine fotografiche
Cristallo	Molto lucido e brillante per la presenza di sali di piombo	Vetri preziosi
Acidato	Vetro abraso con acido fluoridrico	Vetri semi-opachi
Sabbiato	Vetro abraso mediante getti di sabbia ad alte pressioni	Vetri semi-opachi
Specchio	Vetro su un lato del quale è applicato uno strato di ossido metallico di stagno, argento o piombo	Superfici riflettenti
Infrangibile (Di sicurezza)	Vetro costituito da un sandwich di due lastre, con interposta una pellicola di nylon. Non produce schegge se frantumato	Porte e finestre di sicurezza
Temperato	Vetro sottoposto a trattamento di tempera, che ne aumenta la resistenza alle sollecitazioni. Se rotto, si riduce in tanti minuscoli pezzi	Porte e finestre che devono resistere a forti sollecitazioni

Tab. 8

8 I metalli

Al giorno d'oggi siamo circondati da oggetti metallici: le posate, le pentole, gli orologi, le forbici, gli attrezzi da lavoro, le carrozzerie delle auto etc.

Ma nella preistoria non era così, e questo perché in natura sono pochi i metalli che si possono reperire allo stato «nativo» (oro, argento, rame e stagno). Probabilmente i nostri lontani antenati scoprirono i metalli per caso: dalle pietre che erano state messe intorno al fuoco per contenerlo colò fuori qualcosa di liquido (metallo fuso), che poi raffreddandosi divenne estremamente duro e forte. Così iniziarono a lavorare il rame e lo stagno, e poi la loro miscela, il bronzo, prima lega della storia.

a) Proprietà

I metalli sono di **aspetto lucente, solidi** (tranne il mercurio), freddi al tatto in quanto **buoni conduttori di calore** (il calore delle nostre dita passa nel metallo, e per questo sentiamo freddo:

è una prova che potremmo fare su ogni materiale per capire se è un buon conduttore di calore o se, invece, è un buon isolante termico).

Sono anche **buoni conduttori elettrici** e si dilatano quando aumentano di temperatura.

Sono **resistenti a tutte le sollecitazioni**, per questo vengono adoperati in edilizia e nella produzione di attrezzi e macchine da lavoro.

Come il vetro, sono **temperabili** e possono essere ridotti facilmente in lastre e fili (**malleabilità**).

Ultima importante caratteristica dei metalli è la **saldabilità**, ovvero la capacità di due pezzi di metallo analogo di unirsi solidamente e definitivamente.

b) Come si producono e ricavano

I metalli si trovano in particolari rocce (dette metallifere), che costituiscono la risorsa naturale dalla quale vengono estratti. Eccezion fatta per oro, argento, rame e stagno, i metalli in natura, oltre a essere «nascosti» nelle rocce, sono anche mescolati ad altri elementi chimici, formando così dei *composti* (sali e ossidi).

Dunque, per ottenere la materia prima è necessario:

- estrarre il composto dalla roccia;
- separare il metallo dalle altre sostanze del composto (**estrazione**).

Per effettuare quest'ultima operazione si può procedere in due modi:

- per riscaldamento;
- con procedimento chimico (*elettrolisi*).

Una volta separato dagli altri elementi, il metallo è finalmente disponibile come materia prima.

A questo punto, viene fuso e ulteriormente purificato con un processo che va sotto il nome di **raffinazione**: si aggiungono alcune sostanze che, combinandosi con le impurità, le portano a galla o addirittura allo stato aeriforme, comunque riuscendo a separarle dal metallo che adesso è estremamente puro.

Il metallo **fuso** viene quindi colato in stampi di diversa forma, in cui si solidifica dando origine a semilavorati: *lingotti*, *bramme*, *blumi*, *bilette* (Fig. 7).

Questi saranno poi trasportati alle diverse fabbriche, che ne faranno **prodotti finiti**.

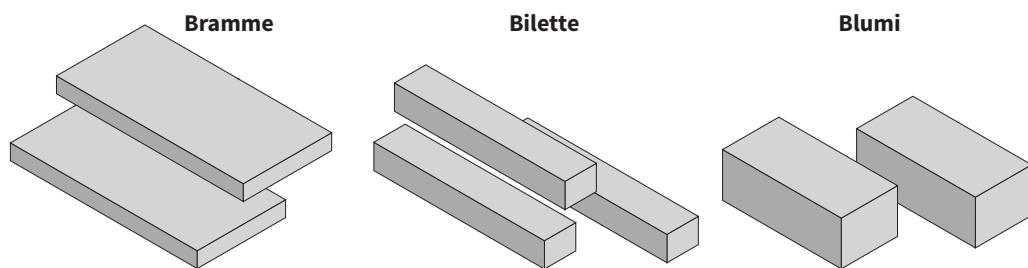


Fig. 7 - I metalli una volta estratti vengono raccolti in *bramme* (lamiere di alto spessore), *bilette* (parallelepipedi con base pressoché quadrata) e *blumi* (simili alle bilette, ma più grandi)

Il metallo fuso si può lavorare in diversi modi così da ottenerne **semilavorati** e prodotti di vario genere.

La **laminazione** è il processo con cui la bramma, passando attraverso diverse gabbie, viene via via schiacciata da grandi cilindri, che alla fine la riducono ad un sottile foglio metallico: la **lamiera**.

L'**estrazione** è il processo per cui il metallo molle viene schiacciato contro una piastra forata, con piccole aperture opportunamente sagomate, da cui poi fuoriescono lunghe strisce di metallo: l'analogia più immediata è con il tubetto della maionese.

c) Ferro, ghisa e acciaio

Il **ferro** si estrae dall'ematite e dalla magnetite, ha una temperatura di fusione di 1.535°C ed una densità di $7,86 \text{ Kg/dm}^3$. Se posto all'interno di un campo magnetico si magnetizza, ossia assume la capacità di attrarre a sé altri oggetti di materiale analogo.

Il ferro da solo è poco usato. Lo si abbina quasi sempre al **carbonio**, unito al quale genera due leghe fondamentali dei tempi moderni: la **ghisa** (con carbonio superiore all'1,8%) e l'**acciaio** (con carbonio inferiore all'1,8%).

La **ghisa** è ottenuta in particolari impianti di produzione, detti **altiforni**. A ferro e carbonio si aggiungono i **fondenti**, ossia sostanze purificanti che si uniscono alle impurità del minerale formando delle scorie facilmente eliminabili.

La ghisa è molto **resistente all'usura** (ed è pertanto usata negli arredi per esterni); ha ottima **conducibilità termica** (con essa si fabbricano stufe e caloriferi); **resiste alle sollecitazioni** ma è **fragile**.

L'**altoforno** viene «caricato» dall'alto con minerali di ferro e **coke** (un particolare tipo di carbone). Per eseguire quest'operazione si adoperano recipienti carrellati, chiamati **Skip**.

Dell'aria calda viene immessa alla base dell'altoforno per permettere la combustione del coke (ricordiamo che nulla brucia senza ossigeno). Bruciando, il coke genera ossido di carbonio che si infila nei minerali di ferro, separando quest'ultimo dal resto ed emettendo anidride carbonica, la quale fuoriesce dal camino in alto.

La ghisa così ottenuta viene raccolta al fondo dell'altoforno, da cui cola, liquida, attraverso un condotto.

Parte della ghisa verrà trasformata in prodotti finiti; parte, invece, sarà impoverita di carbonio in modo da divenire ottimo acciaio.

Le pareti dell'altoforno sono realizzate in mattoni resistenti alle altissime temperature, denominati **mattoni refrattari**.

Tra l'altro, una quantità variabile di acqua fredda viene fatta circolare tra i mattoni stessi per mantenerne la temperatura entro determinate soglie.

L'**acciaio** può essere pensato come una ghisa contenente meno carbonio. Per produrlo si parte, dunque, proprio dalla ghisa (e talvolta da rottami ferrosi), che si impoverisce poi di carbonio, fondendola e mettendola in contatto con aria. Il carbonio presente nella ghisa fusa si combina con l'ossigeno dell'aria formando un

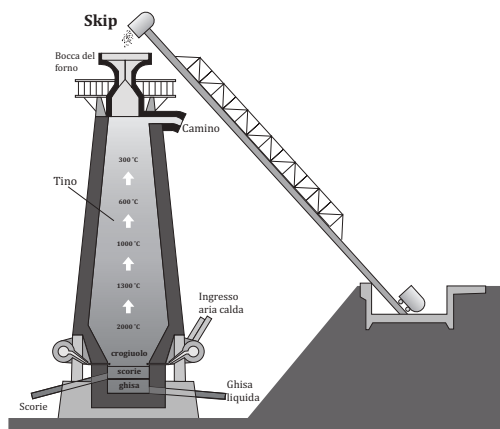


Fig. 8

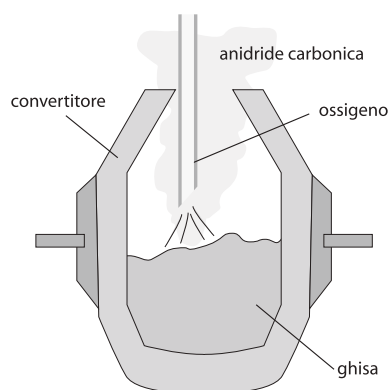


Fig. 9 - Per produrre l'acciaio si procede realizzando una reazione chimica tra ghisa ed aria. Se, infatti, sulla ghisa si soffiava un getto d'aria, questa perde carbonio sotto forma di anidride carbonica. E la ghisa impoverita di carbonio altro non è se non acciaio. Questo tipo di processo, detto «Bessemer», si realizza in contenitori appositi, chiamati «convertitori»

gas che poi viene eliminato. Quindi basta soffiare un po' d'aria sulla ghisa fusa per sottrarle carbonio e trasformarla in acciaio. Detto così sembra semplice ma in realtà è un processo costoso, complesso e spesso, purtroppo, inquinante.

L'acciaio è **duttile, malleabile, buon conduttore di corrente e di calore**; inoltre, **resiste** ottimamente **alle sollecitazioni**, è facilmente saldabile e, a differenza della ghisa, **non è fragile**.

Unici handicap dell'acciaio sono la **scarsa resistenza alla corrosione** e la **facile fusibilità**. Nei grattacieli, ad esempio, le strutture in acciaio devono essere opportunamente trattate e protette per evitare che cedano rapidamente ed improvvisamente in caso di incendio.

Esistono vari tipi di acciaio, riportati nella seguente tabella.

TIPO DI ACCIAIO	PROPRIETÀ	IMPIEGHI
Inox	Acciaio al cromo e nichel, resiste bene alla corrosione	Pentole, posate, arredamento
Acciaio al manganese	Più duro e resistente all'usura	Binari ferroviari
Acciaio al silicio	Molto elastico	Molle

Tab. 9

d) Alluminio

L'**alluminio** è un metallo che in natura si trova, unito all'ossigeno, in un minerale detto **bauxite**. Lo si estrae mediante il processo elettrochimico dell'**elettrolisi**.

Di colore bianco lucido, l'alluminio ha peso specifico pari a $2,7 \text{ kg/dm}^3$ e temperatura di fusione pari a 659°C .

Benché in natura sia il terzo elemento più diffuso dopo ossigeno e silicio, è entrato nell'uso quotidiano solo all'inizio del Novecento. Questo perché la sua produzione è costosa e complessa.

Nonostante ciò, l'alluminio ha alcune qualità per cui il suo utilizzo risulta essere particolarmente efficace in ben determinati ambiti: infatti, è estremamente **malleabile, difficilmente si corrode**, è molto **leggero** ed è un buon **conduttore di corrente**.

La sua malleabilità gli permette di essere ridotto in fogli sottilissimi (come quelli che in cucina si adoperano per avvolgere gli alimenti) e poiché si corrode difficilmente può essere riciclato a più riprese.

Oltre ai già citati usi (fogli e lattine), l'alluminio viene impiegato nei cavi aerei delle linee di alta tensione, nelle apparecchiature chimiche, nelle automobili e nei serramenti. In quest'ultimo caso, in verità, non si tratta di puro alluminio bensì di una lega da esso derivata, l'**anticorodal**, composta da alluminio, silicio, manganese e magnesio.

e) Il rame e le sue leghe

Il **rame** era usato fin dall'antichità dall'uomo poiché facilmente reperibile in natura anche allo stato puro. Di colore rossiccio, ha densità pari a $8,9 \text{ kg/dm}^3$ e temperatura di fusione di 1.083°C .

È un ottimo **conduttore di corrente**, e per questo motivo è con esso che si realizzano tutti i cavi elettrici; è molto **malleabile e duttile**; **resiste bene alla corrosione** in quanto sulla superficie si forma subito una sottile patina verdastra di ossido che lo protegge da ulteriori deterioramenti. Per questo è sempre stato adoperato anche nell'edilizia per la copertura degli edifici.

Il **bronzo** è una **lega di rame e stagno**, la più antica conosciuta dall'uomo. Più duro del rame, viene impiegato nella realizzazione di tubi, rubinetti, lampadari, maniglie e, in campo artistico, nelle sculture.

L'**ottone** è una **lega di rame e zinco**, di colore simile a quello dell'oro. È usatissimo nell'arredamento per fabbricare maniglie, lampadari, rubinetteria; lo si ritrova anche in molte apparecchiature elettriche e nel campo automobilistico.

f) Altri metalli

Nella tabella che segue riportiamo gli altri metalli principali, con le loro proprietà fisiche e gli usi a cui sono destinati.

METALLO	DENSITÀ kg/dm ³	FUSIONE °C	PROPRIETÀ	USI
Stagno	7,28	232	Malleabile ma non duttile, bassa temperatura di fusione, inossidabile	Saldature e copertura di metalli facili a essere corrosi
Zinco	7,14	419	Malleabile e resistente alla corrosione	Protezione di superfici di ferro (zincatura)
Piombo	11,35	327	Duttile e malleabile	Accumulatori, cuscinetti
Nichel	8,9	1.453	Resistente alla corrosione	Finitura superficiale di oggetti metallici, monete
Cromo	7,2	1.875	Duro e resistente	Finitura lucida e protettiva di metalli
Tungsteno	19,35	3.370	Duro	Filamenti delle lampadine a incandescenza
Oro	19,3	1.063	Malleabile e duttile	Oggetti preziosi, circuiti stampati, arte della miniatura
Argento	10,5	960	Duttile	Oggetti preziosi, pellicole fotografiche

Tab. 10

9 I materiali fittili (ceramiche)

Per **materiali fittili** intendiamo le **ceramiche**, ossia le **terre argillose cotte**. Possono essere costituiti da materiale ceramico oggetti come stoviglie, vasi, piastrelle, mattoni.

L'**argilla** è una terra costituita da fini particelle di caolinite che la rendono particolarmente plastica, ossia **plasmabile** nelle forme più svariate.

È reperibile in banchi situati specialmente nelle pianure alluvionali o lungo il corso dei fiumi. L'argilla, infatti, si forma dall'accumulo di particelle nate dalla disgregazione di rocce di montagna e trasportate a valle. A seconda delle altre sostanze che contengono, le argille possono cambiare colore e qualità.

L'estrazione delle argille avviene in cave a cielo aperto, scavate a gradoni mediante ruspe escavatrici.

NOME	PROPRIETÀ	PRINCIPALI USI
Terrecotte	Si ottengono sottoponendo l'argilla ad un'unica cottura	Vasi, brocche, oggetti artistici
Porcellane	Molto pregiate, difficili da scalfire, lucidissime, si ottengono con un'argilla bianchissima: il caolino	Vasi ed oggetti di pregio
Maioliche	Terrecotte raffinate, rifinite con smalto bianco decorato	Piatti e vasellame
Terraglie	Argille cotte a 1.250°C	Stoviglie

Tab. 11

I prodotti di ceramica possono essere realizzati tanto artigianalmente quanto industrialmente. In entrambi i casi la produzione prevede quattro fasi:

- foggatura;
- essiccazione;
- prima cottura;
- smaltatura e seconda cottura.

La **foggatura** è la fase in cui l'oggetto finito prende la sua forma. L'impasto può essere lavorato in diversi modi:

- al tornio;
- per stampaggio;
- per colatura;
- per trafilatura o profilatura.

Il **tornio** è uno strumento adoperato per ottenere forme a sezione costante, come vasi, bicchieri, brocche. È costituito da una ruota inferiore, mossa dai piedi dell'artigiano, che è collegata ad una seconda ruota, superiore, sulla quale si lavora il blocco di argilla. Oggigiorno esistono anche torni elettrici che non necessitano della fatica dell'uomo per ruotare.

Lo **stampaggio** è usato, ad esempio, per creare i piatti. L'impasto molle è pressato tra due matrici in gesso, assumendo così la forma desiderata.

La **colata** si adopera per realizzare pezzi cavi (ad esempio statuette) di forma anche complessa. Lo stampo è di gesso: al suo interno viene colata l'argilla allo stato liquido (barbotina), dopodiché lo strato d'argilla più vicino allo stampo si solidifica mentre il resto rimane liquido e viene tolto. Rimane dunque una scultura cava, leggera e solida.

La **trafilatura** o **profilatura** è lo stesso meccanismo che si usa per formare i serramenti in alluminio o alcuni tipi di travi in acciaio. L'impasto molle viene pressato contro una piastra forata secondo un ben determinato profilo, in modo da creare lunghi pezzi di materiale poi tagliati da cavi di acciaio. Con la trafilatura si ottengono i mattoni forati, i cosiddetti **laterizi**.

Con l'**essiccazione**, il pezzo di argilla formato viene lasciato seccare all'aria per alcuni giorni. Alla fine risulta perfettamente solido, ma ancora molto fragile.

La fase successiva è la **prima cottura**: il pezzo viene posto in un forno la cui temperatura raggiunge, lentamente, i 1.000°C circa. Si ottiene così un prodotto, detto **terracotta** o **biscotto**. I laterizi terminano qui il loro processo di produzione poiché non vengono dipinti.

Infine abbiamo la **smaltatura** e la **seconda cottura**. La terracotta può essere colorata con smalti che, raggiunte elevate temperature, **vetrificano**, ossia si trasformano in una pellicola dura e lucidissima. Dopo la seconda cottura l'oggetto è smaltato, impermeabile, resistente ai graffi e non più fragile come dopo l'essiccazione.

10 Le plastiche

Le **plastiche** sono materiali di recente invenzione (1869) e tuttavia, nell'arco di pochi decenni, hanno invaso la nostra quotidianità andando a formare molti degli oggetti che ci circondano: parti di automobili, televisori, sedie, computer, telefonini, recipienti, mobili etc.

Le plastiche non si trovano in natura, ma sono un'invenzione dell'uomo che le ottiene dal petrolio e, in percentuali minori, da altre risorse (carbone, cellulosa, gas naturale) mediante processi di *distillazione frazionata* e *polimerizzazione*.

Il **petrolio** è un materiale organico, formato da molecole contenenti atomi di carbonio. Una volta estratto viene sottoposto a una **distillazione frazionata**, al termine della quale esso si trasforma da risorsa naturale in materia prima: i **monomeri**. Questi sono molecole organiche

semplici, che attraverso un opportuno processo, detto di **polimerizzazione**, possono unirsi a formare lunghissime catene molecolari: i **polimeri** (Fig. 10). Le materie plastiche sono costituite, appunto, da polimeri.

a) Proprietà

Le plastiche hanno diverse **proprietà eccezionali** che le rendono uno dei materiali più utilizzati nell'*industrial design*.

Innanzitutto, sono **isolanti termici ed elettrici** in quanto non conducono energia elettrica e calore: per questo vengono adoperate come isolante sui cavi elettrici o sugli utensili da lavoro, e come coibente termico nelle abitazioni.

Inoltre, sono **leggere**, ossia hanno bassa densità; **si deformano con l'aumento della temperatura**, dilatandosi notevolmente; e **resistono bene a tutti i tipi di sollecitazione**.

Ancora, sono «plastiche», come suggerisce il loro stesso nome, vale a dire **prendono la forma che gli si attribuisce** e quindi possono essere facilmente modellate. Proprio per quest'ultima proprietà sono di solito fornite alle industrie sotto forma di granulati, ovvero di piccoli chicchi che vengono fusi e resi molli: a questo punto possono essere stampate e modellate a piacere prima di risolidificarsi, raffreddandosi.

Sono, infine, **duttili, malleabili** e assolutamente **impermeabili**.

Esistono due grandi famiglie di plastiche: le **termoindurenti** e le **termoplastiche**.

Le *termoindurenti* una volta modellate e solidificate non possono più essere lavorate. Le *termoplastiche*, invece, se opportunamente scaldate, possono di nuovo essere plasmate.

Questa diversità si deve alla differente composizione molecolare: mentre le *plastiche termoindurenti* hanno le molecole disposte secondo un edificio cristallino, solido e difficilmente modificabile, le *termoplastiche* hanno le molecole disposte in linee facili a rompersi, in maniera tale che è possibile ricondurle a uno stato semisolido di notevole malleabilità.

b) Lavorazione

La **lavorazione** delle plastiche, siano esse termoplastiche o termoindurenti, è analoga fino al trattamento dei semilavorati.

I polimeri, di cui abbiamo già visto in estrema sintesi il processo di formazione, possono essere forniti in lastre, polvere o granuli. A questo punto si passa alla realizzazione dei prodotti finiti, che può avvenire con l'ausilio di diverse tecniche:

- l'**estrusione** (usata anche per formare l'acciaio e l'alluminio) consiste nel comprimere le plastiche in stato semiliquido dentro un tubo, dal quale escono attraverso un'estremità opportunamente formata;
- lo **stampaggio per compressione** (Fig. 11) consiste nell'inserire la plastica «molle» tra due forme che, schiacciandola, le danno la sagoma richiesta;
- lo **stampaggio per iniezione** consiste nell'inserimento della plastica fluida in uno stampo. È un tipo di lavorazione che permette la creazione di forme anche molto complesse;
- lo **stampaggio per decompressione** (o **sottovuoto**) utilizza degli stampi microforati (Fig. 11). Su questi viene posta la plastica semisolido, dopodiché si aspira l'aria da sotto i fori, facendo così aderire il materiale allo stampo;

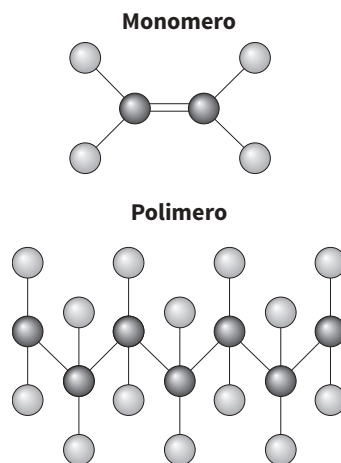


Fig. 10 - Dai monomeri ai polimeri

- la **laminazione** è il processo attraverso il quale la plastica semiliquida viene pressata da due grandi rulli, detti *calandre*, che la appiattiscono fino a farne una sottile pellicola, chiamata anche *film plastico*;
- la **spalmatura** viene adoperata per impermeabilizzare tessuti e cuoi.

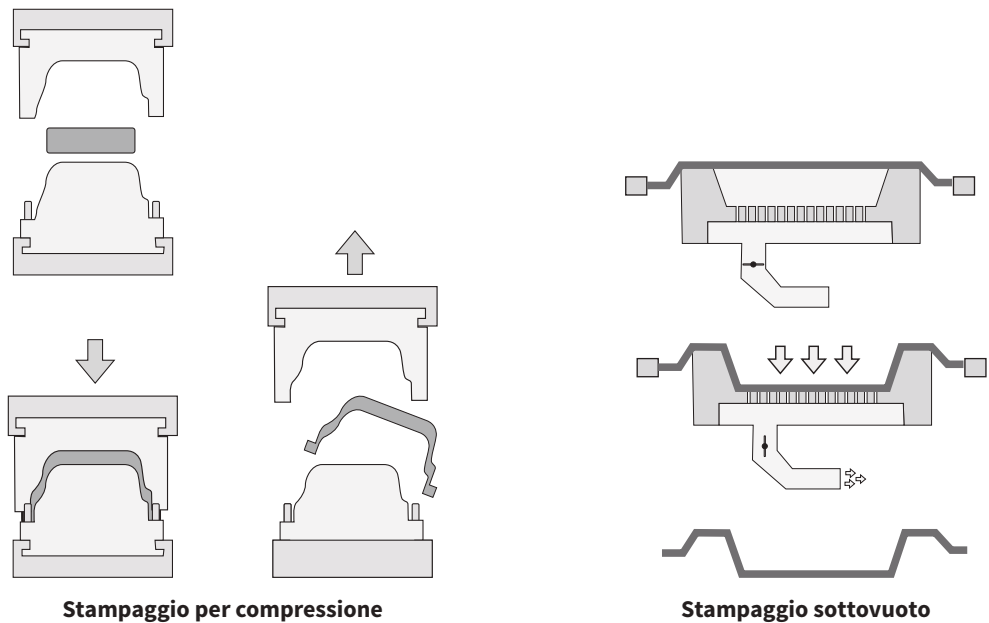


Fig. 11 - Tecniche di stampaggio della plastica

LE RESINE TERMOINDURENTI			
Nome	Derivazione	Caratteristiche	Campo di utilizzo
Bachelite	Derivato dell'idrocarburo del benzene	Isolante	Isolanti elettrici, elettrodomestici, telefoni
Melamina	Composto di carbonio, idrogeno e azoto	Molto resistente	Casalinghi, vasi, laminati
Poliestere	Derivato degli idrocarburi, si ottiene da un acido e un alcol	Resistente e lavorabile	Oggetti d'arredo, scafi di imbarcazioni, coperture

Tab. 12

LE RESINE TERMOPLASTICHE			
Nome	Derivazione	Caratteristiche	Campo di utilizzo
Polietilene (PE)	Derivato dell'etilene	Poco costoso e non tossico	Pellicole, film plastici e sacchetti
Acetato	Derivato della cellulosa	Poco costoso	Occhiali, pellicole fotografiche
Cloruro di polivinile (PVC)	Acetilene trattato	Resistente alle sollecitazioni meccaniche	Tapparelle, pannelli, giochi

(Segue)

LE RESINE TERMOPLASTICHE			
Nome	Derivazione	Caratteristiche	Campo di utilizzo
Plexiglas	Resina acrilica	Trasparente	Coperture, insegne, lenti
Polipropilene	Derivato del propilene	Resistente alle sollecitazioni e al calore	Tubi, recipienti, giochi, arredi
Polietilene tereftalato (PET)	Derivato del polietilene	Economico e resistente	Bottiglie
Polistirene (PSR)	Resina polistirenica	Isolante termico	Isolanti termici, oggetti d'arredo, imballaggi

Tab. 13

11 I materiali compositi da produzione

Sempre più spesso nel mondo del design si sta diffondendo l'uso di **materiali compositi**, ossia ottenuti dall'utilizzo di due o più materiali che «collaborano», facendo valere le proprie caratteristiche migliori.

Un composito è dunque costituito da più materiali che devono essere collocati nel posto corretto, nella giusta quantità e nella geometria adeguata.

Esistono perlomeno *tre tipologie di compositi*:

- quelli costituiti da una matrice e da un rinforzo fibroso;
- quelli costituiti da fogli distanziati;
- quelli con struttura ossa-carne-pelle.

I *compositi costituiti da una matrice e da un rinforzo fibroso* solitamente sfruttano la capacità di certe resine (matrice) di passare dallo stato liquido a quello solido (assumendo dunque la forma del contenitore in cui sono disposte) e di alcune fibre di resistere alle sollecitazioni di trazione. Un esempio classico di questo genere di compositi è la **vetroresina**, con cui si realizzano gli scafi delle moderne imbarcazioni. All'interno della resina vengono «annegate» delle fibre estremamente resistenti, benché molto sottili. Quando la resina solidifica, il composto diventa un unico robustissimo materiale.

I *compositi costituiti da fogli distanziati* hanno un'*anima* e una *pelle*. L'*anima* è leggerissima e serve a dar consistenza alla struttura, senza troppo appesantirla: si tratta solitamente di **nidi d'ape** di materiale plastico. La *pelle* è, invece, costituita da fogli di fibra di vetro, fogli di alluminio o lamiera preverniciata. Questo tipo di compositi è particolarmente usato in edilizia (porte o setti mobili per uffici), ma anche in altri campi (ad esempio nelle tecnologie spaziali).

I *compositi con struttura ossa-carne-pelle* sono molto simili a questi ultimi.

II

Scienze fisiche

Sommario Sezione II

1 I sistemi di unità di misura	18 Luce e rifrazione
2 I vettori	19 La luce: interferenza e diffrazione
3 Le forze	20 Temperatura e dilatazione termica
4 La cinematica	21 Il calore
5 Il moto relativo	22 I gas perfetti
6 La dinamica	23 Il primo principio della termodinamica
7 Moto oscillatorio e molle	24 Il secondo principio della termodinamica
8 Lavoro ed energia	25 Carica elettrica e campi elettrici
9 Impulso e quantità di moto	26 Il potenziale elettrico
10 La dinamica di un corpo rigido	27 La corrente elettrica
11 La gravitazione universale	28 Il campo magnetico
12 La statica dei fluidi	29 Correnti elettriche e campi magnetici: l'elettromagnetismo
13 La dinamica dei fluidi	30 Induzione magnetica ed elettromagnetismo
14 Onde e loro propagazione	31 Quanti, materia, radiazione
15 Il suono	32 La fisica del nucleo e delle particelle
16 Luce e colori	
17 Luce e riflessione	

1 Grandezze fondamentali e grandezze derivate

Un problema di notevole importanza in metrologia è la scelta dei sistemi di unità di misura nei quali vengono definite le unità di un numero limitato di grandezze fondamentali indipendenti tra di loro da cui vengono poi ricavate, in base a relazioni fisiche, le altre grandezze, dette derivate, nell'ambito a cui si intende riferito il sistema di unità considerato.

La distinzione tra grandezze fisiche **fondamentali** e **derivate** è del tutto arbitraria in quanto ogni grandezza può essere considerata fondamentale o derivata a seconda che si faccia riferimento all'una o all'altra relazione fisica.

Una volta stabilito il sistema delle grandezze fondamentali è necessario scegliere le unità di misura da adottare per ognuna di queste grandezze; tale scelta viene effettuata tenendo conto che le unità fondamentali devono possedere le caratteristiche di precisione, accessibilità, riproducibilità e invariabilità.

Un campione ideale deve essere, innanzitutto, preciso in modo da poter costituire un riferimento perfetto per ogni sperimentatore che ad esso volesse eventualmente ricorrere per controllare la taratura dei propri strumenti di laboratorio o l'esattezza delle unità usate; deve essere facilmente disponibile per chiunque intenda accedervi per motivi scientifici; deve essere riproducibile qualora dovesse andare accidentalmente distrutto e quindi devono essere precisati nei minimi dettagli i principi costruttivi; infine deve mantenere costante il proprio valore senza risentire minimamente dell'azione di fattori esterni quali la temperatura, la pressione, l'umidità, la corrosione e l'ossidazione.

Attualmente, in sede internazionale, la sperimentazione di nuovi metodi atti a migliorare la precisione con cui sono realizzati i campioni primari, i confronti internazionali, il coordinamento delle tecniche di misura adottate dai vari laboratori sono affidate al Bureau International des Poids et Mesures (B.I.P.M.). Tale organo è controllato dalla Conference Generale des Poids et Mesures (C.G.P.M.) convocata a Sevres di norma ogni quattro anni.

2 Il Sistema Internazionale (S.I.)

Il **Sistema Internazionale di unità di misura** (S.I.) è stato introdotto nel 1960 dalla XI Conferenza Generale dei Pesì e Misure e perfezionato dalle Conferenze successive.

a) Grandezze fondamentali

Il S.I. prevede 7 grandezze fondamentali e ne definisce le unità di misura.

Grandezza	Unità di misura	Definizione	Simbolo
Intervallo di tempo	secondo	durata di 9 192 631 770 periodi della radiazione corrispondente alla transizione tra i livelli iperfini dello stato fondamentale dell'atomo di cesio -133	s
Lunghezza	metro	tragitto percorso dalla luce nel vuoto in un tempo di 1/299 792 458 di secondo	m

(Segue)

Grandezza	Unità di misura	Definizione	Simbolo
Massa	kilogrammo	massa del campione platino-iridio, conservato nel Museo Internazionale di Pesi e Misure di Sèvres (Parigi)	<i>kg</i>
Temperatura	kelvin	valore corrispondente a 1/273,16 della temperatura termodinamica del punto triplo dell'acqua	<i>K</i>
Quantità di sostanza	mole	quantità di materia di una sostanza tale da contenere tante particelle elementari quante ne contengono 0,012 <i>kg</i> di carbonio -12. Tale valore corrisponde al numero di Avogadro	<i>mol</i>
Intensità di corrente elettrica	ampere	quantità di corrente che scorre all'interno di due fili paralleli e rettilinei, di lunghezza infinita e sezione trascurabile, immersi nel vuoto ad una distanza di un metro, induce in loro una forza di attrazione o repulsione di $2 \cdot 10^{-7} N$ per ogni metro di lunghezza	<i>A</i>
Intensità luminosa	candela	intensità luminosa di una sorgente che emette una radiazione monocromatica con frequenza $540 \cdot 10^{12} Hz$ e intensità energetica di $1/683 W/sr$	<i>cd</i>

Tab. 1

b) Grandezze derivate

In tabella sono riassunte le unità di misura dotate di nome proprio.

Grandezza	Unità di misura	Simbolo	Conversione
Angolo piano	radiante	<i>rad</i>	
Angolo solido	steradiano	<i>sr</i>	
Frequenza	hertz	<i>Hz</i>	$1 Hz = 1 s^{-1}$
Forza	newton	<i>N</i>	$1 N = 1 kgms^{-2}$
Pressione	pascal	<i>Pa</i>	$1 Pa = 1 Nm^{-2}$
Lavoro, energia	joule	<i>J</i>	$1 J = 1 Nm$
Potenza	watt	<i>W</i>	$1 W = 1 Js^{-1}$
Temperatura Celsius	grado Celsius	$^{\circ}C$	$T (^{\circ}C) = T (K) + 273,15$
Carica elettrica	coulomb	<i>C</i>	
Differenza di potenziale elettrico	volt	<i>V</i>	
Capacità elettrica	farad	<i>F</i>	$1 F = 1 CV^{-1}$
Resistenza elettrica	ohm	Ω	$1 \Omega = 1 VA^{-1}$
Conduttanza elettrica	siemens	<i>S</i>	$1 S = 1 W^{-1}$
Flusso d'induzione magnetica	weber	<i>Wb</i>	$1 Wb = 1 Vs$
Induzione magnetica	tesla	<i>T</i>	$1 T = 1 Wbm^{-2}$
Induttanza	henry	<i>H</i>	$1 H = 1 WbA^{-1}$
Flusso luminoso	lumen	<i>lm</i>	$1 lm = 1 cdsr$

(Segue)

Grandezza	Unità di misura	Simbolo	Conversione
Illuminamento	lux	<i>lx</i>	$1lx = 1lm\cdot m^{-2}$
Attività (di un radionuclide)	becquerel	<i>Bq</i>	$1Bq = 1s^{-1}$
Dose assorbita, kerma	gray	<i>Gy</i>	$1Gy = 1Jkg^{-1}$
Dose equivalente	sievert	<i>Sv</i>	$1Sv = 1Jkg^{-1}$

Tab. 2

Vediamo più in dettaglio quali sono le grandezze derivate nei vari ambiti della fisica.

- Angoli

Grandezza	Unità di misura	Definizione	Simbolo
Angolo piano	radiante	angolo al centro di una circonferenza che sottende un arco di lunghezza pari al raggio $1rad = 180^\circ/\pi$	<i>rad</i>
Angolo solido	steradiane	angolo che su di una sfera con centro nel vertice dell'angolo intercetta una calotta di area uguale a quella di un quadrato avente lato uguale al raggio della sfera stessa	<i>sr</i>

Tab. 3

- Unità definite in meccanica

Grandezza	Unità di misura	Definizione	Simbolo	Conversione
Frequenza	hertz	La frequenza ν di un fenomeno periodico è l'inverso del suo periodo T : $\nu = 1/T$. La frequenza misura il numero di volte che un fenomeno periodico si ripete in un secondo	<i>Hz</i>	$1Hz = 1s^{-1}$
Forza	newton	La forza unitaria di $1N$ è la forza che imprime alla massa di $1kg$ un'accelerazione di $1ms^{-2}$	<i>N</i>	$1N = 1kg \cdot ms^{-2}$
Pressione	pascal	La pressione unitaria di $1Pa$ è la pressione esercitata su una superficie di $1m^2$ dalla forza di $1N$ esercitata perpendicolarmente alla superficie	<i>Pa</i>	$1Pa = 1Nm^{-2}$
Lavoro, energia, quantità di calore	joule	Il lavoro unitario di $1J$ è il lavoro della forza di $1N$ per uno spostamento di $1m$ nella direzione della forza	<i>J</i>	$1J = 1N \cdot m$
Potenza	watt	La potenza unitaria di $1W$ corrisponde al lavoro di $1J$ svolto nell'intervallo di tempo di $1s$	<i>W</i>	$1W = 1Js^{-1}$

Tab. 4

• Temperatura

Grandezza	Unità di misura	Definizione	Simbolo	Conversione
Temperatura Celsius	grado Celsius	La scala Celsius è definita in modo che i valori 0 e 100 corrispondano rispettivamente al punto di fusione e al punto di ebollizione dell'acqua a pressione atmosferica. La scala Celsius corrisponde esattamente alla scala Kelvin a meno di un termine additivo pari a 273,15. Entrambe le scale Kelvin e Celsius sono centigrade, in quanto l'intervallo tra punto di fusione e punto di ebollizione dell'acqua è diviso in 100 parti uguali	°C	$T(^{\circ}\text{C}) = T(\text{K}) - 273,15$

Tab. 5

• Unità derivate dell'elettromagnetismo

Grandezza	Unità di misura	Definizione	Simbolo	Conversione
Carica elettrica	coulomb	1C è la carica elettrica trasportata in 1s dalla corrente di 1A	C	
Differenza di potenziale elettrico	volt	1V è la differenza di potenziale elettrico tra due punti di un conduttore che, percorso dalla corrente di 1A, dissipa per effetto Joule la potenza di 1W	V	
Capacità elettrica	farad	1F è la capacità di un condensatore su cui la carica di 1C provoca una differenza di potenziale di 1V	F	$1F = 1\text{C}V^{-1}$
Resistenza elettrica	ohm	1Ω è la resistenza elettrica tra due punti di un conduttore ai quali è applicata la differenza di potenziale di 1V quando scorre la corrente di 1A	Ω	$1\Omega = 1\text{V}A^{-1}$
Conduttanza elettrica	siemens	1 S è la conduttanza di un conduttore avente resistenza di 1Ω	S	$1S = 1\Omega^{-1}$
Flusso d'induzione magnetica	weber	1Wb = 1Vs è il flusso magnetico che, concatenato con una spira, induce una forza elettromotrice di 1V, annullandosi in 1s a velocità costante	Wb	$1\text{Wb} = 1\text{Vs}$
Induzione magnetica	tesla	1T è l'induzione magnetica che, attraversando una superficie piana di 1m ² , produce un flusso magnetico di 1Wb s	T	$1T = 1\text{Wbm}^{-2}$
Induttanza	henry	1H è l'induttanza di una spira nella quale la variazione uniforme di intensità di corrente di 1A/s produce una forza elettromotrice di 1V	H	$1H = 1\text{Wb}A^{-1}$

Tab. 6

• Unità definite in fotometria

Grandezza	Unità di misura	Simbolo	Conversione
Flusso luminoso	lumen	<i>lm</i>	$1\text{lm} = 1\text{cdsr}$
Illuminamento	lux	<i>lx</i>	$1\text{lx} = 1\text{lm}\text{m}^{-2}$

Tab. 7

c) Prefissi moltiplicativi

Il S.I. codifica l'uso dei prefissi moltiplicativi secondo le potenze di 1000.
Sono previsti anche i prefissi per multipli e sottomultipli per fattori 10 e 100.

Fattore	Prefisso	Simbolo
10^{24}	yotta	<i>Y</i>
10^{21}	zetta	<i>Z</i>
10^{18}	exa	<i>E</i>
10^{15}	peta	<i>P</i>
10^{12}	tera	<i>T</i>
10^9	giga	<i>G</i>
10^6	mega	<i>M</i>
10^3	chilo	<i>k</i>
10^2	etto	<i>h</i>
10	deca	<i>da</i>

Fattore	Prefisso	Simbolo
10^{-24}	yocto	<i>y</i>
10^{-21}	zepto	<i>z</i>
10^{-18}	atto	<i>a</i>
10^{-15}	femto	<i>f</i>
10^{-12}	pico	<i>p</i>
10^{-9}	nano	<i>n</i>
10^{-6}	micro	μ
10^{-3}	milli	<i>m</i>
10^{-2}	centi	<i>c</i>
10^{-1}	deci	<i>d</i>

Tab. 8

d) Regole di scrittura

Il S.I. codifica le norme di scrittura dei nomi e dei simboli delle grandezze fisiche.

Le più importanti sono:

I nomi delle unità di misura vanno sempre scritti in carattere minuscolo, privi di accenti o altri segni grafici.

Es.: **ampere**, non **Ampère**.

I nomi delle unità non hanno plurale.

Es.: **3 ampere**, non **3 amperes**.

I simboli delle unità di misura vanno scritti con l'iniziale minuscola, tranne quelli derivanti da nomi propri.

Es.: **mol** per la mole, **K** per il kelvin.

I simboli non devono essere seguiti dal punto (salvo che si trovino a fine periodo).

I simboli devono sempre seguire i valori numerici.

Es.: **1 kg**, non **kg 1**.

Il prodotto di due o più unità va indicato con un punto a metà altezza o con un piccolo spazio tra i simboli.

Es.: $N \cdot m$ oppure Nm .

Il quoziente tra due unità va indicato con una barra obliqua o con esponenti negativi.

Es.: J/s oppure $J s^{-1}$.

3 Il sistema C.G.S.

Nei **sistemi c.g.s.** le unità fondamentali della meccanica sono il centimetro, il grammo e il secondo. Per quanto riguarda la meccanica, quindi, la differenza tra S.I. e c.g.s. si limita a fattori potenze di 10 nei valori delle grandezze fondamentali e derivate.

La differenza sostanziale tra i sistemi c.g.s. e il Sistema Internazionale riguarda le grandezze elettromagnetiche. Mentre il S.I. introduce una grandezza fondamentale per l'elettromagnetismo (l'intensità di corrente), nei sistemi c.g.s. le grandezze elettromagnetiche sono tutte derivate da quelle meccaniche.

Nella Tabella seguente riportiamo un confronto tra alcune unità c.g.s. di Gauss e le corrispondenti unità S.I.:

Grandezza	Unità	Simbolo	Conversione S.I.
Forza	dina	<i>dyn</i>	$1dyn = 10^{-5}N$
Lavoro, energia	erg	<i>erg</i>	$1erg = 10^{-7}J$
Carica elettrica	statcoulomb	<i>statC</i>	$1statC = 3.333 \cdot 10^{-10}C$
Corrente elettrica	statampere	<i>statA</i>	$1statA = 3.333 \cdot 10^{-10}A$
Potenziale elettrico	statvolt	<i>statV</i>	$1statV = 300V$
Induzione magnetica B	gauss	<i>G</i>	$1G = 10^{-4}T$
Campo magnetico H	oersted	<i>Oe</i>	$1Oe = (1/4\pi) \cdot 10^3 A/m$

Tab. 9

2 I vettori

1 Vettori e scalari

Per **scalare** si intende la grandezza fisica totalmente determinata da un numero.

Per **vettore** intendiamo invece la grandezza fisica totalmente determinata da una lunghezza, da una direzione e da un verso.

Un vettore si indica con V o con $\vec{V}$ ed è rappresentato graficamente da una freccia.

La lunghezza del vettore, detta anche modulo, si indica con V o $|\vec{V}|$.

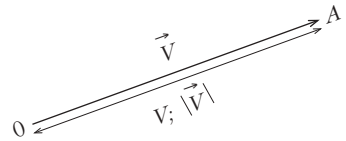


Fig. 1

2 I versori

Un **versore** è un vettore avente lunghezza unitaria.

In figura 2 è indicato il versore $\vec{u}$ del vettore V :

$$\vec{V} = |\vec{V}| \vec{u} = V \vec{u}$$

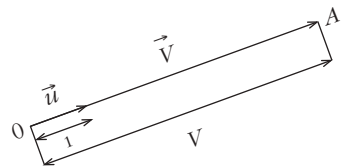


Fig. 2

3 Il vettore opposto

Il **vettore opposto** di un vettore è un vettore avente lo stesso modulo e la stessa direzione del vettore dato ma verso opposto.

In figura il vettore V_1 è l'opposto del vettore V .

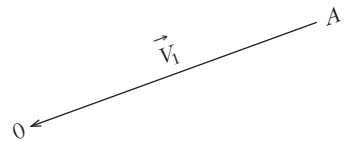


Fig. 3

4 La somma di vettori

Dati due vettori $\vec{A}$ e $\vec{B}$ la loro somma $\vec{R}$ è un vettore i cui parametri (modulo, direzione e verso) **dipendono** da quelli di $\vec{A}$ e $\vec{B}$ ma **non sono** la loro somma.

a) Metodo geometrico

- 1° metodo (metodo del parallelogramma)

La somma di due vettori è rappresentata dalla diagonale maggiore del parallelogramma costruito sui due vettori.

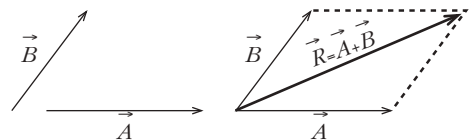


Fig. 4

• 2° metodo (metodo punta-coda)

Per calcolare la somma di due vettori con il *metodo punta-coda* si fa coincidere la coda del secondo vettore con la punta del primo. Il vettore somma è il vettore che unisce la coda del primo vettore con la punta del secondo.

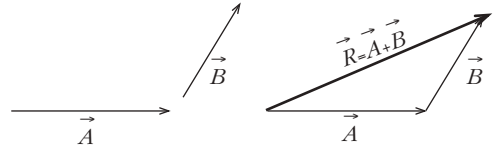
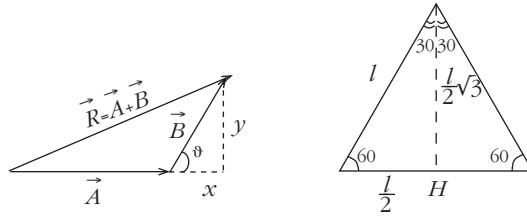


Fig. 5

Come si può notare i due metodi sono del tutto equivalenti.

b) Metodo analitico

Dopo aver tracciato il vettore $\vec{R}$, somma dei vettori $\vec{A}$ e $\vec{B}$, possiamo calcolarne la lunghezza e la direzione con il *metodo analitico*.



Figg. 6.a e 6.b

Dalla figura 6.a:

$$R = \sqrt{(A + x)^2 + y^2}$$

Ricordando le relazioni tra il lato e l'altezza di un triangolo equilatero (Fig. 6.b), i valori di x e y della figura 6.a per particolari valori di ϑ sono riassunti in tabella:

ϑ	x	y
0	B	0
30	$\frac{B\sqrt{3}}{2}$	$\frac{B}{2}$
45	$\frac{B\sqrt{2}}{2}$	$\frac{B\sqrt{2}}{2}$
60	$\frac{B}{2}$	$\frac{B\sqrt{3}}{2}$
90	0	B
120	$-\frac{B}{2}$	$\frac{B\sqrt{3}}{2}$
135	$-\frac{B\sqrt{2}}{2}$	$\frac{B\sqrt{2}}{2}$
150	$-\frac{B\sqrt{3}}{2}$	$\frac{B}{2}$

Tab. 1

In generale, dati due vettori $\vec{A}$ e $\vec{B}$ il vettore somma $\vec{R}$ ha modulo:

$$R = \sqrt{A^2 + B^2 + 2AB\cos\vartheta}$$

e direzione:

$$\frac{R}{\sin \vartheta} = \frac{A}{\sin \beta} = \frac{B}{\sin \alpha}$$

DIMOSTRAZIONE

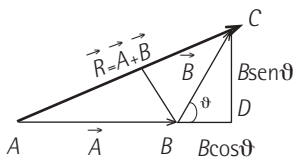


Fig. 7

Modulo di R

Dal triangolo ADC:

$$R = \overline{AC} = \sqrt{(\overline{AD})^2 + (\overline{DC})^2}$$

Ma: $\overline{AD} = \overline{AB} + \overline{BD} = A + B \cos \vartheta$ e $\overline{DC} = B \sin \vartheta$. Quindi:

$$R = \sqrt{(A + B \cos \vartheta)^2 + (B \sin \vartheta)^2} = \sqrt{A^2 + B^2 \cos^2 \vartheta + 2AB \cos \vartheta + B^2 \sin^2 \vartheta} = \sqrt{A^2 + B^2 + 2AB \cos \vartheta}$$

Direzione di R

Dal triangolo ABC $\overline{AB} \sin \alpha = \overline{BC} \sin \beta$ e $A \sin \alpha = B \sin \beta$. Da cui:

$$\frac{A}{\sin \beta} = \frac{B}{\sin \alpha} \quad (4.1)$$

Dal triangolo ACD $\overline{CD} = \overline{AC} \sin \alpha = R \sin \alpha$. Dal triangolo BCD $\overline{CD} = \overline{BC} \sin \vartheta = B \sin \vartheta$.

Da cui:

$$\frac{R}{\sin \vartheta} = \frac{B}{\sin \alpha} \quad (4.2)$$

Confrontando la (4.1) e la (4.2):

$$\frac{A}{\sin \beta} = \frac{B}{\sin \alpha} = \frac{R}{\sin \vartheta}$$

c) Proprietà della somma

- La somma di due vettori è indipendente dall'ordine degli addendi: essa è quindi **commutativa**.
- Due vettori di uguale intensità e direzione ma con verso opposto hanno come risultante un vettore di lunghezza nulla, detto **vettore nullo**, per il quale è impossibile stabilire direzione e verso.
- Il vettore nullo sommato a qualunque altro vettore ha come risultante il vettore sommato. Per questo motivo è detto **elemento neutro della somma**.
- Due vettori con risultante nulla si dicono **opposti**.
- La somma dei vettori è **associativa**: $(\vec{A} + \vec{B}) + \vec{C} = \vec{A} + (\vec{B} + \vec{C})$.

5 La differenza di vettori

Come la somma anche la differenza $\vec{D}$ tra due vettori $\vec{A}$ e $\vec{B}$ è un vettore i cui parametri (modulo, direzione e verso) **dipendono** da quelli di $\vec{A}$ e $\vec{B}$ ma **non sono** la loro differenza.

In particolare se:

$$A = |\vec{A}|; \quad B = |\vec{B}|; \quad D = |\vec{D}|$$

a parte il caso in cui i vettori $\vec{A}$ e $\vec{B}$ sono paralleli non vale la relazione:

$$D = A - B$$

a) Metodo geometrico

• 1° metodo (metodo del parallelogramma)

La differenza di due vettori è rappresentata dalla diagonale minore del parallelogramma costruito sui due vettori.

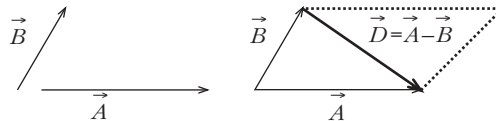


Fig. 8

• 2° metodo (metodo punta-coda)

La differenza tra due vettori $\vec{A}$ e $\vec{B}$ può essere pensata come la somma di $\vec{A}$ con il vettore opposto di $\vec{B}$:

$$\vec{A} - \vec{B} = \vec{A} + (-\vec{B})$$

Quindi possiamo applicare il metodo analitico già visto per la somma.

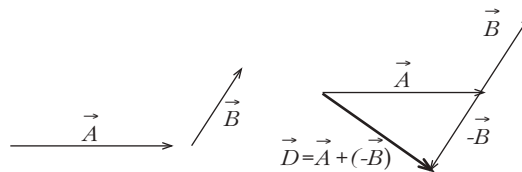


Fig. 9

Anche in questo caso i due metodi sono del tutto equivalenti.

b) Metodo analitico

Dopo aver tracciato il vettore $\vec{D}$, differenza dei vettori $\vec{A}$ e $\vec{B}$, possiamo calcolarne la lunghezza e la direzione con il metodo analitico.

Dalla figura 10:

$$R = \sqrt{(A - x)^2 + y^2}$$

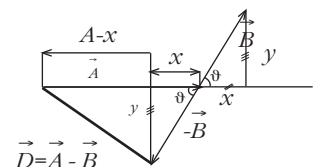


Fig. 10

Ricordando le relazioni tra il lato e l'altezza di un triangolo equilatero (Fig. 6.b) i valori di x e y sono riassunti in tabella 1.

In generale, dati due vettori $\vec{A}$ e $\vec{B}$ il vettore differenza $\vec{R}$ ha modulo:

$$R = \sqrt{A^2 + B^2 - 2AB\cos\vartheta}$$

e direzione:

$$\frac{R}{\sin\vartheta} = \frac{A}{\sin\beta} = \frac{B}{\sin\alpha}$$

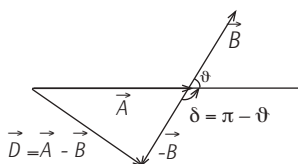
DIMOSTRAZIONE

Fig. 11

Modulo di R

Riprendiamo la formula della somma:

$$R = \sqrt{A^2 + B^2 + 2AB\cos\delta}$$

ma $\delta = \pi - \vartheta$, quindi:

$$R = \sqrt{A^2 + B^2 + 2AB\cos(\pi - \vartheta)} = \sqrt{A^2 + B^2 - 2AB\cos\vartheta}$$

Direzione di R

Riprendiamo la formula della somma:

$$\frac{A}{\sin\beta} = \frac{B}{\sin\alpha} = \frac{R}{\sin\delta}$$

ma:

$$\frac{R}{\sin\delta} = \frac{R}{\sin(\pi - \vartheta)} = \frac{R}{\sin\vartheta}$$

quindi vale ancora la relazione:

$$\frac{A}{\sin\beta} = \frac{B}{\sin\alpha} = \frac{R}{\sin\vartheta}$$

c) Proprietà della differenza

La differenza di due vettori è **anticommutativa**:

$$\vec{A} - \vec{B} = -(\vec{B} - \vec{A})$$

6 La scomposizione di un vettore

Così come, dati due vettori, è possibile calcolarne la somma, è sempre possibile pensare ad un vettore come alla somma di due vettori qualsiasi.

Dato un sistema di assi cartesiani, è possibile supporre un qualsiasi vettore $\vec{V}$ come il risultato della somma delle componenti sugli assi xy .

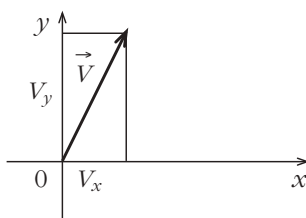


Fig. 12

Introducendo i versori $\vec{i}$ e $\vec{j}$ degli assi cartesiani:

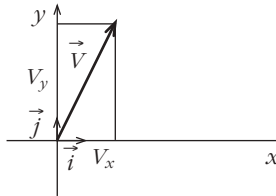


Fig. 13

$$\vec{V} = V_x \vec{i} + V_y \vec{j}$$

$$V = \sqrt{V_x^2 + V_y^2}$$

Indicando con α l'angolo di incidenza del vettore $\vec{V}$:

$$V_x = V \cos \alpha; \quad V_y = V \sin \alpha$$

$$\operatorname{tg} \alpha = \frac{V_y}{V_x}$$

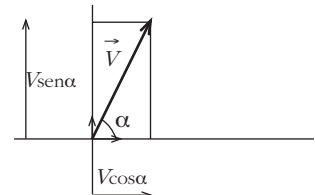


Fig. 14

7 La somma di più vettori

Dati due o più vettori espressi mediante le loro componenti ortogonali, il vettore somma è un vettore avente come componenti la somma delle singole componenti.

Dati i vettori:

$$\vec{V}_1 = V_{1x} \vec{i} + V_{1y} \vec{j}$$

$$\vec{V}_2 = V_{2x} \vec{i} + V_{2y} \vec{j}$$

$$\vec{V}_3 = V_{3x} \vec{i} + V_{3y} \vec{j}$$

$$\dots\dots\dots$$

$$\vec{V}_n = V_{nx} \vec{i} + V_{ny} \vec{j}$$

il vettore somma è:

$$\vec{V} = \sum_i V_{ix} \vec{i} + \sum_i V_{iy} \vec{j}$$

DIMOSTRAZIONE

$$\begin{aligned} \vec{V} &= \vec{V}_1 + \vec{V}_2 + \vec{V}_3 + \dots + \vec{V}_n = V_{1x} \vec{i} + V_{1y} \vec{j} + V_{2x} \vec{i} + V_{2y} \vec{j} + V_{3x} \vec{i} + V_{3y} \vec{j} + \dots + V_{nx} \vec{i} + V_{ny} \vec{j} = \\ &= (V_{1x} + V_{2x} + V_{3x} + \dots + V_{nx}) \vec{i} + (V_{1y} + V_{2y} + V_{3y} + \dots + V_{ny}) \vec{j} = \vec{V} \\ &= \sum_i V_{ix} \vec{i} + \sum_i V_{iy} \vec{j} \end{aligned}$$

8 Il prodotto scalare o prodotto interno

Si definisce **prodotto scalare** s di due vettori $\vec{A}$ e $\vec{B}$ l'area del rettangolo che ha per lati il modulo del vettore e la lunghezza della proiezione del vettore $\vec{B}$ sul vettore $\vec{A}$.

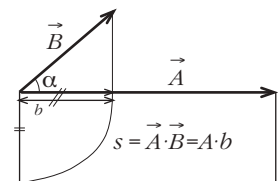


Fig. 15

Indicando con b la lunghezza della proiezione del vettore $\vec{B}$ su $\vec{A}$, dalla figura 15 si ha $s = A \cdot b$.

1. Se $\alpha < 90^\circ \rightarrow b > 0 \rightarrow A \cdot b > 0$ il prodotto scalare è positivo
2. Se $\alpha > 90^\circ \rightarrow b < 0 \rightarrow A \cdot b < 0$ il prodotto scalare è negativo
3. Se $\alpha = 0^\circ \rightarrow b = B \rightarrow A \cdot b = A \cdot B$
4. Se $\alpha = 90^\circ \rightarrow b = 0 \rightarrow A \cdot b = 0$ il prodotto scalare è nullo

Il prodotto scalare s dei vettori $\vec{A}$ e $\vec{B}$ è dato da $s = AB \cos \alpha$.

a) Prodotto scalare tra i versori degli assi cartesiani

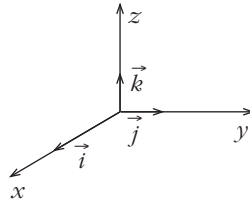


Fig. 16

Essendo $\vec{i}, \vec{j}, \vec{k}$ perpendicolari fra loro e di misura unitaria, dalla 3 e dalla 4 risulta che:

$$\begin{aligned}\vec{i} \cdot \vec{i} &= \vec{j} \cdot \vec{j} = \vec{k} \cdot \vec{k} = 1 \\ \vec{i} \cdot \vec{j} &= \vec{i} \cdot \vec{k} = \vec{j} \cdot \vec{k} = 0\end{aligned}$$

b) Proprietà del prodotto scalare

Il prodotto scalare è **commutativo**:

$$\vec{A} \cdot \vec{B} = \vec{B} \cdot \vec{A}$$

Il prodotto scalare gode della **proprietà distributiva** rispetto alla somma:

$$\vec{C} \cdot (\vec{A} + \vec{B}) = \vec{C} \cdot \vec{A} + \vec{C} \cdot \vec{B}$$

c) Prodotto scalare di due vettori in funzione delle loro componenti

Dati due vettori:

$$\vec{A} = A_x \vec{i} + A_y \vec{j} \quad \text{e} \quad \vec{B} = B_x \vec{i} + B_y \vec{j}$$

il loro prodotto scalare s è:

$$s = \vec{A} \cdot \vec{B} = A_x B_x + A_y B_y$$

DIMOSTRAZIONE

$$s = \vec{A} \cdot \vec{B} = (A_x \vec{i} + A_y \vec{j}) \cdot (B_x \vec{i} + B_y \vec{j}) = A_x B_x \vec{i} \cdot \vec{i} + A_x B_y \vec{i} \cdot \vec{j} + A_y B_x \vec{j} \cdot \vec{i} + A_y B_y \vec{j} \cdot \vec{j} = A_x B_x + A_y B_y$$

9 Il prodotto vettoriale o prodotto esterno

Due vettori applicati allo stesso punto individuano un parallelogramma di lati corrispondenti ai moduli dei due vettori.



Fig. 17

Il **prodotto vettoriale** $\vec{p}$ tra due vettori $\vec{A}$ e $\vec{B}$ si indica con $\vec{A} \times \vec{B}$ e si legge *A esterno B* ed è un vettore che ha:

- direzione perpendicolare al piano formato dai due vettori;
- modulo pari all'area del parallelogramma definito dai due vettori;
- verso definito dalla **regola della mano destra** (indice = primo vettore; medio = secondo vettore; pollice = risultato) o della **vite destrorsa**.

a) Regola della mano destra

Disponendo le dita della mano destra come una terna di assi cartesiani e facendo corrispondere l'indice e il medio con le direzioni di $\vec{A}$ e $\vec{B}$, il momento ha la direzione del pollice.

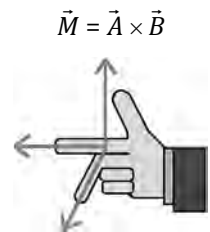


Fig. 18

b) Regola della vite destrorsa

Il verso del prodotto vettoriale è quello di avanzamento di una vite destrorsa che ruota da $\vec{A}$ e $\vec{B}$, ed è:

- negativo se la rotazione avviene in senso orario;
- positivo se la rotazione avviene in senso antiorario.

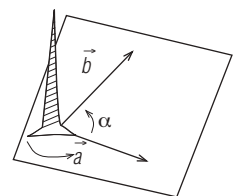


Fig. 19

Consideriamo la figura 17 in cui il parallelogramma è definito dai vettori $\vec{A}$ e $\vec{B}$.

Se indichiamo con b l'altezza del parallelogramma e con α l'angolo compreso tra i due vettori notiamo che il modulo del prodotto vettoriale $\vec{p}$ dei vettori $\vec{A}$ e $\vec{B}$, essendo uguale all'area del parallelogramma, è uguale a $p = A \cdot b$

Dalla trigonometria si ha che:

$$b = B \sin \alpha \text{ e quindi}$$

$$p = AB \sin \alpha$$

Inoltre:

- Se $\alpha = 0^\circ \rightarrow b = 0 \rightarrow A \cdot b = 0$
- Se $\alpha = 90^\circ \rightarrow b = B \rightarrow p = A \cdot B$

c) Prodotto tra i versori degli assi cartesiani

Dalla figura 16 si ha che, essendo $\vec{i}, \vec{j}, \vec{k}$ perpendicolari fra loro e di misura unitaria il loro prodotto esterno è dato da:

$$\begin{aligned}\vec{i} \times \vec{i} = \vec{j} \times \vec{j} = \vec{k} \times \vec{k} &= 0 \\ \vec{i} \times \vec{j} &= \vec{k} \\ \vec{j} \times \vec{k} &= \vec{i} \\ \vec{k} \times \vec{i} &= \vec{j}\end{aligned}$$

d) Proprietà del prodotto vettoriale

Il prodotto scalare è **anticommutativo**:

$$\vec{A} \times \vec{B} = -\vec{B} \times \vec{A}$$

Il prodotto vettoriale gode della **proprietà distributiva** rispetto alla somma:

$$\vec{C} \times (\vec{A} + \vec{B}) = \vec{C} \times \vec{A} + \vec{C} \times \vec{B}$$

e) Prodotto vettoriale di due vettori in funzione delle loro componenti

Dati due vettori:

$$\vec{A} = A_x \vec{i} + A_y \vec{j} \quad \text{e} \quad \vec{B} = B_x \vec{i} + B_y \vec{j}$$

il loro **prodotto vettoriale** $\vec{p}$ è:

$$\alpha \vec{p} = \vec{A} \times \vec{B} = (A_x B_y - A_y B_x) \vec{k}$$

DIMOSTRAZIONE

$$\begin{aligned}\vec{A} \times \vec{B} &= (A_x \vec{i} + A_y \vec{j}) \times (B_x \vec{i} + B_y \vec{j}) = A_x B_x (\vec{i} \times \vec{i}) + A_x B_y (\vec{i} \times \vec{j}) + A_y B_x (\vec{j} \times \vec{i}) + A_y B_y (\vec{j} \times \vec{j}) = \\ &= A_x B_y \vec{k} - A_y B_x \vec{k} = (A_x B_y - A_y B_x) \vec{k}\end{aligned}$$

3 Le forze

1 La composizione di forze concorrenti

La forza risultante di N **forze concorrenti**, cioè tutte applicate nel medesimo punto, è data dalla loro somma vettoriale:

$$\vec{R} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \dots + \vec{F}_n = \sum_i \vec{F}_i$$

Se le forze sono complanari, cioè appartengono tutte al piano XY , possiamo esprimere $\vec{R}$ in funzione delle sue componenti rispetto a una coppia di assi xy :

$$\vec{R} = R_x \vec{i} + R_y \vec{j} = \sum_i F_{ix} \vec{i} + \sum_i F_{iy} \vec{j}$$

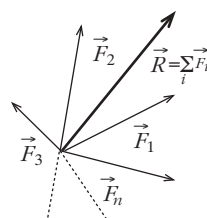


Fig. 1

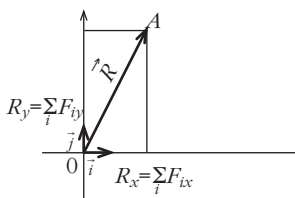


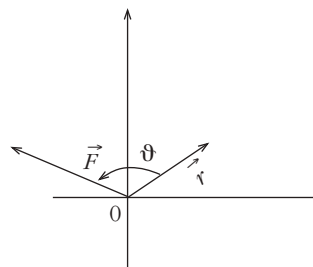
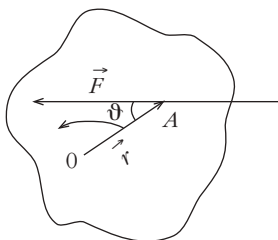
Fig. 2

R ha modulo: $R = \sqrt{(R_x^2 + R_y^2)}$

e direzione: $\text{tg} \alpha = \frac{R_y}{R_x}$

2 Il momento di una forza

Consideriamo una forza $\vec{F}$ applicata in un punto A su un corpo libero di ruotare attorno ad un punto O (Fig. 3.a).



Figg. 3.a e 3.b

L'effetto risultante sarà la rotazione del corpo attorno al punto O .

Il senso di rotazione del corpo è quello che va da $\vec{r}$ verso $\vec{F}$ (Fig. 3.b).

Si definisce **momento di una forza** $\vec{M}$ rispetto ad un punto O il prodotto vettoriale $\vec{M} = \vec{r} \times \vec{F}$ dove $\vec{r}$ è il vettore che unisce il punto O con il punto di applicazione della forza.

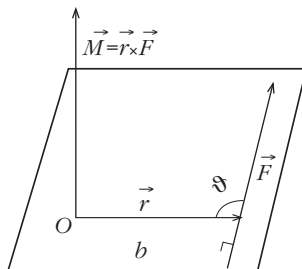


Fig. 4

Essendo il momento di una forza definito da un prodotto vettoriale si ha che:

- **Modulo di $\vec{M}$** : area del parallelogramma costruito sui vettori $\vec{r}$ e $\vec{F}$.
- **Direzione di $\vec{M}$** : perpendicolare al piano definito dai vettori $\vec{r}$ e $\vec{F}$.
- **Verso di $\vec{M}$** : definito dalla regola della mano destra o dalla regola della vite destrorsa (cfr. Cap. I vettori).

Per calcolare il modulo del momento di una forza è necessario prima definire il *braccio di una forza*.

a) Braccio di una forza

Si definisce **braccio di una forza** $\vec{F}$ rispetto ad un punto O il segmento di perpendicolare condotta da O rispetto alla retta di azione di $\vec{F}$.

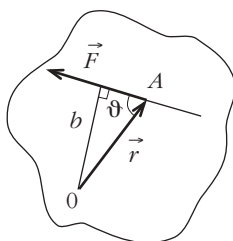


Fig. 5

b) Calcolo del modulo di $\vec{M}$

Dalla definizione di modulo di un prodotto vettoriale, per calcolare il modulo di $\vec{M}$ si deve prima costruire il parallelogramma sui vettori $\vec{r}$ e $\vec{F}$ e poi calcolarne l'area. Prendendo il vettore $\vec{F}$ come base e tracciando l'altezza $\overline{BH}$ si nota che questa coincide con il braccio della forza. Quindi:

$$M = Fb$$

Dalla figura 6 si ha che $b = r \sin \theta$ e quindi:

$$M = Fr \sin \theta$$

L'unità di misura di M nel Sistema Internazionale è newton × metro.

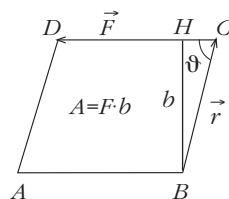


Fig. 6

c) Proprietà del momento di una forza

- Il momento di una forza non varia se si sposta il punto di applicazione della forza lungo la sua retta d'azione (b rimane invariato).
- Il momento di una forza rispetto ad un punto della sua retta d'azione è nullo ($b = 0$).

d) Momento di una forza in funzione delle componenti di $\vec{r}$ e $\vec{F}$

Se:

$$\vec{r} = r_x \vec{i} + r_y \vec{j} \quad \text{e} \quad \vec{F} = F_x \vec{i} + F_y \vec{j}$$

applicando le regole del prodotto vettoriale si ha che:

$$\vec{M} = (r_x F_y - r_y F_x) \vec{k}$$

DIMOSTRAZIONE

$$\begin{aligned} \vec{M} = \vec{r} \times \vec{F} &= (r_x \vec{i} + r_y \vec{j}) \times (F_x \vec{i} + F_y \vec{j}) = r_x F_x (\vec{i} \times \vec{i}) + r_x F_y (\vec{i} \times \vec{j}) + r_y F_x (\vec{j} \times \vec{i}) + r_y F_y (\vec{j} \times \vec{j}) = \\ &= r_x F_y \vec{k} - r_y F_x \vec{k} = (r_x F_y - r_y F_x) \vec{k} \end{aligned}$$

e) Momento di più forze concorrenti

Il **momento di più forze concorrenti** rispetto ad un punto è uguale al momento della forza risultante rispetto a quel punto.

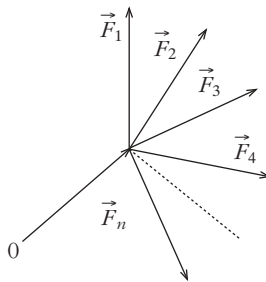


Fig. 7

$$\begin{aligned} \vec{M} &= \vec{r} \times \vec{R} \\ \vec{R} &= \sum \vec{F}_i \end{aligned}$$

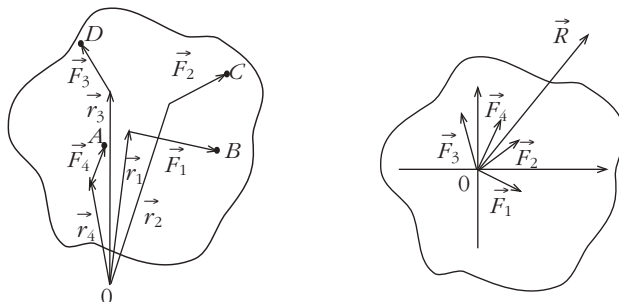
DIMOSTRAZIONE

$$\begin{aligned} M &= \sum_i \vec{M}_i = \vec{M}_1 + \vec{M}_2 + \vec{M}_3 + \vec{M}_4 + \dots + \vec{M}_n = r \times \vec{F}_1 + r \times \vec{F}_2 + r \times \vec{F}_3 + r \times \vec{F}_4 + \dots + r \times \vec{F}_n = \\ &= r \times (\vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4 + \dots + \vec{F}_n) = r \times \sum_i \vec{F}_i = \vec{r} \times \vec{R} \end{aligned}$$

3 Le forze applicate ad un corpo rigido

Quando su un corpo sono applicate più forze si hanno due effetti:

- la **traslazione** del corpo dovuta alla risultante delle forze applicate;
- la **rotazione** del corpo dovuta alla risultante dei momenti applicati.



Figg. 8.a e 8.b

La forza risultante $\vec{R}$, così come una qualsiasi forza, non ha un punto di applicazione definito e quindi si può supporre di applicarla in un punto qualsiasi del corpo (Fig. 8.b).

Se questo concetto sembra un po' astruso immaginate di spingere un carrello. La forza che applicate sul carrello non dipende dal punto di applicazione della forza.

Invece i momenti delle forze, e quindi il momento risultante $\vec{M}$, sono calcolati rispetto a un punto ben determinato.

Sarebbe quindi logico definire come punto di applicazione di $\vec{R}$ quello rispetto al quale il momento di $\vec{R}$ è uguale a $\vec{M}$. Tale punto, definito dal vettore $\vec{r}_c$, è detto **centro delle forze**. Per esso vale la relazione:

$$\vec{M} = \vec{r}_c \times \vec{R}$$

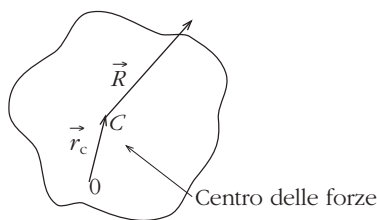


Fig. 8.c

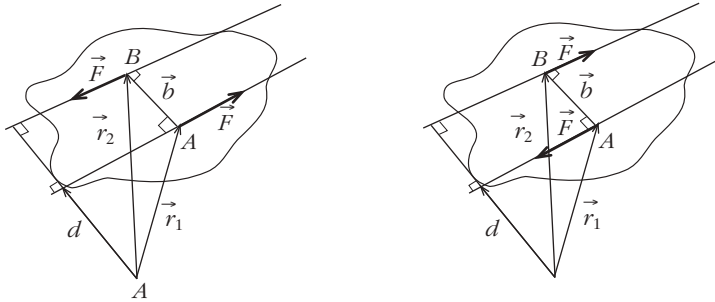
a) Coppia di forze

Per **coppia di forze** si intendono due forze della stessa intensità, agenti su rette parallele ma con verso opposto.

Se b è la distanza tra le due rette d'azione (detta braccio della coppia), il momento di una coppia di forze è il prodotto dell'intensità di una delle due forze per il braccio della coppia:

$$M = Fb$$

Consideriamo la figura 9.a e calcoliamo il momento della coppia di forze rispetto al punto A.



Figg. 9.a e 9.b

b) Momento della forza che agisce lungo la retta r_1 : $\vec{M}(r_1) = \vec{r}_1 \times \vec{F}$

Il modulo di tale forza è $M(r_1) = dF$, la direzione è perpendicolare al piano, e il verso, per la regola della mano destra (indice = r_1 ; medio = F) è entrante nel foglio (negativo).

c) Momento della forza che agisce lungo la retta r_2 : $\vec{M}(r_2) = \vec{r}_2 \times \vec{F}$

Il modulo di tale forza è $M(r_2) = (d + b)F$, la direzione è perpendicolare al piano, e il verso, per la regola della mano destra, è uscente dal foglio (positivo).

Quindi il modulo del momento della coppia di forze rispetto al punto A è:

$$M = M(r_1) - M(r_2) = Fd - F(d + b) = Fd - Fd - Fb = -Fb$$

In questo caso il momento risultante è negativo; se le forze avessero avuto il verso opposto (Fig. 9.b), il momento risultante sarebbe stato positivo:

$$M = M(r_1) - M(r_2) = -Fd + F(b + d) = F(-d + b + d) = +Fb$$

In ogni caso il momento della coppia è indipendente dal punto rispetto a cui si calcolano i momenti.

4 Composizione di forze parallele e centro delle forze parallele

Consideriamo un sistema di N forze parallele $\vec{F}_1, \vec{F}_2, \vec{F}_3, \dots, \vec{F}_N$ che agiscono su un corpo in direzione perpendicolare all'asse x .

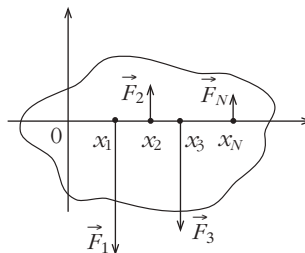


Fig. 10

La forza e il momento risultante sono dati da:

$$M = \sum_i M_i = \sum_i x_i F_i \quad \text{e} \quad R = \sum_i F_i$$

Quindi il centro delle forze parallele ha coordinata:

$$x_c = \frac{M}{R} = \frac{\sum_i x_i F_i}{\sum_i F_i}$$

Se le forze parallele sono distribuite in una regione tridimensionale (Fig. 11.a), le coordinate del punto C saranno:

$$x_c = \frac{\sum_i x_i F_i}{\sum_i F_i}; \quad y_c = \frac{\sum_i y_i F_i}{\sum_i F_i}; \quad z_c = \frac{\sum_i z_i F_i}{\sum_i F_i}$$

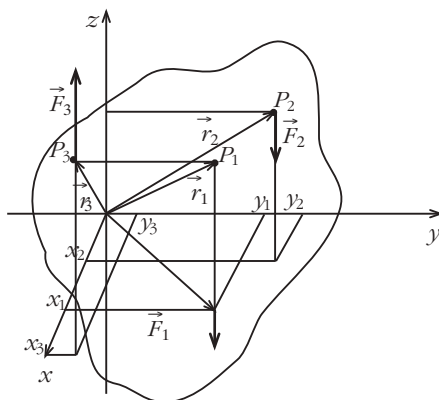


Fig. 11.a

Quindi, in forma vettoriale, il centro di forza sarà descritto dal vettore:

$$\vec{r}_c = \frac{\sum_i \vec{r}_i F_i}{\sum_i F_i}$$

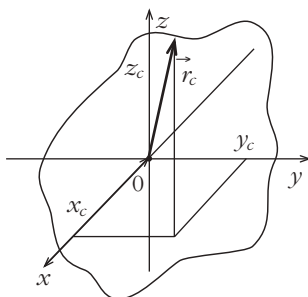


Fig. 11.b

a) Baricentro di un corpo

Il **baricentro**, o centro di massa di un corpo di massa M , è il centro delle sue forze gravitazionali.

In altre parole, tutto il peso di un corpo di massa M è applicato nel suo baricentro di coordinate:

$$\vec{r}_c = \frac{\sum_i \vec{r}_i m_i}{M}$$

DIMOSTRAZIONE

Un corpo di massa M è soggetto alla forza gravitazionale $\vec{F} = M\vec{g}$ diretta verso il basso.

$\vec{g}$ è l'accelerazione gravitazionale e può essere considerata costante.

Possiamo immaginare il corpo di massa M come suddiviso in un numero N di sotto-corpi aventi massa m_i . Su ognuno di questi sotto-corpi agisce la forza gravitazionale $\vec{F}_i = m_i \vec{g}$.

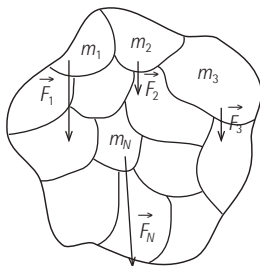


Fig. 12

Quindi possiamo pensare ad un corpo soggetto alla forza gravitazionale come ad un sistema su cui agiscono N forze parallele $\vec{F}_1, \vec{F}_2, \vec{F}_3, \dots, \vec{F}_N$.

Il centro di forza in questo caso è dato da:

$$\vec{r}_c = \frac{\sum_i \vec{r}_i \vec{F}_i}{\sum_i \vec{F}_i} = \frac{\sum_i \vec{r}_i (m_i \vec{g})}{\sum_i m_i \vec{g}} = \frac{\sum_i \vec{r}_i m_i}{\sum_i m_i} = \frac{\sum_i \vec{r}_i m_i}{M}$$

Le componenti di $\vec{r}_c$ sono:

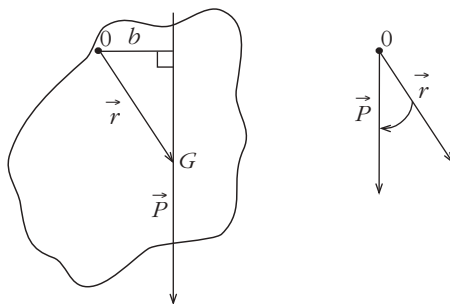
$$x_c = \frac{\sum_i x_i m_i}{M}; \quad y_c = \frac{\sum_i y_i m_i}{M}; \quad z_c = \frac{\sum_i z_i m_i}{M}$$

Se un corpo ha un centro di simmetria, il baricentro coincide con il centro di simmetria.

b) Equilibrio di un corpo

Abbiamo visto che la **traslazione** del corpo è dovuta alla risultante $\vec{R}$ delle forze applicate e che la **rotazione** è dovuta alla risultante $\vec{M}$ dei momenti applicati. Quindi se $\vec{R} = \vec{M} = 0$ il corpo è in equilibrio.

Consideriamo un corpo libero di ruotare attorno ad un perno 0.



Figg. 13.a e 13.b

La forza agente sul corpo è la sua forza peso $\vec{P}$ applicata nel baricentro G (Fig. 13.a).

A tale forza è associato il momento $\vec{M}$ dato da $\vec{M} = \vec{r} \times \vec{P}$ di modulo $M = bP$ il quale è responsabile della rotazione del corpo attorno al punto 0. Traslando il vettore $\vec{P}$ parallelamente a se stesso fino a far coincidere la sua coda con il punto 0, il corpo ruoterà fino a far coincidere $\vec{r}$ con il vettore $\vec{P}$ (Fig. 13.b).

La condizione di equilibrio in questo caso è $M = bP = 0$ per cui $b = 0$.

Quindi il corpo è in equilibrio se il perno attorno a cui ruota si trova sulla verticale condotta per il suo baricentro.

c) Equilibrio stabile, instabile e indifferente

Si possono distinguere tre tipi di equilibrio a seconda se il baricentro del corpo:

- è sotto l'asse di rotazione;
- è sopra l'asse di rotazione;
- coincide con l'asse di rotazione.

• Baricentro sotto l'asse di rotazione

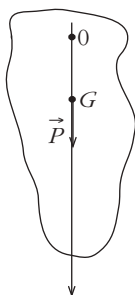


Fig. 14.a



Fig. 15.a

In questo caso si ha **equilibrio stabile**. Infatti, spostando il corpo dalla sua posizione di equilibrio (Fig. 14.a) si genera un momento $\vec{M}$ che tende a riportare il corpo nella posizione di partenza (Fig. 15.a).

• Baricentro sopra l'asse di rotazione

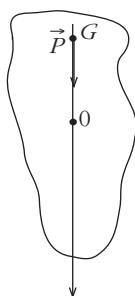


Fig. 14.b

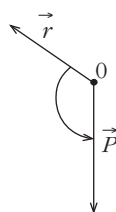


Fig. 15.b

In questo caso si ha **equilibrio instabile**. Infatti, spostando il corpo dalla sua posizione di equilibrio (Fig. 14.b) si genera un momento $\vec{M}$ che tende a spostare ulteriormente il corpo dalla posizione di partenza per portarlo nella posizione in cui il baricentro del corpo si trova sotto l'asse di rotazione (Fig. 15.b).

- Baricentro coincidente con l'asse di rotazione

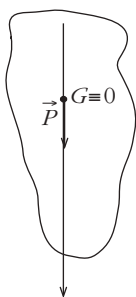


Fig. 14.c

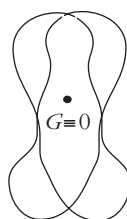


Fig. 15.c

In questo caso si ha **equilibrio indifferente**. Infatti, pur spostando il corpo dalla sua posizione di equilibrio (Fig. 14.c), il baricentro continuerà a coincidere con l'asse vincolare e quindi il corpo continua ad essere in equilibrio (Fig. 15.c).

5 Le leve

La leva è un'asta rigida con un asse fisso detto fulcro (F).

Ad essa vengono applicate due forze: la forza motrice $\vec{M}$ e la forza resistente $\vec{R}$.

A seconda della posizione del fulcro, della forza motrice e della forza resistente si possono distinguere tre tipi di leve.

a) Leva di primo genere: Forza resistente - Fulcro - Forza motrice

La condizione di equilibrio si ha quando il momento della forza motrice è uguale al momento della forza resistente.

Dalla figura 16 si ha che questo succede quando $M \cdot a = R \cdot b$ da cui segue che la condizione d'equilibrio è data da:

- se $a > b \rightarrow M < R$ e la leva è *vantaggiosa*. È possibile bilanciare una forza resistente applicando una forza di intensità minore;
- se $a < b \rightarrow M > R$ e la leva è *svantaggiosa*. Per bilanciare una forza resistente si deve applicare una forza di intensità maggiore;
- se $a = b \rightarrow M = R$ e la leva è *indifferente*. Per bilanciare una forza resistente si deve applicare una forza di uguale intensità.

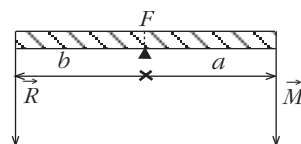


Fig. 16

b) Leva di secondo genere: Fulcro - Forza resistente - Forza motrice

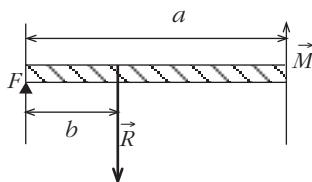


Fig. 17

In questo caso (Fig. 17) il braccio motrice è sempre maggiore del braccio resistente ($a > b$) e la leva è sempre vantaggiosa.

c) Leva di terzo genere: Fulcro - Forza motrice - Forza resistente

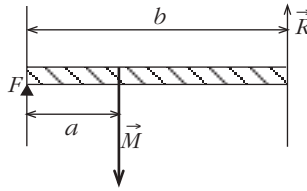


Fig. 18

In questo caso (Fig. 18) il braccio motrice è sempre minore del braccio resistente ($a < b$) e la leva è sempre svantaggiosa.

4

La cinematica

1 I termini della cinematica

Iniziamo ad analizzare alcuni termini della cinematica che utilizzeremo in questo capitolo e che saranno applicati ai diversi tipi di moto.

- **Traiettoria** è la linea che unisce le successive posizioni acquisite dal corpo stesso.
- **Spazio percorso** è la lunghezza della traiettoria.
- **Velocità media** è il rapporto fra lo spazio Δs percorso da un corpo in un intervallo di tempo Δt e questo stesso intervallo di tempo:

$$\vec{v}_m = \frac{\Delta s}{\Delta t}$$

Dimensionalmente: $[v] = [l t^{-1}] = m/s$

- **Velocità istantanea** è la velocità media del punto materiale relativa ad un intervallo di tempo piccolissimo, al limite tendente a zero:

$$\vec{v} = \frac{\Delta \vec{s}}{\Delta t} \quad \Delta t \cong 0$$

$$\vec{v} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \vec{s}}{\Delta t} = \frac{d\vec{s}}{dt}$$

da cui deriva:

$$\vec{s} = \vec{s}_0 + \int_{t_0}^t \vec{v} dt \quad (1.1)$$

- **Accelerazione media** è il rapporto fra la variazione di velocità nell'intervallo di tempo Δt e questo stesso intervallo di tempo:

$$\vec{a}_m = \frac{\Delta \vec{v}}{\Delta t}$$

Dimensionalmente: $[a] = \frac{[v]}{[t]} = \frac{[l t^{-1}]}{[t]} = [l t^{-2}] = m/s^2$

- **Accelerazione istantanea** è l'accelerazione media relativa ad un intervallo di tempo piccolissimo, al limite tendente a zero:

$$\vec{a} = \frac{\Delta \vec{v}}{\Delta t} \quad \Delta t \cong 0$$

$$\vec{a} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \vec{v}}{\Delta t} = \frac{d\vec{v}}{dt}$$

da cui deriva che:

$$\vec{v} = \vec{v}_0 + \int_{t_0}^t \vec{a} dt \quad (1.2)$$

A seconda della traiettoria, il **moto** può essere:

- **rettilineo**: la traiettoria del moto è una linea retta;
- **circolare**: la traiettoria del moto è una circonferenza;
- **curvilineo**: la traiettoria del moto è una curva.

2 Il moto rettilineo uniforme

Un corpo si muove di **moto rettilineo uniforme** quando percorre una linea retta con velocità costante:

$$v = \text{cost}$$

a) Accelerazione media e accelerazione istantanea

Nel moto rettilineo uniforme l'**accelerazione** è **nulla**.

DIMOSTRAZIONE

Nel moto rettilineo uniforme il corpo mantiene sempre la stessa velocità v_0 e quindi $\Delta v = 0$. Dalla definizione di accelerazione $a = \frac{\Delta v}{\Delta t}$ segue che, se $\Delta v = 0$, $a = 0$.

b) Velocità media e velocità istantanea

Nel moto uniforme la **velocità media** coincide con la **velocità istantanea**:

$$v_m = v = \frac{x - x_0}{t - t_0}$$

DIMOSTRAZIONE

Nel moto uniforme la velocità è costante sempre, indipendentemente dall'intervallo di tempo considerato.

Per questo la velocità istantanea, calcolata su un intervallo di tempo piccolissimo, è uguale alla velocità media calcolata su un intervallo di tempo finito:

$$v_m = v = \frac{x - x_0}{t - t_0} \quad (2.1)$$

c) Spazio percorso

Lo **spazio percorso** nel moto rettilineo uniforme è dato da:

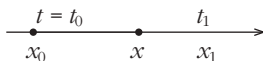


Fig. 1

$$x = v(t - t_0) + x_0$$

DIMOSTRAZIONE

Dalla (2.1):

$$x = v(t - t_0) + x_0 \quad (2.2)$$

Dalla (1.1) siccome il moto rettilineo avviene in una sola direzione, possiamo far coincidere tale direzione con l'asse x :

$$\vec{x} = \vec{x}_0 + \int_{t_0}^t v dt$$

Essendo v costante:

$$x = x_0 + \int_{t_0}^t v dt = x_0 + v \int_{t_0}^t dt = x_0 + v(t - t_0)$$

GRAFICO VELOCITÀ-TEMPO

Il grafico velocità-tempo di un moto rettilineo uniforme è rappresentato in figura 2.

Nell'intervallo di tempo Δt lo spazio percorso è dato da:

$$s = v_0 \Delta t = v_0(t - t_0)$$

che è l'area del rettangolo che ha per base Δt e per altezza v_0 (tratteggiata in figura).

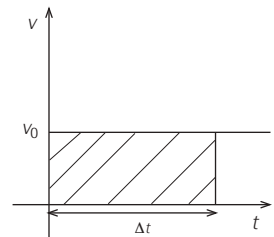


Fig. 2

GRAFICO SPAZIO-TEMPO

Il grafico spazio-tempo del moto rettilineo uniforme è rappresentato in figura 3.

La pendenza della retta è data dal rapporto $\frac{\Delta x}{\Delta t}$ che è proprio la definizione di velocità.

Quindi la velocità coincide con la pendenza della retta del grafico spazio-tempo.

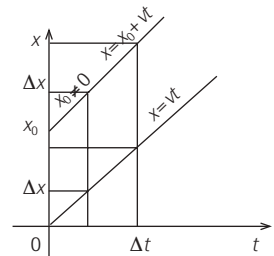


Fig. 3

3 Il moto rettilineo uniformemente accelerato

Un corpo si muove di **moto uniformemente accelerato/decelerato** se la sua accelerazione a è costante.

Se:

- $a > 0$ il moto è uniformemente accelerato;
- $a < 0$ il moto è uniformemente decelerato.

a) Accelerazione media e accelerazione istantanea

Nel moto uniformemente accelerato l'**accelerazione media** coincide con l'accelerazione istantanea:

$$a_m = a = \frac{v_1 - v_0}{t_1 - t_0}$$

DIMOSTRAZIONE

Essendo l'accelerazione costante essa avrà lo stesso valore sia in ogni istante (accelerazione istantanea) che in un intervallo di tempo Δt (accelerazione media).

Quindi:

$$a_m = a = \frac{v_1 - v_0}{t_1 - t_0}$$

b) Velocità media e velocità istantanea

Nel moto uniformemente accelerato la velocità media **non** coincide con la velocità istantanea:

$$v_m = \frac{x - x_0}{t - t_0} \quad \text{e} \quad v = v_0 + a(t - t_0)$$

DIMOSTRAZIONE**VELOCITÀ MEDIA**

Per il moto uniformemente accelerato continuerà a valere la definizione di velocità media:

$$v_m = \frac{\Delta x}{\Delta t} = \frac{x_1 - x_0}{t_1 - t_0}$$

VELOCITÀ ISTANTANEA

Dalla definizione di accelerazione $a = \frac{v - v_0}{t - t_0}$, segue che $v = v_0 + a(t - t_0)$.

Dalla (1.2) ed essendo a costante:

$$v = v_0 + \int_{t_0}^t a dt = v_0 + a \int_{t_0}^t dt = v_0 + a(t - t_0)$$

c) Spazio percorso

Nel moto uniformemente accelerato, si ha:

$$x = x_0 + v_0(t - t_0) + \frac{at}{2}(t - t_0)$$

DIMOSTRAZIONE

Per calcolare lo spazio percorso nel moto uniformemente accelerato possiamo partire dalla formula:

$$x = x_0 + v_m(t - t_0)$$

che vale per qualunque tipo di moto e calcolare la velocità media v_m .

La velocità media può essere considerata come la media aritmetica fra la velocità iniziale e la velocità finale del corpo:

$$v_m = \frac{v_1 + v_0}{2}$$

Nel moto uniformemente accelerato $v_1 = v_0 + at$, quindi:

$$v_m = \frac{v_0 + (v_0 + at)}{2} = v_0 + \frac{at}{2}$$

e sostituendo otteniamo:

$$x = x_0 + \left(v_0 + \frac{at}{2}\right)(t - t_0) = x_0 + v_0(t - t_0) + \frac{at}{2}(t - t_0)$$

Se $t_0 = 0$, allora:

$$x = x_0 + v_0 t + \frac{1}{2}at^2$$

Se il moto è uniformemente decelerato ($a < 0$):

$$s = v_1 t - \frac{1}{2}at^2$$

Dalla (1.1) ed essendo $v = v_0 + at$:

$$x = x_0 + \int_{t_0}^t v dt = x_0 + \int_{t_0}^t (v_0 + at) dt = x_0 + v_0 \int_{t_0}^t dt + a \int_{t_0}^t t dt = v_0(t - t_0) + \frac{1}{2}at^2 \Big|_{t_0}^t = \\ = v(t - t_0) + \frac{1}{2}a(t^2 - t_0^2)$$

GRAFICO ACCELERAZIONE-TEMPO

In figura 4 è rappresentato il grafico accelerazione-tempo per un moto uniformemente accelerato.

$$a(t) = a_0$$

Se il corpo parte all'istante t_0 , la velocità all'istante t è rappresentata dall'area tratteggiata di lati a_0 e $(t - t_0)$.

Infatti dalla formula $a = \frac{v - v_0}{t - t_0}$ ed essendo $v_0 = 0$:

$$v = a_0(t - t_0)$$

v quindi è l'area del rettangolo di base $t - t_0$ e altezza a_0 .

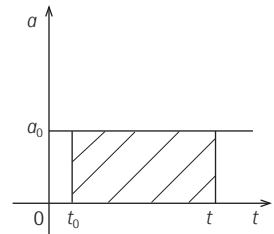


Fig. 4

GRAFICO VELOCITÀ-TEMPO

In figura 5 è visualizzato il grafico velocità-tempo nel moto uniformemente accelerato.

$$v(t) = at$$

La pendenza della retta coincide con l'accelerazione del corpo.

La pendenza della retta è data da $\frac{\Delta v}{\Delta t}$.

Tale valore coincide con la definizione stessa di accelerazione.

La funzione $v(t) = at$, che definisce in ogni istante la velocità del corpo, è una retta di coefficiente angolare a .

Lo spazio percorso tra gli istanti t e t_0 è definito dalla sezione di area sottesa dalla retta tra gli istanti t e t_0 (area tratteggiata in figura).

L'area tratteggiata è data cioè dalla differenza fra l'area del triangolo ACK e l'area del triangolo ABH :

$$\text{Area } ACK = \frac{1}{2}(t \times at) = \frac{1}{2}at^2 \quad \text{e} \quad \text{Area } ABH = \frac{1}{2}(t_0 \times at_0) = \frac{1}{2}at_0^2$$

Sottraendo tra loro le due aree:

$$x = \frac{1}{2}a(t^2 - t_0^2)$$

che è proprio la formula che definisce lo spazio percorso.

La sezione di area della funzione $v = at$ tra gli istanti t e t_0 è definita dall'integrale definito tra i suddetti istanti:

$$x = \int_{t_0}^t v dt = \int_{t_0}^t at dt = a \int_{t_0}^t t dt = \frac{1}{2}at^2 \Big|_{t_0}^t = \frac{1}{2}a(t^2 - t_0^2)$$

GRAFICO SPAZIO-TEMPO

In figura 6 è visualizzato il grafico spazio-tempo nel moto uniformemente accelerato nel caso in cui $v_0 = 0$, $t_0 = 0$, $x_0 = 0$.

$$x(t) = \frac{1}{2}at^2$$

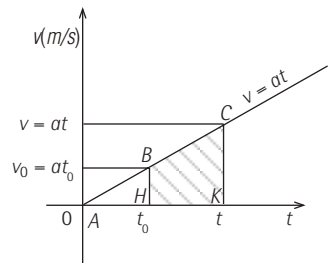


Fig. 5

La pendenza della curva coincide con la velocità del corpo.

Consideriamo gli istanti t_1 e t_2 . All'istante t_1 il corpo si trova in x_1 e all'istante t_2 il corpo si trova in x_2 . La velocità media tra tali istanti:

$$v_m = \frac{x_2 - x_1}{t_2 - t_1}$$

coincide con la pendenza della secante $\overline{P_1P_2}$ alla curva.

Diminuendo l'intervallo di tempo $\Delta t = t_2 - t_1$ si vede che tale secante tende a coincidere con la tangente alla curva e la velocità media tende alla velocità istantanea. Quindi la pendenza della tangente ad una curva in un punto coincide con la velocità del corpo in quel punto.

La derivata ad una curva in un punto indica la tangente alla curva in quel punto.

Derivando la funzione:

$$\frac{d}{dt}x(t) = \frac{d}{dt}\left(\frac{1}{2}at^2\right) = \frac{1}{2}2at = at$$

che definisce la velocità nel moto rettilineo uniforme.

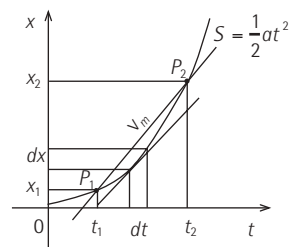


Fig. 6

• Riassumendo

Tipo di moto rettilineo	Caratteristica	Spazio	Velocità	Accelerazione
Uniforme	$v = \text{costante}$	$x = x_0 + v(t - t_0)$	costante	0
Uniformemente accelerato	$a = \text{costante}$	$x = x_0 + v(t - t_0) + \frac{at}{2}(t - t_0)$	$v = v_0 + a(t - t_0)$	costante

Tab. 1

4 La caduta dei gravi

Tutti i corpi soggetti al campo gravitazionale terrestre cadono con accelerazione costante g detta **accelerazione di gravità**.

Alla nostra latitudine $g \approx 9,8\text{m/s}^2$.

Quindi il moto di un corpo in caduta libera è uniformemente accelerato con accelerazione g e se introduciamo un sistema di assi cartesiani con l'asse y diretto verso il basso le sue equazioni del moto sono:

$$y = y_0 + v_0(t - t_0) + \frac{gt}{2}(t - t_0)$$

$$v = v_0 + g(t - t_0)$$

— Se il corpo è lasciato cadere all'istante $t_0 = 0$ esse diventano:

$$y = y_0 + v_0 t + \frac{gt^2}{2}$$

$$v = v_0 + gt$$

— Se si fa coincidere l'origine della caduta con l'origine degli assi cartesiani ($y_0 = 0$):

$$y = v_0 t + \frac{gt^2}{2}$$

$$v = v_0 + gt$$

— Se la velocità iniziale del corpo è nulla ($\mathbf{v}_0 = \mathbf{0}$):

$$y = \frac{gt^2}{2}$$

$$v = gt$$

5 Il moto curvilineo

Lo spostamento, la velocità e l'accelerazione di un corpo sono dei vettori e quindi definiti da una lunghezza, da una direzione e da un verso.

Nel moto rettilineo non ci siamo preoccupati dell'aspetto vettoriale di queste grandezze in quanto il moto avveniva in una sola direzione.

Nel **moto curvilineo** si deve invece tener conto di questo aspetto perché il moto avviene su una traiettoria curvilinea in cui lo spostamento, la velocità e l'accelerazione variano sia in modulo sia in direzione.

Consideriamo una particella che percorre una traiettoria curvilinea. In ogni istante la sua posizione è definita dal vettore posizione $\vec{r}$.

a) Velocità

Supponiamo che il corpo si trovi all'istante t_1 nel punto A, definito dal vettore $\vec{r}_1$ e all'istante t_2 nel punto B, definito dal vettore $\vec{r}_2$.

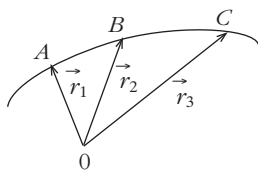


Fig. 7

La variazione di posizione è definita dal vettore $\Delta\vec{r} = \vec{r}_2 - \vec{r}_1$.

La **velocità media** è rappresentata dal vettore:

$$\vec{v}_m = \frac{\Delta\vec{r}}{\Delta t}$$

parallelo al vettore posizione.

La **velocità istantanea** è data da:

$$\vec{v} = \lim_{\Delta t \rightarrow 0} \vec{v}_m = \lim_{\Delta t \rightarrow 0} \frac{\Delta\vec{r}}{\Delta t} = \frac{d\vec{r}}{dt}$$

Quando $\Delta t \rightarrow 0$ il punto A si avvicina sempre di più al punto B, il vettore $\vec{r}_2$ si avvicina sempre di più al vettore $\vec{r}_1$ (Fig. 9.a) e $\Delta\vec{r}$ tende ad assumere la direzione della tangente alla curva nel punto A.

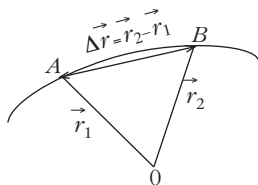
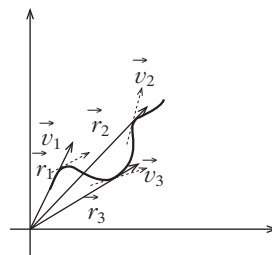
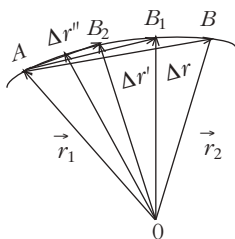


Fig. 8



Figg. 9.a e 9.b

Quindi in ogni istante la velocità nel moto curvilineo è un vettore tangente la curva in quell'istante (Fig. 9.b).

b) Vettore spostamento e velocità in funzione delle componenti

Nel capitolo sui vettori abbiamo visto che è sempre possibile esprimere un vettore come somma dei suoi vettori componenti.

Possiamo quindi scomporre i vettori $\vec{r}$ e $\vec{v}$ (Fig. 10) in funzione delle loro componenti: $\vec{r} = r_x \vec{i} + r_y \vec{j}$; $\vec{v} = v_x \vec{i} + v_y \vec{j}$, e calcolarne i moduli:

$$r = |\vec{r}| = \sqrt{r_x^2 + r_y^2}; \quad v = |\vec{v}| = \sqrt{v_x^2 + v_y^2}$$

Tra le componenti del vettore velocità e del vettore posizione vale la relazione: $v_x = \frac{dr_x}{dt}$; $v_y = \frac{dr_y}{dt}$

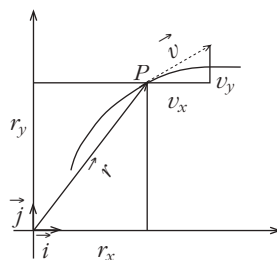


Fig. 10

DIMOSTRAZIONE

Infatti, scomponendo il vettore $\vec{v}$ nelle sue componenti si ha che:

$$\vec{v} = v_x \vec{i} + v_y \vec{j} \text{ ma:}$$

$$\vec{v} = \frac{d\vec{r}}{dt} = \frac{d}{dt}(r_x \vec{i} + r_y \vec{j}) = \frac{dr_x}{dt} \vec{i} + \frac{dr_y}{dt} \vec{j}$$

confrontando i secondi membri si ha che:

$$v_x = \frac{dr_x}{dt}, \quad v_y = \frac{dr_y}{dt}.$$

c) Moto curvilineo e accelerazione

Il vettore velocità è, in ogni istante, tangente alla traiettoria, quindi, da istante a istante, essa varierà sia in modulo sia in direzione.

L'**accelerazione** è definita come la variazione del **vettore** velocità nel tempo.

Nel moto rettilineo l'accelerazione è legata soltanto alla variazione di velocità in modulo in quanto il moto avviene su una traiettoria rettilinea.

Nel moto curvilineo, invece, la velocità può variare sia in direzione sia in modulo e quindi l'accelerazione dovrà tener conto di entrambe queste variazioni. Quindi sarà del tipo $\vec{a} = \vec{a}_1 + \vec{a}_2$, dove:

- $\vec{a}_1$ è l'accelerazione legata alla variazione di velocità in modulo, detta **accelerazione tangenziale**;
- $\vec{a}_2$ è l'accelerazione legata alla variazione di velocità in direzione, detta **accelerazione normale**.

Supponiamo, dunque, che una particella si trovi all'istante t_1 nel punto P_1 con velocità $\vec{v}_1$ e all'istante t_2 nel punto P_2 con velocità $\vec{v}_2$ (Fig. 11).

La differenza di velocità $\Delta\vec{v} = \vec{v}_2 - \vec{v}_1$ è un vettore diretto sempre verso la concavità della curva.

L'**accelerazione media** è definita da:

$$\vec{a}_m = \frac{\Delta\vec{v}}{\Delta t}$$

ed è un vettore parallelo a $\Delta\vec{v}$.

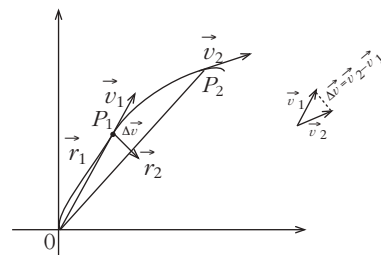


Fig. 11

L'**accelerazione istantanea** è definita da:

$$\vec{a} = \frac{d\vec{v}}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \vec{v}}{\Delta t}$$

d) Vettore accelerazione in funzione delle componenti

Possiamo scomporre il vettore $\vec{a}$ in $\vec{a} = a_x \vec{i} + a_y \vec{j}$:

$$a = |\vec{a}| = \sqrt{a_x^2 + a_y^2}$$

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d}{dt}(v_x \vec{i} + v_y \vec{j}) = \frac{dv_x}{dt} \vec{i} + \frac{dv_y}{dt} \vec{j} = a_x \vec{i} + a_y \vec{j}$$

Quindi:

$$\frac{dv_x}{dt} = a_x; \quad \frac{dv_y}{dt} = a_y$$

e) Moto piano con accelerazione costante (moto uniformemente accelerato)

Un **moto piano con accelerazione costante** è ancora detto uniformemente accelerato. In questo caso le equazioni del moto ricavate per il moto rettilineo continuano a valere per le componenti del moto.

	Componente x	Componente y
$\vec{v} = \vec{v}_0 + \vec{a}t$	$v_x = v_{0x} + a_x t$	$v_y = v_{0y} + a_y t$
$\vec{s} = \vec{s}_0 + \vec{v}t + \frac{1}{2}\vec{a}t^2$	$x = x_0 + v_{0x} t + \frac{1}{2}a_x t^2$	$y = y_0 + v_{0y} t + \frac{1}{2}a_y t^2$

Tab. 2

COME STUDIARE IL MOTO PIANO UNIFORMEMENTE ACCELERATO DI UN CORPO

Per lo studio del moto di un corpo si deve:

1. Definire la velocità del corpo all'istante iniziale e scomporla lungo gli assi cartesiani.
2. Definire le forze che agiscono sul corpo, e quindi le accelerazioni, e scomporle lungo gli assi cartesiani.
3. Definire il tipo di moto lungo ogni asse e scriverne le equazioni. Si ricordi che i moti lungo i due assi sono indipendenti fra di loro.
4. Ricomporre le equazioni del moto.

f) Moto di un proiettile

Il **moto di un proiettile** è un esempio di moto curvilineo uniformemente accelerato. Esso è infatti caratterizzato da un'accelerazione $\vec{g}$ costante. Studieremo due casi particolari:

- il proiettile ha velocità iniziale **parallela** all'asse x ;
- il proiettile ha velocità iniziale **incidente** l'asse x .

In entrambi i casi applicheremo i punti definiti nel paragrafo precedente.

• Moto di un proiettile con velocità iniziale parallela all'asse x

Studiamo questo moto sviluppando i 4 punti definiti in precedenza.

1. Definizione della velocità del corpo all'istante iniziale e scomposizione lungo gli assi

All'istante iniziale il proiettile ha velocità parallela all'asse x :

$$\vec{v}_0 = v_0 \vec{i}$$

Quindi: $v_{0x} = v_0$ $v_{0y} = 0$.

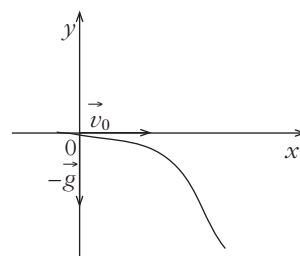


Fig. 12

2. Definizione delle forze che agiscono sul corpo, e quindi delle accelerazioni, e scomposizione lungo gli assi cartesiani

L'unica forza agente è quella di gravità agente verso il basso che provoca la caduta del proiettile con accelerazione g :

$$\vec{g} = -g\vec{j}$$

Quindi: $a_x = 0$ e $a_y = -g$.

3. Definizione del tipo di moto lungo ogni asse e relative equazioni

Lungo l'asse x non esiste accelerazione e quindi il moto è uniforme con velocità v_0 :

$$v_x(t) = v_0$$

$$x(t) = v_0 t$$

Lungo l'asse y il moto è uniformemente decelerato (g negativa) con accelerazione $-g$:

$$v_y(t) = -gt$$

$$y(t) = -\frac{1}{2}gt^2$$

4. Ricomposizione dei risultati

$$\vec{v}(t) = v_x(t)\vec{i} + v_y(t)\vec{j} = v_0\vec{i} - gt\vec{j}$$

$$\vec{s}(t) = x(t)\vec{i} + y(t)\vec{j} = v_0 t\vec{i} - \frac{1}{2}gt^2\vec{j}$$

• Moto di un proiettile con velocità iniziale incidente l'asse x

Studiamo questo moto sviluppando i 4 punti soliti.

1. Definizione della velocità del corpo all'istante iniziale e scomposizione lungo gli assi

All'istante iniziale il proiettile ha velocità:

$$\vec{v}_0 = v_{0x}\vec{i} + v_{0y}\vec{j}$$

Quindi: $v_x = v_{0x}$ e $v_y = v_{0y}$.

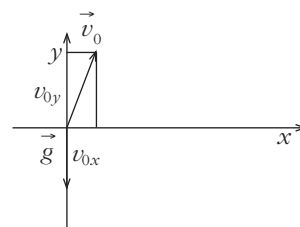


Fig. 13

2. Definizione delle forze che agiscono sul corpo, e quindi delle accelerazioni, e scomposizione lungo gli assi cartesiani

L'unica forza agente è quella di gravità agente verso il basso che provoca la caduta del proiettile con accelerazione g :

$$\vec{g} = -g\vec{j}$$

Quindi: $a_x = 0$ e $a_y = -g$.

3. Definizione del tipo di moto lungo ogni asse e relative equazioni

Lungo l'asse x non esiste accelerazione e quindi il moto è uniforme con velocità v_{0x} :

$$v_x(t) = v_{0x} \quad \text{e} \quad x(t) = v_{0x}t.$$

Lungo l'asse y esiste l'accelerazione g . Il moto è uniformemente accelerato con accelerazione g :

$$v_y(t) = v_{0y} - gt \quad \text{e} \quad y(t) = v_{0y}t - \frac{1}{2}gt^2.$$

4. Ricomposizione dei risultati

$$\vec{v}(t) = v_x(t)\vec{i} + v_y(t)\vec{j} = v_{0x}\vec{i} + (v_{0y} - gt)\vec{j}$$

$$\vec{s}(t) = x(t)\vec{i} + y(t)\vec{j} = v_{0x}t\vec{i} + \left(v_{0y}t - \frac{1}{2}gt^2\right)\vec{j}$$

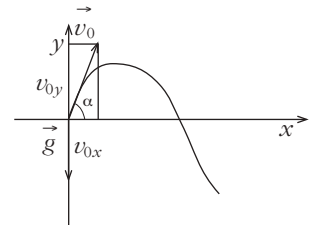
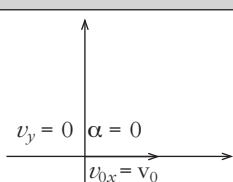
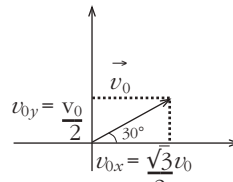
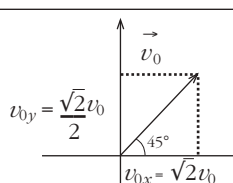


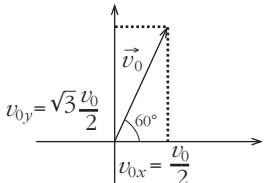
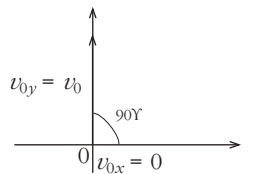
Fig. 14

g) Studio delle componenti di $\vec{v}_0$

Indicando con α l'angolo di incidenza della velocità iniziale, consideriamo i seguenti casi particolari:

α		v_{0x}	v_{0y}
0	Fig. 15.a 	v_0	0
30°	Fig. 15.b 	$\frac{\sqrt{3}}{2}v_0$	$\frac{1}{2}v_0$
45°	Fig. 15.c 	$\frac{\sqrt{2}}{2}v_0$	$\frac{\sqrt{2}}{2}v_0$

(Segue)

60°	Fig. 15.d		$\frac{1}{2}v_0$	$\frac{\sqrt{3}}{2}v_0$
90°	Fig. 15.e		0	v_0

Tab. 3

$$v_{0x} = v_0 \cos \alpha \quad \text{e} \quad v_{0y} = v_0 \sin \alpha.$$

h) Componente tangenziale e normale dell'accelerazione

Consideriamo una particella che descrive una traiettoria curvilinea.

Supponiamo che la particella all'istante t abbia velocità $\vec{v}$ e accelerazione $\vec{a}$ sempre diretta verso la concavità della curva.

Il vettore $\vec{a}$ è sempre **scomponibile** in una componente $\vec{a}_N$ normale alla curva e in una componente $\vec{a}_T$ tangente alla curva:

$$\vec{a} = \vec{a}_N + \vec{a}_T = a_T \vec{u}_T + a_N \vec{u}_N$$

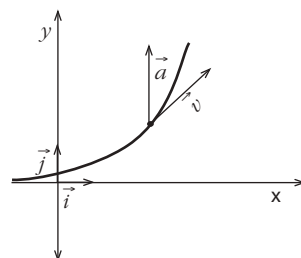


Fig. 16

dove:

- $\vec{a}_N$ è l'accelerazione dovuta alla variazione di velocità in direzione;
- $\vec{a}_T$ è l'accelerazione dovuta alla variazione di velocità in modulo.

DIMOSTRAZIONE

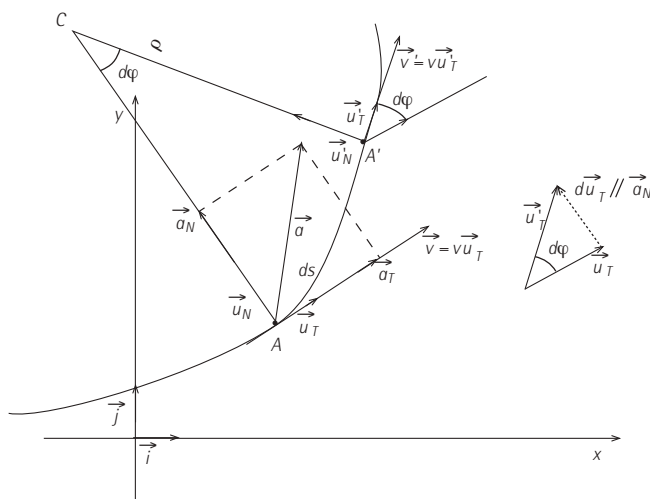


Fig. 17

In figura 17 tracciamo in A il versore tangente alla curva $\vec{u}_\tau$:

$$\vec{v} = v\vec{u}_\tau$$

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d}{dt}(v\vec{u}_\tau) = \vec{u}_\tau \frac{dv}{dt} + v \frac{d\vec{u}_\tau}{dt} \quad (5.1)$$

Il primo termine è legato alla variazione di velocità in modulo.

Il secondo termine è legato alla variazione di velocità in direzione.

— Se la velocità variesse in modulo ma non in direzione (**moto rettilineo**):

$$\frac{d\vec{u}_\tau}{dt} = 0 \Rightarrow \vec{a} = \vec{u}_\tau \frac{dv}{dt} = a\vec{u}_\tau$$

e l'accelerazione sarebbe tangente alla curva.

— Se la velocità variesse in direzione ma non in modulo (**moto uniforme**):

$$\frac{dv}{dt} = 0 \Rightarrow \vec{a} = v \frac{d\vec{u}_\tau}{dt}$$

Per calcolare $\frac{d\vec{u}_\tau}{dt}$ tracciamo un vettore $\vec{u}_n$ perpendicolare alla curva nel punto A .

Supponiamo che la particella si trovi nel punto A' all'istante t' .

Tracciamo il versore tangente $\vec{u}'_\tau$ e il versore $\vec{u}'_n$ normale alla curva in A' .

Le normali in A e A' si intersecano nel punto C (centro di curvatura).

Indicando con $d\varphi$ l'angolo al centro si ha:

$$ds = \rho d\varphi$$

dove ρ è il raggio di curvatura.

Il vettore $d\vec{u}_\tau = \vec{u}'_\tau - \vec{u}_\tau$ è parallelo al versore $\vec{u}_n$ e ha modulo uguale a $d\varphi$:

$$d\vec{u}_\tau = \vec{u}_n d\varphi$$

$$\frac{d\vec{u}_\tau}{dt} = \vec{u}_n \frac{d\varphi}{dt} = \vec{u}_n \left(\frac{d\varphi}{ds} \cdot \frac{ds}{dt} \right)$$

Essendo $\frac{ds}{dt} = v$ (infatti è lo spazio realmente percorso dalla particella diviso per il tempo) e $\frac{d\varphi}{ds} = \frac{1}{\rho}$, si ha:

$$\frac{d\vec{u}_\tau}{dt} = \frac{v}{\rho} \vec{u}_n$$

Sostituendo nella (5.1):

$$\vec{a} = \frac{d\vec{v}}{dt} = \frac{d}{dt}(v\vec{u}_\tau) = \vec{u}_\tau \frac{dv}{dt} + v \frac{d\vec{u}_\tau}{dt} = a_\tau \vec{u}_\tau + \frac{v^2}{\rho} \vec{u}_n$$

Quindi $\vec{a} = a_\tau \vec{u}_\tau + a_n \vec{u}_n$ dove:

$$a_\tau = \frac{dv}{dt} \quad \text{e} \quad a_n = \frac{v^2}{\rho}$$

Questa è la formula più generale dell'accelerazione.

Nel caso del moto rettilineo si ha variazione della velocità in modulo e non in direzione:

$$\vec{a} = a_\tau \vec{u}_\tau$$

ed è un vettore avente la stessa direzione dello spostamento.

Nel caso del moto circolare uniforme si ha variazione della velocità in direzione e non in modulo, ρ coincide con il raggio R della circonferenza:

$$\vec{a} = \frac{v^2}{R} \vec{u}_n$$

che è la formula dell'accelerazione centripeta che vedremo nel paragrafo successivo.

Nel caso più generale l'accelerazione avrà entrambe le componenti.

6 Il moto circolare uniforme

Consideriamo il caso particolare in cui una particella percorre una circonferenza con velocità costante.

Supponiamo che il corpo occupi, all'istante $t = 0$, la posizione P_0 . Dopo aver percorso un giro completo il corpo si troverà di nuovo in P_0 .

a) Periodo T

Il periodo T è l'intervallo di tempo impiegato dalla particella per compiere un giro completo. L'unità di misura, essendo un tempo, è il secondo:

$$[T] = [t]$$

b) Frequenza ν

La frequenza ν è il numero di giri compiuti dalla particella nell'unità di tempo. L'unità di misura è l'hertz (simbolo Hz).

Dimensionalmente si ha che:

$$[\nu] = [t^{-1}]$$

Dalle due definizioni segue che:

$$T = \frac{1}{\nu}$$

c) Velocità tangenziale e velocità angolare

Un corpo che percorre una circonferenza di raggio R con periodo T possiede sia una velocità tangenziale che una velocità angolare.

- Velocità tangenziale

La **velocità tangenziale** è la velocità con cui il corpo percorre la circonferenza. Essa è pari a:

$$v = \frac{2\pi R}{T} = 2\pi R\nu \quad (6.1)$$

DIMOSTRAZIONE

Se all'istante iniziale la particella si trova nel punto P_0 (Fig. 18), dopo un periodo T la particella avrà percorso l'intera circonferenza di perimetro $P = 2\pi R$ ritornando nel punto P_0 .

La velocità con cui ha percorso l'intera circonferenza è data da:

$$v = \frac{\text{spazio percorso}}{\text{tempo impiegato}} = \frac{2\pi R}{T} = 2\pi R\nu$$

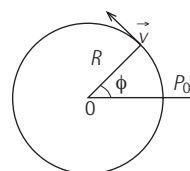


Fig. 18

- Velocità angolare

La **velocità angolare** è la velocità con cui viene spazzato l'angolo al centro della circonferenza.

Essa è data dal rapporto fra l'angolo al centro spazzato dalla circonferenza ed il tempo impiegato a spazzarlo:

$$\omega = \frac{\Delta\phi}{\Delta T}$$

Per il moto circolare uniforme di periodo T :

$$\omega = \frac{2\pi}{T} \quad (6.2)$$

DIMOSTRAZIONE

Se all'istante iniziale la particella si trova nel punto P_0 (Fig. 18), dopo un periodo T la particella avrà spazzato l'intero angolo giro $\phi = 2\pi$ ritornando nel punto P_0 .

La velocità angolare sarà:

$$\omega = \frac{\text{angolo spazzato}}{\text{tempo impiegato}} = \frac{2\pi}{T} = 2\pi\nu$$

Dalla (6.1) e (6.2) segue che:

$$v = \omega R$$

d) Accelerazione centripeta

Nel moto circolare uniforme si ha $\vec{a} = \frac{v^2}{R} \vec{u}_N$.

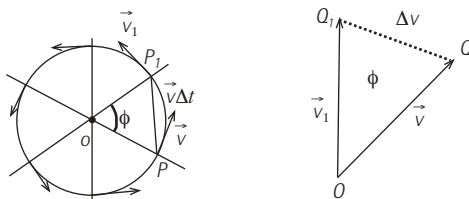
DIMOSTRAZIONE

È già stato notato che l'accelerazione di un corpo è scindibile in due componenti:

$$\vec{a}_T \quad \text{e} \quad \vec{a}_N$$

Nel moto circolare uniforme la velocità in modulo si mantiene costante e quindi $\vec{a} = \vec{a}_N$. L'accelerazione sarà quindi normale alla tangente alla curva.

Per ricavare il modulo di $\vec{a}_N$ suddividiamo la circonferenza in 6 spicchi e tracciamo i rispettivi vettori velocità.



Figg. 19.a e 19.b

Supponiamo che all'istante $t = 0$ la particella si trovi in P e all'istante $t = t_1$ la particella si trovi in P_1 . In figura 19.b sono ridisegnati i vettori $\vec{v}$ e $\vec{v}_1$.

Nella figura 19.a i triangoli POP_1 e QQ_1 sono simili essendo isosceli e avendo lo stesso angolo al vertice (i due angoli al vertice sono uguali in quanto $\vec{v}$ e $\vec{v}_1$ sono perpendicolari a $\overline{OP}$ e $\overline{OP_1}$).

Per $\Delta t \rightarrow 0$ possiamo considerare gli archi PP_1 e vv_1 coincidenti con le corde $\overline{PP_1}$ e $\overline{vv_1}$.

Quindi:

$$\frac{\overline{PP_1}}{\overline{OP}} = \frac{\overline{QQ_1}}{\overline{OQ}}$$

$$\frac{dv}{v} = \frac{v dt}{R}$$

$$\frac{dv}{dt} = \frac{v^2}{R}$$

$$a = \frac{dv}{dt} = \frac{v^2}{R}$$

Vedi quanto trattato in precedenza.

7 Il moto circolare uniformemente accelerato

Il moto circolare è detto **uniformemente accelerato** quando la particella percorre la circonferenza con accelerazione costante.

In questo caso né la velocità tangenziale né quella angolare della particella sono costanti.

È quindi necessario introdurre una nuova accelerazione α , detta **accelerazione angolare**, definita come la variazione della velocità angolare nel tempo:

$$\alpha = \frac{\Delta\omega}{\Delta t}$$

In tabella è stato schematizzato un parallelo fra il moto «tangenziale» ed il moto «angolare» nel moto circolare uniformemente accelerato.

	Moto tangenziale	Moto angolare
Spostamento	x	ϕ
Velocità	$v = \frac{dx}{dt}$	$\omega = \frac{d\phi}{dt}$
Accelerazione	$a = \frac{dv}{dt}$	$\alpha = \frac{d\omega}{dt}$
Moto uniforme	$v = \text{cost}$ $x = x_0 + vt$	$\omega = \text{cost}$ $\phi = \phi_0 + \omega t$
Moto uniformemente accelerato	$a = \text{cost}$ $x = x_0 + v_0 t + \frac{1}{2}at^2$ $v = v_0 + at$	$\alpha = \text{cost}$ $\phi = \phi_0 + \omega_0 t + \frac{1}{2}\alpha t^2$ $\omega = \omega_0 + \alpha t$

Tab. 4

8 La composizione degli spostamenti

Lo spostamento assoluto è dato dalla somma vettoriale dello spostamento relativo e dello spostamento di trascinamento:

$$\vec{S} = \vec{S}_R + \vec{S}_T$$

DIMOSTRAZIONE

Supponiamo che all'istante iniziale ($t = 0$) i due sistemi di riferimento coincidano (Fig. 20) e che il viaggiatore inizi a camminare lungo il corridoio del treno.

Supponiamo che, dopo un intervallo di tempo Δt , il viaggiatore abbia percorso una distanza $S_{\text{Viaggiatore}}$ e il treno, una distanza S_{Treno} .

Per il casellante a terra, il viaggiatore avrebbe percorso la distanza:

$$S_{\text{Casellante}} = S_{\text{Viaggiatore}} + S_{\text{Treno}} \quad (8.1)$$

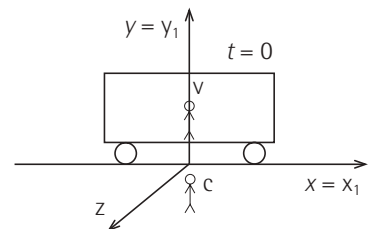


Fig. 20

Se il viaggiatore si fosse mosso in direzione opposta a quella del treno per il casellante egli avrebbe percorso la distanza:

$$S_{\text{Casellante}} = -S_{\text{Viaggiatore}} + S_{\text{Treno}} \quad (8.2)$$

In termini vettoriali la (8.1) e la (8.2) diventano:

$$\vec{S}_{\text{Casellante}} = \vec{S}_{\text{Viaggiatore}} + \vec{S}_{\text{Treno}} \quad (8.3)$$

La (8.3) continua a valere se il viaggiatore si muove in una qualsiasi direzione e non necessariamente nella direzione del treno.

È consuetudine chiamare:

$S_{\text{Viaggiatore}}$: spostamento relativo (S_r) ed è lo spostamento del corpo rispetto al S.R. in movimento;

S_{Treno} : spostamento di trascinamento (S_t) ed è lo spostamento del S.R. in moto rispetto al S.R. fermo;

$S_{\text{Casellante}}$: spostamento assoluto (S) ed è lo spostamento del S.R. in moto rispetto al S.R. fermo.

E la (8.3) diventa:

$$\vec{S} = \vec{S}_r + \vec{S}_t \quad (8.4)$$

Indicando con $\vec{v}_0$ la velocità costante del treno $\vec{S}_t = \vec{v}_0 \Delta t = \vec{v}_0 t$ e introducendo i vettori posizione $\vec{r}$, la (8.4) diventa:

$$\vec{r} = \vec{r}_r + \vec{v}_0 t \quad (8.5)$$

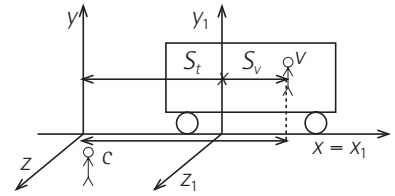


Fig. 21

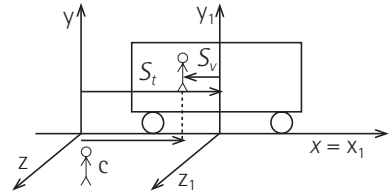


Fig. 22

9 La composizione delle velocità

La **velocità assoluta** è data dalla somma vettoriale della velocità relativa e della velocità di trascinamento:

$$\vec{v} = \vec{v}_r + \vec{v}_t$$

DIMOSTRAZIONE

Dividendo la (8.5) per l'intervallo di tempo Δt in cui sono avvenuti tali spostamenti si ottiene:

$$\frac{\vec{r}}{\Delta t} = \frac{\vec{r}_r}{\Delta t} + \frac{\vec{r}_t}{\Delta t} \quad (9.1)$$

e, al limite per $\Delta t \rightarrow 0$, la (9.1) diventa:

$$\vec{v} = \vec{v}_r + \vec{v}_t$$

10 La composizione delle accelerazioni

Un punto materiale in moto ha uguale accelerazione rispetto a due sistemi di riferimento in moto rettilineo uniforme l'uno rispetto all'altro:

$$\vec{a}_A = \vec{a}_R$$

DIMOSTRAZIONE

Supponiamo che all'istante iniziale ($t = 0$) i due sistemi di riferimento coincidano e che il viaggiatore inizi a correre lungo il corridoio del treno.

Supponiamo che all'istante $t = t_1$ il viaggiatore abbia raggiunto una velocità v_1 e all'istante $t = t_2$ una velocità v_2 . Il treno si muove con velocità costante v_0 . All'istante t_1 la velocità assoluta $\vec{v}_{A1}$ del viaggiatore (rispetto al casellante fermo ai binari) è $\vec{v}_{A1} = \vec{v}_0 + \vec{v}_1$ e all'istante t_2 è $\vec{v}_{A2} = \vec{v}_0 + \vec{v}_2$. Sottraendo membro a membro:

$$\vec{v}_{A2} - \vec{v}_{A1} = \vec{v}_2 - \vec{v}_1$$

Dividendo per l'intervallo di tempo $\Delta t = t_2 - t_1$ si ha:

$$\frac{\vec{v}_{A2} - \vec{v}_{A1}}{\Delta t} = \frac{\vec{v}_2 - \vec{v}_1}{\Delta t}$$

e per $\Delta t \rightarrow 0$: $\vec{a}_A = \vec{a}_R$

1 La composizione degli spostamenti

Lo spostamento assoluto è dato dalla somma vettoriale dello spostamento relativo e dello spostamento di trascinamento:

$$\vec{S} = \vec{S}_R + \vec{S}_T$$

DIMOSTRAZIONE

Supponiamo che all'istante iniziale ($t = 0$) i due sistemi di riferimento (S.R.) coincidano (Fig. 1) e che il viaggiatore inizi a camminare lungo il corridoio del treno.

Supponiamo che, dopo un intervallo di tempo Δt , il viaggiatore abbia percorso una distanza $S_{\text{Viaggiatore}}$ e il treno una distanza S_{Treno} .

Per il casellante a terra, il viaggiatore avrebbe percorso la distanza:

$$S_{\text{Casellante}} = S_{\text{Viaggiatore}} + S_{\text{Treno}} \quad (1.1)$$

Se il viaggiatore si fosse mosso in direzione opposta a quella del treno, per il casellante egli avrebbe percorso la distanza:

$$S_{\text{Casellante}} = -S_{\text{Viaggiatore}} + S_{\text{Treno}} \quad (1.2)$$

In termini vettoriali la (1.1) e la (1.2) diventano:

$$\vec{S}_{\text{Casellante}} = \vec{S}_{\text{Viaggiatore}} + \vec{S}_{\text{Treno}} \quad (1.3)$$

La (1.3) continua a valere se il viaggiatore si muove in una qualsiasi direzione e non necessariamente nella direzione del treno.

È consuetudine chiamare:

- $S_{\text{Viaggiatore}}$: spostamento relativo (S_R), ovvero lo spostamento del corpo rispetto al S.R. in movimento;
- S_{Treno} : spostamento di trascinamento (S_T), ovvero lo spostamento del S.R. in moto rispetto al S.R. fermo;
- $S_{\text{Casellante}}$: spostamento assoluto (S), ovvero lo spostamento del S.R. in moto rispetto al S.R. fermo.

E la (1.3) diventa:

$$\vec{S} = \vec{S}_R + \vec{S}_T \quad (1.4)$$

Indicando con $\vec{v}_0$ la velocità costante del treno, $\vec{S}_T = \vec{v}_0 \Delta t = \vec{v}_0 t$, e introducendo i vettori posizione $\vec{r}$, la (1.4) diventa:

$$\vec{r} = \vec{r}_R + \vec{v}_0 t \quad (1.5)$$

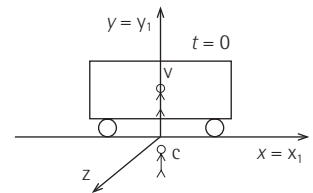


Fig. 1

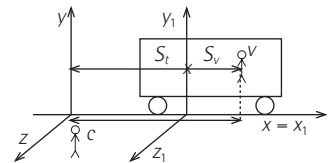


Fig. 2

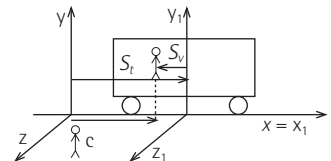


Fig. 3

2 La composizione delle velocità

La **velocità assoluta** è data dalla somma vettoriale della velocità relativa e della velocità di trascinamento:

$$\vec{v} = \vec{v}_R + \vec{v}_T \quad (2.1)$$

DIMOSTRAZIONE

Dividendo la (1.5) per l'intervallo di tempo Δt in cui sono avvenuti tali spostamenti si ottiene:

$$\frac{\vec{r}}{\Delta t} = \frac{\vec{r}_R}{\Delta t} + \frac{\vec{r}_r}{\Delta t} \quad (2.2)$$

e, al limite per $\Delta t \rightarrow 0$, la (2.2) diventa:

$$\vec{v} = \vec{v}_R + \vec{v}_r$$

3 La composizione delle accelerazioni

Un punto materiale in moto ha uguale accelerazione rispetto a due sistemi di riferimento in moto rettilineo uniforme l'uno rispetto all'altro:

$$\vec{a}_A = \vec{a}_R \quad (3.1)$$

DIMOSTRAZIONE

Supponiamo che all'istante iniziale ($t = 0$) i due sistemi di riferimento coincidano e che il viaggiatore inizi a correre lungo il corridoio del treno.

Supponiamo che all'istante $t = t_1$ il viaggiatore abbia raggiunto una velocità v_1 e all'istante $t = t_2$ una velocità v_2 . Il treno si muove con velocità costante v_0 . All'istante t_1 la velocità assoluta $\vec{v}_{A1}$ del viaggiatore (rispetto al casellante fermo ai binari) è $\vec{v}_{A1} = \vec{v}_0 + \vec{v}_1$ e all'istante t_2 è $\vec{v}_{A2} = \vec{v}_0 + \vec{v}_2$. Sottraendo membro a membro:

$$\vec{v}_{A2} - \vec{v}_{A1} = \vec{v}_2 - \vec{v}_1$$

Dividendo per l'intervallo di tempo $\Delta t = t_2 - t_1$ si ha:

$$\frac{\vec{v}_{A2} - \vec{v}_{A1}}{\Delta t} = \frac{\vec{v}_2 - \vec{v}_1}{\Delta t}$$

E per $\Delta t \rightarrow 0$:

$$\vec{a}_A = \vec{a}_R$$

4 Le trasformazioni di Galileo

Nei paragrafi precedenti abbiamo ricavato le equazioni di composizione degli spostamenti (1.5), delle velocità (2.1) e delle accelerazioni (3.1).

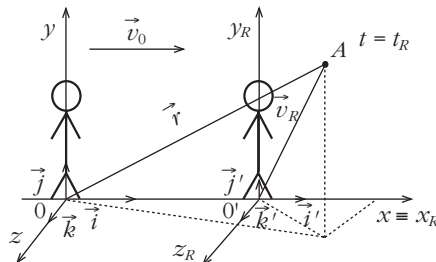


Fig. 4

Esse, scomposte lungo gli assi cartesiani, diventano:

$$\begin{aligned}x &= x_R + v_0 t \\ y &= y_R\end{aligned}\quad (4.1)$$

$$\begin{aligned}z &= z_R \\ v_x &= v_{Rx} + v_0 \\ v_y &= v_{Ry}\end{aligned}\quad (4.2)$$

$$\begin{aligned}v_z &= v_{Rz} \\ a_x &= a_{Rx} \\ a_y &= a_{Ry} \\ a_z &= a_{Rz}\end{aligned}\quad (4.3)$$

A queste equazioni se ne deve aggiungere una **fondamentale**:

$$t = t_R$$

e cioè i due osservatori possono misurare il tempo con lo stesso orologio.

In altre parole, il **tempo** è un **invariante** per sistemi di riferimento in moto rettilineo uniforme.

L'insieme delle equazioni (4.1), (4.2) e (4.3) costituiscono una **trasformazione galileiana**, dal nome del celebre scienziato Galileo Galilei che fu il primo a definire con chiarezza il concetto di **moto relativo**.

a) Il problema della velocità della luce

La velocità del suono **dipende** dalla velocità di spostamento dell'osservatore.

Due osservatori in moto rettilineo uniforme l'uno rispetto all'altro misureranno due diverse velocità del suono. Questo perché il suono si propaga attraverso le vibrazioni dell'aria.

Nel 1881 il fisico americano Albert Michelson iniziò una serie di esperimenti per misurare la velocità della luce in diverse direzioni e nel 1887, con l'aiuto del collega Edward Morley, scoprì che la velocità della luce era la stessa in tutte le direzioni.

Ciò era in disaccordo con le trasformate di Galileo, le quali indicano che **nessun** corpo può avere la stessa velocità rispetto a due sistemi di riferimento in moto rettilineo uniforme.

Questo enigma venne risolto da Einstein nel 1905 quando postulò il suo **principio della relatività**:

«tutte le leggi della Natura sono invarianti per tutti gli osservatori in moto relativo traslatorio uniforme».

Einstein stabilì, dunque, che la velocità della luce è un invariante per sistemi di riferimento in moto rettilineo uniforme.

Indicando con c la velocità della luce si ha $c_R = c$.

Questa condizione è necessaria quando si applica il principio della relatività alle leggi dell'elettromagnetismo.

Poiché la velocità è un rapporto $\frac{\text{spazio}}{\text{tempo}}$, la velocità di un oggetto per due sistemi di riferimento in moto rettilineo uniforme l'uno rispetto all'altro è data da:

$$v = \frac{\Delta x}{\Delta t} \quad \text{e} \quad v_R = \frac{\Delta x_R}{\Delta t_R}$$

Nel caso di un oggetto che viaggia alla velocità della luce si ha $c_R = c$ e quindi:

$$\frac{\Delta x}{\Delta t} = \frac{\Delta x_R}{\Delta t_R} \quad (4.4)$$

Ma, essendo $\Delta x \neq \Delta x_R$, per mantenere il rapporto costante dovrà anche essere $\Delta t \neq \Delta t_R$.

In altre parole, per velocità prossime alla velocità della luce il tempo non è più un invariante, ossia $t \neq t_R$.

5 Le trasformazioni di Lorentz

È necessario, quindi, rivedere le trasformazioni di Galileo per i corpi che viaggiano con velocità prossime alla velocità della luce.

Supponiamo che due osservatori, 0 e 0_R , si muovano su due sistemi di riferimento in moto rettilineo uniforme con velocità v_0 e supponiamo che gli osservatori sincronizzino i loro orologi all'istante $t = 0$, in modo che $t = t_R = 0$.

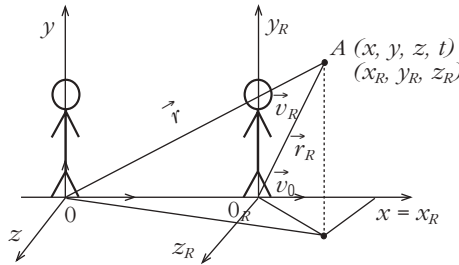


Fig. 5

Supponiamo che all'istante $t = 0$ venga emesso un lampo di luce.

L'osservatore in 0 rileverà che la luce ha raggiunto il punto A dopo un intervallo di tempo t .

L'osservatore in 0_R rileverà che la luce ha raggiunto il punto A dopo un intervallo di tempo t_R .

Essendo la velocità della luce un invariante, per l'osservatore in 0 sarà $r = ct$, mentre per l'osservatore in 0_R sarà $r_R = ct_R$.

Essendo $r^2 = x^2 + y^2 + z^2$ e $r_R^2 = x_R^2 + y_R^2 + z_R^2$, per l'osservatore in 0:

$$c^2 t^2 = x^2 + y^2 + z^2 \quad (5.1)$$

per l'osservatore in 0_R :

$$c^2 t_R^2 = x_R^2 + y_R^2 + z_R^2 \quad (5.2)$$

A questo punto si devono trovare delle nuove trasformazioni che leghino le misure dei due osservatori.

Siccome il sistema (x_R, y_R, z_R) si muove di moto rettilineo uniforme lungo l'asse x , in ogni istante si avrà:

$$y = y_R \quad \text{e} \quad z = z_R$$

Inoltre, poiché il sistema (x_R, y_R, z_R) si muove con velocità v , deve essere $x = vt$ per $x_R = 0$.

Quindi possiamo supporre una relazione del tipo $x_R = k(x - vt)$ per la (4.4) e possiamo pensare ad una relazione analoga per i tempi: $t_R = a(t - bx)$.

Pertanto, le nuove trasformazioni sono:

$$x_R = k(x - vt) \quad (5.3.a)$$

$$y_R = y \quad (5.3.b)$$

$$z_R = z \quad (5.3.c)$$

$$t_R = a(t - bx) \quad (5.3.d)$$

che diventano le trasformate di Galileo per:

$$k = 1; \quad a = 1; \quad b = 0.$$

Sostituendo le (5.3) nella (5.2) otteniamo:

$$c^2[a(t - bx)]^2 = [k(x - vt)]^2 + y^2 + z^2$$

da cui:

$$x^2(k^2 - b^2a^2c^2) - 2(k^2v - ba^2c^2)xt + y^2 + z^2 = \left(a^2 - \frac{k^2v^2}{c^2}\right)c^2t^2$$

Questa equazione deve essere identica alla (5.1), dunque:

$$k^2 - b^2a^2c^2 = 1$$

$$k^2v - ba^2c^2 = 0$$

$$a^2 - \frac{k^2v^2}{c^2} = 1$$

da cui segue che:

$$k = a = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$b = \frac{v}{c^2}$$

Sostituendo nelle (5.3):

$$x_R = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (5.4.a)$$

$$y_R = y \quad (5.4.b)$$

$$z_R = z \quad (5.4.c)$$

$$t_R = \frac{t - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (5.4.d)$$

Le (5.4) costituiscono le **trasformazioni di Lorentz**, dal nome del fisico olandese Henry Lorentz che le formulò nel 1890.

Da notare che per $v \ll c$ (cioè nei casi normali) le (5.4) diventano:

$$x_R = x - vt$$

$$y_R = y$$

$$z_R = z$$

$$t_R = t$$

che sono le **trasformazioni di Galileo**.

In altre parole, le trasformazioni di Galileo sono valide quando si studiano fenomeni che riguardano corpi che viaggiano a velocità molto inferiori a quella della luce.

Quando si studiano fenomeni con particelle che viaggiano a velocità prossime a quelle della luce le trasformazioni di Galileo non sono più valide e si devono utilizzare le trasformazioni di Lorentz.

6 Le conseguenze delle trasformazioni di Lorentz

a) Contrazione delle lunghezze

La lunghezza di un corpo misurata da un osservatore in movimento è inferiore a quella misurata da un osservatore fermo:

$$L_{\text{moto}} < L_{\text{quiete}}$$

DIMOSTRAZIONE

Supponiamo che i due osservatori in O e O' decidano di misurare la lunghezza di una sbarra che si muove in modo solidale con l'osservatore O' . L'osservatore in O' vede la sbarra ferma, mentre l'osservatore O vede la sbarra in movimento.

La lunghezza della sbarra sarà data dalla differenza delle coordinate dei due punti estremi.

Per l'osservatore O la lunghezza della sbarra è data da:

$$L = x_B - x_A$$

Mentre per l'osservatore O' è:

$$L' = x'_B - x'_A$$

Per le trasformazioni di Lorentz:

$$x'_A = \frac{x_A - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \text{e} \quad x'_B = \frac{x_B - vt}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$L' = x'_B - x'_A = \frac{x_B - x_A}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{L}{\sqrt{1 - \frac{v^2}{c^2}}}$$

da cui:

$$L = L' \sqrt{1 - \frac{v^2}{c^2}}$$

Essendo $\sqrt{1 - \frac{v^2}{c^2}} < 1$, si ha $L < L'$, cioè:

$$L_{\text{moto}} < L_{\text{quiete}}$$

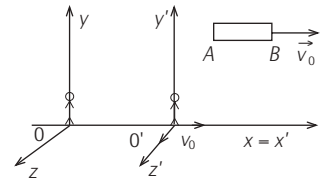


Fig. 6

b) Dilatazione dei tempi

L'intervallo di tempo misurato da un osservatore in moto è maggiore di quello misurato da un osservatore in quiete:

$$\Delta t_{\text{moto}} > \Delta t_{\text{quiete}}$$

DIMOSTRAZIONE

Supponiamo che i due osservatori in O e O' decidano di misurare l'intervallo di tempo che intercorre fra due eventi.

Un evento è un avvenimento che si verifica in un determinato punto e in un determinato istante.

Supponiamo che due eventi si verifichino nello stesso punto quando l'orologio dell'osservatore in movimento segna gli istanti t'_A e t'_B .

Per l'osservatore fermo, l'intervallo di tempo è dato da:

$$\Delta t = t_2 - t_1 = \frac{t'_2 - t'_1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{\Delta t'}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Essendo $\sqrt{1 - \frac{v^2}{c^2}} < 1$, si ha $\Delta t > \Delta t'$, cioè:

$$\Delta t_{\text{moto}} > \Delta t_{\text{quiete}}$$

6

La dinamica

1 Il primo principio della dinamica (principio d'inerzia)

*Il **primo principio della dinamica** (o principio d'inerzia o di Galileo), afferma che un corpo mantiene il suo stato di quiete o di moto uniforme se non interviene una forza esterna a modificarlo.*

Tale principio è anche detto **principio di inerzia** in quanto l'inerzia è proprio la tendenza di un corpo a mantenere lo stato di quiete o di moto in cui si trova. In assenza di forze, dunque, un corpo si muove di moto rettilineo uniforme. Il corpo a cui fa riferimento il principio d'inerzia è un corpo libero non soggetto a forza. In realtà tutti i corpi interagiscono con i corpi circostanti e quindi un tale corpo non esiste.

2 Il secondo principio della dinamica (legge di Newton)

*Il **secondo principio della dinamica** (o legge di Newton), afferma che l'accelerazione acquisita da un oggetto in seguito all'azione di una forza su di esso è direttamente proporzionale alla forza applicata e inversamente proporzionale alla massa dell'oggetto cioè:*

$$\vec{F} = m\vec{a}.$$

a) Massa inerziale

Dal principio di Newton deriva quindi la definizione di **massa inerziale**:

$$m = \frac{\vec{F}}{\vec{a}}$$

La massa inerziale rappresenta il rapporto, costante per un determinato corpo, fra una qualsiasi forza che agisca su di esso e l'accelerazione ad esso impressa. Essa indica la tendenza di un corpo a mantenere lo stato di moto in cui si trova.

La massa inerziale:

- è una grandezza scalare;
- non è legata alle caratteristiche chimiche e morfologiche di un oggetto;
- ha per unità di misura il chilogrammo massa.

Nel Sistema Internazionale la massa è una grandezza fondamentale: il campione unitario di platino iridio è conservato a Sevres ed è stato chiamato chilogrammo-massa (*Kg*).

Dalla definizione dell'unità di misura della massa deriva la definizione dell'unità di misura della forza.

L'unità di misura della forza è il newton (*N*).

1*N* è la forza necessaria per imprimere ad un corpo di massa pari a 1*Kg* un'accelerazione di 1*m/s*².

b) Equazione di Newton: $\vec{F} = m\vec{a}$

Da questa equazione possiamo notare che la forza è un vettore che ha la stessa direzione e lo stesso verso dell'accelerazione.

Possiamo scomporre questa equazione lungo gli assi cartesiani:

$$F_x = ma_x; \quad F_y = ma_y; \quad F_z = ma_z$$

da cui possiamo notare che ogni componente dell'accelerazione è determinata **solo** dalla componente della forza lungo quell'asse.

c) Peso di un corpo

Se si solleva una pietra e la si lascia andare, essa cade a terra.

La Terra attrae tutti i corpi che si trovano in prossimità della sua superficie con una forza:

$$\vec{P} = m\vec{g} \quad (2.1)$$

dove:

g è l'accelerazione di gravità ed è pari a $9,8 \text{ m/s}^2$.

Tale forza è il peso stesso del corpo.

d) Chilogrammo massa e chilogrammo peso

Bisogna fare attenzione a non confondere la massa di un corpo con il suo peso. Questo perché siamo abituati a comprare un chilogrammo di mele, a pesarci in chilogrammi etc.

In realtà dobbiamo ricordarci che:

- la massa è una proprietà intrinseca del corpo, mentre il peso dipende dal luogo in cui un oggetto si trova;
- la massa è uno scalare mentre il peso è un vettore;
- l'unità di misura della massa è il *chilogrammo* (Kg) o chilogrammo-massa, mentre per il peso si può usare:
 - il *newton* (N), unità di misura della forza nel Sistema Internazionale (S.I.);
 - il *chilogrammo-peso* (Kgp) o *chilogrammo-forza* (Kgf), utilizzato nel sistema pratico o degli ingegneri.

Il chilogrammo-forza è la forza con cui la massa di $1Kg$ viene attratta dalla Terra.

Il Kgf ed il N sono legati dalla seguente relazione:

$$1Kgf = 9,8N$$

- la massa di un corpo non dipende dall'ubicazione del corpo mentre il suo peso dipende dal luogo in cui si trova. Sulla Luna l'accelerazione di gravità è un sesto di quella terrestre, quindi un corpo di massa pari a $1Kg$ sulla Luna peserebbe $1,62N$ e sulla Terra $9,8N$.

3 Il terzo principio della dinamica

Quando due corpi interagiscono le forze esercitate da un corpo sull'altro sono uguali in modulo e direzione ma hanno verso opposto.

Se consideriamo due corpi P_1 e P_2 e indichiamo con $\vec{F}_{12}$ la forza che P_1 esercita su P_2 e con $\vec{F}_{21}$ la forza che P_2 esercita su P_1 si ha che:

$$\vec{F}_{12} = -\vec{F}_{21}$$

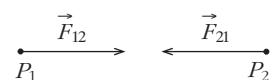


Fig. 1

In parole più semplici, il terzo principio della dinamica afferma che ad ogni azione corrisponde una reazione uguale e contraria e per questo è anche chiamato **principio di azione-reazione**.

Per esempio, se si appoggia un libro su un tavolo, il libro esercita sul tavolo la forza peso e il tavolo esercita sul libro una forza uguale e contraria (Fig. 2).

Le due forze si equilibrano e il libro non cade.

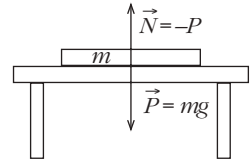


Fig. 2

4 Il moto di un corpo su un piano inclinato

Studiamo ora il **moto di un corpo** di massa m che scivola su un **piano inclinato** di lunghezza l e altezza h nell'ipotesi che l'attrito sia trascurabile.

Le forze che agiscono sul corpo sono la sua forza peso $\vec{P}$ e la reazione vincolare del piano $\vec{R}$.

Per la seconda legge della dinamica dovrà essere:

$$\vec{P} + \vec{R} = m\vec{a}$$

Disegniamo un sistema di assi cartesiani con l'asse x parallelo al piano e scomponiamo le forze agenti sul corpo lungo tali assi (Fig. 4).

a) Asse x

Lungo tale asse agisce la componente della forza peso $\vec{P}$ parallela al piano. Questa forza fa scorrere il corpo lungo il piano inclinato con accelerazione a . Per la seconda legge di Newton:

$$\vec{a} = \frac{\vec{P}_1}{m} \quad (4.1)$$

b) Asse y

Lungo tale asse agiscono la componente della forza peso $\vec{P}_2$ normale al piano inclinato e la reazione vincolare $\vec{R}$ del piano.

Tali forze, per il terzo principio della dinamica, si annullano a vicenda:

$$\vec{P}_2 + \vec{R} = 0$$

A questo punto ricaviamo $\vec{P}_1$.

In figura 4 consideriamo i triangoli MNP e ABC . Essi sono simili, quindi:

$$\frac{\overline{MP}}{\overline{BC}} = \frac{\overline{MN}}{\overline{AB}}$$

da cui segue che $\frac{P_1}{P} = \frac{h}{l}$, ossia $\frac{P_1}{mg} = \frac{h}{l}$ e quindi:

$$P_1 = mg \frac{h}{l}$$

Sostituendo nella (4.1):

$$mg \frac{h}{l} = ma$$

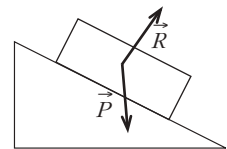


Fig. 3

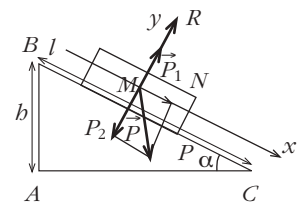


Fig. 4

da cui segue:

$$a = g \frac{h}{l}$$

In generale $P_1 = P \sin \alpha = mg \sin \alpha$. Sostituendo nella (4.1) si ha $mg \sin \alpha = ma$, da cui segue che:

$$a = g \sin \alpha$$

Da notare che l'**accelerazione** del corpo **non dipende** dalla massa ma soltanto dal rapporto fra l'altezza e la lunghezza del piano o dall'angolo di inclinazione del piano.

Nel caso della caduta libera $h = l$ ($\alpha = 90^\circ$) e quindi $a = g$.

Questo dimostra che, in assenza dell'attrito dell'aria, una piuma e un sasso fatti cadere dalla stessa altezza o fatti scivolare su un piano inclinato, arrivano a terra contemporaneamente.

5 L'attrito

L'**attrito** è una forza sempre diretta in direzione opposta al moto che si manifesta ogni qualvolta due corpi sono a contatto.

La forza di attrito ha le seguenti proprietà:

- è proporzionale alla reazione vincolare del piano su cui l'oggetto è appoggiato, ossia $f_a = \mu N$ dove μ è il coefficiente d'attrito;
- non dipende dall'area delle superfici a contatto;
- una volta che il corpo è stato messo in moto, non dipende dalla velocità del corpo.

Esistono diversi tipi di attrito:

- **attrito statico** $f_s = \mu_s P$: è la forza resistente che contrasta l'iniziale movimento del corpo;
- **attrito dinamico** $f_d = \mu_d P$: è la forza resistente che contrasta il moto del corpo una volta messo in movimento;
- **attrito radente** $f_r = \mu_r P$: è l'attrito che si crea fra due oggetti che strisciano l'uno sull'altro;
- **attrito volvente** $f_v = \mu_v P$: è l'attrito che si crea quando un oggetto rotola.

A parità di condizioni $\mu_s > \mu_d$ e quindi la forza di attrito statico è maggiore della forza di attrito dinamico.

Consideriamo un oggetto appoggiato sopra una superficie ed applichiamo una forza F :

- $F_s = f_s P$ è la forza minima necessaria per mettere in moto l'oggetto;
- $F_d = f_d P$ è la forza minima necessaria per mantenere costante la velocità del corpo.

6 Forza centripeta e forza centrifuga

Consideriamo un corpo che si muove su una traiettoria circolare o comunque curvilinea. Il suo vettore velocità è sempre tangente alla curva mentre il vettore accelerazione è sempre diretto verso la concavità della curva.

Per studiarne il moto possiamo introdurre un sistema di assi cartesiani avente, in ogni istante, un asse tangente e uno perpendicolare alla traiettoria e scomporre i vettori velocità e accelerazione secondo tali assi:

$$\vec{v} = \vec{v}_T; \quad \vec{a} = \vec{a}_T + \vec{a}_N$$

$\vec{a}_T$ è l'accelerazione dovuta alla variazione di velocità in modulo;

$\vec{a}_N = \frac{v^2}{R} \vec{u}_N$ è l'accelerazione dovuta alla variazione di velocità in direzione.

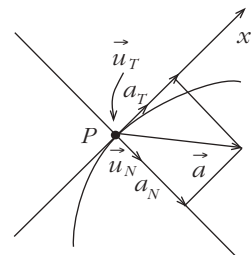


Fig. 5

Pertanto, le leggi di Newton nel sistema di riferimento introdotto diventano:

$$F_T = ma_T$$

$$F_N = ma_N = m \frac{v^2}{R}$$

F_T è la forza che provoca la variazione della velocità in modulo.

F_N è la forza che provoca la variazione della velocità in direzione ed è meglio nota come

forza centripeta.

La forza centripeta è quindi la forza che fa curvare il corpo.

Nel moto circolare uniforme si ha solo una variazione di velocità in direzione e quindi:

$$\vec{a} = \vec{a}_N = \frac{v^2}{R} \vec{u}_N$$

Di conseguenza $F_T = 0$ pertanto l'unica forza agente sul corpo è quella centripeta.

La domanda sorge spontanea: «*Che cosa fa curvare i corpi?*»

A prima vista questa domanda può sembrare stupida ma non lo è affatto nel senso che, mentre è abbastanza intuitivo associare una forza che spinge un oggetto con l'accelerazione dell'oggetto stesso, non è così immediato capire quale sia la forza che fa curvare un oggetto.

Per un sasso che sta ruotando all'estremità di una corda è la tensione della corda, per un pianeta che orbita attorno al Sole è la forza gravitazionale, per un elettrone che ruota attorno al nucleo è la forza elettrica etc.

Nell'istante in cui questa forza venisse a mancare, l'oggetto partirebbe immediatamente per la tangente alla curva (termine poco fisico ma che rende il concetto).

Per il principio di azione-reazione, per reazione alla forza centripeta si genera una forza uguale ed opposta, detta **forza centrifuga**, che tenderà ad allontanare il corpo dal centro di rotazione.

1 Proprietà generali

Dal punto di vista della cinematica possiamo definire il **moto armonico** come il moto della proiezione di un punto che percorre una circonferenza.

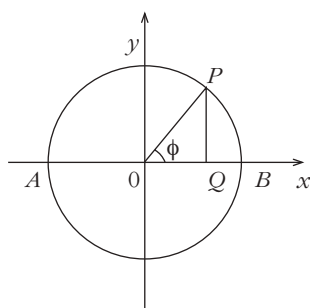


Fig. 1

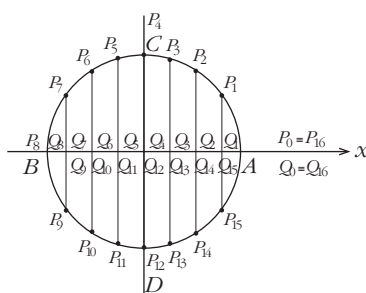


Fig. 2

In figura 1 è mostrato il punto Q, proiezione sull'asse delle x del punto P.

In figura 2 è mostrato come, mentre il punto P descrive la circonferenza, il punto Q oscilla tra i punti A e B.

Si definisce:

- **centro di oscillazione**, il centro della circonferenza;
- **ampiezza**, la massima distanza del punto Q dal centro di oscillazione;
- **oscillazione completa**, il moto da A a B e ritorno in A;
- **elongazione**, la distanza di Q dal centro di oscillazione in un generico istante t ;
- **periodo**, il più piccolo intervallo di tempo dopo il quale il moto assume le stesse caratteristiche. In figura 2 si mostra che, mentre P percorre l'intera circonferenza, il punto Q descrive un'oscillazione completa;
- **pulsazione**, la velocità angolare del moto circolare uniforme.

2 L'equazione oraria del moto armonico semplice

In figura 2 sono stati tracciati sulla circonferenza dei punti a intervalli regolari con le relative elongazioni sull'asse x.

In figura 3 sono state riportate le stesse elongazioni con i relativi istanti t .

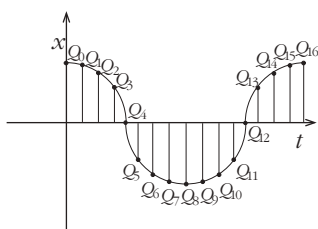


Fig. 3

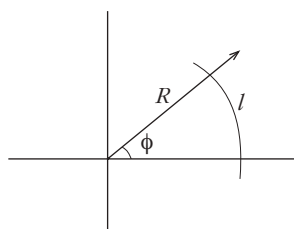


Fig. 4

Per trovare l'equazione oraria del moto dobbiamo riprendere la definizione della misura in radianti di un angolo.

Consideriamo la figura 4. Indichiamo con l la lunghezza dell'arco di circonferenza corrispondente all'angolo al centro ϕ .

La misura in radianti dell'angolo ϕ è il numero reale $t = \frac{l}{R}$ dove l è la lunghezza dell'arco della circonferenza e R il raggio della circonferenza. Tale misura non dipende dal raggio della circonferenza. Dopo un giro completo, $l = 2\pi$ e quindi la misura in radianti corrispondente è $t = \frac{2\pi R}{R} = 2\pi$ ($\varphi = 360^\circ$) mentre mezzo giro ($\varphi = 180^\circ$) corrisponde a π radianti.

Le misure degli angoli vengono fatte in radianti proprio per rendere le grandezze indipendenti dai raggi delle circonferenze considerate.

Dalla figura 1 all'istante t , l'angolo al centro ϕ sarà uguale a $\phi = \omega t$.

Consideriamo il triangolo rettangolo OPQ . All'istante t :

$$\overline{OQ} = \overline{OP} \cos \phi = R \cos \omega t$$

Quindi $x(t) = R \cos \omega t$.

Se il punto P all'istante iniziale si fosse trovato in P_0 (Fig. 5) corrispondente ad un angolo ϕ_0 , nell'istante t esso si sarebbe trovato nel punto $P(t)$ corrispondente ad un angolo al centro $\phi(t) = \omega t + \phi_0$ e in questo caso sarebbe stato:

$$x(t) = R \cos(\omega t + \phi_0) \quad (2.1)$$

che è la **legge oraria del moto armonico semplice**. Nella (2.1):

- $\omega t + \phi_0$ è la fase dell'oscillazione;
- ϕ_0 è la fase iniziale dell'oscillazione o la costante di fase.

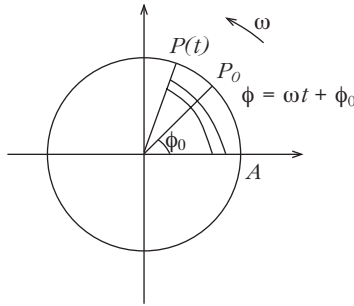


Fig. 5

In figura 6 sono mostrati i grafici relativi alle equazioni orarie:

$$x_1(t) = R \cos(\omega t - \phi_0); \quad x_2(t) = R \cos \omega t; \quad x_3(t) = R \cos(\omega t + \phi_0).$$

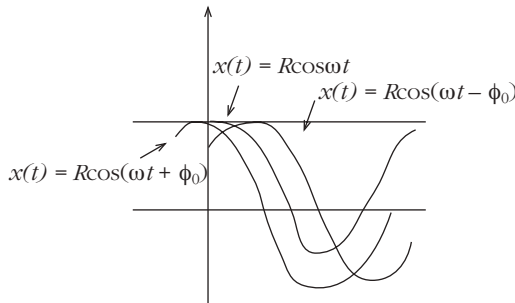


Fig. 6

a) Equazione della velocità

In figura 7 sono stati tracciati i vettori tangenti alla traiettoria della particella che si sta muovendo di moto armonico semplice.

Siccome la velocità della particella è data dalla pendenza del vettore tangente, da tale figura si nota che la velocità è:

- nulla agli estremi di oscillazione quando il punto P si trova in A e in B ;
- massima nel centro di oscillazione quando P si trova in C e in D ;
- negativa tra A e B ;
- positiva tra B e A .

Per ottenere il valore della velocità del moto armonico è sufficiente calcolare la proiezione sul diametro del vettore velocità nel moto circolare uniforme.

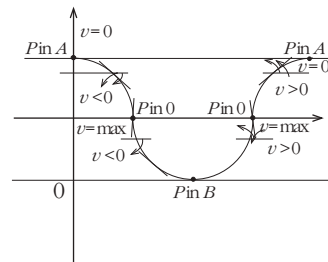


Fig. 7

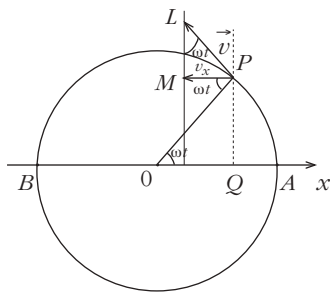


Fig. 8

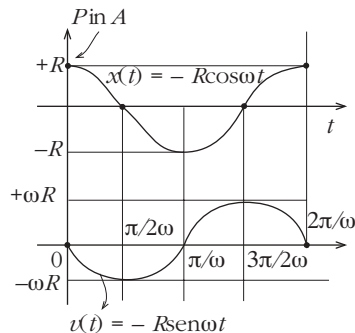


Fig. 9

Dal triangolo PLM (Fig. 8) si ha $v_x(t) = -v \sin \omega t$, ma essendo $v = \omega R$ si ha:

$$v_x(t) = -\omega R \sin \omega t$$

Essendo la velocità la derivata rispetto al tempo della posizione:

$$v_x(t) = \frac{dx(t)}{dt} = \frac{d}{dt}(R \cos \omega t) = -R \omega \sin \omega t \quad (2.2)$$

Dalla relazione $\cos\left(\alpha + \frac{\pi}{2}\right) = -\sin \alpha$ la (2.2) diventa:

$$v_x(t) = R \omega \cos\left(\omega t + \frac{\pi}{2}\right)$$

In figura 9 sono mostrati i grafici dell'equazione oraria e delle velocità del moto armonico semplice.

Come si vede, la velocità risulta in anticipo di $\frac{\pi}{2}$ rispetto alla posizione.

b) Equazione dell'accelerazione

In figura 10 sono stati tracciati i vettori tangenti alla velocità della particella che si sta muovendo di moto armonico semplice.

3 Il moto di un corpo soggetto a una forza elastica: la legge di Hooke

Supponiamo di avere un corpo appoggiato su un piano privo di attrito e fissato ad una molla di costante elastica k . In questa situazione le uniche forze che agiscono sul corpo sono la forza peso e la reazione vincolare del piano, le quali si equilibrano a vicenda (Fig. 13.a).

Se spostiamo il corpo dalla posizione di equilibrio di una distanza x_0 , sul corpo agisce una forza di richiamo F .

In laboratorio si può dimostrare che esiste una relazione lineare tra lo spostamento x_0 della molla, la sua costante elastica k e la forza di richiamo F con cui essa cerca di tornare nella posizione di equilibrio:

$$F = -kx_0 \quad (3.1)$$

Tale legge è nota come **legge di Hooke**.

Il segno $-$ indica che tale forza è diretta sempre in verso opposto allo spostamento.

Sotto l'azione di questa forza il corpo inizierà ad oscillare dal punto A al punto A' e viceversa. Dalla (3.1) e per il secondo principio della dinamica per cui $F = ma$ segue che:

$$a = -\frac{k}{m}x \quad (3.2)$$

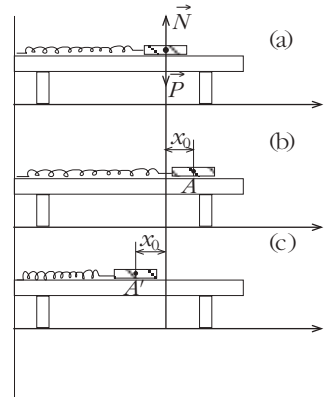
Ponendo $\omega^2 = \frac{k}{m}$, la (3.2) diventa $a = -\omega^2 x$, la quale è formalmente analoga alla (2.4) che descrive la relazione fra spostamento e accelerazione nel moto armonico semplice.

Si può perciò concludere che un punto materiale di massa m , soggetto ad una forza elastica di costante k , si muove di moto armonico con pulsazione:

$$\omega = \sqrt{\frac{k}{m}}$$

La costante elastica k è una proprietà caratteristica della molla che dipende dal materiale di cui è costituita, dalla forma e dalle dimensioni. La legge di Hooke vale solo per piccole deformazioni, cioè per deformazioni che non portano la molla a superare i cosiddetti *limiti di elasticità*, oltre i quali non vale più la semplice legge lineare tra forza e deformazione, e la molla, in genere, non recupera più le proprietà iniziali (*deformazione plastica*). Aumentando ulteriormente la forza esercitata su una molla si può giungere alla sua rottura (*limiti di rottura*).

Vista la semplicità della legge di Hooke, le molle servono a costruire i dinamometri che, a loro volta, possono essere usati come strumenti per misurare le forze. Solo se una molla obbedisce alla legge di Hooke può essere usata per misurare intensità di forza, in quanto le deformazioni sono direttamente proporzionali alla forza applicata. In particolare, noto il valore di k della molla, sarà possibile misurare la forza incognita moltiplicando il valore di k per l'allungamento x prodotto dalla forza stessa.



Figg. 13.a, 13.b e 13.c

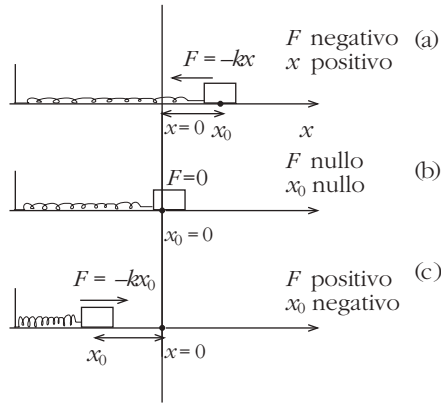
4 Il lavoro della forza elastica

Calcoliamo ora il **lavoro** fatto dalla **forza elastica**:

$$F = -kx$$

quando la molla viene deformata dalla sua posizione di equilibrio.

In figura 14 è mostrato come varia la forza di richiamo della molla in funzione dello spostamento.



Figg. 14.a, 14.b e 14.c

Essendo tale spostamento sempre di segno opposto rispetto alla forza, il lavoro, che è dato dal prodotto tra forza e spostamento, sarà sempre negativo. Per calcolare il lavoro compiuto dalla forza di richiamo, quando il corpo viene spostato dalla posizione di equilibrio ($x = 0$) in $x = x_0$, dobbiamo calcolare l'area sottesa dalla retta $F(x) = -kx$ tra l'origine degli assi e il punto x_0 .

Consideriamo il triangolo ABO in figura 15.

Esso ha base $OA = x_0$ e altezza $AB = kx_0$, da cui:

$$\text{Area}_{ABO} = \frac{1}{2} \overline{OA} \cdot \overline{BO} = \frac{1}{2} kx_0^2$$

Quindi, ricordando il segno negativo:

$$L_{el} = -\frac{1}{2} kx_0^2$$

Il lavoro tra i punti 0 e x è dato dall'integrale definito della funzione F :

$$L_{el} = \int_0^{x_0} -kx dx = -k \int_0^{x_0} x dx = -k \frac{x^2}{2} \Big|_0^{x_0} = -\frac{1}{2} kx_0^2$$

In generale, il lavoro compiuto dalla forza elastica quando un corpo viene spostato da una posizione x_1 alla posizione x_2 è dato da:

$$L_{el} = -\frac{1}{2} k(x_2^2 - x_1^2)$$

Il lavoro compiuto dalla forza applicata è uguale ed opposto al lavoro fatto dalla forza elastica in quanto la forza applicata è opposta alla forza elastica:

$$L = -L_{el} = \frac{1}{2} k(x_2^2 - x_1^2)$$

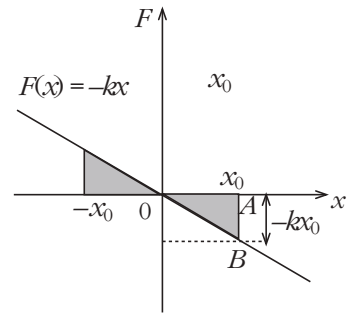


Fig. 15

5 L'energia potenziale elastica

Nel capitolo successivo, quello su *lavoro ed energia*, si suddivideranno le forze in conservative e non conservative.

Il lavoro di una forza conservativa dipende esclusivamente dalla posizione occupata dal corpo e non dal tragitto percorso. Abbiamo visto che il lavoro di una forza elastica dipende soltanto dalla posizione iniziale e finale occupata dal corpo. Quindi la forza elastica è una forza conservativa.

Avendo già calcolato il lavoro fatto per estendere una molla dalla sua posizione di equilibrio a x_0 , possiamo introdurre, così come è stato fatto per la forza peso, un' **energia potenziale elastica** che è l'energia che la molla possiede per il semplice fatto di essere stata allungata o compressa.

L'energia potenziale elastica è un'energia che la molla ha in potenza e che si trasforma in lavoro o in un'altra forma di energia soltanto quando si annulla la forza esterna che ha compresso o allungato la molla.

Infatti, così come un corpo lasciato cadere da un'altezza h raggiungerà il suolo terrestre diminuendo la sua energia potenziale e aumentando la sua energia cinetica, così una molla che viene lasciata libera dopo essere stata compressa raggiungerà la sua posizione di equilibrio diminuendo la sua energia potenziale elastica e aumentando la sua energia cinetica.

Possiamo quindi definire un'energia potenziale elastica come:

$$U_{el} = \frac{1}{2} k x_0^2$$

Per il principio di conservazione dell'energia cinetica (che è sempre valido sia per forze conservative che non conservative):

$$\Delta L = \Delta E_c = E_{cf} - E_{ci}$$

Nel caso della forza elastica abbiamo visto che:

$$\Delta L = -\frac{1}{2} k (x_f^2 - x_i^2) = U_i - U_f$$

da cui segue che:

$$E_{cf} + U_f = E_{ci} + U_i$$

E cioè anche nel caso della forza elastica si conserva l'energia meccanica data dalla somma:

$$E = E_{cf} + U_f = \frac{1}{2} k x^2 + \frac{1}{2} m v^2$$

Quando la molla è compressa oppure allungata aumenta l'energia potenziale (con x) e diminuisce l'energia cinetica, ovvero la velocità del corpo, fino al limite di massima compressione o dilatazione in cui:

$$U = U_{max} \quad \text{e} \quad E_c = 0$$

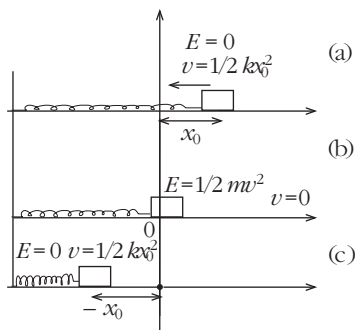
La molla compie lavoro resistente (Figg. 16.a e 16.c).

Quando la molla torna verso la sua posizione di riposo l'energia potenziale si trasforma in energia cinetica: U diminuisce e T aumenta finché nella posizione $x = 0$ si ha $U = 0$ e $E_c = E_{cmax}$.

La molla compie lavoro (Fig. 16.b).

In tale posizione la velocità è massima.

Il lavoro totale compiuto durante un'oscillazione è nullo.



Figg. 16.a, 16.b e 16.c

1 Il lavoro

Il lavoro meccanico fatto da una forza è il prodotto scalare tra la forza $\vec{F}$ e lo spostamento $\vec{s}$ del punto di applicazione della forza: $L = \vec{F} \cdot \vec{s}$.

a) Dimensioni

$$[L] = [Fs] = [ml^2t^{-2}]$$

b) Unità di misura

Nel Sistema Internazionale: 1 Joule = 1N · 1m;

Nel Sistema C.G.S: 1erg = 1dine · 1cm.

In figura 1 è mostrato un ragazzo che tira una slitta con una forza parallela allo spostamento. Il lavoro compiuto dalla forza è:

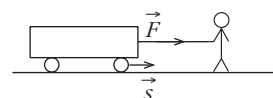


Fig. 1

$$L = \vec{F} \cdot \vec{s} = F \cdot s \quad (1.1)$$



Fig. 2

In figura 2 è mostrato lo stesso ragazzo che tira la stessa slitta ma con una forza non più parallela alla direzione dello spostamento.

In questo caso il lavoro della forza è dato da:

$$L_1 = \vec{F} \cdot \vec{s} = F \cos \vartheta \quad (1.2)$$

Confrontando la (1.1) con la (1.2) si può notare che $L_1 < L$.

In altre parole, la forza «attiva», e cioè la componente della forza nella direzione dello spostamento, è minore nel secondo caso.

c) Lavoro positivo e lavoro negativo

Dalla definizione del **lavoro** $L = F \cos \vartheta$, possiamo notare che, a seconda del valore di ϑ , il lavoro può essere positivo, nullo o negativo.

In tabella sono riassunti i vari casi:

ϑ	$\cos \vartheta$	$L = F \cos \vartheta$	Segno di L
0	1	$L = Fs$	Positivo
$0 < \vartheta < 90^\circ$	$0 < \cos \vartheta < 1$	$L = F \cos \vartheta$	Positivo
$\vartheta = 90^\circ$	0	0	Nulla
$90^\circ < \vartheta < 180^\circ$	$\cos \vartheta < 0$	$L = F \cos \vartheta$	Negativo
$\vartheta = 180^\circ$	$\cos \vartheta = -1$	$L = -Fs$	Negativo

Tab. 1

Supponiamo che un ragazzo voglia fermare con una fune un carrello in moto (Fig. 3.a).



Figg. 3.a e 3.b

In figura 3.b sono mostrate le direzioni dello spostamento del carrello e della forza applicata dal ragazzo. Esse sono discordi, $\vartheta = 180^\circ$, e il lavoro compiuto dal ragazzo è negativo.

d) Lavoro di una forza variabile

Finora ci siamo occupati di calcolare il lavoro compiuto da una forza costante.

Una forza, essendo un vettore, può variare sia in modulo che in direzione. Occupiamoci soltanto della forza variabile in modulo.

In figura 4 è mostrato il grafico di una forza variabile in modulo.

Possiamo suddividere lo spostamento totale $s = x_2 - x_1$ in tanti sottointervalli Δx in modo da considerare costante la forza in ciascuno di essi.

In ogni intervallo Δx il lavoro compiuto dalla forza F_i è $L_i = F_i \Delta x$.

Il lavoro totale sarà dato dalla sommatoria di tutti i lavori L_i :

$$L = \sum_i L_i = \sum_i F_i \Delta x \quad (1.3)$$

Per migliorare ulteriormente l'approssimazione possiamo far tendere $\Delta x \rightarrow 0$ e la (1.3) diventa:

$$L = \int_{x_1}^{x_2} F(x) dx$$

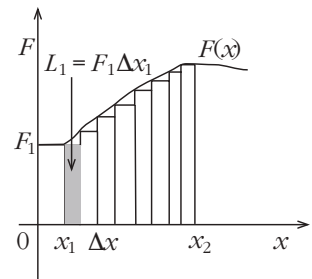


Fig. 4

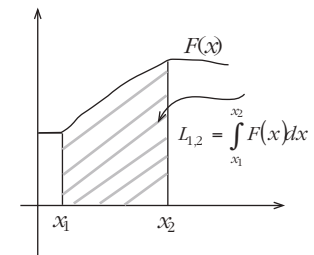


Fig. 5

ed è rappresentato dall'area sottesa dalla curva in figura 5.

2 Energia cinetica e teorema dell'energia cinetica

Vediamo ora che cosa succede quando un corpo si muove con accelerazione costante.

In tal caso la **forza risultante** agente sul corpo è **costante**.

Consideriamo un oggetto che si sta muovendo sotto l'azione di una forza $\vec{F}$ costante in direzione x .

Supponiamo che la forza e l'accelerazione abbiano la direzione di x : $\vec{F} = F_x \vec{i}$ e $a = a_x \vec{i}$.

Supponiamo che il corpo nell'istante t_0 si trovi in $x_0 = 0$ con velocità v_0 . Spostandosi con accelerazione costante il suo moto è uniformemente accelerato e le sue equazioni del moto sono:

$$v = v_0 + at \quad (2.1)$$

$$x = v_0 t + \frac{1}{2} at^2 \quad (2.2)$$

Dalla (2.1) si ha $a = \frac{v - v_0}{t}$ e sostituendo nella (2.2):

$$x = v_0 t + \frac{1}{2} \frac{v - v_0}{t} t^2 = \frac{v + v_0}{2} t$$

Il lavoro compiuto dalla forza F è dato da:

$$L = F \cdot s = ma \cdot x = m \left(\frac{v - v_0}{t} \right) \left(\frac{v + v_0}{2} t \right) = \frac{1}{2} mv^2 - \frac{1}{2} mv_0^2 \quad (2.3)$$

$E_c = \frac{1}{2} mv^2$ è l'energia cinetica del corpo in movimento.

Quindi dalla (2.3) si può concludere che:

$$L = E_c - E_{c_0} = \Delta E_c \quad (2.4)$$

E cioè il lavoro compiuto su un corpo da una forza è pari alla variazione della sua energia cinetica. La (2.4) è anche conosciuta come il **teorema dell'energia cinetica**.

OSSERVAZIONI

- Se il corpo si muove con velocità costante $\Delta E_c = 0$ e quindi il lavoro fatto dalla forza risultante è nullo.
- Soltanto quando la forza ha una componente nella direzione dello spostamento si ha un lavoro. Nel caso del moto circolare uniforme la forza centripeta è sempre perpendicolare alla direzione dello spostamento e quindi non compie alcun lavoro.
- Se l'energia cinetica del corpo diminuisce, il lavoro compiuto dalla forza è negativo; se, invece, l'energia cinetica del corpo aumenta, il lavoro compiuto dalla forza è positivo.
- Le unità di misura del lavoro e dell'energia sono uguali. Anche l'energia si misura in joule o in dine.

3 La potenza

La **potenza (media)** indica la rapidità con cui viene fatto un lavoro. Essa è data dal rapporto fra il lavoro compiuto e il tempo impiegato a compierlo.

Nel Sistema Internazionale l'unità di potenza è il watt $[W]$. La potenza unitaria di $1W$ corrisponde al lavoro di $1J$ svolto nell'intervallo di tempo di $1s$, ossia $1W = \frac{1J}{1s}$.

Un'altra unità di potenza usata è il cavallo-vapore (hp):

$$1hp = 746W$$

a) Kilowattora

Se una forza compie un lavoro a una potenza costante di 1 kilowatt (corrispondente a $1000J/s$), produrrà in 1 ora 1 kilowattora ($1kW \cdot h$) di lavoro, ovvero:

$$1kW \cdot h = 1000W \cdot 3600s = 3,6 \cdot 10^6 J = 3,6MJ$$

4 Forze conservative e forze non conservative

Si dice **conservativa** una forza il cui lavoro non dipende dal cammino percorso; altrimenti si dice **non conservativa** o **dissipativa**.

Abbiamo visto che il lavoro fatto da una forza per spostare un corpo da un punto A ad un punto B è dato da:

$$L = \sum_i F_i \Delta x \quad \text{e} \quad L = \int_{x_1}^{x_2} F(x) dx$$

che rappresenta la somma infinita degli infinitesimi lavori da A a B .

Il problema è che questo integrale dal punto A al punto B , detto **integrale di linea**, può dipendere dal **cammino** (linea) che si sceglie per andare dal punto A al punto B .

Infatti, mentre in una dimensione (retta) per andare da un punto A ad un punto B c'è un solo cammino possibile, in due dimensioni (piano) o in tre dimensioni (spazio) ci sono infiniti cammini possibili.

Sono **forze conservative**:

- la forza peso;
- la forza elastica;
- la forza di gravitazione universale.

Sono **forze non conservative**:

- la forza di attrito statico e dinamico;
- la forza di attrito viscoso.

5 L'energia potenziale gravitazionale

L'energia potenziale di un corpo è il lavoro che esso può compiere grazie alla sua posizione. Essa è pari a $U = mgh$.

Abbiamo visto che il lavoro necessario per sollevare un corpo di massa m ad un'altezza h è pari a $L = mgh$, e che tale valore è indipendente dal percorso seguito. La forza gravitazionale è, quindi, un esempio di forza conservativa.

Questo è anche il lavoro fatto dalla forza gravitazionale quando il corpo viene lasciato cadere dalla stessa altezza.

Ciò significa che il corpo, per il solo fatto di trovarsi all'altezza h , ha la possibilità di compiere un lavoro e quindi possiede una sua energia.

Ecco perché tale energia è detta potenziale, perché il corpo l'ha soltanto in potenza, verrà trasformata in lavoro o in un'altra forma di energia soltanto quando verrà fatto cadere.

È poi detta **energia potenziale gravitazionale** perché è dovuta al campo gravitazionale e cioè alla forza di gravità esercitata dalla terra su tutti i corpi che si trovano in prossimità della sua superficie. Se lo stesso corpo venisse sollevato e posto in un luogo talmente lontano dalla superficie terrestre da non risentire della forza di gravità, esso, non cadendo, non svolgerebbe alcun lavoro e quindi non avrebbe in potenza alcuna energia.

L'energia potenziale gravitazionale si indica con la lettera U ed è misurata in Joule o dine.

In generale, il lavoro fatto dalla forza gravitazionale per portare un corpo da un'altezza iniziale h_i a un'altezza finale h_f è:

$$\Delta L = mg(h_i - h_f) = mgh_i - mgh_f = -(mgh_f - mgh_i) = -(U_f - U_i) = -\Delta U$$

Per il teorema dell'energia cinetica tale lavoro è anche uguale alla variazione dell'energia cinetica del corpo $\Delta L = \Delta E_c$.

Da cui segue che:

$$\Delta E_c = -\Delta U \quad \text{e} \quad E_{cf} - E_{ci} = U_i - U_f \quad \text{e} \quad E_{cf} - U_f = E_{ci} - U_i$$

Sia:

$$E = E_c + U \tag{5.1}$$

$$E_f = E_i \tag{5.2}$$

La (5.1) esprime l'**energia meccanica** del corpo. In questo caso essa è data dalla somma dell'energia cinetica e dell'energia potenziale del corpo. Più in generale, essa è uguale alla somma dell'energia cinetica e di **tutte** le energia potenziali del corpo.

La (5.2) rappresenta il **teorema della conservazione dell'energia meccanica**.

L'energia meccanica di una particella soggetta a sole forze conservative si conserva.

È importante osservare che, mentre il teorema dell'energia cinetica vale sempre, quello della conservazione dell'energia meccanica vale solo se le forze in gioco sono conservative.

Nota In questo capitolo ci siamo limitati a trattare soltanto una forza conservativa: la **forza peso**. Abbiamo tralasciato altre due forze conservative molto importanti in Fisica: la **forza elastica** e la **forza gravitazionale universale**. Questo perché, proprio per la loro importanza, si è dedicato un capitolo a ciascuna di esse. La *forza elastica* è stata approfondita nel capitolo sul *moto oscillatorio*, mentre la *forza gravitazionale universale* verrà analizzata nel capitolo sulla *gravitazione universale*.

Impulso e quantità di moto

1 Quantità di moto e impulso di una forza

Se siete tranquillamente seduti su una panchina in un parco cittadino e un cagnolino vi corre incontro scodinzolando, probabilmente, vi viene da sorridere pensando che voglia giocare con voi. Quasi sicuramente non avreste la stessa reazione se a venirvi incontro con la stessa velocità del cagnolino fosse un elefante.

Questo per una questione di quantità di moto.

Sir Isaac Newton nei suoi *principia* osservò che la massa di un corpo aveva un ruolo importante nel moto di un corpo esattamente come la sua velocità, e definì una nuova grandezza, che chiamò **quantità di moto**:

$$\vec{q} = m\vec{v} \quad (1.1)$$

Essa è un vettore che ha la direzione e il verso della velocità della particella e modulo uguale al prodotto della massa per il modulo della velocità $q = mv$.

a) Quantità di moto e secondo principio della dinamica

Supponiamo che su un corpo di massa m agisca una forza $\vec{F}$. Per il secondo principio della dinamica:

$$\vec{F} = m\vec{a} \quad (1.2)$$

Ricordando che $\vec{a} = \frac{\Delta\vec{v}}{\Delta t}$ la (1.2) diventa:

$$\vec{F} = m \frac{\Delta\vec{v}}{\Delta t} = \frac{\Delta(m\vec{v})}{\Delta t} = \frac{\Delta\vec{q}}{\Delta t} \quad (1.3)$$

Il secondo principio della dinamica venne inizialmente formulato da Newton secondo l'equazione (1.3): «La velocità di variazione della quantità di moto è uguale alla forza netta agente sul corpo».

In altre parole la (1.3) afferma che la **forza è uguale alla variazione della quantità di moto nel tempo**.

Poiché la (1.3) è un'equazione vettoriale essa è equivalente alle tre equazioni lungo gli assi

$$x, y \text{ e } z: F_x = \frac{\Delta q_x}{\Delta t}; \quad F_y = \frac{\Delta q_y}{\Delta t}; \quad F_z = \frac{\Delta q_z}{\Delta t}.$$

b) Impulso di una forza

La (1.3) può anche essere scritta come:

$$\Delta\vec{q} = \vec{F}\Delta t \quad (1.4)$$

La quantità $\vec{F}\Delta t$ è detta **impulso della forza** e si indica con $I_{\Delta t}$.

Dalla (1.4) segue che:

- la variazione della quantità di moto di un punto materiale, durante l'intervallo di tempo Δt , è uguale all'impulso della forza agente sul punto materiale, durante lo stesso intervallo di tempo;

- l'impulso ha la stessa dimensione della quantità di moto:

$$[q] = [I] = [mv] = [m \text{ t}^{-1}]$$

- le unità di misura dell'impulso e della quantità di moto sono: $\text{Kg} \cdot \text{m/s}$ nel Sistema Internazionale; $\text{dine} \cdot \text{g/s}$ nel Sistema c.g.s.

La (1.4) è equivalente alle tre equazioni scalari:

$$\Delta q_x = F_x \Delta t; \quad \Delta q_y = F_y \Delta t; \quad \Delta q_z = F_z \Delta t$$

c) Conservazione della quantità di moto

Dalle (1.4) segue che se su un corpo non agisce alcuna forza, ossia se $\vec{F} = 0$ allora $\Delta \vec{q} = 0$ e cioè $\vec{q}_2 = \vec{q}_1 \rightarrow \vec{q} = \text{cost}$. Possiamo quindi dire che, in assenza di forze agenti su di un punto materiale, la sua quantità di moto si mantiene costante nel tempo (conservazione della quantità di moto).

Tale risultato costituisce una forma alternativa del **principio d'inerzia**.

d) Forze esterne e forze interne in un sistema

Consideriamo un sistema composto da N punti materiali:

$$m_1, m_2, \dots, m_N$$

Su ognuno di questi corpi possono agire sia delle forze dovute all'interazione con altri corpi del sistema sia delle forze esterne.

Consideriamo il caso illustrato in figura 2.

Sul corpo di massa m_1 agisce sia la forza di interazione $\vec{F}_{12}$ dovuta alla massa m_2 che la forza esterna $\vec{F}_1$ e sul corpo di massa m_2 agisce sia la forza di interazione $\vec{F}_{21}$ dovuta alla massa m_1 che la forza esterna $\vec{F}_2$.

La somma delle forze agenti sul sistema è pari a:

$$\sum \vec{F} = \vec{F}_1 + \vec{F}_{12} + \vec{F}_{21} + \vec{F}_2$$

Per il terzo principio della dinamica è $\vec{F}_{12} = -\vec{F}_{21}$, quindi $\sum \vec{F} = \vec{F}_1 + \vec{F}_2$, cioè la somma di tutte le forze agenti sul sistema è uguale alla somma delle forze esterne; in simboli:

$$\sum \vec{F} = \vec{F}_{\text{est}}$$

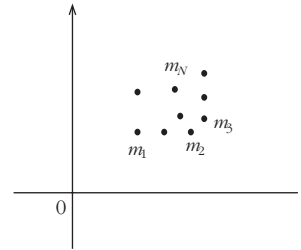


Fig. 1

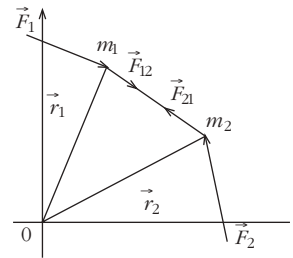


Fig. 2

2 La quantità di moto di un sistema di particelle

Consideriamo nuovamente un sistema di particelle di massa $m_1, m_2, m_3, \dots, m_N$ aventi velocità $\vec{v}_1, \vec{v}_2, \vec{v}_3, \dots, \vec{v}_N$.

La **quantità di moto totale del sistema** è data dalla somma vettoriale delle quantità di moto delle singole particelle:

$$\vec{q} = \vec{q}_1 + \vec{q}_2 + \vec{q}_3 + \dots + \vec{q}_N = m_1 \vec{v}_1 + m_2 \vec{v}_2 + m_3 \vec{v}_3 + \dots + m_N \vec{v}_N$$

Studiamo la variazione della quantità di moto nel tempo:

$$\frac{\Delta \vec{q}}{\Delta t} = m_1 \frac{\Delta \vec{v}_1}{\Delta t} + m_2 \frac{\Delta \vec{v}_2}{\Delta t} + \dots + m_N \frac{\Delta \vec{v}_N}{\Delta t} = m_1 \vec{a}_1 + m_2 \vec{a}_2 + \dots + m_N \vec{a}_N = \vec{F}_1 + \vec{F}_2 + \dots + \vec{F}_N = \sum \vec{F}$$

Ma abbiamo visto che $\sum \vec{F} = \vec{F}_{\text{est}}$ e quindi $\frac{\Delta \vec{q}}{\Delta t} = \vec{F}_{\text{est}}$, ovvero la **variazione** della **quantità di moto** è uguale alla sommatoria delle forze esterne agenti sul sistema.

Da ciò segue che se sul sistema non agiscono forze esterne si ha che $\frac{\Delta \vec{q}}{\Delta t} = 0$ e quindi:

$$\vec{q} = \text{cost} \quad (2.1)$$

La (2.1) è conosciuta come la **legge di conservazione della quantità di moto**. Essa afferma che se su un sistema non agiscono forze esterne la sua quantità di moto si mantiene costante:

$$\vec{F}_{\text{est}} = 0 \rightarrow q = \text{cost}$$

Il principio di conservazione della quantità di moto di un sistema di particelle isolato è il secondo dei principi di conservazione in cui ci siamo imbattuti, in seguito ne incontreremo degli altri.

I principi di conservazione sono molto importanti in Fisica in quanto permettono di ricavare delle informazioni sull'evolversi di un sistema: essi indicano che, mentre il sistema si sta evolvendo in determinate condizioni, c'è una grandezza che rimane invariata.

Abbiamo finora visto che:

- se sul sistema agiscono soltanto forze conservative viene conservata l'energia meccanica del sistema;
- se il sistema è isolato viene conservata la quantità di moto.

a) Centro di massa di un sistema

Vediamo adesso come un sistema di particelle si evolve nel tempo.

Nel capitolo sulle forze abbiamo definito il centro di massa o baricentro di un sistema di particelle:

$$\vec{r}_c = \frac{\sum_i \vec{r}_i m_i}{M}$$

Se calcoliamo la variazione del centro di massa nel tempo scopriamo che:

$$\vec{v}_c = \frac{\Delta \vec{r}_c}{\Delta t} = \frac{\sum_i m_i \frac{\Delta \vec{r}_i}{\Delta t}}{M} = \frac{\sum_i m_i \vec{v}_i}{M} = \frac{\vec{Q}}{M} \quad (2.2)$$

Da cui segue che $\vec{Q} = M\vec{v}_c$ e cioè:

La quantità del moto di un sistema di particelle è dato dal prodotto della massa totale del sistema per la velocità del suo centro di massa.

Studiamo ora l'accelerazione del centro di massa.

Dalla (2.2):

$$\vec{a}_c = \frac{\Delta \vec{v}_c}{\Delta t} = \frac{\sum_i m_i \frac{\Delta \vec{v}_i}{\Delta t}}{M} = \frac{\sum_i m_i \vec{a}_i}{M} = \frac{\sum_i \vec{F}_i}{M}$$

da cui segue che $\sum_i \vec{F}_i = M\vec{a}_c$ ma $\sum_i \vec{F}_i = \vec{F}_{\text{est}}$ e quindi:

$$\vec{F}_{\text{est}} = M\vec{a}_c \quad (2.3)$$

Il centro di massa si sposta come un punto materiale in cui è concentrata tutta la massa del sistema su cui agisce la risultante delle forze esterne.

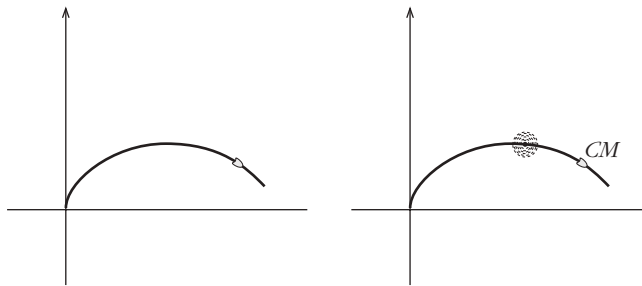
Dalla (2.3) segue anche che $\vec{F}_{\text{est}} = M \frac{\Delta \vec{v}_c}{\Delta t}$.

Se sul sistema non agiscono forze esterne si ha $M \frac{\Delta \vec{v}_c}{\Delta t} = 0$ da cui segue che: $\vec{v}_c = \text{cost.}$

Quindi, in un sistema isolato, la posizione e la velocità delle particelle che lo compongono variano da istante a istante a causa delle forze interne che agiscono su di esse, ma il centro di massa continua a mantenere il suo stato di moto.

Se, invece, sul sistema agiscono delle forze esterne, il centro di massa si muoverà come se avesse concentrata in sé tutta la massa del sistema e come se su di esso agissero tutte le forze esterne del sistema.

Nel capitolo sulla *cinematica* abbiamo studiato il moto parabolico di un proiettile (Fig. 3.a). Supponiamo che il proiettile sia una granata e che esploda in aria ad un certo istante (Fig. 3.b).



Figg. 3.a e 3.b

Sui singoli frammenti della granata agiscono le forze interne del sistema e quindi avranno un moto del tutto aleatorio, ma il centro di massa del sistema continuerà a muoversi come se l'esplosione non fosse avvenuta. Su di esso, dopo lo scoppio, continua ad agire esclusivamente la forza esterna di gravità come se in esso fosse concentrata l'intera massa del sistema e cioè la massa della granata prima dello scoppio.

3 Il momento angolare

Nel capitolo sulle forze è stato definito il momento di una forza $\vec{F}$ come un vettore $\vec{M} = \vec{r} \times \vec{F}$.

Definiamo allo stesso modo il **momento della quantità di moto** $\vec{q}$, detto **momento angolare**, come un vettore:

$$\vec{L} = \vec{r} \times \vec{q} \quad (3.1)$$

Essendo $\vec{q} = m\vec{v}$ la (3.1) diventa:

$$\vec{L} = \vec{r} \times \vec{q} = \vec{r} \times m\vec{v} = m\vec{r} \times \vec{v} \quad (3.2)$$

Il momento angolare è pertanto un vettore avente:

- **direzione** perpendicolare al piano formato dai vettori $\vec{r}$ e $\vec{v}$;
- **verso** definito dalla regola della mano destra o della vite destrorsa che ruota da $\vec{r}$ a $\vec{v}$;
- **modulo**:

$$L = qr = mrv \quad (3.3)$$

Nel moto circolare si ha $v = \omega r$ e la (3.3) diventa $L = qr = mr^2\omega$.

La direzione di $\vec{L}$ coincide con quella di $\vec{\omega}$.

Derivando rispetto al tempo la (3.1) si ottiene:

$$\frac{d\vec{L}}{dt} = \frac{d\vec{r}}{dt} \times \vec{q} + \vec{r} \times \frac{d\vec{q}}{dt}$$

ma $\frac{d\vec{r}}{dt} \times \vec{q} = \vec{v} \times m\vec{v} = m(\vec{v} \times \vec{v}) = 0$, quindi $\frac{d\vec{L}}{dt} = \vec{r} \times \frac{d\vec{q}}{dt}$ e $\frac{d\vec{q}}{dt} = \vec{F}$, quindi $\frac{d\vec{L}}{dt} = \vec{r} \times \vec{F}$ e $\vec{M} = \vec{r} \times \vec{F}$, quindi:

$$\frac{d\vec{L}}{dt} = \vec{M} \quad (3.4)$$

ovviamente $\vec{L}$ e $\vec{M}$ devono essere misurati rispetto allo stesso punto.

La (3.4) è analoga all'equazione:

$$\frac{d\vec{q}}{dt} = \vec{F} \quad (3.5)$$

Nella (3.4) si è sostituito alla forza il momento della forza e alla quantità di moto il momento angolare.

La (3.5) indica che la variazione rispetto al tempo del momento angolare di una particella è uguale al momento della forza da essa applicata quando entrambi sono misurati rispetto allo stesso punto.

4 Gli urti

Studiamo, adesso, cosa succede quando due corpi si urtano.

Siccome l'**urto** avviene in un tempo piccolissimo, le forze d'urto sono molto intense e quindi possiamo trascurare tutte le altre forze agenti sul sistema (forza di gravità, elastica etc.). Quindi, nell'istante dell'urto, il sistema formato dal corpo urtante e da quello urtato può considerarsi isolato e la sua quantità di moto viene conservata dopo l'urto.

Non è invece sempre detto che venga conservata anche l'energia cinetica del sistema.

Un urto in cui l'energia cinetica:

- viene conservata è detto **elastico**;
- non viene conservata è detto **anelastico**.

Inoltre, un urto può essere **centrale** o **obliquo**. In un urto centrale la direzione del centro del corpo urtante è diretta verso il centro del corpo urtato, mentre nell'urto obliquo la direzione del centro del corpo urtato non è diretta verso il centro del corpo urtante.

Noi ci occuperemo soltanto di urti centrali elastici e anelastici.

a) Urti elastici

Negli **urti elastici** viene dunque conservata sia la quantità di moto che l'energia cinetica del sistema.

Consideriamo due particelle che subiscono un urto centrale elastico (Fig. 4).

Prima dell'urto le due particelle di massa m_1 e m_2 hanno velocità $\vec{v}_{1i}$ e $\vec{v}_{2i}$ e dopo l'urto hanno velocità $\vec{v}_{1f}$ e $\vec{v}_{2f}$.

Vediamo come le velocità finali sono legate alle velocità iniziali e alle masse delle particelle.

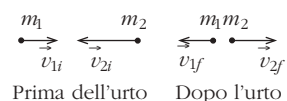


Fig. 4

Siccome viene conservata sia l'energia che la quantità di moto del sistema prima e dopo l'urto si ha $q_i = q_f$:

$$m_1 v_{1i} + m_2 v_{2i} = m_1 v_{1f} + m_2 v_{2f} \quad \text{e} \quad E_i = E_f \quad (4.1a)$$

$$\frac{1}{2} m_1 v_{1i}^2 + \frac{1}{2} m_2 v_{2i}^2 = \frac{1}{2} m_1 v_{1f}^2 + \frac{1}{2} m_2 v_{2f}^2 \quad (4.1b)$$

Mettendo a sistema le (4.1) si ottiene:

$$m_1 (v_{1i} - v_{1f}) = m_2 (v_{2f} - v_{2i})$$

$$m_1 (v_{1i} - v_{1f})(v_{1i} + v_{1f}) = m_2 (v_{2f} - v_{2i})(v_{2i} + v_{2f})$$

Dividendo membro a membro:

$$v_{1i} + v_{1f} = v_{2i} + v_{2f} \quad (4.2a)$$

$$m_1 (v_{1i} - v_{1f}) = (v_{2f} - v_{1f}) \quad (4.2b)$$

Dalla prima equazione si ha $v_{1f} = v_{2i} + v_{2f} - v_{1i}$.

Sostituendo nella seconda si ha $m_1 (v_{1i} - v_{2i} - v_{2f} + v_{1i}) = m_2 (v_{2f} - v_{1f})$.

Da cui segue:

$$v_{1f} = \frac{m_1 - m_2}{m_1 + m_2} v_{1i} + \frac{2m_2}{m_1 + m_2} v_{2i} \quad (4.3a)$$

$$v_{2f} = \frac{2m_1}{m_1 + m_2} v_{1i} + \frac{m_2 - m_1}{m_1 + m_2} v_{2i} \quad (4.3b)$$

Dalle (4.3) segue che:

se $m_1 = m_2$:

$$v_{1f} = v_{2i} \quad \text{e} \quad v_{2f} = v_{1i}$$

se $v_2 = 0$:

$$v_{1f} = \frac{m_1 - m_2}{m_1 + m_2} v_{1i} \quad \text{e} \quad v_{2f} = \frac{2m_1}{m_1 + m_2} v_{1i}$$

se $v_2 = 0$ e $m_1 \gg m_2$:

$$v_{1f} \approx v_{1i} \quad \text{e} \quad v_{2f} \approx 2v_{1i}$$

se $v_2 = 0$ e $m_1 \ll m_2$:

$$v_{1f} \approx -v_{1i} \quad \text{e} \quad v_{2f} \approx 0$$

b) Urti anelastici

Negli **urti anelastici** si conserva la quantità di moto mentre l'energia cinetica non si conserva. Se gli oggetti rimangono uniti dopo l'urto si ha un urto perfettamente anelastico.

Consideriamo due particelle che subiscono un urto centrale perfettamente anelastico (Fig. 5).

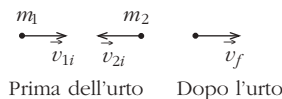


Fig. 5

Prima dell'urto le due particelle di massa m_1 e m_2 hanno velocità $\vec{v}_{1i}$ e $\vec{v}_{2i}$ e dopo l'urto le particelle rimangono unite e viaggiano con velocità $\vec{v}_f$.

In questo caso viene soltanto conservata la quantità di moto del sistema:

$$q_i = q_f \quad m_1 v_{1i} + m_2 v_{2i} = (m_1 + m_2) v_f$$

da cui segue che:

$$v_f = \frac{m_1 v_{1i} + m_2 v_{2i}}{m_1 + m_2}$$

c) Coefficiente di restituzione

Per una generica collisione tra due oggetti che si muovono lungo la medesima retta esiste un coefficiente di restituzione definito da:

$$e = \frac{v_{2f} - v_{1f}}{v_{1i} - v_{2i}}$$

dove $\vec{v}_{1i}$ e $\vec{v}_{2i}$ sono le velocità degli oggetti prima dell'urto e $\vec{v}_{1f}$ e $\vec{v}_{2f}$ le velocità dopo l'urto.

Se l'urto è:

- elastico, allora $e = 1$;
- anelastico, allora $e < 1$;
- totalmente anelastico, $v_{2f} = v_{1f}$ e quindi $e = 0$.

1 Il momento angolare di un corpo rigido

Possiamo distinguere due tipi di moto di un corpo rigido:

- il **moto traslazionale**, ossia quando tutte le particelle descrivono traiettorie parallele (Fig. 1);

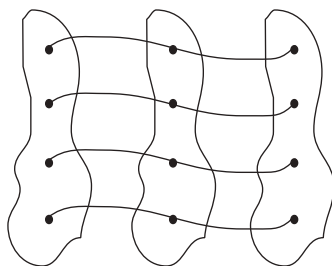


Fig. 1

- il **moto rotazionale**, ovvero quando tutte le particelle descrivono traiettorie circolari attorno a una retta, chiamata asse di rotazione (Fig. 2). L'asse può essere fisso o variare da istante a istante.

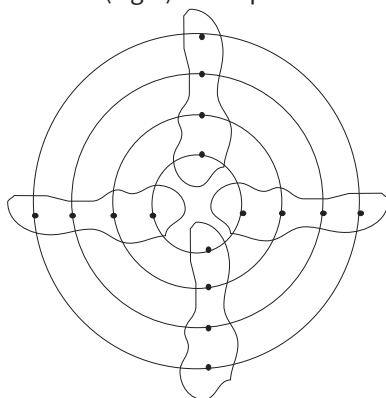


Fig. 2

Il moto più generale di un corpo è detto **moto rototraslazionale** ed è dato dalla combinazione di una traslazione e di una rotazione attorno ad un asse opportuno. Per esempio, il moto del corpo in figura 3, che passa dalla posizione 1 alla posizione 2, può essere considerato come una traslazione e una rotazione attorno al centro di massa.

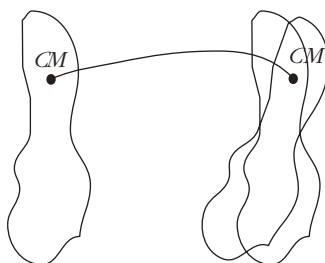


Fig. 3

a) Definizione di momento angolare

Avendo già analizzato il moto traslatorio nella trattazione sulla dinamica, in questo capitolo esamineremo soltanto il moto rotatorio di un corpo rigido.

Studiamo quindi il **momento angolare di un corpo rigido** che ruota attorno ad un determinato asse. Definiamo un sistema di assi cartesiani con origine in un punto 0 e supponiamo che il corpo ruoti attorno all'asse z con velocità angolare $\vec{\omega}$. Ciò vuol dire che tutte le particelle che compongono il corpo ruotano attorno all'asse z con velocità angolare $\vec{\omega}$ (Fig. 4).

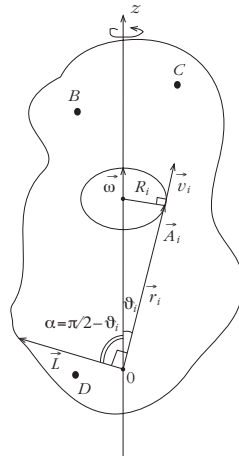


Fig. 4

Consideriamo il moto di una particella qualsiasi, che indichiamo con A_i . Essa descrive attorno all'asse z un cerchio di raggio R_i .

Nel capitolo sulla cinematica abbiamo ricavato la sua velocità tangenziale $\vec{v}_i$:

$$v_i = \omega R_i \quad (1.1)$$

perpendicolare al piano definito dai vettori $\vec{v}_i$ ed $\vec{R}_i$.

Definendo con $\vec{r}_i$ il vettore posizione di A_i rispetto all'origine degli assi, dalla figura 4 possiamo notare che:

$$R_i = r_i \sin \theta_i \quad (1.2)$$

Sostituendo la (1.2) nella (1.1) otteniamo che:

$$v_i = \omega r_i \sin \theta_i \quad (1.3)$$

Ricordando che il prodotto vettoriale tra due vettori $\vec{A}$ e $\vec{B}$ è un vettore $\vec{C}$ di modulo $C = AB \sin \alpha$, con direzione perpendicolare al piano definito dai vettori $\vec{A}$ e $\vec{B}$, e verso dato dalla regola della vite destrorsa, possiamo scrivere la (1.3) come:

$$\vec{v}_i = \vec{\omega} \cdot \vec{r}_i \quad (1.4)$$

Nel capitolo sulla quantità di moto abbiamo definito il momento angolare o momento della quantità di moto di una particella come un vettore $\vec{L} = \vec{r} \times \vec{q}$; quindi, il momento angolare di A_i è il vettore $\vec{L}_i = \vec{r}_i \cdot \vec{q}_i = \vec{r}_i \cdot m_i \vec{v}_i = m_i \vec{r}_i \cdot \vec{v}_i$ avente:

- **direzione** perpendicolare al piano definito dai vettori $\vec{r}_i$ e $\vec{v}_i$;
- **verso** definito dalla regola della vite destrorsa;
- **modulo** $L = m_i r_i v_i$.

Infatti, dalla (1.4), $\vec{r}_i$ e $\vec{v}_i$ sono perpendicolari e dunque:

$$|\vec{r}_i \times \vec{v}_i| = r_i v_i \sin 90^\circ = r_i v_i$$

- Componente di $\vec{L}_i$ lungo l'asse di rotazione

Il vettore $\vec{L}_i$ forma con l'asse di rotazione un angolo $\alpha = 90 - \vartheta$ (Fig. 4), quindi:

$$L_{iz} = L_i \cos(90 - \vartheta) = L_i \sin \vartheta = m_i r_i v_i \sin \vartheta = m_i (r_i \sin \vartheta) (v_i) = m_i R_i \omega R_i = m_i R_i^2 \omega$$

- Momento angolare totale

Il **momento angolare totale** del corpo rigido è dato dalla sommatoria dei momenti angolari delle singole particelle che lo compongono:

$$\vec{L} = \vec{L}_1 + \vec{L}_2 + \vec{L}_3 + \dots = \sum_i \vec{L}_i \quad (1.5)$$

e la componente del momento angolare lungo l'asse di rotazione è data da:

$$L_z = L_{1z} + L_{2z} + L_{3z} + \dots = \sum_i L_{iz} = \left(\sum_i m_i R_i^2 \right) \omega \quad (1.6)$$

b) Momento d'inerzia

La grandezza:

$$I = \sum_i m_i R_i^2 \quad (1.7)$$

è detta **momento d'inerzia** del corpo rispetto all'asse di rotazione z.

Dalla (1.7) possiamo ricavare che le dimensioni del momento di inerzia sono $[I] = [ml^2]$ e le sue unità di misura sono:

- $kg \cdot m^2$ nel Sistema Internazionale;
- $g \cdot cm^2$ nel Sistema c.g.s.

Sostituendo la (1.7) nella (1.6) si ottiene:

$$L_z = I \omega \quad (1.8)$$

c) Differenza tra momento d'inerzia e massa inerziale

Dal confronto della (1.8) con la quantità di moto di un corpo $q = mv$ possiamo dedurre che come la massa inerziale m di un corpo rappresenta la sua inerzia a un movimento traslazionale, così il momento inerziale I rappresenta l'inerzia del corpo ad un movimento rotazionale.

Una differenza sostanziale tra la massa inerziale e il momento inerziale di un corpo rigido è che mentre la massa è una proprietà intrinseca del corpo stesso, e quindi indipendente dalla sua posizione, il momento d'inerzia dipende dal particolare asse attorno a cui avviene la rotazione.

Inoltre, il momento d'inerzia non dipende soltanto dalla massa del corpo ma, soprattutto, da come essa viene distribuita.

In figura 5 sono mostrati tre diversi corpi rigidi aventi la stessa massa ma forma differente.

Si può dimostrare che $I_1 < I_2 < I_3$.

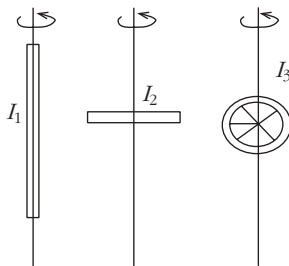


Fig. 5

d) Calcolo del momento d'inerzia

Vediamo ora di calcolare il momento d'inerzia di un corpo rigido qualsiasi, di massa M . Siccome il numero di particelle che compongono un corpo è infinito, la (1.7) diventa:

$$I = \int_0^M R^2 dm \quad (1.9)$$

dove R è la distanza dell'elemento di massa dm dall'asse di rotazione.

Se ρ è la densità del corpo, $dm = \rho dV$ e quindi:

$$I = \int_0^V R^2 \rho dV \quad (1.10)$$

Se il corpo è omogeneo ($\rho = \text{cost}$), la (1.10) diventa:

$$I = \rho \int_0^V R^2 dV \quad (1.11)$$

• Momento d'inerzia rispetto all'asse z

Proiettando la posizione di dm sul piano xy , dalla figura 6.a osserviamo che la distanza di dm dall'asse z è $R_z = \sqrt{x^2 + y^2}$ e dunque la (1.11) diventa:

$$I_z = \rho \int_0^V (x^2 + y^2) dV$$

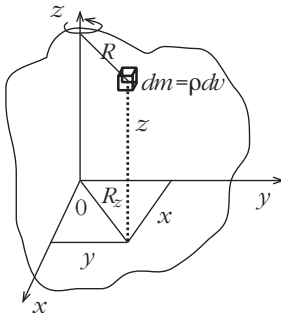


Fig. 6.a

- Momento d'inerzia rispetto all'asse x

Proiettando la posizione di dm sul piano yz (Fig. 6.b), si ottiene che la distanza di dm dall'asse x è $R_x = \sqrt{y^2 + z^2}$ e pertanto la (1.11) diventa:

$$I_x = \rho \int_0^V (y^2 + z^2) dV$$

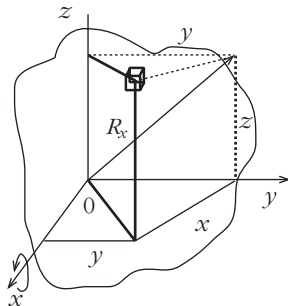


Fig. 6.b

- Momento d'inerzia rispetto all'asse y

Proiettando la posizione di dm sul piano xz (Fig. 6.c), si ottiene che la distanza di dm dall'asse y è $R_y = \sqrt{x^2 + z^2}$ e quindi la (1.11) diventa:

$$I_y = \rho \int_0^V (x^2 + z^2) dV$$

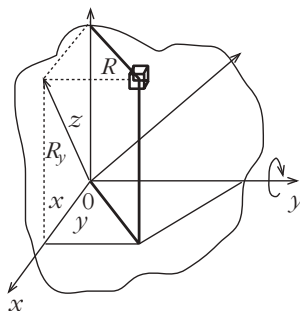
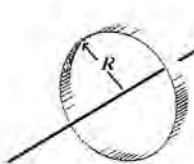
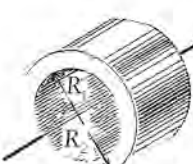
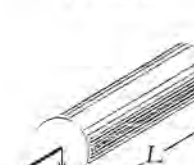
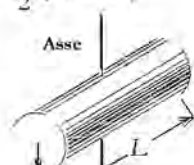
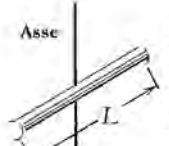
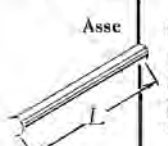
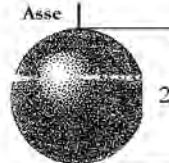
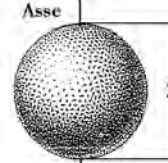
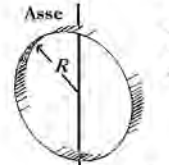
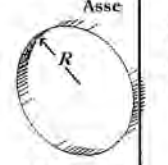


Fig. 6.c

In Tabella 1 sono mostrati alcuni momenti di inerzia.

 Anello (rispetto all'asse del cilindro) $I = MR^2$	 Cilindro cavo (rispetto all'asse del cilindro) $I = \frac{M}{2}(R_1^2 + R_2^2)$
 Cilindro pieno (rispetto all'asse del cilindro) $I = \frac{MR^2}{2}$	 Cilindro pieno o disco (rispetto a un diametro centrale) $I = \frac{MR^2}{4} + \frac{ML^2}{12}$
 Sbarra sottile (rispetto ad un asse passante per il centro perpendicolare alla sua lunghezza) $I = \frac{ML^2}{12}$	 Sbarra sottile (rispetto ad un asse passante per un estremo perpendicolare alla sua lunghezza) $I = \frac{ML^2}{3}$
 Sfera piena (rispetto ad un diametro qualunque) $I = \frac{2MR^2}{5}$	 Sottile crosta sferica (rispetto ad un diametro qualunque) $I = \frac{2MR^2}{3}$
 Anello (rispetto ad un diametro qualunque) $I = \frac{MR^2}{2}$	 Anello (rispetto ad un qualunque asse tangente) $I = \frac{3MR^2}{2}$

Tab. 1

e) Teorema di Steiner

Il **teorema di Steiner** afferma che se $\vec{z}$ è un asse qualsiasi, $\vec{z}_{CM}$ è l'asse passante per il centro di massa del corpo rigido e d è la distanza tra i due assi, vale la relazione $I = I_{CM} + Md^2$, dove:

- I è il momento di inerzia rispetto all'asse $\vec{z}$;
- I_{CM} è il momento di inerzia rispetto all'asse $\vec{z}_{CM}$;
- M è la massa del corpo rigido.

DIMOSTRAZIONE

Definiamo il sistema xyz , avente centro qualsiasi, e il sistema $x_{CM}y_{CM}z_{CM}$, avente centro nel centro di massa del corpo rigido (Fig. 7), tali per cui $\vec{y} \equiv \vec{y}_{CM}$, $\vec{x}$ parallelo a $\vec{x}_{CM}$ e $\vec{z}$ parallelo a $\vec{z}_{CM}$, e indichiamo con d la distanza fra i centri dei due sistemi.

Definiamo un punto P appartenente al corpo rigido e indichiamo con P' la sua proiezione sui piani xy .

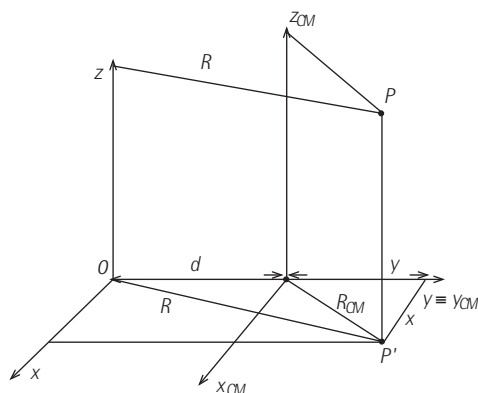


Fig. 7

Indicando con R e R_{CM} le distanze di P' dalle origini degli assi, si ha che:

$$R_{CM}^2 = x^2 + y^2$$

$$R^2 = x^2 + (y + d)^2 = x^2 + y^2 + 2dy + d^2 = R_{CM}^2 + 2dy + d^2$$

Il momento di inerzia rispetto all'asse $\vec{z}$ dà $I = \sum mR^2$. Quindi:

$$I = \sum m(R_{CM}^2 + 2dy + d^2) = \sum mR_{CM}^2 + 2d \sum my + d^2 \sum m \quad (1.12)$$

Analizziamo i singoli termini della (1.12).

1° termine

$\sum mR_{CM}^2$ è il momento di inerzia rispetto al centro di massa $\sum mR_{CM}^2 = I_{CM}$.

2° termine

Per analizzare il termine $2d \sum my$ ricordiamo che la coordinata y_{CM} del centro di massa è data da:

$$y_{CM} = \frac{\sum my}{\sum m}, \text{ dunque } \sum my = y_{CM} \sum m = My_{CM}, \text{ ma in questo caso } y_{CM} = 0 \text{ e pertanto } \sum my = 0.$$

3° termine

$$d^2 \sum m = d^2 M$$

La (1.12) diventa $I = I_{CM} + Md^2$, come volevasi dimostrare.

2 La dinamica rotazionale di un corpo rigido

Vediamo ora più in dettaglio che cosa succede quando un corpo rigido ruota attorno ad un asse dal punto di vista **energetico**.

Nel capitolo sulla dinamica abbiamo studiato il moto traslatorio di un corpo e abbiamo trovato una relazione tra la forza applicata, l'energia prodotta, il lavoro effettuato e la sua potenza; in questo paragrafo faremo altrettanto per un corpo che si muove di moto rotatorio.

a) Energia cinetica rotazionale

Consideriamo un corpo rigido che ruota attorno all'asse z con velocità angolare ω . Ciascuna particella del corpo, per il solo fatto di essere in movimento, possiede una sua energia cinetica.

Focalizziamo l'attenzione sulla particella A_i (Fig. 4).

La sua velocità tangenziale v_i è legata alla velocità angolare ω dalla relazione $v_i = \omega R_i$. Quindi la sua energia cinetica sarà data da:

$$E_{c_i} = \frac{1}{2} m v_i^2 = \frac{1}{2} m_i (\omega R_i)^2 = \frac{1}{2} m_i R_i^2 \omega^2$$

L'energia cinetica totale del corpo in movimento sarà data dalla somma delle energie cinetiche di tutte le particelle che compongono il corpo rigido:

$$E_c = \sum_i E_{c_i} = \sum_i \left(\frac{1}{2} m_i R_i^2 \omega^2 \right) = \frac{1}{2} \left(\sum_i m_i R_i^2 \right) \omega^2 \quad (2.1)$$

Ma, tenendo conto della (1.7), la (2.1) diventa:

$$E_c = \frac{1}{2} I \omega^2 \quad (2.2)$$

La (2.2) è simile alla formula dell'energia cinetica di un corpo in moto traslatorio $\left(E_c = \frac{1}{2} m v^2 \right)$ se sostituiamo la velocità angolare con la velocità della particella e il momento d'inerzia con la massa della particella.

In altre parole, il momento angolare I è l'analogo della massa m così come la velocità angolare è l'analogo della velocità nel caso del moto rotatorio.

b) Lavoro e potenza rotazionale

Abbiamo finora osservato che esistono delle corrispondenze fra grandezze e formule che definiscono il moto rettilineo e quelle che definiscono il moto rotatorio.

	Moto rettilineo	Moto rotatorio
1	Spostamento lineare x	Spostamento angolare ϑ
2	Velocità rettilinea v	Velocità angolare ω
3	Accelerazione rettilinea a	Accelerazione angolare α
4	Massa (inerzia traslazionale) m	Momento d'inerzia (inerzia rotazionale) I
5	Forza $\vec{F}$	Momento di una forza $\vec{M}$
6	Energia cinetica traslazionale $E_c = \frac{1}{2} m v^2$	Energia cinetica rotazionale $E_c = \frac{1}{2} I \omega^2$
7	Quantità di moto $q = m v$	Momento angolare $L = I \omega$

Tab. 2

In particolare, nelle formule delle righe 5 e 6, la massa nel moto rettilineo gioca il ruolo del momento d'inerzia nel moto rotatorio e la velocità nel moto rettilineo gioca il ruolo della velocità angolare nel moto rotatorio.

Quindi, dovremmo aspettarci che esistano delle altre analogie tra formule applicabili nel moto rettilineo e nel moto rotatorio.

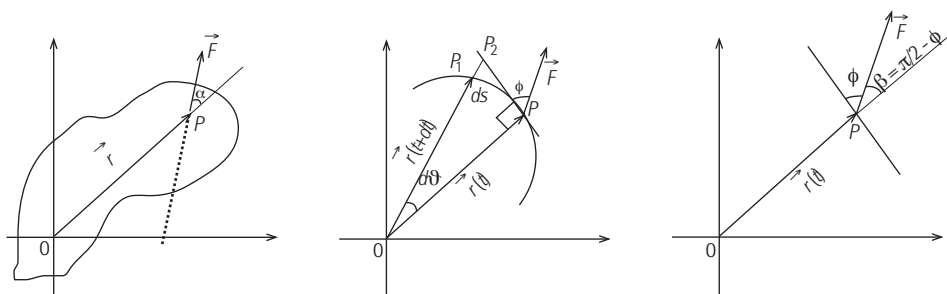
In effetti, si può dimostrare che esistono ulteriori analogie fra moto rettilineo e moto rotatorio.

	Moto rettilineo	Moto rotatorio
8	Lavoro $L = F \cdot s$	Lavoro $L = M\theta$
9	Potenza $P = qv$	Potenza $P = M\omega$
10	Forza $\vec{M} = \vec{r} \cdot \vec{F}$	Momento di una forza $M = Ia$

Tab. 3

DIMOSTRAZIONE

Consideriamo un corpo rigido libero di ruotare attorno ad un asse fisso e supponiamo di applicare una forza $\vec{F}$ in un punto P del corpo definito dal vettore posizione $\vec{r}$ (Fig. 8.a).



Figg. 8.a, 8.b e 8.c

Il momento della forza, $\vec{M} = \vec{r} \cdot \vec{F}$, è un vettore avente:

- direzione perpendicolare al piano definito dai vettori $\vec{r}$ ed $\vec{F}$, e quindi in questo caso parallela all'asse z ;
- verso definito dalla regola della vite destrorsa o della mano destra

(indice $\equiv \vec{r}$, medio $\equiv \vec{F} \rightarrow$ pollice $\equiv \vec{M}$);

- modulo $M = rF \sin \alpha$, dove α è l'angolo compreso fra $\vec{r}$ ed $\vec{F}$.

Sotto l'azione della forza F , nell'intervallo di tempo infinitesimo dt , il corpo rigido ruoterà di un angolo $d\theta$ e il punto P si sposterà di un tratto infinitesimo ds .

Consideriamo la Fig. 8.b. Per valori di $d\theta$ molto piccoli, possiamo sostituire la lunghezza dell'arco di cerchio PP_1 con il segmento PP_2 staccato sulla tangente alla curva nel punto P : $ds = \overline{PP_2}$.

Consideriamo il triangolo rettangolo OPP_2 in figura 8.b.

Dalla trigonometria si ha che $\overline{PP_2} = r \sin(d\theta)$. Per angoli molto piccoli vale la relazione $\sin \alpha = \alpha$ e quindi $ds = \overline{PP_2} = r d\theta$.

Calcoliamo il lavoro fatto dalla forza $\vec{F}$ per effettuare questa piccola rotazione:

$$dL = \vec{F} \cdot d\vec{s} = F ds \cos \phi = F \cos \phi (r d\theta) = (F r \cos \phi) d\theta$$

Consideriamo la Fig. 8.c: $\cos \phi = \sin\left(\frac{\pi}{2} - \phi\right) = \sin \beta$ $dL = (F r \sin \beta) d\theta$

Ma $F_{rsen\beta}$ è il modulo del momento della forza $\vec{F}$ rispetto al punto O . Quindi:

$$dL = Md\vartheta \quad (2.3)$$

Questa espressione del lavoro compiuto da una forza per una rotazione infinitesimale $d\vartheta$ è analoga a quella del lavoro compiuto dalla stessa forza per uno spostamento infinitesimale ds , ossia $dL = Fds$ (**Riga 8 della tabella 3**).

Per calcolare la potenza della forza, ricordando che $\text{potenza} = \frac{\text{lavoro}}{\text{tempo}}$, possiamo dividere la (2.3) per il tempo dt , ottenendo $P = \frac{dL}{dt}$, da cui:

$$P = \frac{Md\vartheta}{dt} = M \frac{d\vartheta}{dt} = M\omega \quad (2.4)$$

La (2.4) è l'analogo rotazionale della formula $P = Fv$ del moto rettilineo (**Riga 9 della tabella 3**).

Dal punto di vista energetico, se il corpo è rigido non ci può essere dissipazione di energia per moti interni al corpo. Quindi il lavoro fatto da una forza sul corpo rigido in rotazione, nell'istante dt , deve essere uguale alla variazione della sua energia cinetica nello stesso istante, cioè $\frac{dL}{dt} = \frac{dE_c}{dt}$.

Dalla (2.4): $\frac{dL}{dt} = M\omega$

Dalla (2.2): $\frac{dE_c}{dt} = \frac{d}{dt} \left(\frac{1}{2} I \omega^2 \right) = I \omega \frac{d\omega}{dt} = I \omega \alpha$

Da cui segue che $M\omega = I \omega \alpha$ e semplificando:

$$M = I \alpha \quad (2.5)$$

che è l'analogo rotazionale della formula $F = ma$ (**Riga 10 della tabella 3**).

3 Il moto rototraslatorio di un corpo rigido

Finora abbiamo studiato il moto rotatorio di un corpo e cioè il moto di un corpo che ruota attorno ad un asse fisso.

Consideriamo adesso il moto di una pallina che rotola su una superficie (Fig. 9).

Tale moto può essere considerato come una successione di moti rotatori, in cui l'asse di rotazione è rappresentato dalla retta passante per il punto di contatto tra la pallina ed il piano, e perpendicolare alla direzione di rotolamento. Tale retta è detta **asse di istantanea rotazione**.

Consideriamo una particella qualsiasi A della pallina.

In ogni istante:

- il moto di A è un moto rotatorio attorno all'asse di istantanea rotazione con una certa velocità angolare. La sua velocità lineare è perpendicolare alla retta che unisce la sua posizione con il punto di contatto con il piano ed è proporzionale alla distanza da tale punto;
- tutte le particelle che compongono il corpo rigido hanno la stessa velocità angolare e la loro velocità lineare è inversamente proporzionale al loro vettore posizione rispetto al punto P ;
- l'energia cinetica di rotazione del corpo è pari a:

$$E_c = \frac{1}{2} I_p \omega^2 \quad (3.1)$$

dove I_p è il momento di inerzia del corpo rispetto all'asse di istantanea rotazione passante per P .

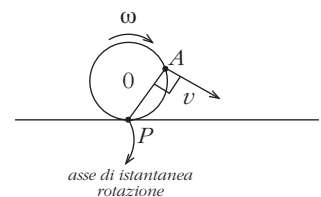


Fig. 9

Il centro di massa della pallina coincide con il suo centro O (Fig. 10): e quindi la distanza tra il centro di massa e il punto P è uguale al raggio della pallina.

I momenti di inerzia della pallina rispetto al punto P e al centro di massa sono legati dal teorema di Steiner:

$$I_P = I_{CM} + MR^2 \quad (3.2)$$

Sostituendo la (3.2) nella (3.1):

$$E_c = \frac{1}{2}(I_{CM} + MR^2)\omega^2 = \frac{1}{2}I_{CM}\omega^2 + \frac{1}{2}MR^2\omega^2 = \frac{1}{2}I_{CM}\omega^2 + \frac{1}{2}M(R\omega)^2$$

Ma $R\omega = v_{CM}$, dunque:

$$E_c = \frac{1}{2}I_{CM}\omega^2 + \frac{1}{2}Mv_{CM}^2 \quad (3.3)$$

Nella (3.3):

- il primo termine rappresenta l'**energia cinetica di rotazione** del corpo rispetto all'asse passante per il suo centro di massa;
- il secondo termine rappresenta l'**energia cinetica di traslazione** del centro di massa.

In altre parole, l'energia cinetica di un corpo che rotola su un piano orizzontale senza strisciare è data dalla somma dell'energia cinetica della traslazione del centro di massa e dell'energia cinetica della rotazione del centro di massa.

4 La conservazione del momento angolare

Nel capitolo sull'impulso e la quantità di moto abbiamo trovato la relazione:

$$\frac{d\vec{L}}{dt} = \vec{M} \quad (4.1)$$

che lega la variazione del momento angolare di una particella con il momento delle forze esercitate sulla particella.

Consideriamo ora un **sistema di particelle**.

Il loro momento angolare totale è dato dalla somma dei momenti angolari delle singole particelle, calcolati rispetto a uno stesso punto: $\vec{L} = \vec{L}_1 + \vec{L}_2 + \vec{L}_3 + \vec{L}_4 + \dots + \vec{L}_N$.

$\vec{L}$ può variare da istante a istante e tale variazione, per la (4.1), può essere dovuta a momenti di forze esterne o a momenti di forze interne, agenti sul sistema.

Per il terzo principio della dinamica, le forze interne agenti su una qualsiasi coppia di particelle sono uguali, opposte e agiscono lungo la retta congiungente le due particelle: di conseguenza, il momento totale di tali forze è nullo.

La variazione del momento angolare totale di un sistema è, quindi, dovuta esclusivamente ai momenti delle forze esterne agenti sul sistema. Pertanto, la (4.1) diventa:

$$\frac{d\vec{L}}{dt} = \vec{M}_{est} \quad (4.2)$$

dove $\vec{M}_{est}$ è la somma di tutti i momenti delle forze esterne agenti sul sistema.

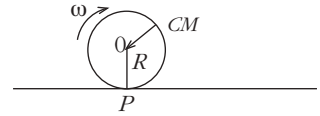


Fig. 10

La (4.2) afferma che, per un sistema di particelle, la variazione del momento angolare totale del sistema è uguale alla somma dei momenti delle forze esterne agenti sul sistema. Se sul sistema non agiscono forze esterne, $\vec{M}_{\text{est}} = 0$, da cui segue che $\frac{d\vec{L}}{dt} = 0$ e quindi:

$$\vec{L} = \text{cost} = \vec{L}_0 \quad (4.3)$$

La (4.3) esprime il **principio di conservazione del momento angolare**.

Allora, se il momento risultante delle forze esterne applicate ad un sistema è nullo, il vettore momento angolare totale del sistema rimane costante nel tempo. In altre parole, se sul sistema non agiscono forze esterne, il momento angolare delle singole particelle può variare, ma il momento totale del sistema rimane costante. La (4.3) corrisponde alle tre equazioni scalari: $L_{0x} = \text{cost}$, $L_{0y} = \text{cost}$ e $L_{0z} = \text{cost}$.

Abbiamo visto che se il sistema ruota attorno ad un qualsiasi asse, per esempio l'asse z , con velocità angolare ω , la componente del momento angolare lungo tale asse è data da $L_z = I\omega$, dove I è il momento d'inerzia del sistema rispetto all'asse z (asse di rotazione).

Quindi, se un sistema su cui non agiscono forze esterne ruota attorno all'asse z , si ha che $L_z = \text{cost}$, da cui segue che se indichiamo con I_0 e ω_0 , rispettivamente, il momento d'inerzia e la velocità angolare del sistema in un certo istante t_0 , e con I e ω il momento d'inerzia e la velocità angolare del sistema in un qualsiasi istante, si ha che $I_0\omega_0 = I\omega$.

Tale principio è utilizzato normalmente dai pattinatori su ghiaccio e dai tuffatori. La pattinatrice che ruota più o meno velocemente, alzando le braccia sopra la testa o aprendole all'altezza della vita, sfrutta questa legge così come il tuffatore che si raggomitola per girare più in fretta.

1 Un po' di storia

All'inizio del 1600 il matematico e astronomo Johannes Kepler iniziò ad analizzare matematicamente i dati raccolti in diversi anni di studio dal suo maestro, l'astronomo danese Thyco Brahe, sul **moto dei pianeti**.

Egli giunse a formulare tre leggi, dette per l'appunto leggi di Keplero, che descrivono il moto e le traiettorie dei pianeti attorno al Sole. Egli intuì che la traiettoria ellittica dei pianeti fosse causata dalla loro interazione con il sole ma non riuscì a spiegare perché questo avvenisse.

Newton riprese il lavoro di Keplero e arrivò ad elaborare la teoria della gravitazione universale. Secondo la leggenda, l'intuizione gli venne da una mela che gli cadde in testa... Non si sa se questo fatto sia realmente accaduto ma sta di fatto che egli riuscì a trovare una relazione tra il moto dei corpi celesti (l'orbita della Luna e dei pianeti) e il moto dei corpi terrestri (la caduta di una mela dall'albero). Nei paragrafi successivi vedremo più in dettaglio le leggi di Keplero e il ragionamento fatto da Newton per arrivare alla sua teoria della gravitazione universale.

2 Le leggi di Keplero

a) Prima legge

I pianeti descrivono intorno al Sole orbite ellittiche di cui il Sole occupa uno dei fuochi.

In figura 1 è mostrato un modo molto semplice per tracciare un'ellisse.

Basta prendere un cartoncino, piantare due spilli e farci passare attorno un pezzo di filo; ruotando una matita all'interno del filo si traccia un'**ellisse**.

Traducendo il disegno appena fatto in termini matematici possiamo dire che l'ellisse è il luogo geometrico dei punti di un piano per i quali la somma delle distanze da due punti fissi, detti fuochi, è costante. Infatti la lunghezza del filo e la distanza fra gli spilli (i fuochi) è sempre la stessa.

Se facciamo coincidere i due spilli la traiettoria descritta dalla matita diventa circolare.

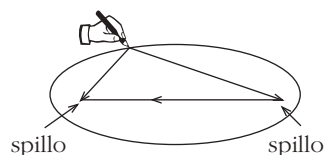


Fig. 1

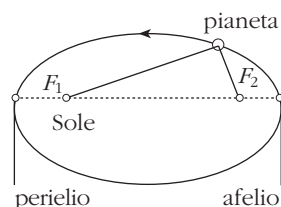


Fig. 2

In figura 2 è mostrata l'orbita ellittica descritta da un pianeta intorno al Sole. Il punto di minima distanza dal Sole è detto *perielio*, quello di massima distanza è detto *afelio*.

Riprendendo la descrizione matematica dell'ellisse, se indichiamo con P un punto generico dell'ellisse e con $2a$ la somma delle distanze di P dai due fuochi F_1 e F_2 , si ha che:

$$\overline{PF_1} + \overline{PF_2} = 2a$$

Consideriamo il punto P in figura 3.

Tra i parametri a , b e c esiste la relazione $a^2 = b^2 + c^2$.

Il segmento:

$\overline{A'A} = 2a$ è l'**asse maggiore** dell'ellisse;

$\overline{A'C} = \overline{CA} = a$ è il **semiasse maggiore** dell'ellisse;

$\overline{BB'} = 2b$ è l'**asse minore** dell'ellisse;

$\overline{B'C} = \overline{CB} = b$ è il **semiasse minore** dell'ellisse;

$\overline{C'C} = 2c$ è la **distanza focale** dell'ellisse.

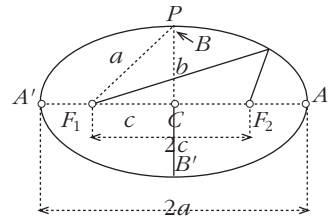


Fig. 3

Si definisce **eccentricità** dell'ellisse il rapporto $e = \frac{c}{a}$.

L'eccentricità misura lo scostamento dell'ellisse dalla forma circolare e il suo valore è compreso fra 0 e 1. Se l'eccentricità è nulla, la traiettoria è circolare. Infatti nella traiettoria circolare i due fuochi coincidono e $c = 0$, da cui segue che $e = 0$.

b) Seconda legge

*Le aree descritte dal raggio vettore dal Sole ai pianeti sono proporzionali ai tempi impiegati a descriverle (**legge delle aree**).*

Questa legge dà un'indicazione sulla velocità con cui i pianeti percorrono le loro orbite attorno al Sole, essa cambia continuamente ed è inversamente proporzionale alla distanza dal Sole. Se però si tracciasse una linea che congiungesse il centro del pianeta con il centro del Sole, essa spazzerebbe aree uguali in tempi uguali.

Per la seconda legge di Keplero, le aree dei triangoli $OP_1P'_1$ e $OP_2P'_2$ sono uguali.

Queste aree sono date da: $\frac{\text{base} \cdot \text{altezza}}{2}$.

Possiamo approssimare la base dei triangoli con la distanza percorsa del pianeta nell'intervallo di tempo Δt . Siccome tale base aumenta al diminuire della distanza Terra-Sole, il pianeta dovrà percorrere il tratto di orbita con velocità maggiore quando si allontana dal Sole e con velocità minore quando si avvicina, in modo da mantenere costante il valore dell'area spazzata.

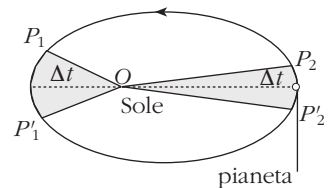


Fig. 4

c) Terza legge

I quadrati dei tempi impiegati dai pianeti a descrivere le proprie orbite sono proporzionali ai cubi dei semiassi maggiori delle ellissi.

Se T_1 e T_2 sono i periodi di rivoluzione di due pianeti che descrivono intorno al Sole ellissi di semiassi maggiori a_1 e a_2 , si ha $\frac{T_1^2}{T_2^2} = \frac{a_1^3}{a_2^3}$.

A controprova di questa legge, nella tabella seguente è stato riportato il periodo, la distanza media della Terra e di Marte dal Sole, e il rapporto:

$$\frac{(\text{periodo})^2}{(\text{distanza})^3}$$

Pianeta	Periodo (s)	Distanza media (m)	$T^2/R^3(s^2/m^3)$
Terra	$3.156 \cdot 10^7s$	$1.4957 \cdot 10^{11}$	$2.977 \cdot 10^{-19}$
Marte	$5.93 \cdot 10^7s$	$2.278 \cdot 10^{11}$	$2.975 \cdot 10^{-19}$

Tab. 1

Possiamo notare che tale rapporto è praticamente costante e lo stesso si potrebbe dimostrare per gli altri pianeti.

Una cosa da non sottovalutare di questa legge è che essa è valida anche per un qualsiasi satellite (Luna compresa) che ruota attorno ad un qualunque pianeta.

3 Da Keplero a Newton

In uno scritto di Newton è stato ritrovato un disegno simile a quello in figura 5 insieme al ragionamento con il quale collegò la caduta dei corpi sulla Terra al moto dei pianeti.

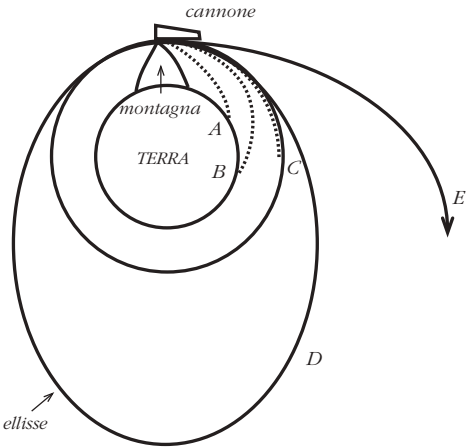


Fig. 5

Se facessimo cadere una bomba da una certa altezza essa cadrebbe verticalmente, proprio come una mela dall'albero; se invece la bomba venisse sparata da un cannone, posto sulla cima di una montagna, con velocità iniziale orizzontale, essa cadrebbe al suolo ad una certa distanza dalla cima della montagna (traiettoria A).

Se la bomba venisse sparata con una velocità maggiore essa cadrebbe ad una distanza maggiore (traiettoria B) e, aumentando la velocità, si arriverebbe ad un punto per cui la traiettoria della bomba coinciderebbe con la curvatura della Terra. In questo caso la bomba inizierebbe a ruotare attorno alla Terra invece che caderci sopra, diventandone un suo satellite (traiettoria C) proprio come la Luna. Aumentando ulteriormente la velocità, la traiettoria della bomba diventerebbe ellittica (traiettoria D e E), come la traiettoria dei Pianeti attorno al Sole, e si arriverebbe ad una velocità per cui la bomba sfuggirebbe al campo gravitazionale allontanandosi tangenzialmente.

Con il ragionamento appena fatto, che è proprio quello che fece Newton, possiamo quindi unificare la caduta di una mela con il moto dei satelliti e affermare che la stessa forza che fa cadere una mela fa ruotare i pianeti attorno al Sole.

Vediamo come Newton affrontò il problema in modo matematico.

Egli approssimò, innanzi tutto, le orbite dei pianeti a orbite circolari. In questo caso le leggi di Keplero diventano:

- **prima legge:** i pianeti descrivono intorno al Sole orbite circolari aventi tutte al centro il Sole;
- **seconda legge:** il moto dei pianeti è uniforme;

— **terza legge:** i quadrati dei tempi impiegati dai pianeti a descrivere le orbite sono proporzionali ai cubi dei raggi delle orbite.

Considero quindi un pianeta di massa m , che ruota con velocità angolare ω e raggio R attorno al Sole.

Il suo periodo è legato alla velocità angolare dalla relazione:

$$T = \frac{2\pi}{\omega}$$

Esso è soggetto a un'accelerazione centripeta pari a:

$$a_c = \omega^2 R$$

e a una forza centripeta:

$$F_{cs} = ma_c = m\omega^2 R = m \frac{4\pi^2}{T^2} R \quad (3.1)$$

diretta verso il Sole.

Dalla terza Legge di Keplero ricavò che:

$$T^2 = KR^3$$

con K costante per tutti i pianeti; quindi riscrisse la (3.1) come:

$$F_{cs} = m_{\text{pianeta}} \frac{4\pi^2}{T^2} R = m_{\text{pianeta}} \frac{4\pi^2}{KR^3} R = m_{\text{pianeta}} \frac{4\pi^2}{KR^2} = c \frac{m_{\text{pianeta}}}{R^2}$$

dove $c = \frac{4\pi^2}{K}$ e quindi scoprì che la forza di attrazione del Sole è direttamente proporzionale alla massa del pianeta che gli gira attorno.

Ma, per il principio di azione e reazione, anche il Sole viene attratto dal pianeta con una forza uguale ed opposta, proporzionale alla massa del Sole: $F_{cp} = c' m_{\text{Sole}}$.

La forza di attrazione reciproca fra il pianeta e il Sole sarà data dal prodotto delle due forze di interazione appena trovate:

$$F = cc' \frac{m_{\text{Sole}} m_{\text{pianeta}}}{R^2} = G \frac{m_{\text{Sole}} m_{\text{pianeta}}}{R^2}$$

in cui G è una costante universale

a) Generalizzazione della legge di gravitazione universale

Successivamente, Newton estese la legge di attrazione tra il Sole e i pianeti a due masse puntiformi qualsiasi m_1 e m_2 , formulando, in tal modo, la **legge di gravitazione universale**:

$$F = G \frac{m_1 m_2}{R^2}$$

in cui m_1 e m_2 sono le masse dei corpi e R la loro distanza.

A conclusione di questo ragionamento matematico dimostriamo che, effettivamente, la forza centripeta esercitata dalla Terra sulla Luna è uguale alla forza gravitazionale Terra-Luna.

A tal fine, dimostriamo che la forza centripeta a cui è soggetta una massa da 1kg posta sulla superficie lunare nel ruotare attorno alla Terra è pari alla forza gravitazionale Terra-Luna.

Supponiamo, con buona approssimazione, che la Luna descriva un'orbita circolare intorno alla Terra. La forza centripeta è data dalla (3.1).

In questo caso: $m_{\text{Luna}} = 1\text{kg}$, $R_{\text{Terra-Luna}} = 384.000\text{km}$ (distanza media della Luna dalla Terra), $T = 27,3$ giorni (tempo impiegato dalla Luna a compiere un giro completo attorno alla Terra).

Sostituendo tali valori nella (3.1) si ottiene:

$$F_{\text{centripeta}} = 1 \cdot \frac{4 \cdot (3,14)^2}{(27,3 \cdot 24 \cdot 3600)^2} (384 \cdot 10^6) = 0,00272N \quad (3.2)$$

$F_{\text{centripeta}}$ è la forza centripeta esercitata dalla Terra su un $1kg$ di Luna. Essa è diretta verso il centro della Terra esattamente come la forza gravitazionale Terra-Luna.

Supponiamo ora che questa forza sia uguale alla forza gravitazionale Terra-Luna cioè che:

$$F_{\text{gravitazionale}} = G \frac{M_{\text{Terra}} m_{\text{Luna}}}{R_{\text{Terra-Luna}}^2} = F_{\text{centripeta}} \quad (3.3)$$

dove:

- M_{Terra} rappresenta la massa della Terra;
- m_{Luna} indica la massa del pezzetto di Luna = $1kg$;
- $R_{\text{Terra-Luna}}$ è la distanza Terra-Luna.

Dalla (3.2) e (3.3) segue che, essendo la massa della Luna pari a $1kg$:

$$F_{\text{gravitazionale}} = G \frac{M_{\text{Terra}}}{R_{\text{Terra-Luna}}^2} = 0,00272N \quad (3.4)$$

La forza gravitazionale con cui la Terra attrae un corpo di massa pari a $1kg$ posto sulla sua superficie è invece data da:

$$F_0 = G \frac{M_{\text{Terra}}}{R_{\text{Terra}}^2} \quad (3.5)$$

Confrontando la (3.4) con la (3.5) si ottiene:

$$\frac{F_0}{F_{\text{gravitazionale}}} = \frac{R_{\text{Terra-Luna}}^2}{R_{\text{Terra}}^2}$$

Da cui segue che:

$$F_0 = F_{\text{gravitazionale}} \frac{R_{\text{Terra-Luna}}^2}{R_{\text{Terra}}^2} = \frac{0,00272 \cdot (384 \cdot 10^6)^2}{(6380 \cdot 10^3)^2} N = 9,8N$$

che è proprio la forza di attrazione gravitazionale con cui la Terra attrae un corpo di massa $1kg$ posto sulla superficie.

Questo significa che la forza centripeta che fa ruotare la Luna è la stessa forza gravitazionale Terra-Luna.

4 Massa inerziale e massa gravitazionale

Nel capitolo sulla *dinamica* abbiamo introdotto il concetto di massa inerziale: essa è la proprietà che hanno i corpi di opporsi ad una variazione del loro stato di quiete o di moto.

Supponiamo di dover spostare un libro su un tavolo.

La forza di gravità agente sul libro viene bilanciata dalla reazione vincolare del tavolo e quindi non entra in gioco: noi faremmo lo stesso sforzo per spostare il libro sia in presenza che in assenza di gravità. Infatti, la forza che applichiamo per spostare il libro serve soltanto per vincere l'inerzia del libro ed è quindi legata unicamente alla sua massa inerziale.

Se invece teniamo il libro sollevato da terra dobbiamo applicare una forza che vada a bilanciare quella di gravità. In questo caso non c'entra l'inerzia del libro ma entra in gioco la proprietà che hanno i corpi materiali di essere attratti da altri oggetti come, in questo caso, il libro dalla Terra. Questa forza di attrazione è definita dalla formula di gravitazione universale

$$F = G \cdot \frac{m_{\text{libro}} m_{\text{Terra}}}{R_{\text{Terra}}^2}, \text{ dove } m_{\text{libro}} \text{ è la massa gravitazionale del libro.}$$

Quindi la massa gravitazionale e la massa inerziale sono due cose ben distinte:

- la massa inerziale rappresenta la proprietà che hanno i corpi di opporsi alla variazione del loro stato di moto o di quiete;
- la massa gravitazionale è la proprietà posseduta dal corpo in virtù della quale nascono le forze gravitazionali.

Sperimentalmente è stato provato che il rapporto fra la massa inerziale e quella gravitazionale di un corpo è costante e non dipende dalla natura e dalle dimensioni del corpo e quindi, pur essendo di natura diversa, esse sono rappresentate da uno stesso numero. Possiamo generalizzare sia la massa inerziale che quella gravitazionale con il termine **massa** così come abbiamo fatto finora.

5 Il campo gravitazionale

In Fisica, il **campo** è una regione dello spazio in cui si verificano dei fenomeni.

Riprendiamo la legge di gravitazione universale considerando come masse la Terra di massa M e una massa m posta in prossimità della superficie terrestre.

La massa m risente della forza di attrazione gravitazionale $F = \frac{GM}{R^2}m$, dove R è la distanza tra la massa m ed il centro della Terra.

Tale forza esiste per il semplice fatto che esiste la Terra. Se m fosse totalmente isolata ai confini dell'Universo essa non risentirebbe di alcuna forza gravitazionale. In altre parole, il fenomeno per cui m risente della forza gravitazionale è dovuto al semplice fatto che m si trova ad una certa distanza dalla superficie terrestre. La Terra infatti genera un campo detto **campo gravitazionale** che ne circonda la superficie. Tutti i corpi che si trovano in questo campo risentono della forza gravitazionale.

La massa m è anche detta **massa esploratrice** e serve per testare l'intensità del campo gravitazionale nel punto in cui essa viene posta.

Si definisce **intensità di campo** o semplicemente **campo** il rapporto: $\vec{g} = \frac{\vec{F}}{m}$. Esso è un vettore avente la direzione ed il verso di $\vec{F}$, ed è quindi diretto da m verso il centro della Terra, e modulo $g = \frac{GM}{R^2}$, dove $\vec{g}$ è l'accelerazione di gravità del corpo.

a) Variazioni di accelerazione dovute alla distanza dal centro della Terra

Consideriamo un corpo di massa m posto sulla superficie terrestre.

La sua accelerazione di gravità è:

$$g_0 = \frac{GM}{R_0^2} \quad (5.1)$$

dove R_0 è il raggio terrestre.

Se portiamo lo stesso corpo ad un'altezza h , la sua accelerazione di gravità sarebbe:

$$g = \frac{GM}{(R_0 + h)^2} \quad (5.2)$$

Confrontando la (5.1) con la (5.2) possiamo scrivere: $g = \frac{g_0 R_0^2}{(R_0 + h)^2}$.

Quindi possiamo notare che g diminuisce con l'aumentare di h .

In figura 6 è rappresentata la variazione dell'accelerazione di gravità con la distanza dal suolo.

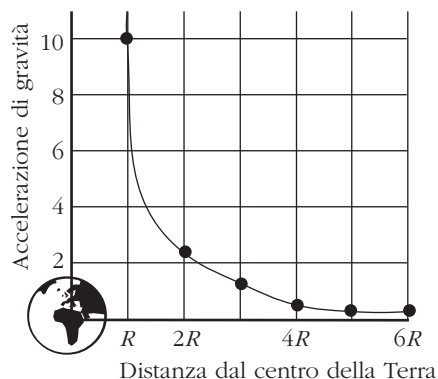


Fig. 6

In particolare, sulla Luna l'accelerazione di gravità è 1/7 di quella terrestre.

6 Lavoro ed energia potenziale gravitazionale

Tutte le volte che abbiamo studiato la caduta di un corpo abbiamo sempre supposto che esso cadesse da punti prossimi alla superficie terrestre e quindi non ci siamo mai dovuti preoccupare di considerare la variazione del peso del corpo con l'altezza.

In realtà, abbiamo visto che la forza gravitazionale, e quindi il peso di un corpo, è inversamente proporzionale al quadrato della distanza dal centro della Terra, infatti $P = G \frac{mM}{R^2}$.

Vediamo ora come la dipendenza del peso di un corpo dalla distanza influisce sull'energia potenziale del corpo.

a) Lavoro della forza gravitazionale

Supponiamo di far cadere un corpo da un'altezza h e calcoliamo il lavoro fatto dalla forza gravitazionale nel passaggio dal punto A al punto B in figura 7.

Possiamo dividere il tratto AB in tanti sottointervalli, all'interno dei quali l'accelerazione di gravità può essere considerata costante.

Consideriamo la caduta dell'oggetto dal punto A al punto P_1 .

Passando dal punto A al punto P_1 , l'energia potenziale gravitazionale del corpo passa dal valore $U_A = mgh$ al valore $U_{P_1} = mgh_1$.

Il lavoro compiuto dal campo gravitazionale è uguale alla differenza di energia potenziale del corpo:

$$\Delta L = U_A - U_{P_1} = mg(h - h_1) \quad (6.1)$$

A questo punto potremmo sostituire a g il valore $g = G \frac{M}{R^2}$, ma il problema è che g varia con la distanza R e quindi dobbiamo definire quale valore di R utilizzare in quanto il valore di R^2 varia da h^2 a h_1^2 .

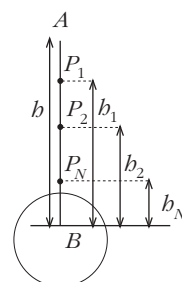


Fig. 7

Possiamo approssimare R^2 con la media geometrica tra h e h_1 cioè $R^2 = hh_1$ e la (6.1) diventa:

$$\Delta L_{A-P_1} = U_A - U_{P_1} = mG \frac{M}{R^2} g(h - h_1) = GmM \frac{(h - h_1)}{hh_1} = GmM \left(\frac{1}{h_1} - \frac{1}{h} \right)$$

In modo analogo si ricava il lavoro fatto dal campo gravitazionale per gli altri spostamenti:

$$\Delta L_{P_1-P_2} = GmM \left(\frac{1}{h_2} - \frac{1}{h_1} \right)$$

$$\Delta L_{P_2-P_3} = GmM \left(\frac{1}{h_3} - \frac{1}{h_2} \right)$$

.....

$$\Delta L_{P_N-B} = GmM \left(\frac{1}{h_N} - \frac{1}{h_B} \right)$$

Il lavoro totale compiuto dal campo gravitazionale è dato dalla somma di tutti i lavori appena trovati:

$$\begin{aligned} L &= \Delta L_{A-P_1} + \Delta L_{P_1-P_2} + \Delta L_{P_2-P_3} + \Delta L_{P_N-B} = \\ &= GmM \left[\left(\frac{1}{h_1} - \frac{1}{h} \right) + \left(\frac{1}{h_2} - \frac{1}{h_1} \right) + \left(\frac{1}{h_3} - \frac{1}{h_2} \right) + \dots + \left(\frac{1}{h_B} - \frac{1}{h_N} \right) \right] \end{aligned}$$

da cui segue che il lavoro fatto dal campo gravitazionale per portare un corpo da un'altezza $h_A = h$ a un'altezza h_B è dato da:

$$L = GmM \left(\frac{1}{h_B} - \frac{1}{h_A} \right) \quad (6.2)$$

b) Energia potenziale gravitazionale

Il lavoro fatto dal campo gravitazionale è uguale alla variazione dell'energia potenziale dell'oggetto, ossia $\Delta L = U_A - U_B$ e, considerando la (6.2), è pari a:

$$\Delta L = U_A - U_B = -GmM \left(\frac{1}{h_A} - \frac{1}{h_B} \right)$$

da cui segue che la grandezza $U = -G \frac{mM}{r}$ rappresenta l'energia potenziale della massa m nel campo gravitazionale terrestre.

Per la definizione data nel capitolo sul *Lavoro* dell'energia potenziale ($U = mgh$), il livello 0 dell'energia potenziale corrispondeva al livello 0 dell'altezza $h = 0 \Rightarrow U = 0$.

In questo caso, invece, l'energia potenziale è nulla all'infinito ($r \rightarrow \infty$).

Il diverso livello assunto per l'energia potenziale non influenza minimamente i calcoli in quanto nei problemi concreti si considera sempre la differenza di potenziale tra due punti.

Variazione di energia potenziale gravitazionale di un corpo che cade da un'altezza h

Indicando con R il raggio terrestre, quando il corpo si trova ad un'altezza h , la sua distanza dal centro della Terra è $R_1 = R + h$ e quindi possiede un'energia potenziale $U_1 = -G \frac{mM}{R+h}$, mentre quando si trova sulla superficie terrestre la sua energia potenziale è $U_2 = -G \frac{mM}{R}$.

La variazione di energia potenziale è quindi:

$$\begin{aligned}\Delta U &= U_1 - U_2 = -G \frac{mM}{R+h} - \left(-G \frac{mM}{R} \right) = -G \frac{mM}{R+h} + G \frac{mM}{R} = GmM \left(\frac{1}{R} - \frac{1}{R+h} \right) = \\ &= GmM \left(\frac{R+h-R}{R(R+h)} \right) = GmM \frac{h}{R(R+h)}\end{aligned}$$

Introducendo l'accelerazione di gravità $g = G \frac{M}{R^2}$, da cui segue che $GM = gR^2$, ed essendo $R+h \approx R$, si ottiene:

$$\Delta U = GmM \frac{h}{R(R+h)} \approx mgR^2 \frac{h}{R^2} = mgh$$

che è la variazione di energia potenziale già ottenuta nel capitolo su *Lavoro ed energia*.

c) Conservazione dell'energia meccanica

Essendo il campo gravitazionale un campo conservativo, vale il teorema di conservazione dell'energia meccanica: $E_1 + U_1 = E_2 + U_2$.

7 Il lancio di un satellite terrestre

Abbiamo già visto che Newton, oltre ad occuparsi di gravitazione universale, dimostrò anche che un qualsiasi grave lanciato con una certa velocità può diventare un satellite terrestre.

In questo paragrafo ci occuperemo di dare una descrizione matematica al fenomeno, oltre ad una terminologia tecnica.

La velocità con cui la bomba di cui al paragrafo 3 diventa un satellite terrestre è detta **prima velocità cosmica**.

La velocità con cui la bomba sfugge al campo gravitazionale è detta **seconda velocità cosmica** o **velocità di fuga**.

Prima e seconda velocità cosmica di un satellite nell'ipotesi che venga lanciato da un'altezza h **Prima velocità cosmica**

Quando il razzo raggiunge la prima velocità cosmica si muove con moto circolare uniforme.

La velocità del razzo deve essere tale che la forza centripeta del razzo deve essere uguale alla forza gravitazionale.

Se il razzo viene lanciato da un'altezza h , la sua orbita ha raggio $R+h$, dove R è il raggio terrestre.

A tale altezza la forza centripeta è data da $F_c = \frac{mv^2}{R+h}$ e la forza gravitazionale è pari a $F_g = G \frac{mM}{(R+h)^2}$.

Uguagliando queste due forze si ottiene:

$$\frac{mv^2}{R+h} = G \frac{mM}{(R+h)^2}$$

da cui segue che:

$$v = \sqrt{\frac{GM}{R+h}}$$

che rappresenta la velocità che un satellite deve avere per diventare un satellite terrestre.

Da notare che tale velocità non dipende dalla massa del satellite ma dal raggio e dalla massa della Terra. Questo vuol dire che la prima velocità cosmica varia a seconda del corpo celeste considerato.

Inoltre essendo $v = \omega(R + h) = \frac{2\pi}{T}(R + h)$ possiamo anche scrivere:

$$\sqrt{\frac{GM}{R + h}} = \frac{2\pi}{T}(R + h)$$

da cui segue che:

$$\frac{T^2}{(R + h)^3} = \frac{4\pi^2}{GM}$$

Per la Terra: $K_T = \frac{4\pi^2}{GM} = 0,99 \cdot 10^{-13} \frac{Kg}{Nm^2}$.

Quindi: $\frac{T^2}{(R + h)^3} = K_T$.

Pertanto, il rapporto tra il quadrato del periodo e il cubo del raggio è una costante come risulta dalla terza legge di Keplero.

Seconda velocità cosmica o di fuga

La velocità di fuga può essere definita più formalmente come la velocità iniziale necessaria per andare da un punto in un campo potenziale gravitazionale fino all'infinito con velocità residua nulla, relativamente al campo stesso. Da ciò segue che, all'infinito, un corpo oltre ad avere energia potenziale nulla, come abbiamo visto in questo paragrafo, ha anche energia cinetica nulla. Quindi la sua energia totale è nulla.

Essendo il campo gravitazionale un campo conservativo, vale il teorema di conservazione dell'energia.

Quindi, essendo nulla l'energia del corpo all'infinito, deve essere nulla l'energia totale del corpo sulla superficie terrestre, ovvero $\frac{1}{2}mv^2 - G\frac{Mm}{R} = 0$, e quindi possiamo ricavare la velocità di fuga: $v = \sqrt{\frac{2GM}{R}}$.



Espansione Web

Cavendish e la costante G

12

La statica dei fluidi

1 Densità e pressione

La materia, dal punto di vista macroscopico, può essere suddivisa in **solida** e **fluida**; i fluidi, dal latino «fluere», ossia «scorrere», sono tutte quelle sostanze presenti in natura allo stato liquido o gassoso.

I fluidi, a differenza dei solidi, non hanno forma propria; infatti, un liquido tende ad espandersi sotto l'influenza del proprio peso, mentre un gas, molto più leggero, tende a occupare tutto lo spazio disponibile. Questo diverso comportamento è determinato dai legami intermolecolari che caratterizzano le varie sostanze, i quali sono molto forti nei solidi, deboli nei liquidi, quasi inesistenti nei gas. Si dice che una porzione di fluido, liquido o gas, è in **quiete** (rispetto a un determinato sistema di riferimento), quando la sua posizione globale non muta al variare del tempo, benché, all'interno della porzione stessa, le molecole continuino a muoversi in ogni senso.

Le variabili che caratterizzano i fluidi in quiete sono principalmente la **densità** e la **pressione**.

a) Densità

La **densità** $\rho = \frac{m}{V}$ è definita dal rapporto fra la massa m del fluido e il volume V occupato da esso.

La densità può dipendere da molti fattori, quali la pressione e la temperatura. Nel caso dei liquidi questa dipendenza è irrilevante mentre per i gas è notevole.

L'equazione dimensionale della densità è $[\rho] = [m \cdot l^{-3}]$ e l'unità di misura è il $\frac{Kg}{m^3}$.

b) Pressione

La **pressione** esercitata da una forza su una superficie S è definita dalla relazione $P = \frac{F_1}{S}$, dove F_1 è la componente della forza perpendicolare alla superficie S .

L'equazione dimensionale della pressione è:

$$[P] = [F \cdot S^{-1}] = [m \cdot a \cdot l^{-2}] = [m \cdot lt^{-2} \cdot l^{-2}] = [m \cdot l^{-1}t^{-2}]$$

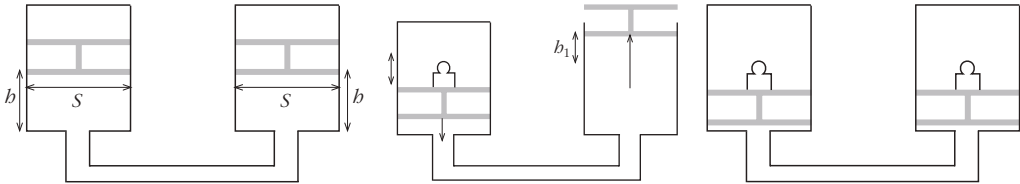
L'unità di misura nel Sistema Internazionale è il **pascal** (Pa), dove $1Pa = \frac{1N}{1m^2}$, nel Sistema C.G.I. è il **baria**, dove $1baria = \frac{1dine}{1cm^2}$.

Vale la trasformazione:

$$1Pa = \frac{1N}{1m^2} = \frac{10^5dine}{10^4cm^2} = 10barie$$

2 La trasmissione delle forze nei fluidi

Consideriamo la figura 1.a in cui due recipienti dotati di pistone contenenti un liquido sono messi in collegamento. I recipienti sono dotati di un pistone e hanno la stessa sezione S .



Figg. 1.a, 1.b e 1.c

In stato di quiete i due pistoni si trovano alla medesima altezza.

Se appoggiamo un peso sul pistone di sinistra, il pistone di destra viene automaticamente sollevato e questo significa che la forza applicata sul pistone di sinistra viene trasmessa al pistone di destra (Fig. 1.b).

Per riportare i due pistoni allo stesso livello è necessario appoggiare su quello di destra lo stesso peso che avevamo posato su quello di sinistra (Fig. 1.c).

L'esperienza dimostra che se il cilindro di destra avesse avuto una superficie doppia del cilindro di sinistra, per ristabilire l'equilibrio avremmo dovuto appoggiare sul pistone di destra un peso doppio di quello appoggiato sul pistone di sinistra (Fig. 2).

Quindi, quello che si trasmette da una parte all'altra del liquido non è la forza applicata ma la pressione.

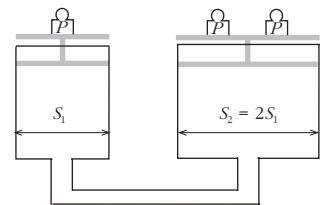


Fig. 2

a) Principio di Pascal

Una pressione esercitata in un punto di una massa fluida si trasmette in ogni altro punto e in tutte le direzioni con la stessa intensità.

Consideriamo due cilindri muniti di stantuffi e comunicanti fra loro (Fig. 3).

Se sullo stantuffo di sinistra di sezione S_1 si applica una forza F_1 , la pressione risultante è $p_1 = \frac{F_1}{S_1}$. Per il principio di Pascal deve essere $p_2 = p_1$ e quindi la forza trasmessa sarà pari a:

$$F_2 = p_2 S_2 = p_1 S_2 = \frac{F_1}{S_1} \cdot S_2 = F_1 \cdot \frac{S_2}{S_1}$$

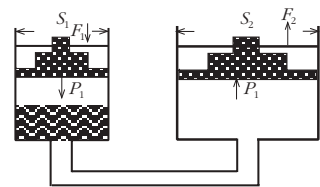


Fig. 3

3 Variazione di pressione in un fluido a riposo e legge di Stevino

Se p_0 è la pressione esercitata sulla superficie di un fluido di densità ρ , la pressione alla profondità h è data da:

$$p = p_0 + \rho gh$$

DIMOSTRAZIONE

Consideriamo una porzione cilindrica del fluido di area S e di spessore dy (Fig. 4).

Indichiamo con dp la differenza di pressione fra la base superiore e quella inferiore del cilindro, quindi, se p è la pressione sulla faccia superiore, su quella inferiore sarà $p + dp$.

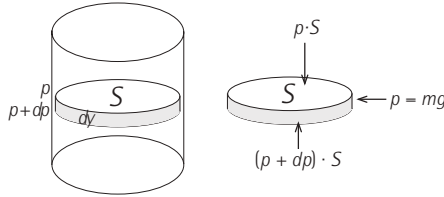


Fig. 4

Essendo il fluido in quiete anche la porzione considerata non è in movimento e quindi la somma delle forze agenti su di essa è media.

Analizziamo le pressioni che si esercitano sulle superfici del cilindro e, siccome una pressione è definita dal rapporto fra la forza applicata *perpendicolarmente* a una superficie e la superficie stessa, vediamo quali sono le forze che agiscono perpendicolarmente alle superfici del cilindro:

- **Superficie laterale**

Le forze che agiscono perpendicolarmente alla superficie laterale sono forze orizzontali, le quali, essendo dovute unicamente alla pressione, hanno tutte lo stesso valore e si annullano per ragioni di simmetria.

- **Basi del cilindro**

Le forze che agiscono perpendicolarmente alle basi del cilindro sono:

- la forza di pressione sulla faccia superiore del cilindro $F_1 = S \cdot p$ diretta verso il basso;
- la forza di pressione sulla faccia inferiore del cilindro $F_2 = S \cdot (p + dp)$ diretta verso l'alto;
- la forza peso del cilindro $F_3 = mg = (\rho V)g = (\rho S dy)g = \rho S g dy$ diretta verso il basso.

Siccome la forza risultante deve essere nulla, si ha che:

$$-F_1 + F_2 - F_3 = 0$$

e quindi:

$$-S \cdot p + S \cdot (p + dp) - \rho S g dy = 0$$

da cui segue che:

$$\frac{dp}{dy} = \rho g \quad (3.1)$$

La (3.1) ci dice come varia la pressione con la profondità.

Per un aumento di profondità di dy si ha un aumento di pressione pari a $dp = \rho g dy$, dove ρg è il peso specifico del fluido e ne rappresenta il peso per unità di volume.

Infatti il peso di un fluido di massa m è dato da:

$$p = mg = \rho Vg \quad (3.2)$$

Se il volume del fluido è unitario ($V = 1$) la (3.2) diventa $p = \rho g$ che è proprio il peso specifico.

Dalla (3.1) si ha che se p_1 è la pressione alla profondità y_1 e p_2 è la pressione alla profondità y_2 , la differenza di pressione è data da:

$$p_2 - p_1 = \rho g (y_2 - y_1) \quad (3.3)$$

Dalla (3.3) segue che se conosciamo la pressione p_0 al livello superficiale di un liquido, la pressione ad una profondità $h = y_2 - y_1$ è data da:

$$p = p_0 + \rho g h \quad (3.4)$$

che è nota come **legge di Stevino**.

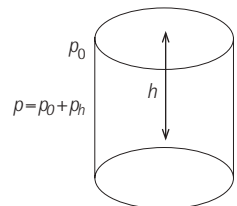


Fig. 5

a) Pressione idrostatica

Possiamo riscrivere la (3.4) come $p = p_0 + p_h$, dove il termine $p_h = \rho gh$ rappresenta la *pressione idrostatica del fluido*.

Esso indica infatti la pressione alla profondità h dovuta soltanto al peso del liquido sovrastante.

Per chiarire ciò consideriamo nuovamente la figura 5.

Se indichiamo con m la massa di liquido contenuta al suo interno e quindi tra la superficie e la profondità h , il peso di tale massa è pari a:

$$P = mg = (\rho V)g = (\rho Sh)g$$

La pressione esercitata da tale liquido è pari a:

$$p_h = \frac{P}{S} = \rho gh$$

che è proprio la **pressione idrostatica** prima definita.

b) Conseguenza della legge di Stevino: il paradosso idrostatico

Consideriamo la figura 6 in cui dell'acqua viene fatta scendere attraverso una cannula in una botte.

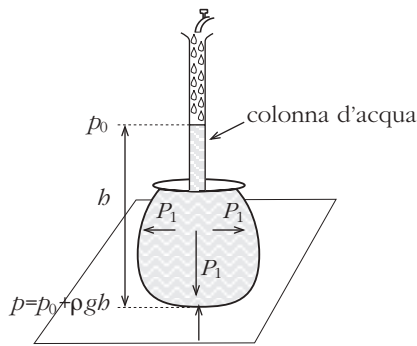


Fig. 6

Per la legge di Stevino la pressione sul fondo della botte è pari a:

$$p = p_0 + p_h = p_0 + \rho gh$$

dove p_h è la pressione idrostatica e dipende *soltanto* dall'altezza della colonna d'acqua contenuta nella cannula.

Quindi anche se la cannula è molto stretta, la pressione nel vaso aumenterà linearmente con l'altezza dell'acqua presente nella cannula.

Per il principio di Pascal, questa pressione si trasmette in ogni altro punto e in **tutte le direzioni** con la stessa intensità.

Quindi, la pressione idrostatica si eserciterà sia sulla base che sulla superficie laterale della botte con la stessa intensità e, di conseguenza, anche la forza esercitata dall'acqua contro la botte, essendo il prodotto della pressione per la superficie, aumenterà proporzionalmente con l'altezza della colonna d'acqua nella canna.

La forza esercitata sulla base della botte viene controbilanciata dalla reazione vincolare del piano su cui la botte è appoggiata, mentre, continuando a versare l'acqua, ad un certo punto la

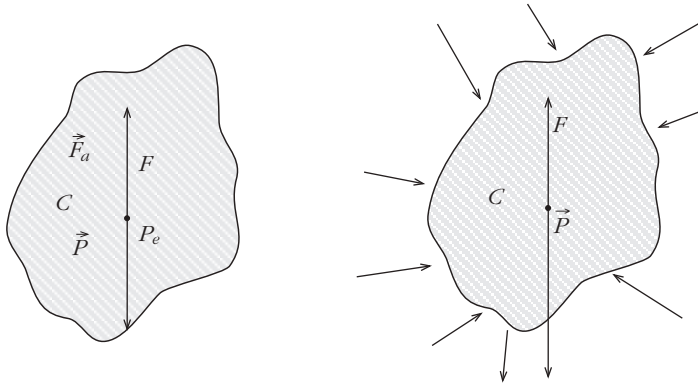
superficie laterale della botte non sarà più in grado di contrastare la forza «premente» dell'acqua e si romperà.

4 Il principio di Archimede

a) Enunciato

Un corpo immerso in un fluido riceve una spinta dal basso verso l'alto pari al peso del volume del liquido spostato.

Consideriamo una porzione C di un fluido in equilibrio in un recipiente (Fig. 7.a).



Figg. 7.a e 7.b

Le forze agenti su C sono:

- la forza peso $\vec{P}$;
- la risultante delle forze di pressione $\vec{F}_a$, detta **forza di Archimede**, esercitata dal restante liquido sulla superficie di C .

Siccome C è in equilibrio, la somma di queste due forze deve essere uguale a zero:

$$\vec{P} + \vec{F}_a = 0$$

da cui segue che:

$$\vec{F}_a = -\vec{P}$$

Pertanto la forza di Archimede deve avere la stessa intensità ma verso opposto alla forza del peso ed è, quindi, rivolta verso l'alto, e di valore pari al peso di C e cioè al peso del volume del liquido in C .

Sostituiamo ora C con un corpo K avente le stesse dimensioni (e quindi lo stesso volume) di C . Anche in questo caso, le forze agenti su K sono la forza peso e la forza di Archimede, ma, mentre il peso di K dipende, ovviamente, da K stesso, la forza di Archimede, essendo esercitata dal liquido esterno, è uguale a quella che si esercitava su C .

Se sostituissimo K con un altro oggetto K' avente le stesse dimensioni, la forza di Archimede continuerebbe ad essere invariata.

FORZE AGENTI SU UN CORPO IMMERSO IN UN FLUIDO

Calcolare la risultante delle forze agenti su un corpo immerso in un fluido nel caso che il corpo abbia la forma di un parallelepipedo.

Per la legge di Stevino, la pressione in un liquido è la stessa alla medesima altezza, quindi sulle facce laterali del parallelepipedo le forze di pressione sono uguali in modulo e si annullano reciprocamente a due a due.

Le due basi, invece, si trovano a profondità diverse e quindi le forze di pressione saranno differenti.

Se p è la pressione a livello della base superiore, per la legge di Stevino, alla base inferiore la pressione è $p + \rho gh$.

Se S è la sezione del solido, la forza agente sulla base superiore è $F_1 = p \cdot S$ ed è diretta verso il basso; la forza agente sulla base inferiore è $F_2 = pS + \rho ghS$. La forza risultante agente sul corpo è quindi data da: $F = F_2 - F_1 = pS + \rho ghS - pS = \rho ghS$. Ma $V = hS$ è il volume del parallelepipedo e $P_s = \rho g$ è il peso specifico del liquido. Quindi $F = P_s V$ ed è proprio il peso del liquido spostato come affermato nella legge di Archimede.

Spesso nei problemi si crea un po' di confusione su questa formula: la forza di Archimede è data dal prodotto tra il **volume V del corpo** immerso e il **peso specifico P_s del fluido** in cui il corpo è immerso.

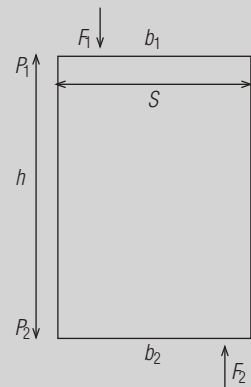


Fig. 8

b) Corpi che affondano e corpi che galleggiano

L'esperienza ci insegna che non tutti i corpi affondano. Vediamo perché un corpo **affonda** e perché un altro **galleggia**.

Abbiamo visto che su un corpo immerso in un liquido agiscono la sua forza peso diretta verso il basso e la spinta di Archimede diretta verso l'alto, quindi:

- (i) se il peso del corpo è maggiore della spinta di Archimede il corpo affonda;
- (ii) se il peso del corpo è uguale alla spinta di Archimede il corpo è in equilibrio in ogni posizione;
- (iii) se il peso del corpo è minore della spinta di Archimede il corpo galleggia.

Indicando con P_s il peso specifico del corpo e con P_L il peso specifico del liquido, il **peso del corpo** è dato da $P = P_s V$ e la **spinta di Archimede** è data da $F = P_L V$.

Quindi dalle (i), (ii) e (iii) deriva che:

- (i*) se $P_s > P_L$ il corpo affonda;
- (ii*) se $P_s = P_L$ il corpo è in equilibrio in ogni posizione;
- (iii*) se $P_s < P_L$ il corpo galleggia.

Per l'equilibrio dei corpi ha anche grande importanza la forma che, se opportunamente sagomata, può spostare un liquido maggiore e ricevere così una superiore spinta verso l'alto. In tali condizioni un corpo può galleggiare anche se il suo peso specifico è maggiore di quello del fluido che lo contiene.

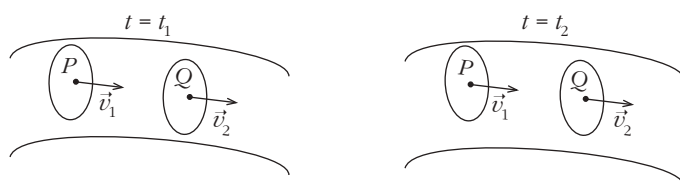
1 Fluidi stazionari e non stazionari

Il moto di un fluido può essere stazionario o non stazionario.

Consideriamo una sezione normale al fluido; il moto di un fluido è **stazionario** quando la velocità di ogni elemento del fluido che passa per quella sezione è costante nel tempo.

La velocità del fluido varia a seconda della sezione che sta attraversando, ma il fluido passa in una stessa sezione con la medesima velocità.

In figura 1 sono state evidenziate le sezioni P e Q nell'istante t_1 e nell'istante t_2 .



Figg. 1.a e 1.b

Tutti gli elementi del fluido che passano per P e Q hanno velocità $\vec{v}_1$ nel punto P e $\vec{v}_2$ nel punto Q .

Un flusso è detto **non stazionario** quando la velocità del fluido passante per una determinata sezione varia nel tempo.

La traiettoria di ogni elemento del fluido è detta **linea di flusso** e risulta in ogni istante tangente alla sua velocità (Fig. 2).

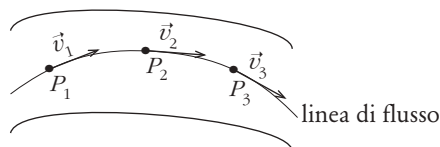


Fig. 2

Due linee di flusso non possono mai intersecarsi in quanto nel punto di intersezione gli elementi del fluido potrebbero seguire l'una o l'altra linea di flusso, variando la propria velocità in modo casuale, e questo è contrario alla definizione di flusso stazionario.

Per ogni punto di un flusso stazionario possiamo tracciare una linea di flusso. In figura 3 sono state evidenziate le linee di flusso che passano per i punti di una linea chiusa.

In questo modo si ottiene una superficie tubolare, detta **tubo di flusso**.

Siccome le linee di flusso non possono intersecarsi, nessun elemento del fluido può attraversare il tubo di flusso e questo vuol dire che, indicando con S_1 e S_2 due sezioni normali alle linee, tutti gli elementi che entrano nel tubo di flusso da S_1 devono uscire da S_2 .

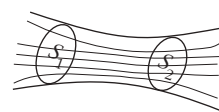


Fig. 3

2 L'equazione di continuità

L'**equazione di continuità** esprime la conservazione della massa nel fluido.

Consideriamo un fluido in moto stazionario e due sezioni S_1 e S_2 normali ad esso, di area A_1 e A_2 . Indichiamo con $\vec{v}_1$ e $\vec{v}_2$ le velocità con cui le particelle attraversano le due sezioni e con ρ_1 e ρ_2 le densità del fluido in prossimità delle sezioni.

La quantità di massa di fluido che attraversa la sezione S_1 nell'intervallo di tempo Δt è data da:

$$\Delta m_1 = \rho_1 \Delta V = \rho_1 A_1 v_1 \Delta t \quad (2.1)$$

Nello stesso intervallo di tempo, la massa che attraversa la superficie S_2 è data da:

$$\Delta m_2 = \rho_2 \Delta V = \rho_2 A_2 v_2 \Delta t \quad (2.2)$$

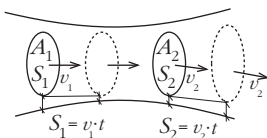


Fig. 4

Siccome il flusso è stazionario non possiamo avere elementi che attraversano il tubo di flusso. Questo vuol dire che tutta la massa che entra dalla superficie S_1 deve uscire dalla superficie S_2 e cioè deve essere $\Delta m_1 = \Delta m_2$, ovvero $\rho_1 A_1 v_1 \Delta t = \rho_2 A_2 v_2 \Delta t$, da cui segue che:

$$\rho_1 A_1 v_1 = \rho_2 A_2 v_2 \quad (2.3)$$

la quale rappresenta l'equazione di continuità.

La (2.3) esprime il principio di conservazione della massa nella dinamica dei fluidi e può anche essere scritta come:

$$\rho A v = \text{cost} \quad (2.4)$$

Se il fluido è incomprimibile, la densità è costante e quindi:

$$A v = \text{cost} \quad (2.5)$$

Av è detto **flusso di volume** o **portata volumetrica** ed è il volume di fluido che attraversa una sezione trasversale nell'unità di tempo. Esso si misura in m^3/s nel S.I. e in cm^3/s nel Sistema c.g.s.

3 L'equazione di Bernoulli

Consideriamo un fluido che si muove di moto stazionario lungo il tubo di flusso disegnato in figura 5.

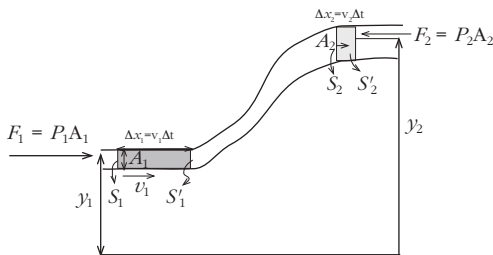


Fig. 5

Supponiamo che il tubo di flusso coincida con le pareti del contenitore in cui si muove il fluido.

Il fluido entra nel tubo da una sezione S_1 ad un'altezza y_1 . S_1 rimane costante per un certo tratto per poi allargarsi ad un valore S_2 e alzarsi ad un'altezza y_2 .

Studiamo il moto della porzione di fluido tratteggiato in figura.

Durante un breve intervallo di tempo Δt , il fluido che attraversa inizialmente la superficie S_1 si è portato fino a S'_1 ad una distanza $\Delta x_1 = v_1 \Delta t$, mentre il fluido in S_2 si è portato fino a S'_2 ad una distanza $\Delta x_2 = v_2 \Delta t$. Dato che il resto del volume tra le due superfici resta invariato, consideriamo soltanto questi due volumi, tratteggiati in figura, che sono uguali per il principio di continuità.

Per il teorema dell'energia cinetica:

il lavoro compiuto dalla forza risultante, agente sul sistema, uguaglia la variazione di energia cinetica del sistema.

Sulla superficie S_1 agisce la forza di pressione $F_1 = p_1 A_1$ del fluido che entra nella sezione ed è diretta verso destra.

Sulla superficie S_2 agisce la forza di pressione $F_2 = p_2 A_2$ del fluido alla destra della sezione ed è diretta verso sinistra.

Su tutto il fluido agisce la forza di gravità, in quanto la porzione di fluido di massa m viene sollevata dalla quota y_1 alla quota y_2 .

Il lavoro compiuto sul sistema dalla forza F_1 è: $L_1 = p_1 A_1 \cdot \Delta x_1$.

Il lavoro compiuto sul sistema dalla forza F_2 è: $L_2 = -p_2 A_2 \cdot \Delta x_2$.

Il lavoro compiuto sul sistema dalla forza F_3 è: $L_3 = -mg(y_2 - y_1)$.

Il lavoro totale compiuto sul sistema è quindi:

$$L = L_1 + L_2 + L_3 = p_1 A_1 \cdot \Delta x_1 - p_2 A_2 \cdot \Delta x_2 - mg(y_2 - y_1) \quad (3.1)$$

Ma:

$$A_1 \cdot \Delta x_1 = A_2 \cdot \Delta x_2 = \Delta V \quad (3.2)$$

è il volume di fluido tratteggiato in figura.

Se la densità del fluido è costante, si ha:

$$\Delta V = \frac{m}{\rho} \quad (3.3)$$

dove m è sempre la porzione di fluido tratteggiata.

Per la (3.2) e la (3.3), la (3.1) diventa:

$$L = p_1 \Delta V - p_2 \Delta V - mg(y_2 - y_1) = p_1 \frac{m}{\rho} - p_2 \frac{m}{\rho} - mg(y_2 - y_1)$$

La variazione dell'energia cinetica del fluido è $\Delta E_c = \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2$.

Per il teorema dell'energia cinetica, $L = \Delta E_c$, quindi:

$$p_1 \frac{m}{\rho} - p_2 \frac{m}{\rho} - mg(y_2 - y_1) = \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2$$

da cui segue, dopo alcuni passaggi matematici:

$$p_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g y_2 \quad (3.4)$$

Poiché sono state prese due sezioni qualsiasi del fluido, la (3.4) può anche essere scritta come:

$$p + \frac{1}{2}\rho v^2 + \rho g y = \text{cost} \quad (3.5)$$

che è l'**equazione di Bernoulli**.

Quest'ultima è applicabile soltanto ai moti stazionari, in quanto le sezioni considerate appartengono ad un tubo di flusso.

Da notare che, in condizioni statiche (idrostatica), l'equazione di Bernoulli continua a valere. Infatti, in questo caso, si ha che $v_1 = v_2 = 0$ e l'equazione di Bernoulli diventa:

$$p_1 + \rho g y_1 = p_2 + \rho g y_2$$

da cui:

$$p_1 - p_2 = -\rho g (y_2 - y_1)$$

che è la **legge di Stevino**.

Il termine $p_1 + \rho g y_1$, che ha le dimensioni di una pressione, è detto **pressione statica**, mentre $p + \frac{1}{2} \rho v^2 + \rho g y$ è detto **pressione dinamica**.

Per condotti orizzontali, la legge di Bernoulli diventa:

$$p_1 + \frac{1}{2} \rho v_1^2 = p_2 + \frac{1}{2} \rho v_2^2$$

e cioè $p + \frac{1}{2} \rho v^2 = \text{cost.}$

In particolare, si può osservare che anche nella legge di Bernoulli troviamo il paradosso idrodinamico. Infatti, dovendo essere la somma $p + \frac{1}{2} \rho v^2$ costante, ad un aumento di velocità deve corrispondere una diminuzione di pressione e viceversa.

4 Il teorema di Torricelli

Un'applicazione interessante del teorema di Bernoulli si ha nel caso di un recipiente contenente del liquido che fuoriesce da un foro praticato in una parete e di superficie ΔS piccola rispetto alla superficie libera del liquido. Alla superficie del liquido, la velocità è praticamente nulla e la pressione è quella atmosferica esterna p_0 ; al di là del foro, la pressione è nuovamente p_0 e l'equazione di Bernoulli (3.5) diventa $p_0 + 0 + y_1 \rho g = p_0 + \frac{1}{2} \rho v_2^2 + y_2 \rho g$, da cui segue che:

$$y_1 \rho g = \frac{1}{2} \rho v_2^2 + y_2 \rho g$$

ossia:

$$v_2 = \sqrt{2g(y_1 - y_2)}$$

la quale, essendo $y_1 - y_2 = h$, diviene:

$$v_2 = \sqrt{2gh}$$

che è l'**equazione di Torricelli**.

Secondo il Teorema di Torricelli:

se un liquido fuoriesce da un piccolo foro praticato in un recipiente, la velocità di efflusso è uguale alla velocità acquistata da un corpo che cade sotto l'azione del suo peso per un dislivello pari a quello esistente tra la superficie libera del liquido e il foro.

1 Onde e moto ondulatorio

Consideriamo una molla in quiete:

- se la comprimiamo o la allunghiamo essa inizia ad oscillare;
- se non la comprimiamo o non la allunghiamo essa non inizia ad oscillare.

Le affermazioni appena fatte definiscono una proprietà fondamentale delle onde: **le oscillazioni e quindi le onde si creano in seguito ad una perturbazione.**

Consideriamo l'onda che si crea dopo che si è lanciato un sasso in uno stagno. Essa è rappresentata da una serie di cerchi concentrici che nascono nel punto in cui il sasso ha toccato la superficie e si propagano con una certa velocità.

Se alcuni galleggianti posti sulla superficie dell'acqua venissero investiti dall'onda inizierebbero a oscillare verticalmente ma non si sposterebbero lungo la direzione di propagazione dell'onda.

Analogamente, se comprimiamo una molla per poi rilasciarla, le sue spire iniziano ad oscillare attorno alla loro posizione di equilibrio trasportando la perturbazione all'altro capo della molla, ma la molla non si sposta.

Questa è un'altra proprietà delle onde: **la propagazione avviene senza trasporto di materia.**

È da sottolineare che l'elasticità di un'onda è dovuta al mezzo in cui essa si propaga e non all'onda in sé.

In altre parole, è il mezzo elastico che rende l'onda elastica.

Un mezzo è elastico quando lo possiamo considerare formato da tante particelle collegate tra loro con delle molle.

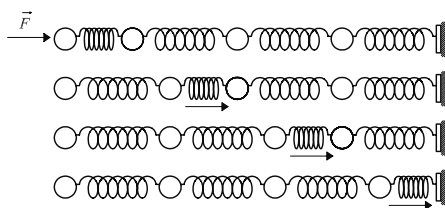


Fig. 1

L'onda che si genera in un mezzo elastico perturbato è un'onda elastica.

Consideriamo il mezzo in figura 1 composto da una sola fila di molle e vediamo cosa accade dal punto di vista energetico quando viene investito da una perturbazione.

Inizialmente il sistema è in quiete.

Supponiamo di comprimere la prima molla. Essa acquista un'energia potenziale pari al lavoro che abbiamo fatto per comprimerla. Quando rilasciamo la molla essa trasmette l'oscillazione alla molla vicina per poi tornare nella posizione di equilibrio.

La prima molla trasferisce la sua energia alla seconda molla la quale, a sua volta, la trasferisce alla terza molla, etc. fino ad arrivare all'ultima molla.

All'interno del mezzo c'è stato un trasporto di energia senza trasporto di materia e questa è un'altra caratteristica delle onde: **in un'onda l'energia si propaga senza trasporto di materia.**

L'affermazione appena fatta definisce completamente un'onda e la distingue da un qualsiasi altro fenomeno.

Si consideri un giocatore di biliardo nell'istante in cui sta per colpire una pallina. Egli trasmette, mediante la stecca, un'energia cinetica alla pallina, la quale si mette in moto trasportando l'energia. In questo caso si ha quindi un trasporto di energia con trasporto di materia: **il fenomeno non è elastico**.

2 Onde trasversali e longitudinali

Le onde possono essere **longitudinali** o **trasversali** a seconda della direzione di propagazione dell'energia trasportata.

a) Onde trasversali

Un'**onda trasversale** è un'onda in cui le particelle del mezzo si muovono **in direzione perpendicolare alla direzione di propagazione dell'onda** (Fig. 2).

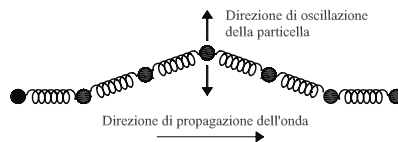


Fig. 2

b) Onde longitudinali

Un'**onda longitudinale** è un'onda in cui le particelle del mezzo si muovono **nella direzione di propagazione dell'onda** (Fig. 3).

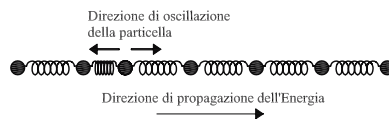


Fig. 3

Le onde che viaggiano in un mezzo solido possono essere sia trasversali che longitudinali mentre in un fluido (liquido o gassoso) si hanno soltanto onde longitudinali.

3 Le grandezze che caratterizzano un'onda

Vediamo ora quali sono le grandezze che caratterizzano un'onda.

a) Creste, gole, ampiezza e lunghezza

• Onda trasversale

Abbiamo visto che cos'è un'**onda trasversale**. Ne possiamo creare facendo oscillare verticalmente una corda avente un estremo fissato a un muro.

La sua forma in un certo istante è mostrata in figura 4.

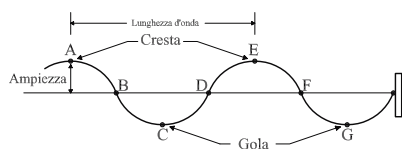


Fig. 4

La linea retta al centro della figura indica la posizione che la corda assume in assenza di disturbi esterni ed è detta **posizione di riposo** o **di equilibrio**.

Quando la corda viene fatta oscillare tutte le sue particelle iniziano a vibrare in direzione verticale:

- i punti di massimo spostamento al di sopra della linea di equilibrio sono detti **creste**;
- i punti di massimo spostamento al di sotto della linea di equilibrio sono detti **gole**.

Per definire un'onda bisogna conoscerne l'ampiezza e la lunghezza d'onda.

L'**ampiezza** rappresenta la massima distanza del mezzo dalla posizione di equilibrio, ossia la distanza tra una cresta o una gola dalla linea di riposo.

La **lunghezza d'onda** è la distanza percorsa dall'onda in un periodo e rappresenta la distanza fra cresta e cresta oppure fra gola e gola.

• Onde longitudinali

Come abbiamo già chiarito, in un'**onda longitudinale** le particelle del mezzo vengono spostate in una direzione **parallela** a quella della direzione di propagazione. Ne consegue che, a differenza delle onde trasversali, la cresta e le onde non possono essere visualizzate.

Supponiamo però di creare un'onda longitudinale comprimendo e rilasciando una molla. Ad un certo istante avremo la situazione raffigurata in figura 5.

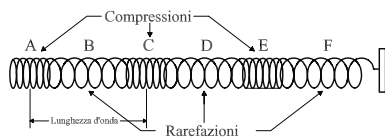


Fig. 5

Le spire della molla oscillano longitudinalmente creando delle regioni in cui esse sono più compresse e delle regioni in cui sono più rarefatte. Esse sono dette, rispettivamente, di **compressione** (i punti A, C, E) e di **rarefazione** (i punti B, D, F).

Quindi:

- in un'onda trasversale si alternano creste e gole;
- in un'onda longitudinale si alternano zone di compressione e di rarefazione.

Di conseguenza:

- in un'onda **trasversale** la lunghezza d'onda è data dalla distanza fra due gole o fra due creste;
- in un'onda **longitudinale** la lunghezza d'onda è data dalla distanza fra due zone di compressione o di rarefazione (cioè fra i punti C e E o fra i punti B e D di figura 5).

b) Frequenza, periodo e velocità

La **frequenza** di un'onda indica il numero di oscillazioni complete fatte in un secondo, mentre il **periodo** indica in quanto tempo un'onda compie un'oscillazione completa.

La frequenza e il periodo sono quindi legate dalla relazione:

$$\text{periodo} = \frac{1}{\text{frequenza}}$$

La **velocità** di un'onda indica la distanza percorsa da un punto (per esempio una cresta o una gola) in un certo periodo di tempo.

c) Energia trasportata e ampiezza

Un'onda trasporta energia attraverso un mezzo senza trasportare materia. L'**energia trasportata** è proporzionale al quadrato dell'ampiezza dell'onda:

$$E \propto A^2$$

*L'**energia** è legata esclusivamente all'ampiezza dell'onda. Essa non dipende dalla velocità, dalla frequenza o dalla lunghezza d'onda.*

d) Equazione d'onda

Per generare un'onda è sufficiente spostare la prima particella di un mezzo elastico dalla posizione di equilibrio.

Il periodo e la frequenza di vibrazione di ogni singola particella del mezzo sono uguali a quelli della sorgente che ha provocato la perturbazione.

La velocità di un'onda è uguale al rapporto tra spazio percorso e tempo impiegato a percorrerlo e poiché la lunghezza d'onda è pari alla distanza percorsa in un periodo, avremo che:

$$\text{velocità} = \frac{\text{spazio}}{\text{tempo}} = \frac{\text{lunghezza d'onda}}{\text{periodo}} \quad (3.1)$$

Essendo il periodo l'inverso della frequenza, la (3.1) può anche essere scritta come:

$$\text{velocità} = \text{frequenza} \cdot \text{lunghezza d'onda} \quad (3.2)$$

Indicando con:

- v la velocità dell'onda;
- λ la lunghezza d'onda;
- f la frequenza dell'onda;
- T il periodo dell'onda;

possiamo scrivere la (3.1) e la (3.2) come:

$$v = \frac{\lambda}{T} = f \cdot \lambda \quad (3.3)$$

La (3.3) rappresenta l'equazione che lega tra loro le grandezze caratteristiche di un'onda e cioè l'**equazione d'onda**.

4 Riflessione, rifrazione e diffrazione delle onde

In questo paragrafo studieremo il comportamento di onde che si propagano in due dimensioni (come le onde del mare) o in tre dimensioni (come le onde sonore o le onde luminose).

A tal fine dobbiamo innanzitutto definire alcuni termini che ci saranno molto utili in seguito.

a) Fronte o superficie d'onda

Il **fronte d'onda** è la linea immaginaria che unisce tutti i punti di un'onda che si trovano nelle stesse condizioni (per esempio, le creste e le gole).

b) Raggi d'onda

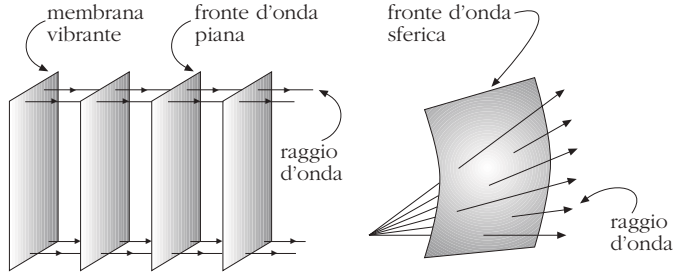
I **raggi d'onda** sono l'insieme delle semirette uscenti dalla sorgente e perpendicolari al fronte d'onda.

A seconda del suo fronte d'onda, un'onda può essere:

- **rettilinea**, se le superfici d'onda sono rette parallele;
- **circolare** (Fig. 6), se le superfici d'onda sono circonferenze concentriche come nel caso delle onde prodotte da un sasso lanciato nell'acqua;
- **piana** (Fig. 7), se le superfici d'onda sono piani fra loro paralleli. Un'onda sferica a grande distanza dalla sorgente può essere considerata piana;
- **sferica** (Fig. 8), se le superfici d'onda sono onde sferiche concentriche.



Fig. 6



Figg. 7 e 8

c) Riflessione

Nella **riflessione** un'onda rimbalza contro una barriera modificando la sua direzione di propagazione.

Consideriamo la figura 9 che schematizza un'onda che si riflette contro una barriera.

Indichiamo con $\hat{i}$ l'angolo di incidenza, ossia l'angolo che il raggio incidente forma con la normale N alla superficie di incidenza, e con $\hat{r}$ l'angolo di riflessione, ossia l'angolo che il raggio riflesso forma con la normale N .

Nella riflessione valgono le seguenti leggi:

- *il raggio incidente, il raggio riflesso e la normale alla superficie di incidenza giacciono nello stesso piano;*
- *l'angolo di incidenza è uguale all'angolo di riflessione, ossia $\hat{i} = \hat{r}$.*

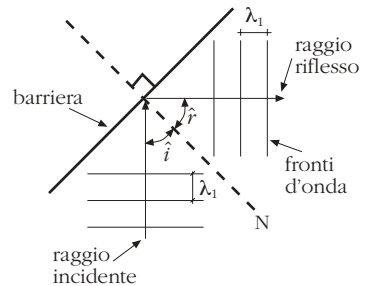


Fig. 9

d) Rifrazione

Nella **rifrazione** l'onda passa da un mezzo di propagazione (M_1) ad un altro (M_2) modificando sia la sua direzione di propagazione sia la sua velocità.

Nella rifrazione valgono le seguenti leggi:

- *il raggio incidente, il raggio rifratto e la normale alla superficie di incidenza giacciono nello stesso piano;*
- *il rapporto tra il seno dell'angolo di incidenza e il seno dell'angolo di rifrazione è costante al variare dell'angolo di incidenza:*

$$\frac{\sin \hat{i}}{\sin \hat{r}} = n_{12}$$

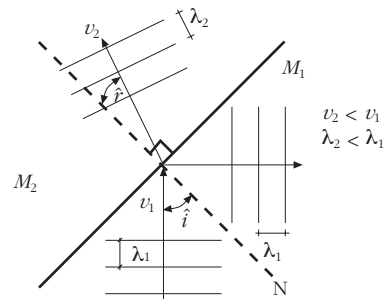


Fig. 10

dove n_{12} è l'**indice di rifrazione** del secondo mezzo rispetto al primo; esso è un valore costante e dipende dalla coppia di mezzi considerati.

Se $n_{12} > 1$ il secondo mezzo è più rifrangente del primo.

Tale rapporto è anche direttamente proporzionale alla velocità dell'onda nei due mezzi:

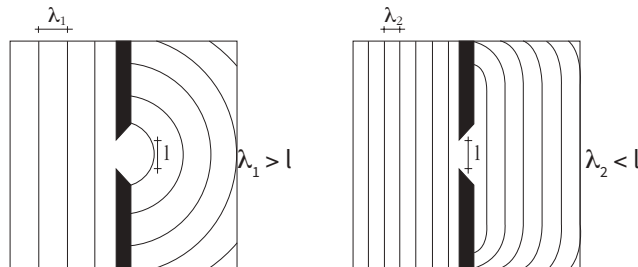
$$\frac{\sin \hat{i}}{\sin \hat{r}} = \frac{v_1}{v_2}$$

e) Diffrazione

Nella **diffrazione** l'onda passa o attraverso una fenditura o attorno a un ostacolo modificando la sua direzione di propagazione.

Se la lunghezza d'onda è maggiore dell'apertura, l'onda si propagerà, al di là dell'apertura, in tutte le direzioni, divenendo un'onda sferica (Fig. 11.a).

Se la lunghezza d'onda è minore dell'apertura, l'onda si propagerà, al di là dell'apertura, in linea retta nella zona centrale mentre si deformerà ai lati (Fig. 11.b).



Figg. 11.a e 11.b

5 Il principio di Huygens

La riflessione, la rifrazione e la diffrazione delle onde possono essere spiegate per via teorica grazie al modello di **Huygens**. Esso si basa sul principio che il fisico olandese Huygens formulò circa tre secoli fa.

Esso afferma che:

ogni punto di un mezzo colpito da un'onda qualsiasi diventa sorgente di onde sferiche.

Questo principio spiega la propagazione di fronti d'onda in un mezzo. Ogni punto colpito da un'onda diventa a sua volta sorgente di onde sferiche le quali si propagano nel mezzo con la stessa velocità, per cui inviluppano un'altra onda piana se i fronti d'onda erano onde piane, oppure un'onda sferica se le onde erano sferiche (con il termine **inviluppo di un treno d'onde** si intende la superficie che è tangente in ogni punto ad un'onda della famiglia).

Nella figura 12 è mostrato l'inviluppo di onde piane e di onde sferiche.

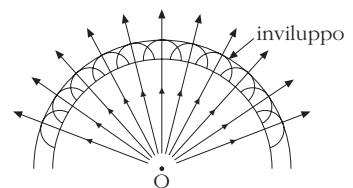


Fig. 12

6 Il principio di sovrapposizione

Vediamo ora cosa accade quando due onde che stanno attraversando lo stesso mezzo si incontrano nella stessa regione di spazio.

La figura 13 mostra le creste di due onde che si incontrano durante la loro propagazione in una corda.

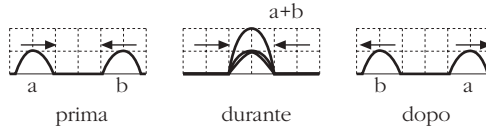


Fig. 13

Come si può osservare dalla sequenza delle immagini, negli istanti in cui le due onde si sovrappongono l'onda risultante (seconda immagine) ha come ampiezza la somma delle ampiezze delle onde incidenti; dopo l'interazione le onde incidenti continuano a mantenere la forma che avevano prima di incontrarsi.

L'effetto della sovrapposizione di due onde è quindi uguale alla somma degli effetti delle due onde incidenti e può essere espresso mediante il **principio di sovrapposizione**, il quale afferma che:

lo spostamento prodotto da più moti ondulatori in un punto in un certo istante è pari alla somma vettoriale degli spostamenti prodotti dalle onde componenti in quel punto e in quell'istante.

Dal punto di vista fisico il principio di sovrapposizione è molto importante in quanto consente di analizzare un'onda molto complicata come sovrapposizione di onde semplici.

In generale, come fu dimostrato dal matematico J. Fourier all'inizio del 1800, una qualsiasi forma d'onda periodica può essere costruita come somma di onde armoniche semplici cioè:

$$y(t) = A_0 + A_1 \sin \omega t + A_2 \sin 2\omega t + A_3 \sin 3\omega t + \dots + B_1 \cos \omega t + B_2 \cos 2\omega t + B_3 \cos 3\omega t + \dots$$

Questa espressione è detta **serie di Fourier**.

7 L'interferenza

Con il termine **interferenza** si indica l'effetto fisico della sovrapposizione di due o più onde.

Non esiste alcuna distinzione fisica fra l'interferenza e la diffrazione delle onde. L'interferenza è la sovrapposizione di poche onde, mentre la diffrazione è la sovrapposizione di molte onde.

Abbiamo visto nel paragrafo precedente che ciascuna onda agisce indipendentemente dalle altre e che l'effetto risultante è dato dalla somma vettoriale degli spostamenti prodotti sulle particelle del mezzo da ogni singola onda. Il fatto che vengano considerati gli spostamenti in termini vettoriali indica che in ogni istante l'ampiezza risultante può essere maggiore o minore delle ampiezze delle onde interagenti e, in casi particolari, potrebbe essere nulla.

Nei casi in cui l'ampiezza risultante assume il valore massimo l'interferenza è detta **costruttiva**, nei casi in cui si annulla l'interferenza è detta **distruttiva**.

8 Le onde stazionarie

Supponiamo di legare ad un palo una corda elastica e di generare un impulso incidente che attraversi tutta la corda e che venga riflesso quando l'onda raggiunge il palo.

Supponiamo di generare un secondo impulso nell'istante esatto in cui il primo raggiunge il palo (Fig. 14).

Lungo la corda si propagano quindi due impulsi aventi la stessa ampiezza e direzione opposta che andranno ad interferire lungo la corda dando origine ad un'onda la cui ampiezza è uguale alla somma delle ampiezze dei due impulsi.

Se gli impulsi creati sono rappresentati da due creste (come in figura) o da due gole, essi interferiranno distruttivamente nel centro della corda.

Nel centro della corda si avrà quindi un **nodo**, cioè un punto in cui la corda non oscilla.

Se sulla corda si invia un treno di impulsi a una determinata frequenza che interferendo creano un'onda limitata nello spazio come quella disegnata in figura 15, l'onda risultante è detta **stazionaria**.

Nella figura sono stati evidenziati i punti in cui non si ha oscillazione, detti **nodi** (i punti B) e i punti in cui l'oscillazione è massima, detti **ventri** (i punti A).

I nodi e i ventri in un'onda stazionaria sono quindi generati dall'interferenza di due onde.

I nodi sono prodotti nei punti in cui si ha un'interferenza distruttiva, i ventri nei punti di interferenza costruttiva.

Calcoliamo ora quali sono le frequenze possibili di un'onda stazionaria.

Supponiamo di avere una corda lunga L e consideriamo alcuni casi particolari rappresentati in figura.

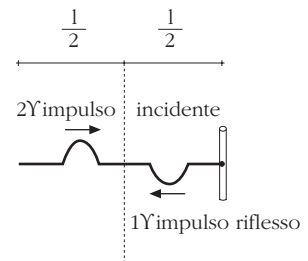


Fig. 14

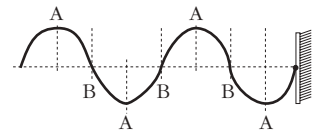


Fig. 15

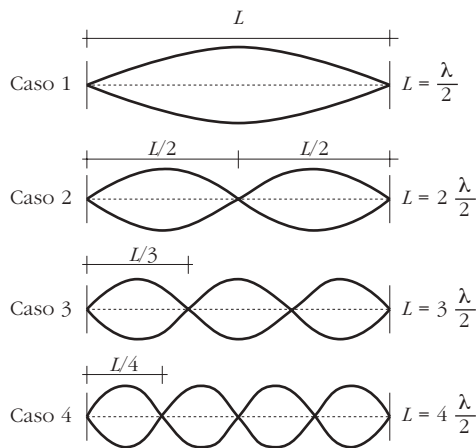


Fig. 16

Caso 1

L'impulso arriva all'estremo fisso della corda dopo **mezza lunghezza d'onda** e, in seguito alla sua riflessione, si forma un'onda riflessa simmetrica alla precedente.

La distanza tra due nodi consecutivi è, in questo caso, uguale a mezza lunghezza d'onda:

$$L = \frac{\lambda_1}{2}$$

da cui segue che:

$$f_1 = \frac{v}{\lambda_1} = \frac{v}{2L}$$

Tutti i punti che si trovano fra due nodi passano contemporaneamente per la posizione di equilibrio raggiungendo la loro massima elongazione, anche se questa è diversa da punto a punto.

Caso 2

L'impulso arriva all'estremo fisso della corda dopo **una lunghezza d'onda**.

In questo caso la lunghezza d'onda è legata alla lunghezza della corda dalla relazione:

$$L = \lambda_2 \rightarrow f_2 = \frac{v}{\lambda_2} = \frac{v}{L}$$

Caso 3

L'impulso arriva all'estremo fisso dopo 3/2 di lunghezza d'onda.

In questo caso, si ha:

$$\lambda_3 = \frac{2}{3}L$$

da cui:

$$L = \frac{3}{2}\lambda_3 \rightarrow f_3 = \frac{v}{\lambda_3} = \frac{3v}{2L}$$

Caso 4

L'impulso arriva all'estremo fisso dopo 2 lunghezze d'onda.

Si ha:

$$\lambda_4 = \frac{L}{2}$$

da cui:

$$L = 2\lambda_4 \rightarrow f_4 = \frac{v}{\lambda_4} = 2\frac{v}{L} = 4\frac{v}{2L}$$

In generale, se n è un numero intero, le possibili frequenze di onde stazionarie su una corda lunga L sono date dalla formula:

$$f_n = \frac{v}{\lambda_n} = n\frac{v}{2L}$$

La più piccola frequenza, corrispondente a $n=1$, e cioè:

$$f_1 = \frac{v}{2L}$$

è detta **frequenza fondamentale** o **prima armonica**, mentre le frequenze successive sono dette **seconda, terza, quarta armonica** etc.

1 Le onde sonore

a) Suono come onda elastica

Nel capitolo precedente abbiamo studiato le onde elastiche e abbiamo dedotto le caratteristiche essenziali che un'onda deve possedere per essere considerata tale e cioè:

- deve propagarsi in **un mezzo**;
- la propagazione deve avvenire **solo** con **trasporto di energia** e non di materia.

Il suono possiede questa proprietà, per cui può essere considerato un'onda elastica.

b) Suono come onda longitudinale

Nel capitolo precedente le onde sono state suddivise in due categorie:

- le **onde longitudinali**;
- le **onde trasversali**.

Le onde sonore sono longitudinali perché le particelle del mezzo attraverso il quale il suono viene trasportato vibrano parallelamente alla direzione di propagazione dell'onda sonora.

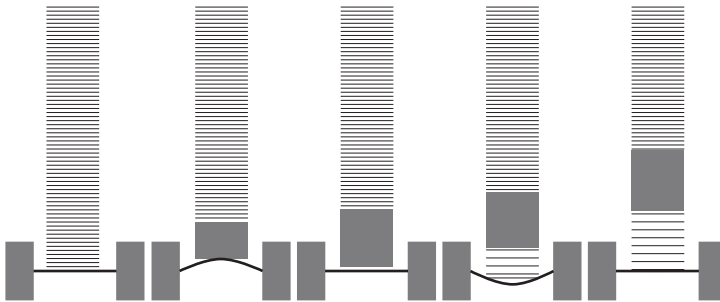


Fig. 1

Nella figura 1 è schematizzato come si crea un'onda dalla vibrazione di una corda.

Quando la corda oscilla essa sposta le molecole d'aria vicine verso l'alto comprimendo la porzione d'aria che si trova in alto rispetto alla corda. Quando la corda oscilla in direzione opposta la pressione diminuisce e le particelle compresse si espandono nella regione lasciata libera dalla corda.

Le molecole d'aria vicine alla corda iniziano quindi a vibrare nella direzione di oscillazione della corda trasmettendo lo stesso moto alle particelle adiacenti.

Le onde sonore generate sono quindi **longitudinali**.

Nel capitolo precedente abbiamo visto che le onde longitudinali creano nel mezzo onde di **compressione** e di **rarefazione**.

La figura 2 mostra le zone di compressione e di rarefazione in un tubo dovute a un'onda sonora generata da un diapason.

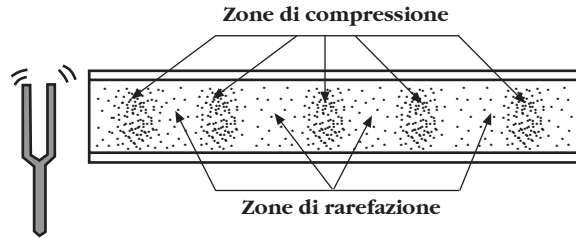


Fig. 2

c) Suono come onda di pressione

La **lunghezza d'onda** è la distanza percorsa dalla perturbazione in un periodo.

Per le onde trasversali la lunghezza d'onda si misura da cresta a cresta o da gola a gola. Per le onde longitudinali, non essendoci né creste né gole ma zone di compressione e di rarefazione, la lunghezza d'onda è misurata tra due zone di compressione o di rarefazione adiacenti.

Poiché la pressione delle particelle nel mezzo passa da un valore minimo nelle zone di rarefazione a un massimo nelle zone di compressione, possiamo tracciare una funzione sinusoidale, detta **onda di pressione**, che descriva la variazione di pressione nel tempo. Le creste della curva corrispondono alle zone di compressione, gli zeri a quelle di equilibrio e i ventri a quelle di rarefazione (Fig. 3).

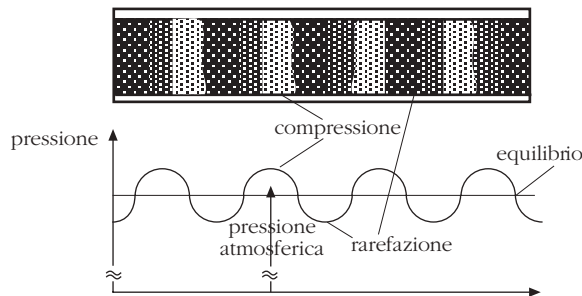


Fig. 3

2 La velocità di propagazione delle onde sonore

Vediamo ora qual è, e da cosa dipende, la velocità con cui le onde sonore si propagano. Riprendiamo il tubo in cui viaggiano le onde sonore delle figure 2 e 3.

Abbiamo già detto più volte che un impulso sonoro viaggia in un mezzo senza trasporto di materia.

Supponiamo di metterci in un sistema di riferimento in cui l'impulso stia fermo mentre il fluido all'interno del tubo si muove e seguiamo il moto di una porzione di fluido che si sposta da sinistra verso destra con velocità v (Fig. 4). Tale elemento di fluido passerà da una zona

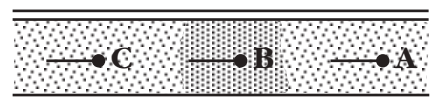


Fig. 4

di rarefazione in A (con pressione p) a una zona di compressione in B (con pressione $p + \Delta p$) e, successivamente, a una zona di rarefazione in C (con pressione p).

Il fluido nel punto A viaggia con velocità v .

Quando arriva nella zona di compressione esso rallenta a causa dell'aumento di pressione e quindi nella zona di compressione avrà velocità $v - \Delta v$.

Una volta superata la zona di compressione, riacquista la stessa velocità v che aveva in A per la diminuzione di pressione dalla zona di compressione alla zona di rarefazione.

Per vedere come varia la velocità dell'elemento di fluido applichiamo la legge fondamentale della dinamica:

$$F = ma \quad (2.1)$$

al suo passaggio dalla zona di rarefazione alla zona di compressione, cioè dal punto A al punto B.

Calcoliamo innanzitutto la forza esercitata sull'elemento di fluido in seguito al passaggio da A a B.

Tale forza è dovuta esclusivamente alla differenza di pressione fra la zona di rarefazione e quella di compressione.

Se S è la sezione del tubo (Fig. 5), la forza che agisce sull'elemento di fluido nel passaggio da A a B è:

$$F = (p_B - p_A)S = [(p + \Delta p) - p]S = \Delta p S \quad (2.2)$$

diretta verso destra.

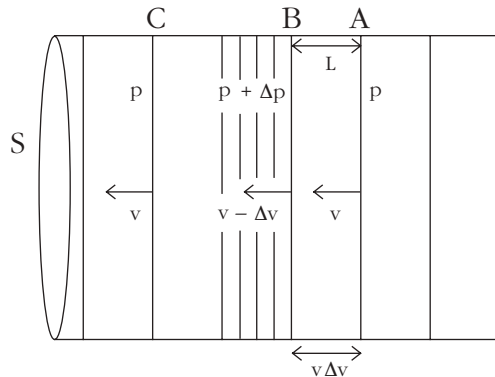


Fig. 5

Se Δt è il generico intervallo di tempo in cui un elemento di fluido di densità ρ attraversa un tratto del tubo, la massa di fluido corrispondente è data da:

$$m = V\rho = S \cdot l \cdot \rho = S \cdot (v\Delta t) \cdot \rho = S\rho v\Delta t \quad (2.3)$$

La sua decelerazione nel passaggio dalla zona di rarefazione a quella di compressione è pari a:

$$a = \frac{v_B - v_A}{\Delta t} = \frac{(v - \Delta v) - v}{\Delta t} = -\frac{\Delta v}{\Delta t} \quad (2.4)$$

Dalla (2.1), (2.2), (2.3), (2.4) si ha che:

$$\Delta p S = (S\rho v\Delta t) \left(-\frac{\Delta v}{\Delta t} \right)$$

da cui segue:

$$\Delta p = -\rho v \Delta v$$

che può essere scritto come:

$$\rho v^2 = \frac{-\Delta p}{\Delta v/v} \quad (2.5)$$

La generica porzione di flusso nella zona di rarefazione ha lunghezza:

$$l_A = v\Delta t$$

e volume:

$$V_A = Sl_A = Sv\Delta t$$

mentre nella zona di compressione ha lunghezza:

$$l_B = (v - \Delta v)\Delta t$$

e volume:

$$V_B = Sl_B = S(v - \Delta v)\Delta t$$

Quindi, nel passaggio dalla zona di rarefazione a quella di compressione, si ha una diminuzione di volume pari a:

$$\Delta V = S(l_A - l_B) = S[v\Delta t - (v - \Delta v)\Delta t] = S\Delta v\Delta t$$

e rispetto al volume iniziale:

$$\frac{\Delta V}{V} = \frac{S\Delta v\Delta t}{Sv\Delta t}$$

da cui segue che:

$$\frac{\Delta v}{v} = \frac{\Delta V}{V} \quad (2.6)$$

Sostituendo la (2.6) nella (2.5) si ottiene:

$$\rho v^2 = \frac{-\Delta p}{\Delta V/V} \quad (2.7)$$

La variazione relativa di volume $\left(-\frac{\Delta V}{V}\right)$ in seguito a una variazione di pressione Δp è detta **modulo di compressione B** del corpo:

$$B = \frac{-\Delta p}{\Delta V/V}$$

Quindi la (2.7) può essere scritta come:

$$\rho v^2 = B$$

da cui segue che:

$$v = \sqrt{\frac{B}{\rho}}$$

Da notare che v dipende sia dalle proprietà elastiche del corpo (B) che dalle proprietà intrinseche del mezzo (ρ).

Infatti, un'onda viaggia più velocemente in un mezzo meno denso, in quanto la densità aumenta l'inerzia del corpo, e in un mezzo più elastico in quanto l'elasticità di un mezzo è legata alla sua tendenza a conservare la propria forma una volta passata la perturbazione.

La velocità di un'onda si può anche ricavare dalla lunghezza d'onda e dalla frequenza secondo la formula:

$$\text{velocità} = \text{lunghezza d'onda} \cdot \text{frequenza}$$

Tale relazione è molto utile per risolvere i problemi matematici relativi alla lunghezza d'onda, alla frequenza e alla velocità. Si ricorda che la velocità di un'onda non dipende dalla lunghezza

d'onda o dalla frequenza. Infatti, variando la lunghezza d'onda, varia anche la frequenza in modo inverso e la velocità dell'onda non cambia.

Essa dipende esclusivamente dalle proprietà di elasticità e di inerzia del mezzo. Solo variando queste proprietà si modifica la velocità di un'onda.

a) Velocità del suono nei gas

Se il mezzo in cui il suono si propaga è un gas, per esempio l'aria, è possibile esprimere B in funzione della pressione p_0 di equilibrio del gas:

$$v_{\text{gas}} = \sqrt{\frac{\gamma p_0}{\rho}} = \sqrt{\frac{\gamma RT}{M}}$$

dove:

- $\gamma = \frac{c_p}{c_v}$ è il rapporto tra i calori specifici del gas;
- M è la massa molecolare del gas;
- T è la temperatura assoluta;
- R è la costante dei gas.

b) Velocità del suono nei solidi

Se il mezzo in cui il suono si propaga è un solido, per esempio una sbarra sottile, B deve essere rimpiazzato dal **modulo di Young**:

$$v_{\text{solidi}} = \sqrt{\frac{\text{modulo di Young}}{\text{densità}}}$$

c) Velocità del suono nei liquidi

Se il mezzo in cui il suono si propaga è un liquido, per esempio l'acqua, B deve essere rimpiazzato dal **modulo di elasticità cubica**:

$$v_{\text{liquidi}} = \sqrt{\frac{\text{modulo di elasticità cubica}}{\text{densità}}}$$

3 Le proprietà del suono

a) Frequenza

La **frequenza** di un'onda sonora è data dal numero di vibrazioni longitudinali che le particelle compiono in un secondo quando sono investite dalla perturbazione.

Abbiamo visto che l'onda sonora è un'onda **di pressione**, cioè un'onda in cui si ha un'oscillazione di pressione con la stessa frequenza del suono prodotto. Quindi per rilevare un'onda sonora si può anche utilizzare un sensore in grado di rilevare variazioni di pressione (come in Fig. 6). Se il sensore rileva 700 compressioni in un secondo, vuol dire che l'onda ha una frequenza di 700Hz.

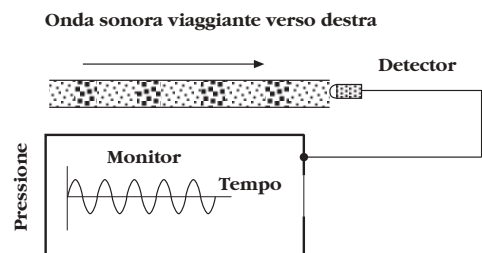


Fig. 6

b) Tono

La sensazione che abbiamo quando captiamo un suono è detto **tono**.

A un tono alto corrisponde un'alta frequenza e a un tono basso corrisponde una bassa frequenza.

c) Intensità

Un'onda sonora trasporta l'energia che è stata fornita dalla sorgente della perturbazione.

L'energia di un'onda è proporzionale al quadrato dell'ampiezza cioè:

$$E \propto A^2$$

In figura 7 sono mostrate, in due dimensioni, le superfici di onda sferica di una sorgente puntiforme.

Mentre l'onda avanza l'energia si distribuisce su superfici sferiche sempre maggiori, la cui area è proporzionale al quadrato della distanza dalla sorgente.

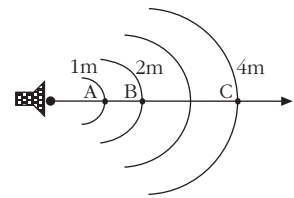


Fig. 7

Per tenere conto di questo fatto è stata introdotta una nuova grandezza: l'**intensità dell'onda**. Essa è definita come la quantità di energia che, nell'unità di tempo, attraversa l'unità di superficie normale al raggio di propagazione e posta a una distanza r dalla sorgente.

Ricordando che la potenza è definita come

$$\text{Potenza} = \frac{\text{Energia}}{\text{Tempo}}$$

l'intensità è data dal rapporto:

$$\text{Intensità} = \frac{\text{Energia}}{\text{Area} \times \text{Tempo}} = \frac{1}{\text{Area}} \cdot \frac{\text{Energia}}{\text{Tempo}} = \frac{1}{\text{Area}} \cdot \text{Potenza} = \frac{\text{Potenza}}{\text{Area}} = \frac{W}{4\pi r^2} \quad (3.1)$$

e la sua unità di misura è il W/m^2 .

Essa viene anche detta **flusso di energia** e si indica con la lettera Φ .

Dalla (3.1) e dalla figura 7 possiamo dedurre che:

$$I_A = \frac{W}{4\pi r_A^2} \quad \text{e} \quad I_B = \frac{W}{4\pi r_B^2}$$

da cui segue che:

$$\frac{I_A}{I_B} = \frac{r_B^2}{r_A^2}$$

Per cui l'**intensità di un'onda** è inversamente proporzionale al quadrato della distanza dalla sorgente.

L'intensità sonora più debole che l'orecchio umano è capace di percepire si chiama **soglia di udibilità**, e vale $1 \cdot 10^{-12} W/m^2$, mentre il suono più forte che l'orecchio può sopportare senza provare dolore ha intensità pari a $1 \cdot 10^{-2} W/m^2$ ed è anche detta **soglia del dolore**.

Essendo il campo di udibilità così ampio, si è pensato di introdurre una scala logaritmica in base 10 per misurare l'intensità del suono.

La soglia di udibilità (pari a $1 \cdot 10^{-12} \text{ W/m}^2$) corrisponde al **livello 0** o **fondamentale**; un suono 10 volte più intenso ($1 \cdot 10^{-11} \text{ W/m}^2$) corrisponde al livello 10; un suono 100 volte più intenso ($1 \cdot 10^{-10} \text{ W/m}^2$) corrisponde al livello 20; un suono 1000 volte più intenso ($1 \cdot 10^{-9} \text{ W/m}^2$) corrisponde al livello 30 e così via.

In generale la corrispondenza tra il valore s del livello assegnato e l'intensità del suono è data dalla relazione: $s = \log_{10} I$.

Quindi la differenza fra il livello fondamentale e un qualsiasi altro livello è pari a:

$$s - s_0 = \log_{10} I - \log_{10} I_0 = \log_{10} \frac{I}{I_0}$$

Introduciamo ora un'unità di misura relativa, denominata **bel** (simbolo B), mediante la quale è possibile paragonare un qualsiasi livello di intensità con quello fondamentale. La misura espressa in *bel* diventa, in questo caso, pari a:

$$\left(\frac{I}{I_0} \right)_B = \log_{10} \frac{I}{I_0}$$

Il valore di $1B$ corrisponde a un suono la cui intensità è 10 volte maggiore della soglia di udibilità.

Normalmente si preferisce utilizzare il **decibel** (dB), pari a $1/10$ di bel. Una misura espressa in *decibel*, nel nostro caso, è pari a:

$$\left(\frac{I}{I_0} \right)_{dB} = 10 \log_{10} \left(\frac{I}{I_0} \right)$$

Nella tabella sottostante è data una lista degli ordini di grandezza dei suoni più comuni.

Tipo di suono	Intensità	Livello di Intensità	Numero di volte maggiore della soglia di udibilità
Soglia di udibilità	$1 \cdot 10^{-12} \text{ W/m}^2$	$0dB$	10^0
Fruscio delle foglie	$1 \cdot 10^{-11} \text{ W/m}^2$	$10dB$	10^1
Bisbiglio	$1 \cdot 10^{-10} \text{ W/m}^2$	$20dB$	10^2
Conversazione ad alta voce	$1 \cdot 10^{-6} \text{ W/m}^2$	$60dB$	10^6
Strada rumorosa	$1 \cdot 10^{-5} \text{ W/m}^2$	$70dB$	10^7
Aspirapolvere	$1 \cdot 10^{-4} \text{ W/m}^2$	$80dB$	10^8
Orchestra	$6,3 \cdot 10^{-3} \text{ W/m}^2$	$98dB$	$10^{9,8}$
Walkman al massimo volume	$1 \cdot 10^{-2} \text{ W/m}^2$	$100dB$	10^{10}
Prima fila di un concerto rock	$1 \cdot 10^{-1} \text{ W/m}^2$	$110dB$	10^{11}
Soglia di dolore (notevole fastidio)	$1 \cdot 10^1 \text{ W/m}^2$	$130dB$	10^{13}
Decollo di un Jet militare	$1 \cdot 10^2 \text{ W/m}^2$	$140dB$	10^{14}
Perforazione immediata del timpano dell'orecchio	$1 \cdot 10^4 \text{ W/m}^2$	$160dB$	10^{16}

Tab. 1

d) Rimbombo ed eco

L'**eco** è un fenomeno a cui tutti noi siamo abituati ed è dovuto alla riflessione delle onde sonore da parte di un ostacolo.

Se lanciamo un grido di fronte a una montagna, le onde sonore si infrangono contro la parete e vengono riflesse. Perché il nostro orecchio senta distintamente sia la parola gridata da noi sia quella riflessa, il tempo impiegato dal suono riflesso per raggiungere le nostre orecchie deve essere almeno un decimo di secondo. In questo lasso di tempo il suono percorre circa 34 metri.

Quindi, perché si verifichi l'eco, l'ostacolo deve trovarsi almeno a diciassette metri di distanza da noi (Fig. 8). Se la distanza è minore si udirà solo un fastidioso **rimbombo**.

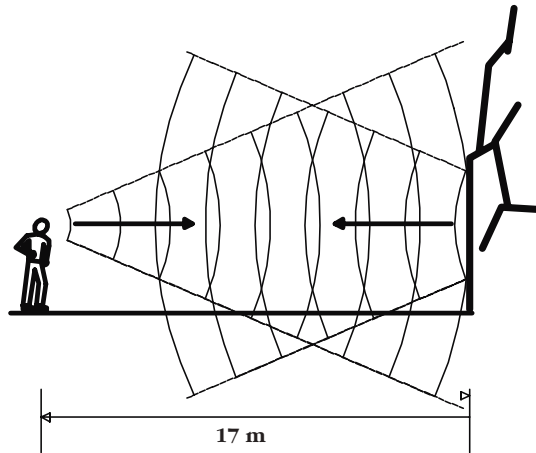


Fig. 8

4 L'interferenza

Abbiamo già visto che l'**interferenza** è il fenomeno che si verifica quando due onde che viaggiano nello stesso mezzo si incontrano. Esse interferiscono costruttivamente quando c'è la sovrapposizione di due creste o di due gole e distruttivamente quando una gola si sovrappone a una cresta e viceversa (Fig. 9).

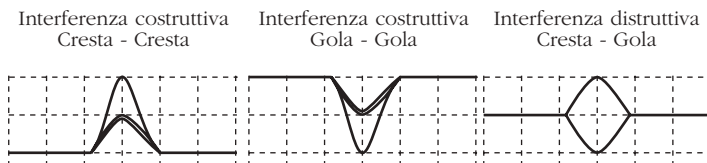


Fig. 9

Ma come possono interferire costruttivamente o distruttivamente due onde sonore non avendo né creste né gole?

Abbiamo visto che il suono può essere considerato un'onda di pressione con delle zone **comprese** e delle zone **rarefatte**.

Per il principio di sovrapposizione il risultato dell'azione di due onde sonore sulle particelle del mezzo è uguale alla somma delle azioni di ogni singola onda sonora.

Se ad esempio una regione di compressione di un'onda incontra una regione di compressione di un'altra onda, le particelle del mezzo subiranno una compressione doppia e questa è una forma di **interferenza costruttiva**.

Così come se due regioni di rarefazione si incontrano, le particelle del mezzo subiranno una rarefazione doppia e anche questa è una forma di interferenza costruttiva.

Se due onde sonore si sovrappongono in una regione dello spazio in modo che una zona di rarefazione della prima onda incontra una zona di compressione della seconda onda il risultato è che le particelle del mezzo in quella regione non si spostano e non viene sentito alcun suono.

5 I battimenti

Il fenomeno dei **battimenti** si verifica quando due onde che si propagano nella stessa direzione ma con frequenze leggermente diverse interferiscono.

Tale fenomeno si verifica, per esempio, quando due tasti adiacenti di un pianoforte sono battuti simultaneamente.

La figura 10 mostra il risultato dell'interferenza fra due onde che si propagano con una frequenza lievemente diversa.

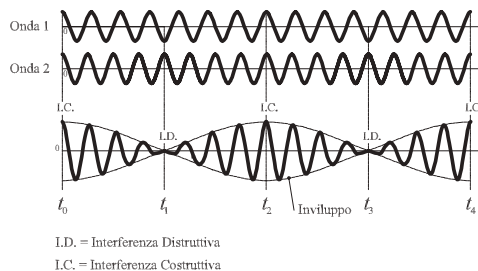


Fig. 10

Le due onde ad un certo istante t_0 sono in fase poi, a causa della loro differenza di frequenza, iniziano a sfasarsi gradualmente fino ad arrivare ad un istante t_1 in cui sono sfasate di 180° (opposizione di fase) e interferiscono distruttivamente e, se le due onde hanno la stessa ampiezza, si ha un'elisione totale. Le due onde continuano a sfasarsi fino ad arrivare all'istante t_2 in cui sono nuovamente in fase e quindi interferiscono costruttivamente. Nell'istante t_3 interferiranno di nuovo distruttivamente e così via.

L'onda risultante avrà una ampiezza che oscilla in modo regolare passando da un valore minimo corrispondente all'interferenza distruttiva delle due onde a un valore massimo negli istanti in cui le due onde interferiscono costruttivamente.

a) Frequenza di battimento

La frequenza dei massimi di inviluppo dell'onda dei battimenti è la **frequenza di battimento**.

Più vicine sono le frequenze delle due onde e maggiore è il tempo che esse impiegano per sfasarsi e per tornare in fase e quindi minore è la frequenza di battimento. Si dimostra che la frequenza di battimento non è altro che la differenza delle due frequenze:

$$f_{\text{batt}} = f_2 - f_1 = \Delta f$$

Quando si ha un'interferenza costruttiva il suono ascoltato è maggiore, quando si ha un'interferenza distruttiva non si ascolta alcun suono. Poiché la sonorità di un'onda è legata all'ampiezza dell'onda, il suono varia con continuità da zero al valore massimo per poi annullarsi di nuovo.

6 L'effetto Doppler

L'**effetto Doppler** è il fenomeno che si verifica quando la sorgente di onde sonore è in movimento rispetto all'ascoltatore e viceversa.

In questo caso la frequenza del suono prodotto dalla sorgente è diverso dalla frequenza percepita dall'ascoltatore.

Per rendere più chiare le idee consideriamo la figura 11 in cui è schematizzata una sorgente sonora ferma rispetto a un ascoltatore.

La frequenza con cui esse arrivano all'orecchio è la stessa con cui vengono prodotte dalla sorgente.

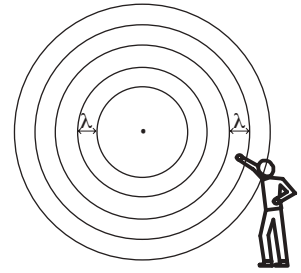


Fig. 11

Nella figura 12 è mostrato cosa accade se la sorgente sonora si avvicina all'ascoltatore.

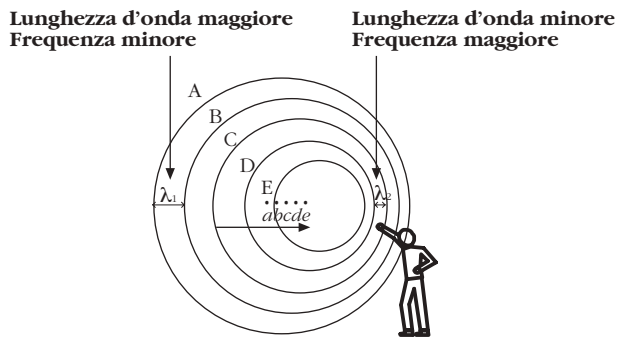


Fig. 12

Poiché la sorgente si sta avvicinando all'orecchio, lo spazio tra due fronti d'onda successivi, ossia la lunghezza d'onda, è minore rispetto a quando la sorgente era ferma. Questo significa che le onde di pressione arrivano all'orecchio con una frequenza maggiore di quella con cui il suono viene prodotto dalla sorgente.

Se, invece, la sorgente si allontana dall'orecchio (Fig. 13), lo spazio tra due fronti d'onda (lunghezza d'onda) successivi è maggiore rispetto a quando la sorgente era ferma.

Questo significa che le onde di pressione arrivano all'orecchio con una frequenza minore di quella prodotta dalla sorgente.

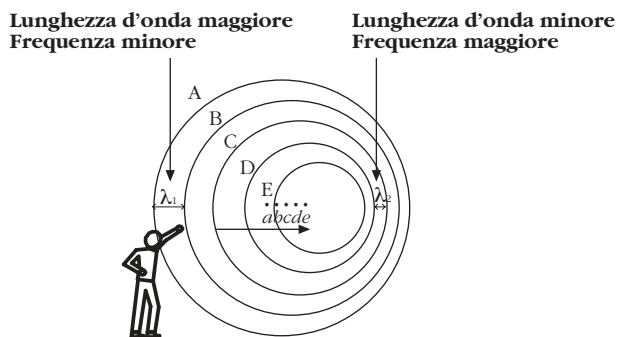


Fig. 13

7 Gli effetti supersonici

Se la sorgente sonora si muove alla stessa velocità del suono essa si trova sempre sul profilo più avanzato dell'onda sonora da essa stessa prodotta.

Se un aereo viaggia ad una velocità V_a pari a quella del suono, la sua posizione rispetto a quella delle onde sonore da esso prodotte è quella tracciata nella figura 14.

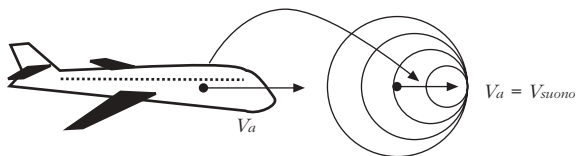


Fig. 14

Possiamo notare che i fronti d'onda sono sovrapposti sulla punta dell'aereo. Questo fenomeno è detto di **onda d'urto**.

Se l'aeroplano viaggia a una velocità maggiore di quella del suono, vuol dire che la sorgente supera le onde sonore da essa prodotte (Fig. 15).

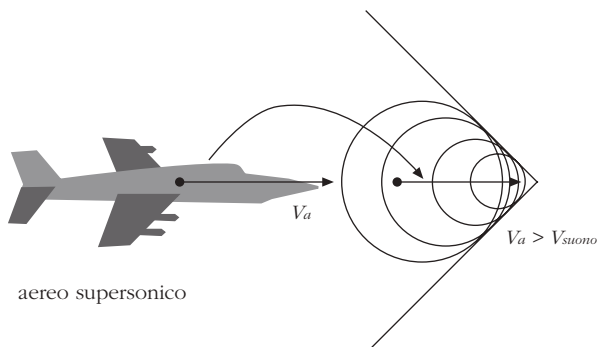


Fig. 15

Nelle vicinanze si sente un rumore fortissimo detto **boom sonico**. Questo fenomeno è dovuto alla sovrapposizione dei fronti d'onda delle onde sonore, la quale crea una regione di alta pressione e cioè di forte compressione, seguita da una regione di rarefazione. I fronti d'onda invece di arrivare uno dopo l'altro, raggiungono l'osservatore simultaneamente e il suono percepito è molto acuto.

Per indicare la velocità di un aereo si usa il numero di **Mach**, definito come il rapporto tra la velocità dell'aereo e la velocità del suono.

Se, ad esempio, un aereo si muove ad una velocità doppia di quella del suono, vuol dire che esso viaggia a Mach 2.

1 La luce: un'onda elettromagnetica

Per trattare i colori dello spettro visibile dobbiamo considerare che la luce è formata da onde elettromagnetiche che viaggiano alla velocità di 300.000km/sec .

Come tutte le onde, anche quelle elettromagnetiche hanno una frequenza propria e una lunghezza d'onda associata a questa frequenza.

Tutte le onde elettromagnetiche sono classificate in base alle loro frequenze.

L'insieme delle frequenze a cui esse possono oscillare è detto **spettro elettromagnetico**.

Esso spazia dalle onde radio ai raggi Gamma; le prime hanno frequenze molto basse (30Hz) e alte lunghezze d'onda (10.000km), i secondi hanno basse lunghezze d'onda (inferiori al milionesimo di millimetro) e frequenze comprese tra i 10^{19} e 10^{22} Hz.

Solo una piccolissima porzione dello spettro appartiene al cosiddetto **spettro visibile** e cioè all'insieme delle lunghezze d'onda percepibili dall'occhio umano. Tale porzione corrisponde a lunghezze d'onda comprese fra i 380 e i 780 nanometri e copre tutti i colori a noi visibili.

Alle lunghezze d'onda minori corrisponde la gamma dei blu-violetto, alla lunghezza d'onda maggiore corrisponde la gamma dei rossi.

Newton circa 3 secoli fa dimostrò che quando un raggio di luce incide su un prisma di vetro viene scomposto nelle onde che lo compongono, e quindi nei diversi colori (Fig. 1).

La suddivisione della luce nelle sue onde componenti è detta **dispersione della luce**.

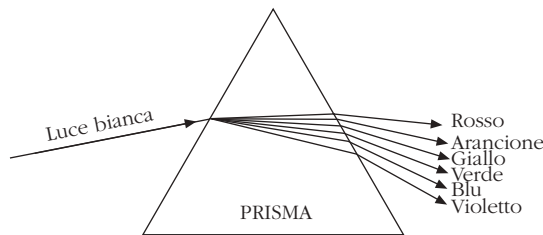


Fig. 1

Maggiore è la lunghezza d'onda del colore e minore è la deviazione subita.

Per esempio il rosso, che è il colore con la maggiore lunghezza, è quello meno deviato.

Sebbene la luce, in teoria, dovrebbe essere dispersa in infiniti colori (tutti quelli corrispondenti alle infinite lunghezze d'onda fra i 380nm e i 780nm) in realtà l'occhio percepisce soltanto sette famiglie di colori, detti **puri o spettrali**, che sono:

- rosso;
- arancione;
- giallo;
- verde;
- blu;
- indaco;
- violetto.

Quando tutte le lunghezze d'onda dello spettro visibile colpiscono i nostri occhi, contemporaneamente percepiamo il colore bianco.

In realtà, il bianco non è un colore ma una combinazione di tutti i colori dello spettro visibile così come il nero non è un colore ma è l'assenza di tutti i colori dello spettro visibile.

In una stanza buia tutti gli oggetti ci appaiono neri perché nessuna onda con frequenza nel campo dello spettro visibile colpisce i nostri occhi.

2 La percezione dei colori

A meno che non si facciano esperimenti di laboratorio, la luce non è monocromatica ma è formata dallo spettro di frequenza del visibile e quando gli oggetti vengono colpiti da un fascio di luce essi assorbono, riflettono e trasmettono la luce in modo selettivo.

Questo vuol dire che può accadere che un oggetto assorba la frequenza del fascio di luce corrispondente al valore verde e non assorba tutte le altre frequenze, mentre un altro assorba il colore giallo.

Il modo con cui la luce interagisce con gli oggetti dipende dalla frequenza della luce e dalla natura degli atomi degli oggetti stessi.

a) Riflessione, assorbimento e trasmissione della luce

Dalla chimica sappiamo che tutti gli oggetti sono formati da atomi e da molecole e che gli atomi sono a loro volta composti da un nucleo centrale intorno al quale orbitano gli elettroni.

Possiamo immaginare il moto rotatorio di un elettrone come quello di una molla oscillante con una frequenza propria: entrambi si muovono di **moto armonico semplice**.

Quando un'onda avente una certa frequenza colpisce un materiale, gli elettroni aventi una frequenza propria uguale a quella dell'onda incidente, **assorbono** completamente l'energia dell'onda incidente aumentando la propria ampiezza di oscillazione.

L'onda incidente viene totalmente assorbita dalla materia.

Quando gli elettroni del materiale su cui incide il raggio luminoso non hanno una frequenza propria uguale a quella dell'onda incidente la luce viene **riflessa o trasmessa**.

Un materiale assorbe, quindi, soltanto quelle onde che hanno una frequenza uguale a quella di oscillazione dei suoi elettroni. Di conseguenza, il colore corrispondente a tale frequenza, essendo assorbito dalla materia, non giunge ai nostri occhi e noi non lo vediamo.

Noi vediamo soltanto le frequenze trasmesse o riflesse da quel materiale.

Quindi, il colore di un oggetto non sta nell'oggetto stesso ma nella luce che lo colpisce: l'oggetto ha semplicemente il ruolo di assorbire o riflettere i raggi incidenti.

b) Addizione dei colori

La percezione dei colori è un fenomeno molto complesso.

Supponiamo che una camicetta ci appaia di colore rosso. Nell'istante in cui guardiamo la camicetta, diversi raggi luminosi colpiscono il nostro occhio con frequenze diverse; il nostro sistema occhio-cervello interpreta tutte le frequenze ricevute in quell'istante e decodifica il colore come rosso.

La complessità del fenomeno della percezione dei colori può essere semplificata se ragioniamo in termini di colori primari.

Abbiamo già detto che il bianco non corrisponde ad una determinata frequenza e quindi a un determinato colore ma all'insieme di tutte le frequenze dello spettro del visibile.

In realtà, per avere il colore bianco non dobbiamo combinare **tutte** le frequenze dello spettro del visibile: **ne bastano tre**.

Ogni terna di colori in grado di produrre il colore bianco è detta terna dei **colori primari**. La più usata è quella formata dal **rosso**, in inglese red, dal **verde**, in inglese green, e dal **blu**, in inglese blue. Questa terna è detta **RGB**.

Un fascio di luce formato da onde luminose di colore rosso, verde e blu con una opportuna intensità sarà di luce bianca.

Per questa ragione, per regolare il colore delle immagini sugli schermi dei computer e delle televisioni si regola l'intensità dei fosfori rossi, verdi e blu.

Supponiamo di avere tre raggi di luce monocromatica aventi la stessa intensità e corrispondenti alle frequenze dei colori primari, modificando le intensità dei raggi luminosi e combinando i diversi colori ottenuti dai colori primari si possono ottenere tutti gli altri colori dello spettro del visibile.

c) Sottrazione dei colori

Nel paragrafo precedente abbiamo visto come sia possibile «sommare» i colori principali per riuscire a percepire tutti quelli dello spettro del visibile. I colori si possono però anche sottrarre.

Se un materiale assorbe la frequenza di un colore noi non vediamo quel colore.

Consideriamo una camicia fatta di un materiale in grado di assorbire il colore blu e di riflettere le altre frequenze dello spettro visibile. Che cosa succede se noi illuminiamo la camicia con la luce visibile?

Per dare una risposta a questa domanda possiamo supporre che la luce visibile sia in realtà formata dai tre colori principali: rosso, verde e blu. Siccome la camicia assorbe il colore blu essa rifletterà il rosso e il verde e noi la vedremo gialla.

Se, invece, la illuminiamo con della luce color ciano, verrà vista verde. Quindi, da questi due esempi, possiamo concludere che una camicia cambia colore a seconda dei raggi a cui è esposta.

Abbiamo già detto che il colore di un oggetto non sta nell'oggetto in sé ma nella sua proprietà di riflettere o assorbire determinate frequenze. Estendendo questo concetto possiamo dire che una camicia appare gialla soltanto se c'è del verde e del rosso nei raggi luminosi che la colpiscono e se il materiale di cui è composta non assorbe tali frequenze. Infatti, il giallo è dato da una combinazione del verde e del rosso e affinché venga riflesso dalla camicia deve essere presente nella radiazione incidente.

d) Pigmenti

Le sostanze capaci di assorbire più colori sono dette **pigmenti**.

Quindi un foglio di carta per assorbire un colore, ad esempio blu, deve essere impregnato di un pigmento giallo o di un qualsiasi colore che contenga il blu.

Se un pigmento assorbe un solo colore è detto **puro**, altrimenti, nel caso assorba più colori, è detto **composto**.

Per comprendere meglio i pigmenti vale la seguente regola:

- i pigmenti assorbono la luce;
- i pigmenti puri assorbono un singolo colore;
- il colore assorbito dal pigmento è il colore complementare del pigmento.

Quindi:

- un pigmento blu assorbe il rosso e il verde e quindi il giallo;
- un pigmento verde assorbe il rosso e il blu e quindi il magenta;
- un pigmento rosso assorbe il verde e il blu e quindi il ciano.

1 Corpi luminosi e corpi illuminati

Gli oggetti che noi vediamo possono essere suddivisi in due categorie: **oggetti luminosi** e **oggetti illuminati**.

- *gli oggetti luminosi generano luce;*
- *gli oggetti illuminati non generano la luce ma la riflettono.*

Il Sole è un tipico esempio di oggetto luminoso mentre la Luna è un oggetto illuminato dal Sole.

a) Linea visiva

Noi possiamo vedere un oggetto soltanto se questo brilla di luce propria o di luce riflessa e se i raggi luminosi provenienti da esso colpiscono i nostri occhi.

Per vedere un oggetto dobbiamo puntare lo sguardo in direzione dell'oggetto e la luce proveniente dall'oggetto deve raggiungere i nostri occhi.

La direzione con cui la luce riflessa da un oggetto raggiunge i nostri occhi è detta **linea visiva**. A seconda della linea visiva, varia la nostra percezione dell'oggetto.

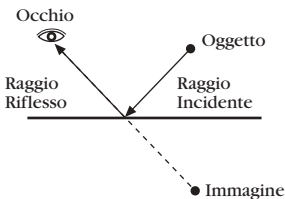


Fig. 1

I nostri occhi, però, non sono in grado di percepire se i raggi da cui sono colpiti sono diretti o riflessi e questo è il motivo di molte illusioni ottiche. Se ci troviamo di fronte ad uno specchio vediamo gli oggetti riflessi come se fossero **dietro** lo specchio.

In questo caso le immagini sono dette **virtuali** in quanto esistono soltanto nel nostro cervello e non nella realtà. Esse dipendono dalla nostra presenza. Se noi non fossimo davanti allo specchio le loro immagini non esisterebbero (Fig. 1).

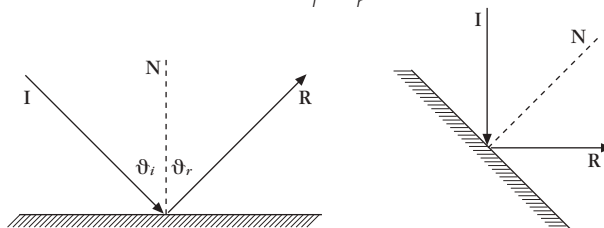
Le immagini reali invece sono quelle che vediamo al cinema; esse esistono effettivamente, indipendentemente dalla nostra presenza.

b) Legge della riflessione

Quando un raggio di luce colpisce una superficie piana viene riflesso e modifica la sua direzione di propagazione.

Indicando come angolo di incidenza ϑ_i quello che il raggio incidente forma con la normale alla superficie e come angolo di riflessione ϑ_r (Fig. 2.a) quello che il raggio riflesso forma con la normale alla superficie, per la **legge della riflessione** deve essere:

$$\vartheta_i = \vartheta_r$$



Figg. 2.a e 2.b

Questa legge è sempre applicabile per qualsiasi superficie orientata in un qualsiasi modo (Fig. 2.b).

Una volta tracciata la normale alla superficie, l'**angolo riflesso** sarà sempre **uguale** all'**angolo incidente**.

Se la superficie riflettente è liscia si ha una **riflessione speculare**, altrimenti si ha una **riflessione diffusa** (Fig. 3). In quest'ultimo caso, per ogni raggio incidente continua a valere la legge della riflessione ma varia l'orientamento della normale alla superficie nel punto di incidenza.

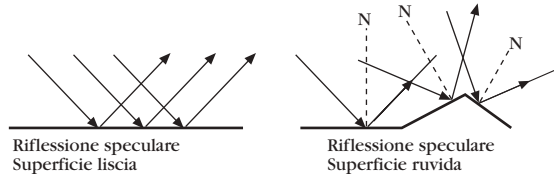
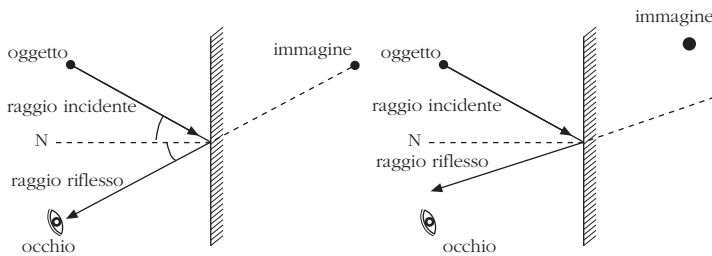


Fig. 3

Nella figura 4 possiamo vedere che i raggi luminosi provenienti dall'oggetto vengono riflessi dallo specchio secondo la legge della riflessione (angolo di incidenza = angolo di riflessione).

Se noi guardassimo in una direzione diversa da quella dell'immagine i raggi luminosi provenienti dall'oggetto, una volta riflessi, non colpirebbero più i nostri occhi e noi non vedremmo l'oggetto.



Figg. 4 e 5

Nella figura 5, infatti, l'occhio guarda in una direzione al di sotto di quella dell'immagine e, in questo caso, per poterla visualizzare, l'angolo di incidenza dovrebbe essere maggiore di quello di riflessione, ma ciò non può avvenire.

Per vedere l'immagine di un oggetto in uno specchio dobbiamo guardare esattamente nella direzione del raggio riflesso dall'oggetto.

2 Gli specchi piani

Il punto in cui viene visualizzata l'immagine di un oggetto si chiama **punto immagine**.

La sua posizione è determinata o dall'intersezione dei raggi riflessi o dall'intersezione dei loro prolungamenti. Nel primo caso si parla di **immagini reali**, nel secondo caso di **immagini virtuali**.

Vediamo cosa accade nel caso di uno specchio piano.

Il punto immagine si trova intersecando i raggi riflessi dalla superficie riflettente. Dalla figura 6 possiamo notare che questi raggi divergono e quindi non si possono intersecare.

Possiamo però prolungare i raggi riflessi dall'altra parte dello specchio e indicare come punto immagine l'intersezione di questi prolungamenti.

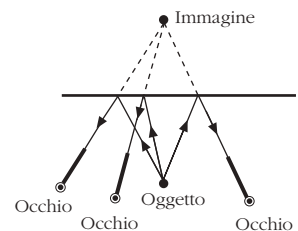


Fig. 6

Un osservatore che dirige lo sguardo in direzione del punto immagine verrà colpito dai raggi riflessi ed effettivamente vedrà l'immagine dell'oggetto.

Il nostro cervello non percepisce la differenza tra i raggi riflessi e quelli diretti; esso istintivamente prolunga la direzione dei raggi incidenti fino a farli intersecare in un punto. Quindi l'osservatore vedrà i raggi divergere dal punto immagine.

Questa immagine è però detta **virtuale** in quanto in essa non si intersecano i raggi riflessi ma i loro prolungamenti.

Se in un punto immagine convergono i raggi riflessi, e non i loro prolungamenti, l'immagine è detta **reale**.

a) Caratteristiche delle immagini formate dagli specchi piani

- Le immagini sono virtuali.
- Le immagini sono speculari sinistra-destra.
- La distanza dell'oggetto dallo specchio è uguale alla distanza dell'immagine dallo specchio.
- La dimensione dell'immagine è uguale alla dimensione dell'oggetto.

3 Gli specchi concavi

Vediamo ora come avviene la riflessione delle immagini su specchi aventi come superfici uno specchio concavo.

Uno **specchio concavo** è uno specchio ottenuto da una calotta sferica in cui è stata annerita la superficie esterna.

Consideriamo lo specchio concavo rappresentato in figura 7 ed introduciamo alcuni termini che utilizzeremo nel corso di questo capitolo:

- il **centro di curvatura** C è il centro della sfera;
- il **vertice** V è il centro geometrico dello specchio;
- l'**asse principale** è la retta che congiunge il centro di curvatura con il vertice dello specchio;
- il **punto focale** o **fuoco** F si trova a metà fra il centro di curvatura e il vertice dello specchio;
- il **raggio di curvatura** R è il raggio della sfera da cui è stato ricavato lo specchio;
- la **distanza focale** f è la distanza fra lo specchio e il suo fuoco; essa è pari alla metà del raggio di curvatura.

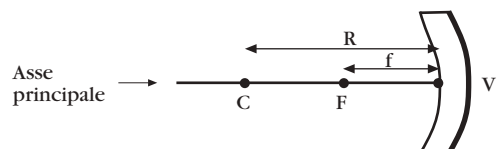


Fig. 7

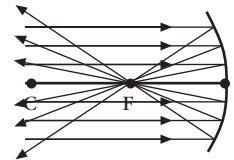


Fig. 8

Il **fuoco** è il punto in cui i raggi incidenti paralleli all'asse principale si intersecano dopo la riflessione (Fig. 8).

a) Formazione di un'immagine

Per definire l'immagine riflessa da uno specchio concavo dobbiamo utilizzare la legge della riflessione in cui gli angoli sono definiti rispetto alla normale alla superficie.

A differenza di quanto succedeva per uno specchio piano, però, la normale a una superficie concava varia a seconda del punto di incidenza del raggio luminoso (Fig. 9) e quindi, trovare l'immagine riflessa tracciando l'intersezione dei raggi riflessi da due **qualsiasi** raggi incidenti, potrebbe sembrare piuttosto complesso.

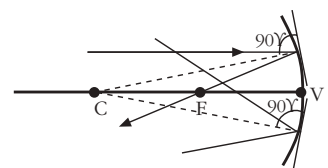


Fig. 9

In realtà, nel prossimo paragrafo vedremo che esiste un metodo semplificato per definire l'immagine di un oggetto riflesso da una superficie concava. Per adesso limitiamoci a uno studio del tutto empirico sulla posizione e sulla natura dell'immagine formata da uno specchio concavo supponendo che l'oggetto si trovi **prima** del centro di curvatura dello specchio (Fig. 10).

Dalla figura 10 possiamo notare che nel punto immagine dell'oggetto si intersecano i raggi riflessi e quindi l'immagine in questo caso è **reale** e si trova **prima** del centro di curvatura.

In figura 11 vediamo che se l'oggetto si fosse trovato dopo il centro di curvatura l'immagine sarebbe sempre stata reale ma si sarebbe trovata **dopo** il centro della curvatura.

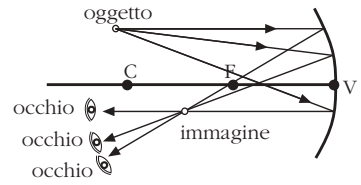


Fig. 10

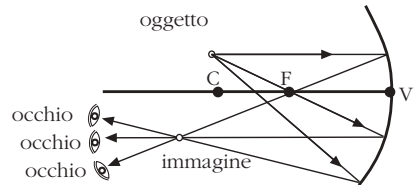
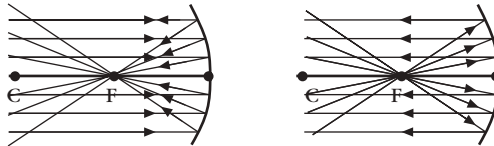


Fig. 11

b) Regole di riflessione

Esiste un metodo semplificato per ricavare l'immagine di un oggetto riflesso da uno specchio concavo. Esso si basa su due principi fondamentali illustrati in figura 12:

- tutti i raggi paralleli all'asse principale di uno specchio concavo una volta riflessi passano per il fuoco dello specchio (Fig. 12.a);
- tutti i raggi passanti per il fuoco dello specchio concavo una volta riflessi hanno direzione parallela all'asse principale (Fig. 12.b).



Figg. 12.a e 12.b

Questi due principi semplificano di molto la determinazione dell'immagine riflessa da uno specchio concavo. Applicando le regole sopra elencate possiamo determinare l'immagine dell'oggetto in figura utilizzando **due soli raggi incidenti**: quello passante per il fuoco e quello diretto parallelamente all'asse principale della lente.

c) Come ricavare l'immagine riflessa

Supponiamo di avere un oggetto come in figura 13 e di volerne determinare l'immagine.

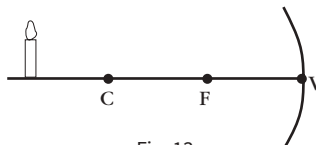


Fig. 13

Il procedimento è il seguente:

1. tracciare due raggi passanti per la sommità dell'oggetto, uno parallelo all'asse principale e l'altro passante per il fuoco dello specchio. Successivamente si indichi con due frecce il verso dei raggi (Fig. 14);

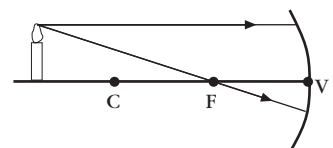


Fig. 14

2. riflettere i due raggi applicando la legge di riflessione per gli specchi concavi, per cui, il raggio riflesso del raggio parallelo all'asse principale deve passare per il fuoco mentre il raggio riflesso del raggio passante per il fuoco deve essere parallelo all'asse principale. Una volta indicata la direzione dei raggi riflessi, si determini il punto di intersezione degli stessi, il quale costituirà il punto immagine (Fig. 15);

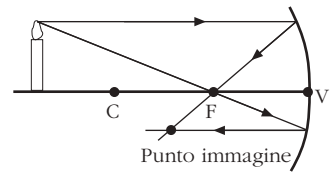


Fig. 15

3. ripetere il procedimento per i punti più significativi dell'oggetto e disegnare l'oggetto immagine. Le immagini dei punti che si trovano sull'asse principale si trovano anch'esse sull'asse principale (Fig. 16).

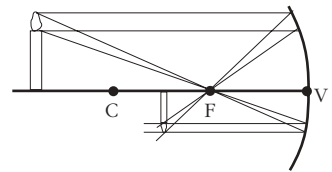
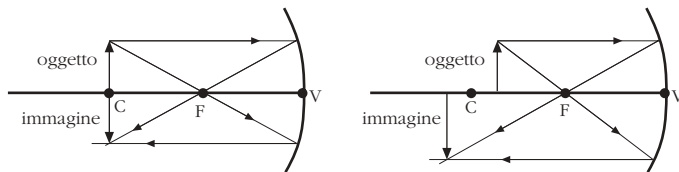


Fig. 16

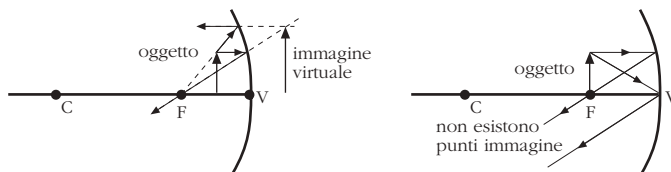
Se l'oggetto si trova prima del centro di curvatura, l'immagine si trova tra il centro di curvatura e il fuoco della lente e risulta rimpicciolita. Se l'oggetto si fosse trovato nel centro di curvatura dello specchio anche l'immagine si sarebbe trovata nello stesso punto (Fig. 17), mentre se fosse stato posto fra il fuoco e il centro di curvatura, la sua immagine si sarebbe trovata oltre il centro di curvatura (Fig. 18).



Figg. 17 e 18

Nelle figure 17 e 18 l'immagine formata è **reale** perché l'oggetto si trova **prima** del punto focale.

In figura 19 possiamo vedere che se l'oggetto si trovasse a una distanza minore della distanza focale la sua immagine sarebbe virtuale.



Figg. 19 e 20

Infine, se l'oggetto si trovasse esattamente nel punto focale non ci sarebbe alcuna immagine in quanto i raggi riflessi sarebbero tutti paralleli e quindi non si intersecherebbero mai (Fig. 20).

d) Posizione e dimensioni dell'immagine di un oggetto

In questo paragrafo determineremo la posizione esatta e le dimensioni dell'immagine di un oggetto.

• Equazione dei punti coniugati

Con il termine **punti coniugati** si intende la coppia **punto oggetto-punto immagine** di un qualsiasi oggetto.

L'equazione dei punti coniugati mette in relazione la posizione dell'immagine con quella dell'oggetto a seconda delle caratteristiche dello specchio concavo riflettente.

Per ricavarla osserviamo la figura 21 in cui il punto oggetto è situato in O , sull'asse principale dello specchio.

Tracciamo dal punto O due raggi: un raggio qualsiasi incidente lo specchio nel punto A e il raggio passante per il centro di curvatura dello specchio. Possiamo notare che anche il punto immagine si trova sull'asse principale. Infatti, per la legge della riflessione, il raggio incidente OA viene riflesso nel punto I e gli angoli OAC e CAI sono uguali.

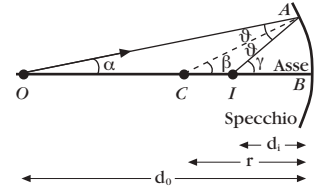


Fig. 21

Dal triangolo OAC possiamo dedurre che $\beta = \alpha + \vartheta$ e, dal triangolo OAI , ricaviamo che $\gamma = \alpha + 2\vartheta$. Da ciò segue che:

$$\gamma + \alpha = 2\beta$$

Possiamo esprimere gli angoli α , β e γ in funzione dell'arco di circonferenza AB :

$$\alpha \cong \frac{AB}{BO} = \frac{AB}{d_o}$$

$$\beta \cong \frac{AB}{CB} = \frac{AB}{r}$$

$$\gamma \cong \frac{AB}{IB} = \frac{AB}{d_i}$$

Sostituendo tali risultati nell'equazione precedente si ottiene:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} = \frac{2}{r} \quad (3.1)$$

che è l'**equazione dei punti coniugati** per uno specchio concavo.

• Equazione dell'ingrandimento trasversale

Determiniamo le dimensioni di un oggetto.

I triangoli ABC e CDE nella figura 23 sono simili quindi:

$$-\frac{h_i}{h_o} = \frac{d_i}{d_o} \quad (3.2)$$

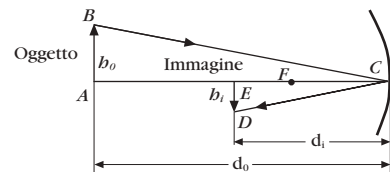


Fig. 22

che è l'**equazione dell'ingrandimento trasversale**.

4 Gli specchi convessi

Possiamo immaginare uno specchio convesso come una calotta sferica in cui è stata annerita la parte interna (Fig. 23).

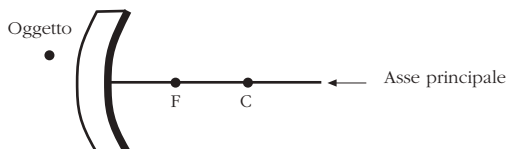


Fig. 23

Il centro della sfera originaria è detto **centro di curvatura** (C) e la linea che unisce il centro di curvatura con il centro della sfera è detta **asse principale**. Lo specchio ha un **punto focale** F situato sull'asse principale a metà fra il centro di curvatura e il centro della sfera. Siccome il punto focale si trova dietro lo specchio la distanza focale è assunta negativa.

Uno specchio convesso è anche detto **divergente** per la sua caratteristica di far divergere i raggi che incidono la sua superficie.

a) Formazione dell'immagine

In figura 24 si può vedere cosa accade quando 4 raggi incidono sulla superficie dello specchio. Dopo la riflessione essi divergono e, per questa ragione, gli specchi convessi producono soltanto immagini virtuali situate da qualche parte dietro lo specchio.

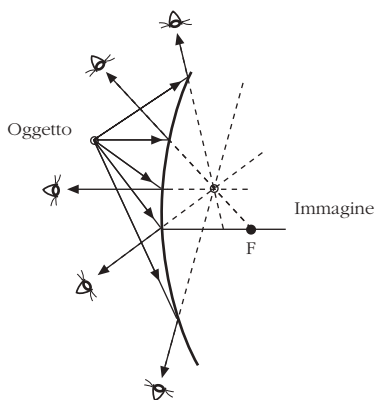


Fig. 24

b) Regole di riflessione

Per gli specchi convessi valgono due regole della riflessione:

- tutti i raggi incidenti paralleli all'asse principale vengono riflessi in modo che i loro prolungamenti passino per il fuoco dello specchio;
- tutti i raggi incidenti su uno specchio convesso i cui prolungamenti passano per il fuoco dello specchio vengono riflessi in direzione parallela all'asse principale.

In figura 25 è stata ripresa la figura 24 in cui è stato evidenziato il raggio parallelo all'asse principale e il raggio il cui prolungamento passa per il fuoco dello specchio.

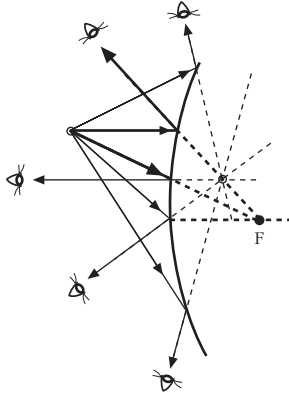


Fig. 25

c) Come ricavare l'immagine riflessa

Ricaviamo l'immagine di una figura riflessa da uno specchio convesso mediante il seguente procedimento:

1. prendere un punto sulla parte più alta dell'oggetto e tracciare un raggio in direzione del fuoco dello specchio fino al punto in cui lo interseca e un raggio parallelo all'asse principale dello specchio. Indicare, inoltre, la direzione dei raggi (Fig. 26);
2. riflettere i due raggi secondo le regole della riflessione per gli specchi convessi. Il raggio in direzione del punto focale viene riflesso parallelamente all'asse principale e il raggio parallelo all'asse principale viene riflesso in modo che il prolungamento del raggio riflesso passi per il fuoco dello specchio. Indicare, altresì, il verso dei raggi riflessi (Fig. 27);
3. poiché i due raggi divergono, tracciare i loro prolungamenti, evidenziare il punto di intersezione e disegnare l'oggetto (Fig. 28);

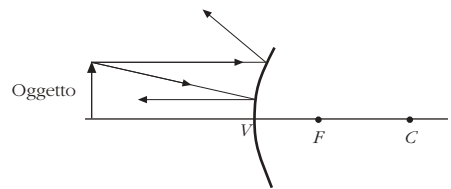


Fig. 26

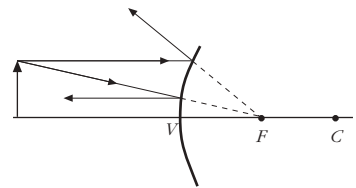


Fig. 27

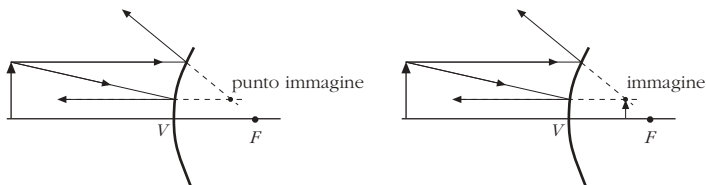


Fig. 28

4. Ripetere poi il procedimento per gli altri punti dell'oggetto.

Analizzando l'immagine ricavata in figura 28 si nota che essa è virtuale, non invertita e rimpicciolita rispetto alle dimensioni dell'oggetto.

In figura 29 sono state costruite le immagini di oggetti posti a distanze diverse dallo specchio e si può notare che anche in questi casi sono virtuali, non invertite e rimpicciolite.

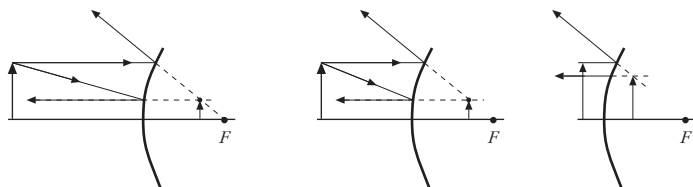


Fig. 29

A differenza degli specchi concavi, le immagini prodotte dagli specchi convessi sono **sempre**:

- virtuali;
- non invertite;
- di dimensioni più piccole rispetto all'oggetto.

d) Posizione e dimensioni dell'immagine di un oggetto

Per ricavare numericamente le dimensioni e la distanza dell'immagine di un oggetto riflesso da uno specchio convesso consideriamo la figura 30.

L'equazione dei punti coniugati (3.1) e dell'ingrandimento trasversale (3.2) continua a valere per gli specchi convessi ma in questo caso dobbiamo fare attenzione ai segni delle distanze. Infatti:

- f è positiva se lo specchio è concavo;
- f è negativa se lo specchio è convesso;
- d_i è positiva se l'immagine è reale e situata dalla parte dell'oggetto rispetto allo specchio;
- d_i è negativa se l'immagine è virtuale e situata dietro lo specchio;
- h_i è positiva se l'immagine non è invertita (e quindi è virtuale);
- h_i è negativa se l'immagine è invertita (e quindi è reale).

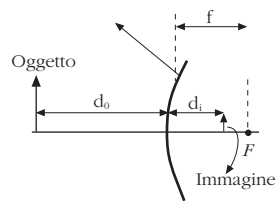


Fig. 30

1 Il fenomeno della rifrazione

Il fenomeno della **rifrazione** si verifica quando un raggio luminoso passa attraverso due materiali diversi deviando la sua traiettoria di incidenza (Fig. 1). Nel momento in cui i fronti d'onda incontrano la superficie di separazione tra i due mezzi, vengono deviati modificando così la loro direzione di propagazione. Questo vuol dire che il fenomeno non avviene all'interno di un mezzo ma esclusivamente al **confine** tra due mezzi.

Per spiegare la rifrazione, così come abbiamo fatto per la riflessione della luce, useremo spesso i **raggi d'onda**. Essi rappresentano la direzione di propagazione dell'onda e sono sempre **perpendicolari** ai fronti d'onda (Fig. 2).

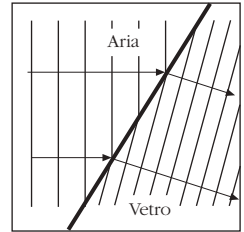


Fig. 1

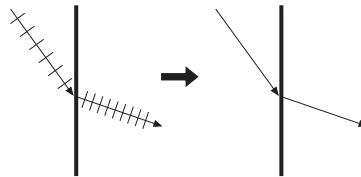


Fig. 2

a) Cause della rifrazione

Nel passaggio fra due mezzi, un'onda:

- viene *rallentata* e *rifratta* se incide *non perpendicolarmente* la superficie di separazione;
- viene soltanto *rallentata* se incide *perpendicolarmente* la superficie di separazione.

ESEMPIO

Consideriamo il modello meccanico rappresentato nelle figure 3a e 3b, le quali mostrano un'auto che passa da una strada asfaltata ad una distesa di sabbia e cioè da un materiale ad un altro.

Se l'auto attraversa perpendicolarmente la linea di separazione tra l'asfalto e la sabbia (Fig. 3.a), manterrà la stessa direzione di moto e subirà solo una diminuzione della propria velocità per il fatto che, a parità di spinta, la velocità è maggiore sull'asfalto che non sulla sabbia.

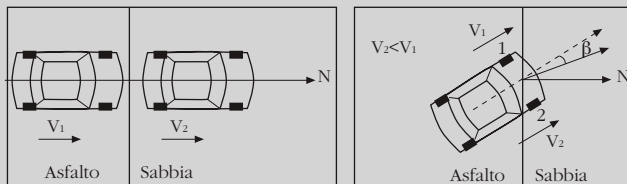


Fig. 3.a e 3.b

Se, invece, l'auto ha una direzione di moto obliqua rispetto alla linea di separazione (Fig. 3b), essa subirà, oltre ad una diminuzione di velocità, anche un certo cambiamento di direzione del proprio moto (supponiamo di un certo angolo β). Ciò è dovuto al fatto che la ruota 2, attraversando prima della ruota 1 la linea di separazione tra la sabbia e l'asfalto, si muoverà, ad una velocità V_2 inferiore rispetto a quella V_1 della ruota 1, nell'intervallo di tempo necessario perché anche la ruota 1 attraversi la linea. La conseguenza sarà che l'auto modificherà la sua direzione di moto di un angolo β .

L'esempio precedente è utile per chiarire alcuni concetti fondamentali:

- se un'onda passa da un mezzo meno denso ad uno più denso diminuisce la propria velocità di propagazione;
- se un'onda passa da un mezzo più denso ad uno meno denso aumenta la propria velocità di propagazione.

Inoltre, se la sua direzione non è perpendicolare alla superficie di separazione tra i due mezzi, l'onda, oltre a subire una variazione di velocità, verrà rifratta, ossia subirà una variazione della propria direzione.

b) Effetti ottici dovuti alla rifrazione

Nel capitolo sulla riflessione abbiamo visto che i nostri occhi possono venir tratti in inganno, creando immagini virtuali esistenti nel nostro cervello ma non nella realtà. Anche nel caso della rifrazione si hanno delle distorsioni ottiche: un esempio tipico è quello della matita spezzata.

Supponiamo di immergere una matita al centro di un bicchiere riempito per metà d'acqua (Fig. 4.a).

Se spostiamo la matita dalla posizione centrale verso il bordo del bicchiere, la porzione di matita che si trova sotto l'acqua ci apparirà spostata rispetto a quella che si trova sopra l'acqua (Fig. 4.b).



Figg. 4.a, 4.b e 4.c

I nostri occhi, infatti, vengono colpiti sia dai raggi luminosi provenienti dalla porzione di matita immersa nell'acqua che da quelli provenienti dalla porzione di matita che si trova fuori dall'acqua.

I primi si propagano nell'aria e colpiscono i nostri occhi direttamente, mentre i secondi subiscono il fenomeno della rifrazione nel passaggio dall'acqua all'aria, ed appaiono agli occhi dell'osservatore nella posizione tratteggiata in figura 5.

Se la matita si trova esattamente nel centro, i raggi luminosi, essendo perpendicolari alla superficie del bicchiere, non vengono rifratti e quindi non c'è una distorsione dell'immagine. Appena si sposta la matita verso il bordo del bicchiere il suo angolo di rifrazione aumenta e la matita sembra spezzata.

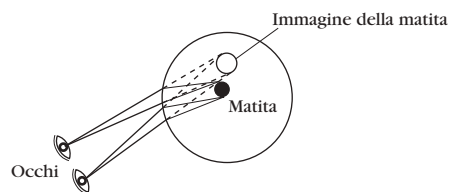


Fig. 5

c) Materiali e rifrazione: densità ottica e indice di rifrazione

Un'onda luminosa è un'onda elettromagnetica che viaggia nel vuoto alla velocità della luce, ossia a $c = 3 \cdot 10^8 \text{ m/s}$.

Quando un'onda luminosa incide sulla superficie di un materiale, viene assorbita dagli elettroni della particella del materiale colpito. Se la sua frequenza non è in risonanza con quella di tali elettroni, viene subito riemessa con la stessa frequenza e riprenderà a viaggiare alla velocità della luce nello spazio interatomico del materiale, fino a colpire un'altra particella.

Con questo ciclo di assorbimento ed emissione, l'onda incidente si propaga nel materiale con una velocità inferiore alla velocità della luce.

Tenendo conto di questo fenomeno, definiamo due nuove *grandezze*: la densità ottica e l'indice di rifrazione del materiale.

La **densità ottica** indica la lentezza con cui gli atomi rimettono le onde luminose.

Più gli atomi sono lenti e maggiore è la densità ottica.

L'indice di rifrazione:

$$n_{\text{materiale}} = \frac{\text{velocità della luce nel vuoto}}{\text{velocità della luce nel materiale}} = \frac{3 \times 10^8 \text{ m/s}}{v_{\text{materiale}}}$$

indica quante volte la velocità della luce nel vuoto è maggiore rispetto a quella nel mezzo e dà un'idea di questa lentezza.

Nella tabella 1 sono indicati alcuni indici di rifrazione di materiali diversi.

Più l'indice di rifrazione si discosta dal valore 1, maggiore è la densità ottica del materiale e minore è la velocità della luce al suo interno.

Materiale	Indice di Rifrazione
Vuoto	1,0000 ← minore densità ottica
Aria	1,0003
Ghiaccio	1,31
Acqua	1,333
Alcool Etílico	1,36
Plexiglas	1,51
Vetro Crown	1,52
Vetro Flint	1,58
Zirconio	1,923
Diamante	2,417
Rutile	2,907 ← maggiore densità ottica

Tab. 1

d) Direzione della rifrazione

Consideriamo un raggio incidente che passa da un mezzo con data densità ottica ad uno con densità ottica maggiore lungo una superficie di separazione (rappresentata in figura 6 dalla linea $x - x'$) obliqua rispetto alla propria direzione. Tale raggio, oltre a subire una **diminuzione** di velocità, subirà quindi anche una rifrazione.

Tracciamo la perpendicolare (normale) alla linea di separazione tra i due mezzi di propagazione dell'onda luminosa.

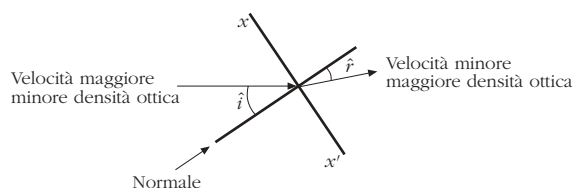


Fig. 6

Se indichiamo con $\hat{i}$ e $\hat{r}$ gli angoli che il raggio incidente e il raggio rifratto formano con la normale alla superficie, possiamo osservare che, nel passaggio da un mezzo con minore densità ottica a uno con maggiore densità ottica, l'angolo rifratto si avvicina alla normale, ossia $\hat{i} > \hat{r}$.

Pertanto, se un raggio luminoso diminuisce la sua velocità nel passaggio fra due mezzi, la sua direzione di propagazione si avvicina alla normale alla superficie.

Consideriamo ora lo stesso raggio incidente che passa però da un mezzo con densità ottica maggiore ad uno con densità ottica minore (Fig. 7). Tale raggio, oltre a subire una diminuzione della propria velocità, subirà anche una rifrazione.

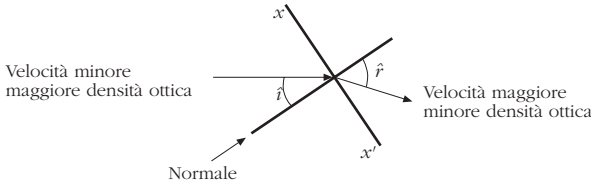


Fig. 7

In questo caso la direzione di propagazione si è allontanata dalla normale alla superficie, ossia $\hat{i} < \hat{r}$.

Pertanto, se un raggio luminoso aumenta la sua velocità nel passaggio fra due mezzi, la sua direzione di propagazione si allontana dalla normale alla superficie.

2 La legge di Snell

Nel paragrafo precedente abbiamo ricavato che un raggio rifratto si avvicina alla normale se passa da un mezzo avente densità ottica minore ad un mezzo avente densità ottica maggiore; in questo paragrafo, invece, studieremo di **quanto** si avvicina alla normale.

Se colpiamo con un raggio laser un contenitore semisferico pieno d'acqua (Fig. 8) e misurassimo l'angolo di rifrazione (ϑ_r) facendo variare l'angolo di incidenza (ϑ_i) ogni volta di 5 gradi, otterremmo i dati in tabella 2.

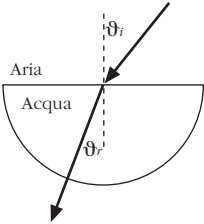


Fig. 8

Angolo di incidenza (gradi)	Angolo di rifrazione (gradi)
0,00	0,00
5,00	3,76
10,0	7,50
15,0	11,2
20,0	14,9
25,0	18,5
...	...

Tab. 2

Dai dati ricavati si può notare che non esiste una relazione lineare fra l'angolo di incidenza e quello di rifrazione.

Se, però, calcoliamo il rapporto tra il seno dell'angolo incidente e quello dell'angolo rifratto, otteniamo sempre il valore 1,33, che è proprio l'indice di rifrazione dell'acqua. In formule si ha:

$$\frac{\sin \vartheta_i}{\sin \vartheta_r} = n_{\text{acqua}} = 1,33 \quad (2.1)$$

La (2.1) può essere riscritta come:

$$\frac{\sin \vartheta_i}{\sin \vartheta_r} = \frac{n_{\text{acqua}}}{n_{\text{aria}}} = \frac{1,33}{1} = 1,33$$

In generale, vale la relazione nota come **legge di Snell**:

$$n_i \cdot \sin \vartheta_i = n_r \cdot \sin \vartheta_r \quad (2.2)$$

dove:

- n_i è l'indice di rifrazione del mezzo in cui si propaga il raggio incidente;
- n_r è l'indice di rifrazione del mezzo in cui si propaga il raggio rifratto.

3 La riflessione totale

Per comprendere la relazione che sussiste tra il raggio riflesso e quello rifratto facciamo il seguente esperimento.

Immergiamo una sorgente di raggi laser in un contenitore pieno d'acqua e inviamo dei raggi luminosi verso la superficie, facendo variare di volta in volta l'angolo di incidenza (Fig. 9).

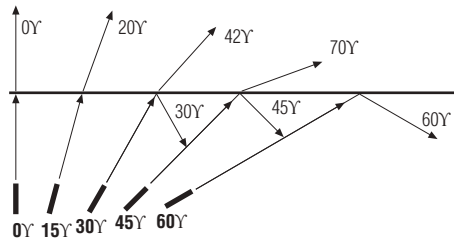


Fig. 9

Possiamo notare che quando il raggio incide perpendicolarmente la superficie del materiale, esso non viene riflesso ma viene trasmesso totalmente nel secondo mezzo.

Aumentando l'angolo di incidenza aumenterà anche l'angolo di rifrazione e, poiché l'indice di rifrazione dell'acqua è maggiore di quello dell'aria, il raggio rifratto si allontanerà sempre di più dalla normale alla superficie. Procedendo in questo modo si arriverà ad un angolo di incidenza, detto **angolo limite**, in corrispondenza del quale l'angolo di rifrazione è di 90° e il raggio rifratto sarà parallelo alla superficie di separazione.

Oltre questo valore non si avrà più rifrazione e tutta l'onda incidente verrà riflessa nel primo mezzo: questo è il fenomeno della **riflessione totale** (Fig. 10).

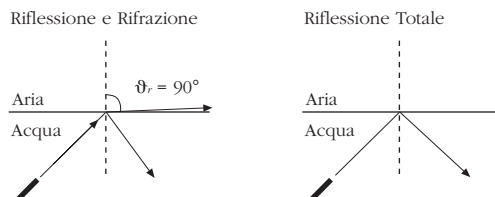


Fig. 10

Il fenomeno della riflessione totale si verifica, quindi, soltanto se:

- la luce passa da un mezzo più denso a un mezzo meno denso;
- l'angolo di incidenza è maggiore dell'angolo limite.

a) Angolo limite

Per ricavare il valore dell'angolo limite basta riscrivere la legge di Snell imponendo che l'angolo di rifrazione sia uguale a un angolo retto. Traducendo in formule, abbiamo:

$$n_i \cdot \sin \vartheta_{\text{limite}} = n_r \cdot \sin 90^\circ$$

Visto che $\sin 90^\circ = 1$, si ha:

$$\sin \vartheta_{\text{limite}} = \frac{n_r}{n_i}$$

da cui segue che:

$$\vartheta_{\text{limite}} = \sin^{-1} \left(\frac{n_r}{n_i} \right) \quad (3.1)$$

b) Legge di Snell e lunghezze d'onda

La legge di Snell è esprimibile anche mediante le lunghezze d'onda della radiazione incidente e di quella rifratta.

Infatti, dalla definizione dell'indice di rifrazione, segue che:

$$n_{\text{materiale}} = \frac{\text{velocità della luce nel vuoto}}{\text{velocità della luce nel materiale}} = \frac{3 \times 10^8 \text{ m/s}}{v_{\text{materiale}}} = \frac{c}{v_{\text{materiale}}} \quad (3.2)$$

Sostituendo nella (3.2) l'equazione d'onda $v = f\lambda$ segue che:

$$n_{\text{materiale}} = \frac{c}{f_{\text{materiale}} \lambda_{\text{materiale}}} \quad (3.3)$$

La legge di Snell ($n_i \cdot \sin \vartheta_i = n_r \cdot \sin \vartheta_r$) può essere, a questo punto, riscritta nella seguente maniera:

$$\frac{c}{f_i \lambda_i} \sin \vartheta_i = \frac{c}{f_r \lambda_r} \sin \vartheta_r \quad (3.4)$$

Siccome quando un'onda passa da un mezzo ad un altro la sua frequenza non cambia, la (3.4) diventa:

$$\frac{\sin \vartheta_i}{\lambda_i} = \frac{\sin \vartheta_r}{\lambda_r} \quad (3.5)$$

che può anche essere scritta come:

$$\frac{\sin \vartheta_i}{\sin \vartheta_r} = \frac{\lambda_i}{\lambda_r} \quad (3.6)$$

Per la legge di Snell:

$$\frac{\sin \vartheta_i}{\sin \vartheta_r} = \frac{n_r}{n_i}$$

e quindi:

$$\frac{n_r}{n_i} = \frac{\lambda_i}{\lambda_r}$$

che è la legge di Snell espressa mediante le lunghezze d'onda.

4 Le lenti

a) Lenti convergenti e lenti divergenti

Una **lente** è delimitata da due calotte sferiche aventi i bordi coincidenti. Ogni lente possiede:

- un **asse principale** che è la linea immaginaria che congiunge il centro della sfera con quello della lente;
- un **asse verticale** che è l'asse perpendicolare a quello principale e che taglia la lente in due parti;
- due **punti focali** – uno per lato – che sono i punti in cui convergono i raggi paralleli all'asse principale;
- un **centro** che è il punto di intersezione fra l'asse principale e l'asse verticale.

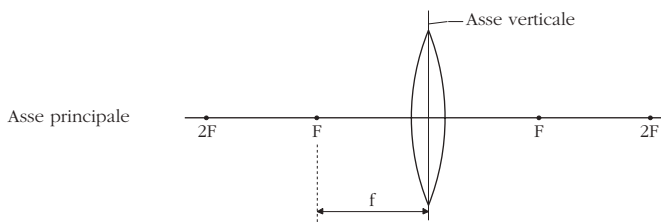


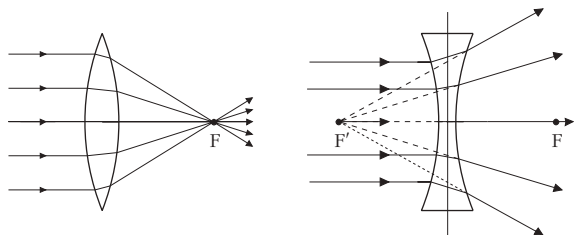
Fig. 11

Le lenti possono essere **convergenti** o **divergenti**.

Nelle prime, i raggi incidenti paralleli all'asse ottico, **convergono nel fuoco** F (Fig. 12.a).

Nelle seconde, gli stessi **divergono all'infinito** e i loro prolungamenti convergono nel fuoco F' (Fig. 12.b).

La **distanza focale**, indicata con la lettera f , è la distanza fra la lente e il punto focale.



Figg. 12.a e 12.b

b) Rifrazione delle lenti

I raggi luminosi che attraversano una lente subiscono una **doppia rifrazione** sulle due facce della lente.

In generale, i raggi luminosi quando colpiscono una lente vengono riflessi in tutte le direzioni a seconda dell'angolo di incidenza, del materiale e della forma della lente.

Esistono, però, tre tipologie di raggi incidenti che hanno un comportamento piuttosto prevedibile:

- quelli che viaggiano parallelamente all'asse principale della lente;
- quelli che passano per il fuoco della lente;
- quelli che passano per il centro della lente.

c) Rifrazione da una lente convergente

• Raggi incidenti paralleli all'asse principale della lente

In questo caso tutti i raggi luminosi, quando colpiscono la prima faccia della lente, vengono rifratti verso la normale alla superficie di incidenza, a causa del passaggio dall'aria a un mezzo più denso.

All'interno della lente essi si propagano in linea retta fino a raggiungere la seconda faccia della lente, dove vengono nuovamente rifratti. Questa volta, però, passando da un mezzo più denso ad un mezzo meno denso, si allontanano dalla normale alla superficie per convergere nel **punto focale** della lente (Fig. 13).

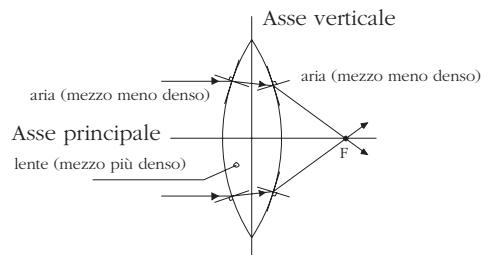


Fig. 13

• Raggi incidenti passanti per il fuoco della lente

Anche in questo caso i raggi incidenti subiscono una doppia rifrazione. Essi prima si avvicinano e poi si allontanano dalla normale alla superficie di incidenza, per poi uscire dalla lente in direzione **parallela** al suo asse principale (Fig. 14).

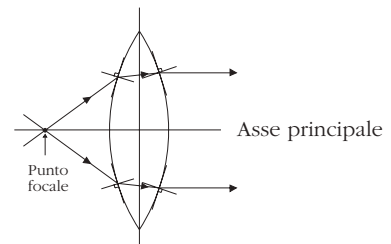


Fig. 14

• Raggi incidenti passanti per il centro della lente

Anche questi raggi subiscono una doppia rifrazione ma si compensano a vicenda nel senso che i raggi incidenti non modificano la loro direzione nel passaggio attraverso la lente (Fig. 15).

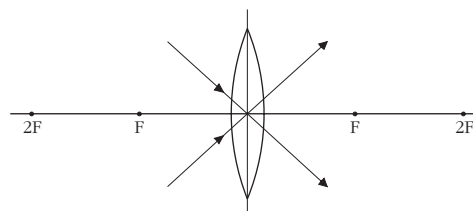


Fig. 15

d) Rifrazione da una lente divergente

• Raggi incidenti paralleli all'asse principale della lente

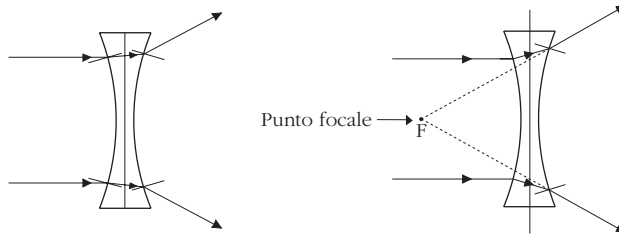
In questo caso tutti i raggi luminosi, quando colpiscono la prima faccia della lente, passando dall'aria a un mezzo più denso, vengono rifratti verso la normale alla superficie di incidenza.

All'interno della lente essi si propagano in linea retta fino a raggiungere la seconda faccia della lente dove vengono nuovamente rifratti. Questa volta, però, passando da un mezzo più denso ad un mezzo meno denso, si allontanano dalla normale alla superficie.

Il processo di rifrazione è quindi identico a quello della lente convergente ma in questo caso, per la diversa conformazione della lente, essi divergono ulteriormente (Fig. 16.a).

Se tracciamo i prolungamenti dei raggi rifratti essi si intersecano in un punto detto **punto focale**.

Da notare che in questo caso i raggi incidenti non producono un'immagine in quanto sono divergenti. Poiché il punto focale non è determinato dai raggi rifratti ma dai loro prolungamenti, la sua distanza focale è negativa (Fig. 16.b).



Figg. 16.a e 16.b

- Raggi incidenti passanti per il fuoco della lente

Abbiamo appena visto che la distanza focale di una lente divergente è negativa, per cui i raggi che viaggiano in direzione del punto focale non raggiungono mai tale punto in quanto vengono prima rifratti dalla superficie della lente.

In questo caso essi subiscono una doppia rifrazione ed escono dalla lente in direzione parallela all'asse principale (Fig. 17).

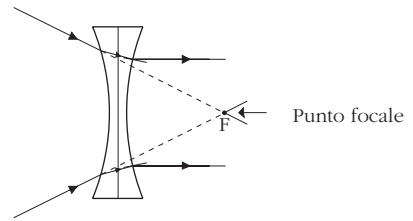


Fig. 17

- Raggi incidenti passanti per il centro della lente

In questo caso i raggi subiscono due rifrazioni che si compensano a vicenda e i raggi incidenti non modificano la loro direzione nel passaggio attraverso la lente (Fig. 18).

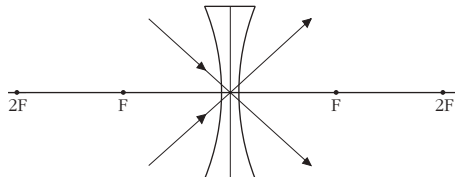


Fig. 18

1 L'interferenza

L'**interferenza** si verifica quando due o più onde dello stesso tipo, della stessa frequenza e aventi una differenza di fase costante si sovrappongono in un punto dello spazio.

Per spiegare il fenomeno dell'interferenza dobbiamo innanzitutto definire il concetto di **onde coerenti**.

a) Onde coerenti

Le onde luminose che arrivano in un dato punto dello spazio possono essere descritte mediante delle oscillazioni armoniche:

$$y(t) = A \cos \varphi(t) = A \cos(\omega t + \varphi_0) = A \cos(2\pi f t + \varphi_0)$$

dove f è la frequenza e φ_0 è la fase iniziale dell'onda.

Se in un dato punto incidono due onde di ampiezza A_1 e A_2 , di uguale frequenza e, quindi, di uguale pulsazione ω :

$$y_1 = A_1 \cos \varphi_1(t) = A_1 \cos(\omega t + \varphi_{01})$$

$$y_2 = A_2 \cos \varphi_2(t) = A_2 \cos(\omega t + \varphi_{02})$$

per il principio di sovrapposizione, l'onda risultante ha la stessa frequenza delle onde componenti e sarà, in ogni punto e in ogni istante, uguale alla somma delle due onde componenti:

$$y(t) = y_1(t) + y_2(t) = A_1 \cos(\omega t + \varphi_{01}) + A_2 \cos(\omega t + \varphi_{02}) = A \cos(\omega t + \varphi) \quad (1.1)$$

Poiché le due onde hanno uguale frequenza, la differenza di fase:

$$\Delta \varphi = (\omega t + \varphi_{01}) - (\omega t + \varphi_{02}) = \varphi_{01} - \varphi_{02} \quad (1.2)$$

non dipendendo dal tempo, si mantiene costante e le onde sono dette **coerenti**.

Dall'uguaglianza fra il secondo e il terzo membro della (1.1):

$$A_1 \cos(\omega t + \varphi_{01}) + A_2 \cos(\omega t + \varphi_{02}) = A \cos(\omega t + \varphi)$$

si ottiene, sviluppando i coseni:

$$\begin{aligned} A_1 \cos \omega t \cos \varphi_{01} - A_1 \sin \omega t \sin \varphi_{01} + A_2 \cos \omega t \cos \varphi_{02} - A_2 \sin \omega t \sin \varphi_{02} = \\ = A \cos \omega t \cos \varphi - A \sin \omega t \sin \varphi \end{aligned}$$

Raccogliamo i termini in $\sin \omega t$ e $\cos \omega t$:

$$\cos \omega t (A_1 \cos \varphi_{01} + A_2 \cos \varphi_{02} - A \cos \varphi) - \sin \omega t (A_1 \sin \varphi_{01} + A_2 \sin \varphi_{02} - A \sin \varphi) = 0 \quad (1.3)$$

La (1.3) è nulla soltanto se i coefficienti dei termini in $\sin \omega t$ e $\cos \omega t$ si annullano contemporaneamente e quindi soltanto se:

$$\begin{cases} A_1 \cos \varphi_{01} + A_2 \cos \varphi_{02} - A \cos \varphi = 0 \\ A_1 \sin \varphi_{01} + A_2 \sin \varphi_{02} - A \sin \varphi = 0 \end{cases}$$

da cui segue che:

$$\begin{cases} A \sin \varphi = A_1 \sin \varphi_{01} + A_2 \sin \varphi_{02} \\ A \cos \varphi = A_1 \cos \varphi_{01} + A_2 \cos \varphi_{02} \end{cases} \quad (1.4)$$

Dividendo membro a membro si ottiene:

$$\varphi = \arctg \frac{A_1 \sin \varphi_{01} + A_2 \sin \varphi_{02}}{A_1 \cos \varphi_{01} + A_2 \cos \varphi_{02}} \quad (1.5)$$

La (1.5) rappresenta la fase iniziale dell'onda risultante.

Quadrando e sommando la (1.4) si ottiene:

$$\begin{aligned} A^2 (\sin^2 \varphi + \cos^2 \varphi) &= A_1^2 \sin^2 \varphi_{01} + A_2^2 \sin^2 \varphi_{02} + 2A_1 A_2 \sin \varphi_{01} \sin \varphi_{02} + \\ &+ A_1^2 \cos^2 \varphi_{01} + A_2^2 \cos^2 \varphi_{02} + 2A_1 A_2 \cos \varphi_{01} \cos \varphi_{02} = \\ &= A_1^2 (\sin^2 \varphi_{01} + \cos^2 \varphi_{01}) + A_2^2 (\sin^2 \varphi_{02} + \cos^2 \varphi_{02}) + \\ &+ 2A_1 A_2 (\sin \varphi_{01} \sin \varphi_{02} + \cos \varphi_{01} \cos \varphi_{02}) \end{aligned}$$

Ricordando la formula fondamentale della trigonometria, $\sin^2 \varphi + \cos^2 \varphi = 1$, si ha che:

$$A^2 = A_1^2 + A_2^2 + 2A_1 A_2 (\sin \varphi_{01} \sin \varphi_{02} + \cos \varphi_{01} \cos \varphi_{02})$$

e, infine, ricordando la formula di sottrazione dei coseni, si ha:

$$A^2 = A_1^2 + A_2^2 + 2A_1 A_2 \cos(\varphi_{01} - \varphi_{02}) \quad (1.6)$$

La (1.6) rappresenta il quadrato dell'ampiezza dell'onda risultante, che è proporzionale all'energia dell'onda stessa.

Essa indica che se due onde coerenti incidono in un punto, in quel punto l'energia risultante non è soltanto pari alla somma delle energie delle onde incidenti in quanto esiste il termine:

$$2A_1 A_2 \cos(\varphi_{01} - \varphi_{02})$$

il quale può essere positivo o negativo.

Questo significa che tra due onde coerenti si verifica il *fenomeno dell'interferenza*, ossia *l'intensità di un'onda, ottenuta sovrapponendo due onde della stessa frequenza che oscillano nello stesso piano e si propagano nella stessa direzione, non è uguale alla somma delle intensità delle singole onde.*

Se le due onde incidenti avessero avuto la stessa ampiezza A_1 la (1.6) sarebbe diventata:

$$A^2 = 2A_1^2 [1 + \cos(\varphi_{01} - \varphi_{02})]$$

e quindi possiamo vedere che, a seconda della differenza di fase, l'onda risultante avrebbe avuto un'**ampiezza massima**:

$$A_{\max}^2 = 4A_1^2$$

per $\cos(\varphi_{01} - \varphi_{02}) = 1 \Rightarrow \varphi_{01} - \varphi_{02} = 2k\pi; \quad k \in N$

e un'**ampiezza minima**:

$$A_{\min}^2 = 0$$

per $\cos(\varphi_{01} - \varphi_{02}) = -1 \Rightarrow \varphi_{01} - \varphi_{02} = (2k+1)\pi; \quad k \in N.$

2 I metodi di osservazione dell'interferenza della luce

Per osservare l'interferenza della luce possiamo utilizzare uno schermo con fenditure dell'ordine della lunghezza d'onda della luce oppure dei corpi rifrangenti.

Le onde luminose vengono prodotte in seguito al processo di eccitazione - diseccitazione degli elettroni.

Ogni elettrone quando si diseccita, cioè quando passa a un livello energetico inferiore, emette un treno di onde luminose per un intervallo di tempo di circa 10^{-8} s.

Siccome le onde luminose viaggiano alla velocità della luce, tali treni sono approssimativamente lunghi 3m:

$$l = v \cdot t = c \cdot t = (3 \times 10^8 \text{ m/s}) \cdot (10^{-8} \text{ s}) = 3 \text{ m}$$

Il processo è del tutto aleatorio e quindi possiamo considerare gli elettroni come delle sorgenti di treni di onde luminose incoerenti.

Quindi l'unico modo per ottenere un'interferenza è facendo interferire un treno d'onde con se stesso.

a) L'esperimento della doppia fenditura di Young

Uno dei metodi con cui è possibile osservare l'interferenza delle onde luminose venne proposto nel 1801 da Thomas Young. Egli riuscì proprio con questo esperimento a dimostrare la natura ondulatoria della luce.

Fece incidere un fascio di luce monocromatica su uno schermo su cui erano incise due fenditure.

Le figure di interferenza si potevano osservare su un secondo schermo posto a una distanza molto maggiore della dimensione delle fenditure (Fig. 1).

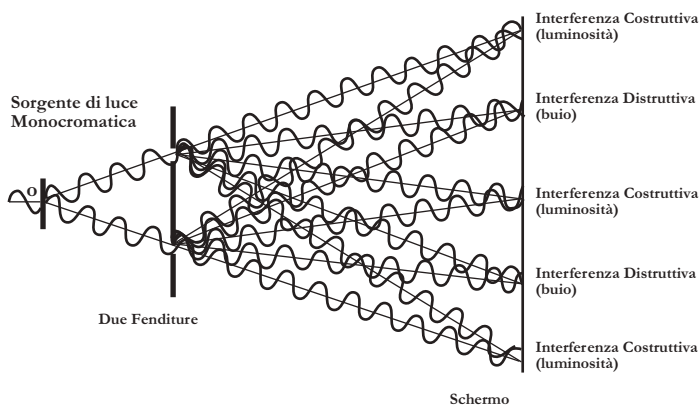


Fig. 1

Per il principio di Huygens, il quale afferma che ogni punto di una superficie può essere considerato come sorgente di onde sferiche secondarie, le due fenditure parallele diventavano, infatti, sorgenti di onde coerenti che potevano quindi interferire fra di loro.

A seconda del cammino ottico, le onde interferivano positivamente o negativamente e Young osservò che sullo schermo si alternavano zone di massima luminosità (interferenza costruttiva) a zone buie (interferenza distruttiva).

Vediamo che cosa accadde nell'esperimento di Young.

Nella figura 2 sono evidenziati due degli infiniti raggi generati dalle fenditure; studiamone l'interferenza.

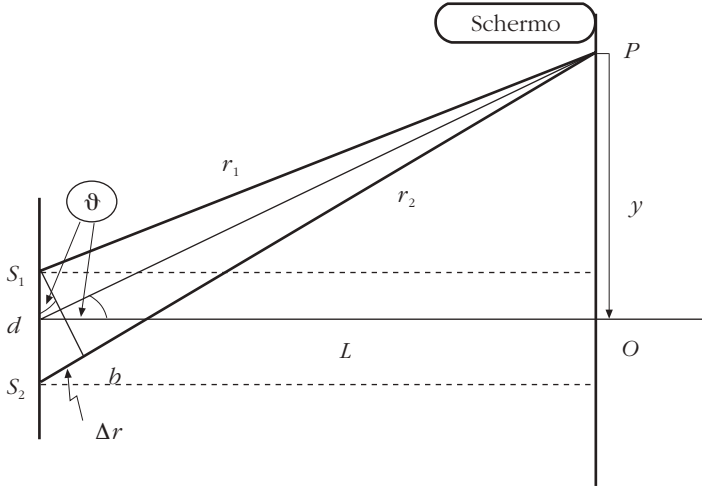


Fig. 2

Dalla figura si ha che:

$$r_1^2 = L^2 + \left(y - \frac{d}{2}\right)^2$$

$$r_2^2 = L^2 + \left(y + \frac{d}{2}\right)^2$$

Sottraendo membro a membro otteniamo:

$$r_2^2 - r_1^2 = L^2 + \left(y + \frac{d}{2}\right)^2 - \left[L^2 + \left(y - \frac{d}{2}\right)^2\right] = L^2 + y^2 + \frac{d^2}{4} + dy - L^2 - y^2 - \frac{d^2}{4} + dy = 2dy$$

Scomponendo il primo membro, si ha che:

$$(r_2 + r_1)(r_2 - r_1) = 2dy$$

Nel dispositivo di Young $L \gg d$ e quindi $r_2 + r_1 \approx 2L$, pertanto:

$$r_2 - r_1 \approx \frac{2dy}{2L} = y \frac{d}{L} \quad (2.1)$$

Al variare della distanza y del punto P varierà la fase:

$$\Delta\phi = 2\pi \left(\frac{r_2 - r_1}{2\lambda} \right)$$

e si avranno dei **massimi** per:

$$\cos 2\pi \left(\frac{r_2 - r_1}{2\lambda} \right) = \pm 1 \quad \text{cioè per: } r_2 - r_1 = k\lambda \quad \text{con } k = 0, \pm 1, \pm 2, \dots$$

e dei **minimi** per:

$$\cos 2\pi \left(\frac{r_2 - r_1}{2\lambda} \right) = 0 \quad \text{cioè per: } r_2 - r_1 = (2k + 1) \frac{\lambda}{2} \quad \text{con } k = 0, \pm 1, \pm 2, \dots$$

Introducendo la (2.1), si avranno dei massimi in:

$$y = k\lambda \frac{L}{d} \quad \text{con } k = 0, \pm 1, \pm 2, \dots$$

e dei minimi in:

$$y = (2k + 1)\lambda \frac{L}{2d} \quad \text{con } k = 0, \pm 1, \pm 2, \dots$$

La distanza fra due massimi o due minimi successivi è:

$$y_{k+1} - y_k = (k+1)\lambda \frac{L}{d} - k\lambda \frac{L}{d} = \lambda \frac{L}{d}$$

Vediamo come variano i massimi e i minimi di interferenza in funzione dell'angolo ϑ formato tra il raggio diretto al punto P e la retta perpendicolare tracciata dalla fenditura allo schermo.

Se lo schermo è posto a una grandissima distanza dalle fenditure (quindi se $d \ll L$), le rette che congiungono le due fenditure con il punto P dello schermo possono essere considerate quasi parallele (Fig. 3) e la differenza di cammino è pari a $d \sin \vartheta$.

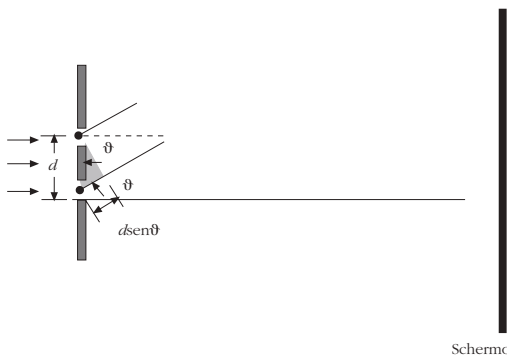


Fig. 3

Quindi si avranno massimi di interferenza per:

$$d \sin \vartheta = k\lambda \quad (2.2)$$

e minimi di interferenza per:

$$d \sin \vartheta = (2k + 1) \frac{\lambda}{2} \quad (2.3)$$

Se nel punto P si osserva un massimo, la relazione tra la coordinata di P e l'angolo ϑ è data da:

$$\tan \vartheta = \frac{y}{L}$$

Per ϑ molto piccoli possiamo approssimare la tangente con il seno di ϑ cioè:

$$\sin \vartheta \cong \tan \vartheta = \frac{y}{L} \quad (2.4)$$

Sostituendo questo risultato nella (2.2), otteniamo:

$$d \sin \vartheta = d \frac{y}{L} = k\lambda \quad (2.5)$$

e quindi:

$$y_m = k\lambda \frac{L}{d}$$

che è il risultato già trovato precedentemente.

Riscrivendo la (2.5) come:

$$\sin \vartheta = \frac{k\lambda}{d} \quad \text{con } k = 0, \pm 1, \pm 2, \dots$$

possiamo tracciare le frange d'interferenza e l'intensità in funzione di ϑ (Fig. 4).

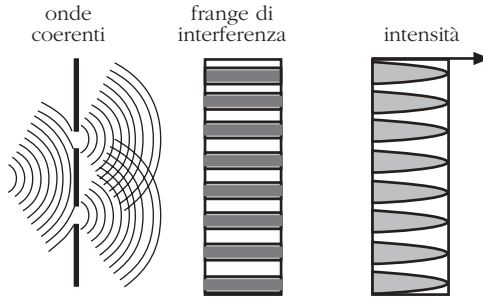


Fig. 4

3 La diffrazione

La **diffrazione** è un particolare fenomeno di interferenza.

Si può osservare sperimentalmente che quando un fascio di luce passa attraverso una fenditura, di dimensioni dell'ordine della lunghezza d'onda della radiazione incidente, viene allargato più di quanto ci si aspetterebbe con una costruzione geometrica del suo percorso.

a) Diffrazione da una singola fenditura (diffrazione di Fraunhofer)

Facciamo passare un fascio di raggi luminosi paralleli attraverso una singola fenditura avente una larghezza dell'ordine della lunghezza d'onda della luce incidente.

Per il principio di Huygens, ogni particella della singola fenditura diventa una generatrice di onde sferiche, le quali interferiscono formando una figura di diffrazione.

Consideriamo la fenditura della figura 5 di larghezza a .

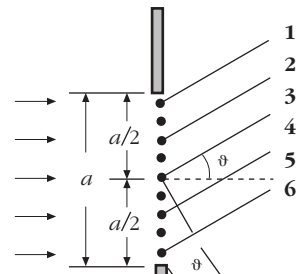


Fig. 5

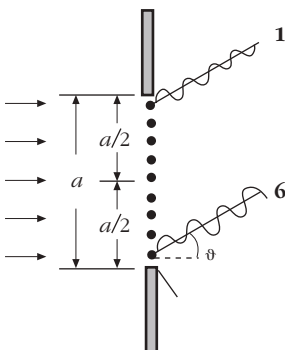


Fig. 6

In figura sono state disegnate 6 sorgenti puntiformi. Osserviamo la fenditura sotto un angolo ϑ per il quale le onde provenienti dall'estremità superiore e dall'estremità inferiore della fenditura sono in fase (Fig. 6).

Immaginiamo di dividere i 6 raggi della figura 6 in due gruppi, composti rispettivamente dai raggi 1, 2, 3 e dai raggi 4, 5, 6.

Il fatto che i raggi 1 e 6 sono in fase vuol dire che fra loro c'è una differenza di cammino di una lunghezza d'onda.

Questo vuol dire che fra i raggi 1 e 4 c'è una differenza di cammino di mezza lunghezza d'onda e le due onde interferiscono negativamente. Lo stesso succede fra i raggi 2 e 5, 3 e 6.

In altre parole, la luce che proviene da ogni coppia di sorgenti distanti tra loro $\frac{a}{2}$ si elidono e quindi, in corrispondenza di questo angolo, si ha una zona buia.

La prima frangia di interferenza distruttiva si ha per:

$$\frac{a}{2} \sin \vartheta = \frac{\lambda}{2}$$

e cioè per $\vartheta = \lambda$.

L'espressione più generale per gli zeri della figura di diffrazione è:

$$a \sin \vartheta = k\lambda \quad (3.1)$$

Le posizioni dei punti di interferenza costruttiva sono a metà fra le frange scure (Fig. 7).

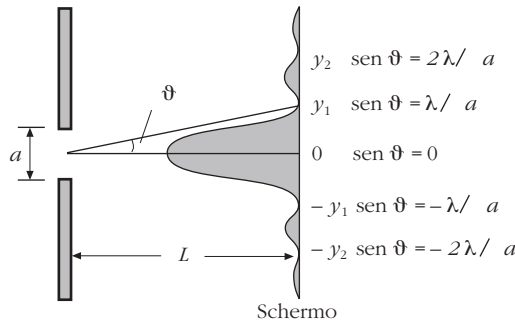


Fig. 7

In genere interessa soltanto la comparsa del primo minimo di intensità della luce perché quasi tutta l'energia è contenuta nel massimo centrale.

Dalla figura 7 possiamo osservare che la distanza y_1 tra il massimo centrale e il primo minimo di diffrazione è legata all'angolo ϑ dalla relazione:

$$\operatorname{tg} \vartheta = \frac{y_1}{L}$$

Poiché questo angolo è molto piccolo, possiamo scrivere:

$$\sin \vartheta = \frac{y_1}{L}$$

e, sostituendo in questa equazione la (3.1) con $k = 1$, si ottiene:

$$\sin \vartheta = \frac{y_1}{L} = \frac{\lambda}{a}$$

da cui segue che:

$$y_1 = \frac{L\lambda}{a}$$

b) Reticolo di diffrazione

Se facessimo passare un'onda piana attraverso uno schermo avente due fenditure, le frange di interferenza tenderebbero ad avvicinarsi e a diventare sempre più marcate soprattutto nella zona centrale.

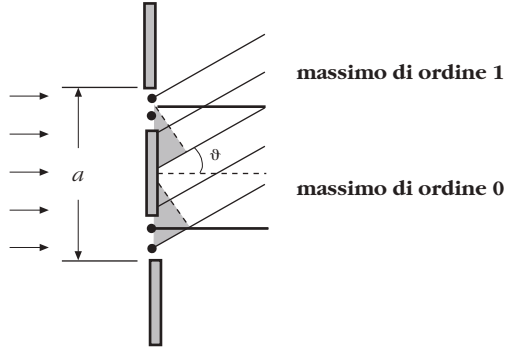


Fig. 8

Si possono poi migliorare ulteriormente le figure di diffrazione utilizzando un sistema di molte fenditure ottenendo così un **reticolo di diffrazione**.

La distanza fra due fenditure è detta **passo o costante** del reticolo.

Quando un'onda, avente lunghezza d'onda λ , incide su un reticolo di diffrazione di passo a , si osservano al di là del reticolo stesso dei massimi di intensità luminosa ad angoli ϑ_m rispetto alla normale tali che:

$$m\lambda = a \sin \vartheta_m$$

dove $m = 1, 2, 3, \dots$

1 Calore e temperatura

Il problema fondamentale che gli scienziati dovettero affrontare per costruire uno strumento in grado di misurare la temperatura era legato al fatto che la temperatura non è una grandezza direttamente misurabile.

Per misurare una lunghezza o un tempo basta avere un metro o un orologio e, a occhio, possiamo dire se un corpo è più o meno lungo di un altro. Per la temperatura non è così.

Due persone possono definire caldo o freddo lo stesso oggetto.

Se immergiamo la mano destra e la mano sinistra in due bacinelle contenenti dell'acqua a diversa temperatura e poi le immergiamo entrambe in una bacinella contenente dell'acqua ad una temperatura intermedia, non possiamo dire se l'acqua sia calda o fredda in quanto a una mano l'acqua sembrerà più calda e all'altra più fredda.

E tanto meno possiamo definire in modo univoco la temperatura dell'acqua nella bacinella.

2 Il termometro

Galileo fu forse il primo scienziato a realizzare nel 1592 uno strumento per misurare la temperatura e lo chiamò **termoscopio**.

Nel termoscopio di Galileo non esisteva però una scala graduata e quindi non era in grado di tradurre la temperatura di un corpo in un numero ma il suo termoscopio fu il padre di tutti i termometri.

Il primo a definire una suddivisione della scala termometrica fu Giovanni Sagredo, un contemporaneo di Galileo costruttore di strumenti scientifici alla corte dei Medici, il quale divise, inizialmente, la scala termometrica in 360 gradi, tanti quanti ce n'erano nell'angolo giro, e da qui deriva il fatto che anche la temperatura ha come **unità di misura il grado**.

Nel 1641 venne costruito il primo termometro ad alcool per il Granduca di Toscana, Ferdinando II, su cui vennero poste 50 tacche. A quei tempi non era ancora stata definita una scala termometrica universale e quindi la temperatura dipendeva dal termometro utilizzato.

Nel 1724, **Daniel Fahrenheit** costruì il primo termometro a mercurio. L'utilizzo del mercurio al posto dell'acqua o dell'alcool fu del tutto casuale ma i fisici compresero subito che avevano scoperto la sostanza giusta da utilizzare nei termometri. Il mercurio, infatti, si espande linearmente con la temperatura, non è trasparente e scivola sul vetro senza lasciare tracce. Egli propose inoltre di considerare come punti fissi la temperatura di una miscela di cloruro d'ammonio e ghiaccio (valore 0) e la temperatura di una persona adulta in buona salute (valore 100). La scala Fahrenheit è tuttora utilizzata in molti paesi anglosassoni.

Nel 1745, **Celsius** propose una scala i cui punti fissi erano la temperatura di fusione del ghiaccio (0°C) e la temperatura di ebollizione dell'acqua (100°C) e divise questa scala in 100 tacche.

Il fisico **Réaumur** utilizzò gli stessi punti fissi introdotti da Celsius. Assegnò il valore 0 alla temperatura di fusione del ghiaccio e il valore 80 alla temperatura di ebollizione dell'acqua e divise questa in 80 parti uguali.

Fahrenheit modificò la sua scala adottando anch'egli i punti fissi della scala Celsius. Per la scala Fahrenheit le temperature dei punti fissi sono: 32°F la temperatura di fusione del ghiaccio e 212°F la temperatura di ebollizione dell'acqua.

Nel 1933 venne scelto come punto fisso il punto triplo dell'acqua. Successivamente venne definita la scala **Kelvin** che fa corrispondere il punto triplo dell'acqua a 273,16K.

Ricapitolando quanto detto le scale termometriche utilizzate sono:

Scala Celsius

Valore 0: punto di congelamento dell'acqua
 Valore 100: punto di ebollizione dell'acqua
 Intervalli uguali: 100

Scala Réamur

Valore 0: punto di congelamento dell'acqua
 Valore 80: punto di ebollizione dell'acqua
 Intervalli uguali: 80

Scala Fahrenheit

Valore 32: punto di congelamento dell'acqua
 Valore 212: temperatura corporea del corpo umano
 Intervalli uguali: 180

Scala Kelvin

Valore 273,15: punto di congelamento dell'acqua
 Valore 373,15: temperatura corporea del corpo umano
 Intervalli uguali: 100

Le relazioni che intercorrono tra le scale sono:

$$\frac{C}{100} = \frac{R}{80} = \frac{F - 32}{180} = \frac{K - 273,15}{100}$$

dove con C , R , F e K è stata indicata la stessa temperatura espressa in gradi Celsius, Réamur, Fahrenheit e Kelvin.

3 Le sostanze termometriche

Le **sostanze termometriche** hanno delle proprietà variabili con la temperatura.

Si considerino alcuni esempi:

- molte sostanze si espandono quando vengono riscaldate e si contraggono quando vengono raffreddate;
- in un conduttore freddo scorre meno corrente che in uno caldo;
- alcune reazioni chimiche avvengono soltanto ad alte o a basse temperature;
- i solidi si liquefanno a temperature diverse;
- in alcuni oggetti il colore varia con la temperatura.

Ciascuna di queste proprietà può essere utilizzata per costruire un termometro.

a) Dipendenza della pressione dalla temperatura

La dipendenza della pressione di un gas dalla temperatura venne inizialmente studiata da un certo Amontons il quale nel 1699 scrisse:

«Due diverse masse di aria poste sotto lo stesso peso aumentano di uno stesso valore la loro spinta verso l'alto in seguito allo stesso aumento di temperatura».

L'aumento di una spinta verso l'alto è sinonimo di aumento di pressione ma ai tempi di Amontons il concetto di pressione non era ancora stato introdotto.

Questo vuol dire che se:

- P_0 è il valore della pressione a 0°C ;
- P_1 è il valore della pressione alla temperatura t_1 ;
- P_2 è il valore della pressione alla temperatura t_2 ;

si ha la relazione:

$$\beta = \frac{P_2 - P_1}{P_0(t_2 - t_1)} \quad (3.1)$$

dove β è il **coefficiente di pressione**.

Il valore del coefficiente di pressione può essere determinato misurando la pressione dell'aria alle due diverse temperature mantenendo il volume costante.

Effettuando misure diverse si è trovato che il valore di β è pressoché costante e pari a $\frac{1}{273}$ e questo vale non soltanto per l'aria ma anche per una categoria di gas, detti **gas perfetti**, che studieremo in seguito, i quali si liquefanno a temperature talmente basse che sono, o meglio non erano, raggiungibili.

b) Zero assoluto

L'equazione (3.1) può essere modificata supponendo che la temperatura maggiore sia di t gradi, a cui corrisponde una pressione P_t , e la temperatura minore sia di 0 gradi, a cui corrisponde una pressione P_0 ; allora possiamo scrivere che:

$$\beta = \frac{P_t - P_0}{P_0(t - 0)}$$

da cui segue che:

$$P_t - P_0 = \beta P_0 t$$

e quindi:

$$P_t = P_0(1 + \beta t) \quad (3.2)$$

siccome $\beta = \frac{1}{273}$, la (3.2) diventa:

$$P_t = P_0 \left(1 + \frac{1}{273} t \right)$$

Se poniamo nella (3.2) $t = -273$ si avrebbe che:

$$P_t = P_0(1 - 1) = 0$$

e quindi la pressione si annulla alla temperatura di -273°C .

Questa considerazione ha fatto nascere il concetto di **zero assoluto** sul quale si basa la Scala Kelvin, la quale fissa il proprio valore a -273°C :

$$0\text{K} = -273^\circ\text{C}$$

4 Gli effetti della temperatura sul volume dei gas

Si può dimostrare che se un gas viene riscaldato a pressione costante il suo volume aumenta proporzionalmente con la temperatura, secondo la relazione:

$$\alpha = \frac{V_2 - V_1}{V_0(t_2 - t_1)} \quad (4.1)$$

dove:

- V_0 è il valore del volume del gas a 0°C ;
- V_1 è il valore del volume del gas t_1 ;
- V_2 è il valore del volume del gas t_2 ;

e α è il **coefficiente di dilatazione cubica**, che vale $\frac{1}{273}$ (di esso ci occuperemo nel corso di questo capitolo).

5 Gli effetti della temperatura sul volume e sulla pressione di un liquido

Per i liquidi non ha senso parlare di un coefficiente di pressione in quanto essi, per definizione, sono incompressibili. Ha senso, invece, parlare di coefficiente di dilatazione cubica in quanto la (4.1) vale anche per i liquidi anche se $\alpha_{\text{liquidi}} < \alpha_{\text{gas}}$ e inoltre varia a seconda del liquido utilizzato.

In tabella 1 sono riportati alcuni valori dei coefficienti di dilatazione cubica dei liquidi:

Materiale	$\alpha(^{\circ}\text{C}^{-1})$
Acetone	$14,9 \times 10^{-4}$
Acqua	$2,1 \times 10^{-4}$
Alcool	$11,2 \times 10^{-4}$
Benzolo	$12,4 \times 10^{-4}$
Glicerina	$5,1 \times 10^{-4}$
Mercurio	$1,82 \times 10^{-4}$
Petrolio	$9,5 \times 10^{-4}$
Benzina	$9,6 \times 10^{-4}$
Aria	$3,67 \times 10^{-3}$
Elio	$3,66 \times 10^{-3}$
Olio d'oliva	$0,74 \times 10^{-3}$

Tab. 1

6 Gli effetti della temperatura sul volume e sulla pressione di un solido

Anche per i solidi, essendo incompressibili, non ha senso definire un coefficiente di pressione.

a) Dilatazione lineare

Consideriamo la barretta di figura 1, supponiamo che essa si dilati lungo una sola direzione al variare della temperatura e studiamone la dilatazione lineare.

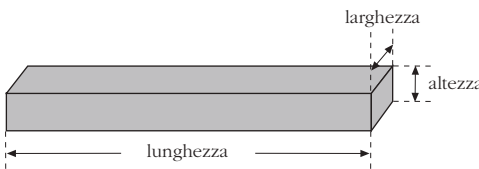


Fig. 1

Supponiamo che la barretta abbia lunghezza:

- L_0 per $t = 0^\circ\text{C}$;
- L_1 per $t = t_1$;
- L_2 per $t = t_2$.

Possiamo definire il coefficiente di dilatazione lineare λ come:

$$\lambda = \frac{L_2 - L_1}{L_0(t_2 - t_1)} \quad (6.1)$$

e supponendo che $t_1 = 0$ e $t_2 = t$, la (6.1) diventa:

$$L_t = L_0(1 + \lambda t)$$

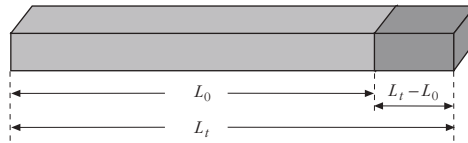


Fig. 2

In tabella 2 sono riportati alcuni valori dei coefficienti di dilatazione lineare dei solidi:

Materiale	$\alpha(^{\circ}\text{C}^{-1})$
Acciaio	$1,2 \times 10^{-5}$
Alluminio	$2,4 \times 10^{-5}$
Cemento	$1,2 \times 10^{-5}$
Ferro	$9,1 \times 10^{-6}$
Ottone	$1,9 \times 10^{-5}$
Piombo	$2,9 \times 10^{-5}$
Rame	$1,7 \times 10^{-5}$
Vetro	$1,0 \times 10^{-5}$
Quarzo	7×10^{-7}

Tab. 2

b) Dilatazione superficiale

Se la barretta di figura 1 si dilata in due direzioni con lo stesso coefficiente di dilatazione lineare possiamo definire un coefficiente di dilatazione superficiale σ come:

$$\sigma = \frac{S_2 - S_1}{S_0(t_2 - t_1)} \quad (6.2)$$

dove:

- S_0 è la superficie della barretta per $t = 0^\circ\text{C}$;
- S_1 è la superficie della barretta per $t = t_1$;
- S_2 è la superficie della barretta per $t = t_2$.

Se $t_2 = t$ e $t_1 = 0$, $S_2 = S_t$ e $S_1 = S_0$, possiamo riscrivere la (6.2):

$$S_t = S_0(1 + \sigma \Delta t)$$

c) Dilatazione cubica

Se la barretta di figura 1 si dilata in tre direzioni con lo stesso coefficiente di dilatazione lineare possiamo definire un **coefficiente di dilatazione cubica** α come:

$$\alpha = \frac{V_2 - V_1}{V_0(t_2 - t_1)}$$

dove:

- V_0 è il volume della barretta per $t = 0^\circ\text{C}$;
- V_1 è il volume della barretta per $t = t_1$;
- V_2 è il volume della barretta per $t = t_2$.

In generale, se $t_1 = 0$ e $V_1 = V_0$, si ha che:

$$V_t = V_0(1 + \alpha\Delta t)$$

Il coefficiente di dilatazione cubica è il triplo del coefficiente di dilatazione lineare cioè:

$$\alpha = 3\lambda$$

1 Calore, temperatura ed equilibrio termico

La **temperatura** di un corpo è legata all'energia cinetica media delle particelle che lo compongono e quindi alla loro velocità. Maggiore è la temperatura di un corpo e maggiore è la velocità delle sue particelle.

Se mettiamo in contatto un corpo più caldo con un corpo più freddo sulla superficie di separazione tra i due corpi le particelle più veloci del corpo caldo urtano quelle più lente del corpo freddo.

Questo provoca da un lato il rallentamento e quindi la diminuzione di energia cinetica media delle particelle del corpo più caldo e dall'altro l'accelerazione e l'aumento dell'energia cinetica media delle particelle del corpo più freddo. Il risultato che si nota a livello macroscopico è un raffreddamento del corpo più caldo e un riscaldamento del corpo più freddo e quindi un trasferimento di calore dal corpo più caldo a quello più freddo.

Al termine di questo passaggio di calore, i due corpi sono detti in **equilibrio termico** (Fig. 1).

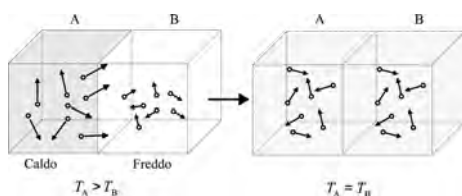


Fig. 1

Da questo semplice esperimento possiamo dedurre che:

- la temperatura è legata all'energia di un corpo;
- il calore è legato al trasferimento di energia fra due corpi in seguito ad una differenza di temperatura.

Quando i due corpi raggiungono la stessa temperatura non si ha più un passaggio di calore.

Un corpo non «possiede» il calore, un corpo **trasferisce** calore e, poiché il trasferimento di calore significa un trasferimento di energia, la sua unità di misura è il **joule**.

Il concetto di equilibrio termico è strettamente connesso con il **principio zero della termodinamica**.

a) Principio zero della termodinamica

Se due corpi A e B sono separatamente in equilibrio termico con un terzo corpo C, allora A e B sono in equilibrio termico fra loro.

La taratura dei termometri si basa su questo principio.

2 Capacità termica e calore specifico di un corpo

Nel corso della Storia sono state scritte diverse teorie sulla relazione tra il calore e la temperatura di un corpo; si è pensato che essa dipendesse dal volume del corpo, poi dal peso del corpo ma queste ipotesi sono state smentite dagli esperimenti.

L'aumento di temperatura di una sostanza in seguito all'assorbimento di una certa quantità di calore dipende dal tipo di sostanza e dalla quantità di sostanza considerata.

Joseph Black, un fisico appartenente al gruppo degli Universitari di Edimburgo, fu il primo a ipotizzare che la relazione tra la quantità di calore che un corpo riesce ad assorbire o a fornire e la conseguente variazione di temperatura è una proprietà intrinseca del corpo stesso e chiamò questa proprietà **capacità termica**.

a) Capacità termica

La **capacità termica** (C) di un corpo è definita come il rapporto tra la quantità di calore fornita al corpo e il corrispondente aumento di temperatura:

$$C = \frac{\Delta Q}{\Delta T}$$

Essa indica la quantità di calore necessaria per far aumentare di un grado la temperatura del corpo.

b) Calore specifico

Il **calore specifico** (c) è la capacità termica riferita all'unità di massa del materiale:

$$c = \frac{\Delta Q}{m\Delta T}$$

e si misura in $\text{J/Kg} \cdot \text{K}$

A volte conviene riferirsi al **calore specifico molare**, in cui la massa viene sostituita dal numero di moli.

Se n è il numero di moli di una certa sostanza, il calore necessario per far aumentare di 1°C 1 mole della sostanza è detto calore specifico molare. Esso è definito da:

$$C_M = \frac{\Delta Q}{n\Delta T}$$

c) Caloria

La **caloria** è l'unità di misura del calore.

Essa è la quantità di calore che deve essere fornita a un grammo d'acqua distillata per elevare la sua temperatura da $14,5^\circ\text{C}$ a $15,5^\circ\text{C}$.

Spesso si utilizza anche la **chilocaloria** che è la quantità di acqua necessaria per innalzare di 1°C un chilogrammo di acqua.

Nella tabella 1 sono stati riportati i calori specifici di alcune sostanze.

Sostanza	cal/g · °C	Sostanza	cal/g · °C
Alluminio	0,215	Argento	0,056
Berillio	0,436	Ottone	0,92
Rame	0,0924	Legno	0,41
Germanio	0,077	Ghiaccio (-5°C)	0,50
Oro	0,0308	Marmo	0,21
Ferro	0,107	Alcool (etilico)	0,58
Piombo	0,0305	Mercurio	0,033
Silicio	0,168	Acqua (15°C)	1,00

Tab. 1

3 Energia, calore e lavoro

Nel capitolo precedente abbiamo visto che la quantità di calore fornita ad un corpo ne fa aumentare la temperatura e quindi l'energia termica. Ciò sta a significare che il calore può essere considerato come una forma di energia.

Per capire meglio la relazione tra calore e energia facciamo il seguente esperimento.

Prendiamo un cilindro pieno di gas e aumentiamone l'energia operando in due modi diversi:

- scaldando il cilindro e cioè fornendo del **calore al gas** (Fig. 2);

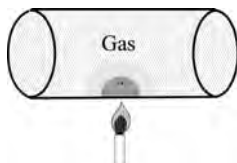


Fig. 2

- comprimendo il gas con un pistone e cioè compiendo del **lavoro** sul gas (Fig. 3).

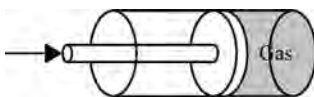


Fig. 3

Più in generale, possiamo dire che un sistema può scambiare energia con l'ambiente mediante:

- il **calore** scambiato;
- il **lavoro** eseguito dal sistema sull'ambiente.

Dopo aver variato la sua energia il sistema infatti *non ricorda* in che modo la sua energia è cambiata.

Essendo calore, lavoro ed energia delle grandezze equivalenti devono avere la stessa unità di misura.

Ma il calore si misura in calorie mentre il lavoro e l'energia si misurano in joule. Come possiamo fare a convertire le calorie in joule?

Per la conversione dell'unità di misura del calore in quella dell'energia dobbiamo ringraziare il fisico inglese Prescott Joule (1818-1889) che trovò sperimentalmente a quanti joule di lavoro corrispondeva una caloria di calore.

Egli trovò sperimentalmente l'**equivalente** tra **calore** e **lavoro meccanico**.

L'apparato di Joule è composto da pesi discendenti che, cadendo, mettono in rotazione delle palette poste in un calorimetro che, ruotando, scaldano l'acqua in cui sono immerse.

Nella figura 4 è mostrato l'esperimento.

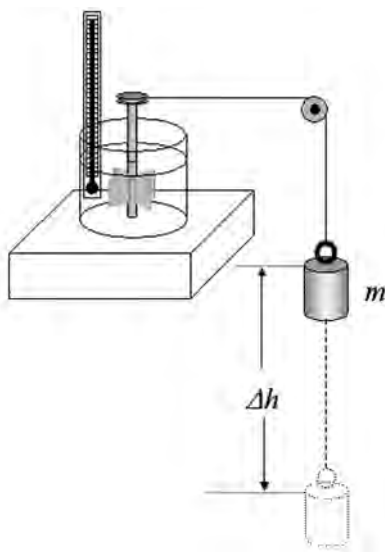


Fig. 4

Il lavoro fatto dalle palette sull'acqua è uguale alla variazione dell'energia potenziale $mg\Delta h$ della massa attaccata al filo:

$$L = \Delta E_p = mg\Delta h \quad \text{in joule}$$

dove:

- m = massa del peso;
- g = accelerazione di gravità;
- Δh = spostamento del peso.

Questo lavoro fa aumentare la temperatura dell'acqua di una certa temperatura ΔT .

Lo stesso aumento di temperatura si sarebbe potuto ottenere fornendo una quantità di calore pari a $Q = cm\Delta T$ in calorie, dove c è il calore specifico (in calorie) dell'acqua e m è la massa d'acqua nel calorimetro.

Ricordando che il calore specifico dell'acqua è unitario, calcolando L e misurando ΔT si ottiene che:

$$1\text{cal} = 4,186\text{J}$$

Questo è l'equivalente meccanico del calore.

La conversione tra calorie e joule è:

- $1\text{cal} = 4,186\text{J}$
- $1\text{kcal} = 4186\text{J} = 4,186\text{kJ}$

4 La propagazione del calore

Il calore si può propagare per convezione, conduzione o irraggiamento:

- la conduzione avviene per contatto tra due corpi;
- la convezione avviene per trasporto di materia fluida;
- l'irraggiamento avviene per mezzo di onde elettromagnetiche.

a) Conduzione

La **conduzione** è l'unico modo con cui il calore si propaga nei solidi.

Prendiamo per esempio la bacchetta in figura 5:

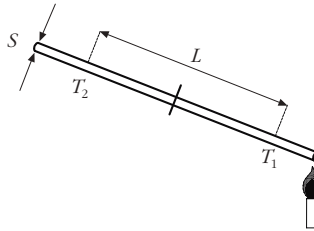


Fig. 5

Le molecole poste all'estremità più vicina alla fiamma, vibrando con un'ampiezza maggiore attorno alla loro posizione di equilibrio, urtano le molecole vicine cedendo parte di questa energia e il calore si propaga dall'estremità più calda verso l'estremità più fredda.

Consideriamone una porzione di lunghezza L , sezione S e con una differenza di temperatura fra gli estremi pari a $\Delta T = T_1 - T_2$.

Si può dimostrare che la quantità di calore Q trasmesso nell'unità di tempo, ossia la **potenza termica specifica**, è definita dalla relazione:

$$\frac{Q}{\Delta t} = \lambda \frac{S(T_1 - T_2)}{L} \quad (4.1)$$

dove λ è detto **coefficiente di conducibilità termica**.

La (4.1) può anche essere scritta come:

$$\frac{Q}{\Delta t} = KS\Delta T$$

dove $K = \frac{\lambda}{L}$ è detto **coefficiente globale di scambio termico** e tiene conto sia del coefficiente di conducibilità λ del materiale sia della geometria che influenza lo scambio termico (generalmente la lunghezza).

- Densità di flusso termico

La **densità di flusso termico** q è il calore trasmesso nell'unità di tempo e di spazio, cioè:

$$q = \frac{Q}{S\Delta t} = K\Delta T = \frac{\lambda}{L}\Delta T$$

La sua unità di misura è espressa in W/m^2 .

b) Convezione

La **convezione** è uno dei modi con cui il calore si propaga nei liquidi e nei gas. A differenza della conduzione, che avviene soltanto con il trasporto di calore, in questo caso il trasporto di calore è associato al trasporto di materia.

Prendiamo come esempio un bollitore e vediamo che cosa succede all'acqua in esso contenuta.

La parte di acqua a contatto con la base del recipiente si scalda e si espande. Espandendosi, diminuisce la propria densità e, per la spinta di Archimede, sale verso l'alto, facendo posto allo strato di acqua fredda sovrastante che, riscaldata a sua volta, diviene più leggera e risale, e così via.

In questo modo si scalda tutta l'acqua contenuta nel bollitore.

c) Irraggiamento

Nell'**irraggiamento** il calore viene trasmesso mediante le onde elettromagnetiche e quindi, in questo caso, la trasmissione avviene anche nel vuoto. Questo è il modo con cui il sole ci riscalda.

Poiché l'irraggiamento avviene mediante onde elettromagnetiche, le stesse che compongono la luce, possiamo studiare la conduzione per irraggiamento riprendendo ciò che è stato detto nei capitoli sulle onde.

Un uomo sta prendendo il sole, viene cioè colpito dalle radiazioni elettromagnetiche emesse dal sole le quali avranno una certa energia proporzionale al quadrato della loro ampiezza di oscillazione.

Quando le radiazioni colpiscono l'uomo in parte verranno riflesse, in parte verranno assorbite e in parte verranno trasmesse.

Egli, quindi, riflettendo le onde elettromagnetiche, irraggia a sua volta l'ambiente circostante con una sua potenza di emissione data dalla **legge di Stefan-Boltzmann**:

$$P = \sigma A e (T_{\text{corpo}} - T_{\text{ambiente}})^4$$

dove:

- P è la sua potenza irradiata;
- $\sigma = 5,6696 \times 10^{-8} \text{ W/m}^2\text{k}^2$ è una costante detta **costante di Boltzmann**;
- A è l'area della superficie del suo corpo;
- e è il suo coefficiente di emissione;
- T_{corpo} è la sua temperatura corporea;
- T_{ambiente} è la temperatura dell'ambiente circostante.

Un corpo in grado di assorbire tutta la radiazione incidente è detto **corpo nero**. Quando un corpo nero viene riscaldato, la sua emissività è unitaria ($e = 1$) e quindi la sua potenza di emissione è:

$$P_{\text{corpo nero}} = \sigma A T^4$$

5 Calore e cambiamenti di stato

I possibili stati di aggregazione della materia sono tre: solido, liquido e gassoso.

Lo stato in cui si trova una certa sostanza è il risultato di una sorta di «competizione» fra due diverse energie che si sviluppano a livello microscopico: l'energia di legame e l'energia cinetica.

Per capire meglio queste due energie dobbiamo ricordarci che la materia è composta da atomi i quali, ad eccezione dei gas nobili, tendono ad aggregarsi tra loro con dei legami chimici. La forza attrattiva di questi legami chimici è detta **energia di legame**.

Questi atomi però non sono fermi ma oscillano intorno alla loro posizione di equilibrio e questo vuol dire che essi posseggono anche un'energia cinetica.

Da una parte l'energia di legame tenderà a compattare le molecole della materia mentre dall'altra l'energia cinetica di oscillazione tenderà ad allontanarle.

Variando opportunamente la pressione, il volume o la temperatura possiamo modificare lo stato della sostanza.

Con il termine **passaggio di stato** si indica il processo attraverso il quale la materia cambia di stato.

Nella figura 6 sono mostrati i possibili cambiamenti di stato:

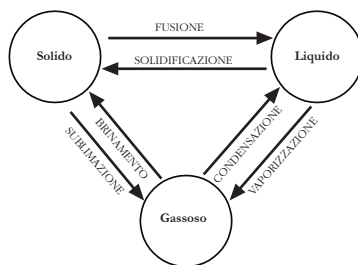


Fig. 6

a) Fusione e solidificazione

La **fusione** indica il passaggio della materia dallo stato solido allo stato liquido.

La **solidificazione** indica il passaggio della materia dallo stato liquido a quello solido.

La temperatura in entrambi i processi si mantiene costante e dipende dalla pressione a cui avviene la trasformazione.

b) Condensazione e vaporizzazione

La **condensazione** indica il passaggio della materia dallo stato gassoso allo stato liquido.

La **vaporizzazione**, invece, indica il passaggio della materia dallo stato liquido allo stato gassoso.

In realtà, possono avvenire due tipi di fenomeni:

- l'**evaporazione** che interessa soltanto la superficie del liquido e si manifesta a qualsiasi temperatura;
- l'**ebollizione** che interessa l'intera massa di liquido e si verifica soltanto a una determinata temperatura.

• Evaporazione e vapore saturo

Ogni liquido tende a passare allo stato gassoso e questa tendenza aumenta con la temperatura.

Se continuiamo a fornire calore ad un liquido contenuto in un sistema aperto esso poco per volta si trasformerà in vapore.

Se invece il liquido è posto in un sistema chiuso non sempre potrà evaporare del tutto; ad un certo punto si creerà una sorta di equilibrio dinamico tra la quantità di molecole che evaporano, passando dallo stato liquido a quello gassoso, e quelle che condensano, che passano cioè dallo stato gassoso a quello liquido.

In questa condizione la quantità di vapore che sovrasta il liquido è la massima compatibile con la temperatura a cui si trova il sistema e pertanto si è in presenza di vapor saturo.

La pressione esercitata dal vapore saturo è detta **pressione (o tensione) di vapor saturo** ed aumenta con la temperatura.

• Ebollizione

Anche l'**ebollizione** indica il passaggio di una sostanza dallo stato liquido a quello gassoso ma, a differenza dell'evaporazione che avveniva soltanto a livello superficiale, l'ebollizione interessa l'intera massa di liquido.

Durante l'ebollizione si formano delle bolle di gas all'interno del liquido.

Supponiamo, infatti, di riscaldare un recipiente mettendolo sul fuoco (Fig. 7).

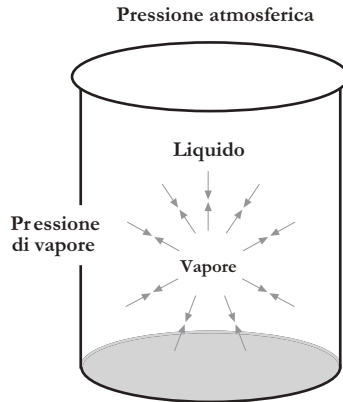


Fig. 7

Appena la tensione di vapore raggiunge quella atmosferica, l'acqua a contatto con il fuoco è in grado di generare del vapore a una pressione tale da «scostare» il liquido circostante formando delle bolle di vapore all'interno del liquido.

La temperatura di ebollizione dipende, quindi, dalla pressione a cui si trova il liquido.

6 Il calore latente

Quando il calore fornito, o sottratto, a una sostanza ne fa variare lo stato, entrano in gioco i **calori latenti** che, a seconda del passaggio di stato, sono detti:

- **calore latente di solidificazione:** è la quantità di calore che bisogna sottrarre a 1 kg del corpo affinché questo passi dallo stato liquido allo stato solido, senza che la sua temperatura diminuisca;
- **calore latente di fusione:** è la quantità di calore che bisogna fornire a 1 kg del corpo affinché questo passi dallo stato solido allo stato liquido, senza che la sua temperatura aumenti. Il calore latente di fusione dell'acqua a 0°C è $\lambda_f = 80 \text{ cal/g}$;
- **calore latente di evaporazione:** è la quantità di calore che bisogna fornire a 1 kg del corpo affinché questo passi dallo stato liquido allo stato gassoso, senza che la sua temperatura aumenti.

Tale quantità varia a seconda della natura del liquido e della sua temperatura.

Il calore latente di evaporazione dell'acqua a 0°C è $\lambda_e = 540 \text{ cal/g}$;

- **calore latente di liquefazione:** è la quantità di calore che bisogna sottrarre a 1 kg del corpo affinché questo passi dallo stato gassoso allo stato liquido, senza che la sua temperatura diminuisca.

Il calore latente di liquefazione è l'opposto del calore latente di vaporizzazione, cioè:

$$\lambda_l = -\lambda_e$$

In tabella sono mostrati alcuni calori latenti di fusione e di vaporizzazione (da notare che sono espressi in J/kg).

Sostanza	Punto di fusione (°C)	Calore latente di fusione (J/kg)	Punto di ebollizione (°C)	Calore latente di vaporizzazione (J/kg)
Elio	-269,65	$5,23 \times 10^3$	-268,93	$2,09 \times 10^4$
Azoto	-209,97	$2,55 \times 10^4$	-195,81	$2,01 \times 10^5$
Ossigeno	-218,79	$1,38 \times 10^4$	-182,97	$2,13 \times 10^5$
Alcool etilico	-114	$1,04 \times 10^5$	78	$8,54 \times 10^5$
Acqua	0,00	$3,33 \times 10^5$	100,00	$2,26 \times 10^6$
Zolfo	119	$3,81 \times 10^4$	444,60	$3,26 \times 10^5$
Piombo	327,3	$2,45 \times 10^4$	1750	$8,70 \times 10^5$
Alluminio	660	$3,97 \times 10^5$	2450	$1,14 \times 10^7$
Argento	960,80	$8,82 \times 10^4$	2193	$2,33 \times 10^6$
Oro	1063,00	$6,44 \times 10^4$	2660	$1,58 \times 10^6$
Rame	1083	$1,34 \times 10^5$	1187	$5,06 \times 10^5$

Tab. 2

1 Atomo, molecola e mole

Per comprendere il comportamento di un gas perfetto è necessario introdurre i seguenti concetti.

- La **molecola** (dal latino, «piccola quantità») è la più piccola unità strutturale di una sostanza.
- La **mole** è la quantità di sostanza contenente $6,02 \times 10^{23}$ **molecole**:

$$1\text{Mol} = 6,02 \times 10^{23} \text{ molecole}$$

dove $6,02 \times 10^{23}$ è un numero costante (uguale al numero di atomi di carbonio contenuti in 0,012 Kg di carbonio 12), detto **numero di Avogadro**, ed è indicato con N_A .

- Il **peso atomico** di una sostanza equivale al peso di un atomo trascurando gli elettroni. Per esempio, l'idrogeno ha peso atomico 1 perché è composto da un protone e da un elettrone (di cui si trascura la massa). È importante sottolineare che il peso atomico non ha niente a che fare con il peso gravitazionale dell'atomo; esso ha più un significato in senso statistico del peso relativo fra atomi diversi.
- Il **peso molecolare** di una sostanza equivale al peso di una molecola trascurando gli elettroni. L'ossigeno è un gas biatomico e ha peso molecolare 16 in quanto ha 8 protoni e 8 neutroni.
- Una **mole di sostanza** è uguale al peso molecolare espresso in grammi di quella sostanza. Una mole di acqua pesa 18g in quanto è composta da due atomi di idrogeno e uno di ossigeno.

2 Le proprietà dei gas perfetti

Un gas, per essere considerato **ideale**, deve avere le seguenti proprietà:

- deve essere costituito da un numero enorme di particelle identiche;
- le particelle del gas devono essere pensate come delle sferette rigide, indeformabili e di dimensioni puntiformi;
- tutte le direzioni in cui si muovono le particelle devono essere equiprobabili;
- non devono esistere forze di interazione fra le particelle;
- gli urti sia fra le particelle del gas che con le pareti del contenitore devono essere perfettamente elastici.

In realtà, i gas perfetti non esistono però esiste una categoria di gas, i cosiddetti **gas nobili**, il cui comportamento è molto simile a quello di un gas perfetto.

Un'ulteriore differenza fra un gas reale e un gas ideale è che al di sotto di una certa temperatura, detta **temperatura critica**, i gas reali cambiano di stato liquefacendosi mentre questo non avviene per un gas ideale.

3 Le variabili termodinamiche

Per conoscere lo stato termodinamico di un oggetto dobbiamo definirne le sue **variabili termodinamiche** e cioè quelle grandezze macroscopiche che globalmente possono rappresentare una certa quantità di gas.

Gli esperimenti effettuati sui gas perfetti hanno dimostrato che, per descrivere completamente il comportamento di un gas, sono sufficienti quattro variabili:

1. la pressione (p);
2. il volume (V);
3. la temperatura (T);
4. il numero di moli (n).

e che in realtà le variabili indipendenti sono soltanto **tre**.

Quindi, se fissiamo la pressione, il numero di moli e la temperatura di un gas, il suo volume sarà definito da una funzione del tipo:

$$V = f(n, p, T) \quad (3.1)$$

Analogamente, se fissiamo la temperatura, il numero di moli e il volume di un gas, la sua pressione sarà definita da una funzione del tipo:

$$p = g(n, T, V) \quad (3.2)$$

La (3.1) e la (3.2) sono dette **equazioni di stato**.

4 Trasformazioni termodinamiche e leggi dei gas perfetti

Un processo termodinamico, detto anche **trasformazione termodinamica**, è un cammino sulla superficie descritto dall'equazione di stato ed è formato da una successione infinita di stati termodinamici.

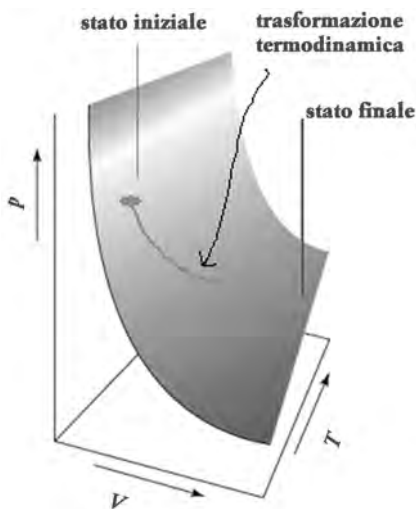


Fig. 1

Il nome di una trasformazione è legato alla variabile termodinamica fissata; analizzeremo di seguito le seguenti trasformazioni:

- trasformazione **isobara**, se p è **costante**;
- trasformazione **isocora**, se V è **costante**;
- trasformazione **isoterma**, se T è **costante**.

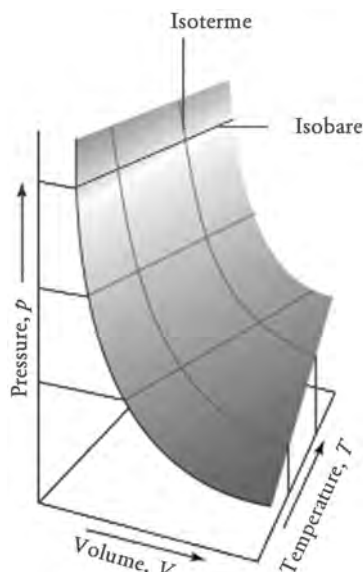


Fig. 2

Supponiamo di riempire un contenitore chiuso da un pistone con gas perfetto.

Come è stato detto, possiamo descrivere lo stato del gas utilizzando le grandezze termodinamiche pressione, volume e temperatura (Fig. 3).

Una volta fissato il volume, il numero di moli, la temperatura e la pressione del gas il suo stato è completamente definito.

Per modificare il suo stato possiamo modificare la variabile:

- pressione: comprimendo o espandendo il gas;
- temperatura: scaldando o raffreddando il gas;
- volume: variando il volume del contenitore in cui si trova.



Fig. 3

a) Trasformazioni isobare e prima legge di Gay-Lussac

In una trasformazione **isobara**, il volume varia linearmente con la temperatura.

In una trasformazione isobara la **pressione** si mantiene **costante**.

Consideriamo il cilindro in figura 3.

Se indichiamo con V_0 il volume iniziale del gas alla temperatura $T_0 = 0^\circ\text{C}$ e con V_t il volume finale alla temperatura t , il volume del gas aumenta linearmente con la temperatura e quindi possiamo scrivere:

$$\frac{V_t - V_0}{t} = k$$

Dividendo entrambi i membri per V_0 e risolvendo rispetto a V_t si ottiene:

$$V_t = V_0 \left(1 + \frac{k}{V_0} t \right) \quad (4.1)$$

Il rapporto $\frac{k}{V_0}$ è costante per tutti i gas, vale $\frac{1}{273} ^\circ\text{C}^{-1}$ e viene indicato con la lettera α .

Sostituendo tale valore nella (4.1) si ottiene:

$$V_t = V_0(1 + \alpha t) \quad (4.2)$$

che è la **prima legge di Gay-Lussac**.

La (4.2) vale per tutti i gas e prevede un volume nullo alla temperatura di $-273,15^\circ\text{C}$ a qualsiasi pressione e per qualsiasi gas.

Se nella (4.2) si passa dalla scala centigrada alla scala Kelvin essa diventa:

$$\frac{V}{T} = \text{cost}$$

in cui si indica la temperatura con T .

b) Trasformazioni isocore e seconda legge di Gay-Lussac

*In una trasformazione **isocora** le variazioni di pressione sono direttamente proporzionali alle variazioni di temperatura.*

In una trasformazione isocora il **volume** si mantiene **costante**.

Nel caso della figura 3, dobbiamo bloccare il pistone e riscaldare o raffreddare il gas.

Se indichiamo con p_0 la pressione iniziale del gas alla temperatura $T_0 = 0^\circ\text{C}$ e con p_t la pressione finale alla temperatura t , la pressione del gas aumenta linearmente con la temperatura e quindi possiamo scrivere:

$$\frac{p_t - p_0}{\Delta t} = k$$

Dividendo entrambi i membri per p_0 e risolvendo rispetto a p_0 si ottiene:

$$p_t = p_0 \left(1 + \frac{k}{p_0} \Delta t \right) \quad (4.3)$$

Il rapporto $\frac{k}{p_0}$ è costante per tutti i gas, vale $\frac{1}{273}^\circ\text{C}^{-1}$ e viene indicato con la lettera α .

Sostituendo tale valore nella (4.3) si ottiene:

$$p_t = p_0(1 + \alpha \Delta t)$$

che è la **seconda legge di Gay-Lussac**.

c) Trasformazioni isoterme e legge di Boyle

*In una trasformazione **isoterma** pressione e volume sono grandezze inversamente proporzionali.*

In una trasformazione isoterma la **temperatura** si mantiene **costante**.

Nel caso del contenitore in figura 3 dobbiamo mantenere costante la temperatura del gas e comprimerlo o espanderlo muovendo il pistone.

Questo fu proprio ciò che fece Robert Boyle nel 1662. Egli notò che ad una temperatura fissata la pressione è direttamente proporzionale all'inverso del volume cioè:

$$p = k \cdot \frac{1}{V}$$

da cui segue che:

$$pV = \text{cost}$$

che è la **legge di Boyle**.

A questo punto citiamo il principio di Avogadro e la legge di Dalton, la cui utilità sarà mostrata nel seguito.

d) Principio di Avogadro

Uguali volumi di gas alla stessa temperatura e pressione, contengono un egual numero di molecole. Il volume molare è lo stesso:

$$V \propto n$$

Sperimentalmente è stato trovato che alla pressione di 1 atm e alla temperatura di 273,16 °K (cioè a 0°C) una mole di un qualunque gas occupa un volume di 22,4 litri.

e) Legge di Dalton

La pressione esercitata da un miscuglio gassoso è pari alla somma delle pressioni parziali che ciascun gas eserciterebbe da solo occupando lo stesso volume del miscuglio.

5 L'equazione di stato dei gas perfetti

Sfruttando le leggi dei gas appena introdotte è possibile determinare l'equazione di stato dei gas perfetti:

$$pV = nRT \quad (5.1)$$

DIMOSTRAZIONE

Per determinare la (5.1) facciamo eseguire due trasformazioni termodinamiche, prima a pressione costante e poi a volume costante, al gas contenuto nel cilindro di figura 3.

— Trasformazione a pressione costante (**Trasformazione isobara**)

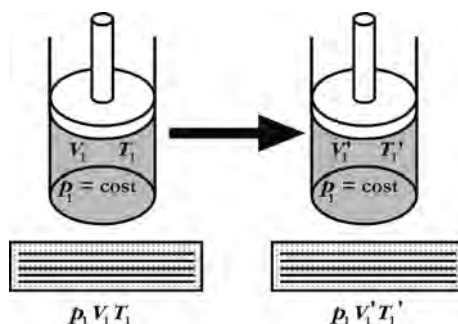


Fig. 4

Per la legge di Gay-Lussac:

$$\frac{V'}{V_1} = \frac{T'}{T_1}$$

da cui si ricava che:

$$T' = \frac{T}{V_1} V' \quad (5.2)$$

— Trasformazione a volume costante (**Trasformazione isocora**)

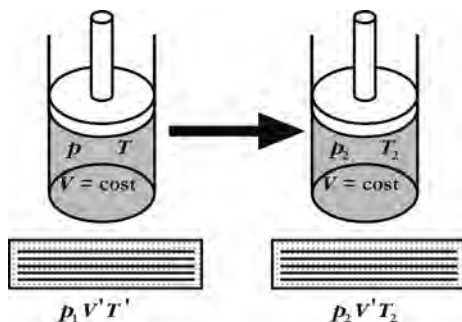


Fig. 5

Per la legge di Gay-Lussac:

$$\frac{p_2}{p_1} = \frac{T_2}{T'} \quad (5.3)$$

ma per la (5.2) ed essendo $V' = V_2$, si ha che:

$$T' = \frac{T_1}{V_1} V' = \frac{T_1}{V_1} V_2 \quad (5.4)$$

Sostituendo la (5.4) nella (5.3) si ottiene:

$$\frac{p_2}{p_1} = \frac{T_2}{T'} = \frac{T_2}{T_1} \frac{V_1}{V_2}$$

da cui segue che:

$$\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2} \quad (5.5)$$

Essendo gli stati 1 e 2 due stati qualsiasi del gas, il rapporto $\frac{pV}{T}$ è un invariante per gli stati del gas perfetto.

Per il principio di Avogadro, il volume occupato da un gas a una data temperatura e a una data pressione è proporzionale al numero di moli, per cui la (5.5) deve anche essere proporzionale al numero n di moli del gas, cioè:

$$\frac{pV}{T} = nR$$

da cui segue che:

$$pV = nRT$$

che è proprio l'**equazione di stato dei gas perfetti**.

Dalla definizione di mole, se m è la massa del gas e M la sua massa molecolare:

$$n = \frac{m}{M}$$

La costante R che compare nella (5.1) è detta **costante universale dei gas**.

Essa è infatti indipendente dalla pressione, dal volume, dalla temperatura e dal numero di moli del gas, e dipende esclusivamente dalle unità di misura utilizzate.

Siccome una mole di gas occupa un volume di 22,4 l alla pressione di 1 atm e alla temperatura di 273,16°C, possiamo scrivere:

$$R = \frac{pV}{nT} = \frac{(1 \text{ atm})(22,414 \text{ l})}{(1 \text{ mol})(273,16^\circ \text{K})} = 0,0821 \text{ l} \cdot \text{atm} / ^\circ \text{Kmol}$$

Il primo principio della termodinamica

1 Calore, lavoro ed energia

Supponiamo di mettere in movimento il pendolo della figura 1 sollevandolo e lasciandolo poi libero di oscillare.

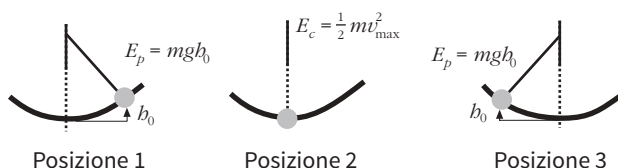


Fig. 1

1. All'istante iniziale (Posizione 1), il **lavoro** fatto per sollevare il pendolo è uguale all'**energia potenziale** acquistata dal pendolo:

$$L = E_p = mgh_0$$

2. Quando il pendolo raggiunge il punto più basso (Posizione 2), tutta l'**energia potenziale** acquisita inizialmente si trasforma in **energia cinetica** $E_c = \frac{1}{2}mv_{\max}^2$, dove la velocità è stata indicata con $v_{\max}$ in quanto è la massima velocità che il pendolo può raggiungere. In questo caso si verifica che:

$$E_p = mgh_0 \Rightarrow E_c = \frac{1}{2}mv^2$$

3. Dopo aver compiuto una mezza oscillazione, il pendolo si ferma per un istante (Posizione 3). In questo istante, tutta l'**energia cinetica** acquisita nel punto più basso si trasforma nuovamente in **energia potenziale**:

$$E_c = \frac{1}{2}mv^2 \Rightarrow E_p = mgh_0$$

4. Il pendolo raggiunge poi nuovamente la posizione più bassa per poi risalire alla posizione 1 e ricominciare così l'oscillazione.

Pertanto, quando il pendolo è lasciato libero di oscillare, una parte di questa energia potenziale si trasforma in energia cinetica. Tuttavia, l'energia totale del sistema viene conservata (**teorema della conservazione dell'energia**).

L'energia potenziale e l'energia cinetica possedute dal pendolo varieranno da istante a istante ma la loro somma sarà sempre uguale al lavoro fatto inizialmente per mettere il pendolo in oscillazione:

$$E_p(t) + E_c(t) = L = mgh_0$$

dove $E_p(t)$ e $E_c(t)$ indicano la dipendenza dell'energia dal tempo.

In assenza di attrito questo moto continuerebbe all'infinito.

In presenza di attrito e, quindi, anche semplicemente in presenza di aria, il moto del pendolo si smorzerebbe progressivamente fino a fermarsi. Questo potrebbe far pensare che l'energia

totale del sistema non si conservi; in realtà, non è proprio così; se misurassimo la temperatura del pendolo prima e dopo la sua oscillazione scopriremmo che è aumentata.

Per studiare il moto del pendolo nella sua globalità è necessario introdurre una nuova grandezza: la temperatura. Il pendolo non è più quindi un semplice sistema dinamico ma un sistema termodinamico.

Con questo termine viene definito un sistema in grado di scambiare energia con l'ambiente esterno. Le trasformazioni che avvengono in un sistema termodinamico sono delle trasformazioni termodinamiche.

Per studiare una trasformazione termodinamica è necessario sempre, infatti, definire sia il sistema che si intende studiare con relativo stato iniziale e finale, sia l'ambiente esterno.

Nel caso del pendolo il sistema è la pallina legata al filo, lo stato iniziale e finale sono rappresentati, rispettivamente, dal pendolo in quiete prima e dopo l'oscillazione; infine, l'ambiente esterno è l'aria circostante.

In figura 2 è schematizzata una trasformazione termodinamica:

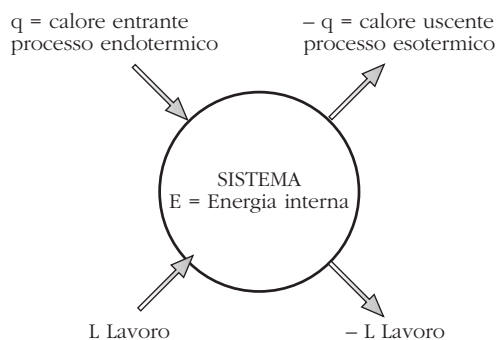


Fig. 2

Durante la trasformazione termodinamica il sistema può scambiare energia con l'ambiente esterno sia mediante del calore scambiato sia mediante del lavoro eseguito.

Nel caso del pendolo, il lavoro delle forze di attrito dell'aria (ambiente esterno) hanno fatto aumentare l'energia cinetica delle molecole della pallina (sistema).

Lo stesso risultato si sarebbe potuto ottenere riscaldando la pallina su una fonte di calore.

Questo per sottolineare il fatto che il sistema non distingue se la sua energia è dovuta al passaggio di calore o al lavoro.

Né il lavoro né il calore possono essere immagazzinati dal sistema. Non si può parlare di calore o di lavoro contenuto in un corpo ma soltanto di **energia** contenuta nel corpo.

Quindi il calore e il lavoro servono soltanto ad aumentare o a diminuire l'energia contenuta in un corpo.

Per convenzione il lavoro è:

- **positivo**, se è fatto **sul** sistema;
- **negativo**, se è fatto **dal** sistema.

Mentre il calore è:

- **positivo**, se è **assorbito** dal sistema;
- **negativo**, se è **emesso** dal sistema.

Questa energia contenuta nel corpo è detta **energia interna** di un sistema ed è indicata con il simbolo U .

2 L'energia interna di un sistema termodinamico

In generale, l'energia interna di un sistema può essere definita come la parte di energia posseduta dal sistema che dipende dalla temperatura:

$$E_{\text{interna}} = E_{\text{totale}} - E_{\text{indipendente dalla temperatura}}$$

Spesso un sistema termodinamico viene definito partendo dalla definizione di energia interna. Infatti, si definisce termodinamico un sistema per il quale può variare l'energia interna.

Il pendolo in oscillazione nell'aria è, quindi, un sistema termodinamico in quanto durante l'oscillazione aumenta la sua temperatura e quindi la sua energia interna.

Le caratteristiche dell'energia interna sono le seguenti:

- può essere immagazzinata nel sistema (a differenza del calore e della temperatura);
- esiste un'energia interna iniziale del sistema (prima del processo) e un'energia interna finale (dopo il processo);
- è una **funzione di stato**, ossia è una proprietà del sistema che dipende soltanto dallo stato in cui si trova il sistema e non dal modo in cui il sistema vi è arrivato (Fig. 3);
- è una sorta di «banca di energia». Se si esegue del lavoro sul sistema o il sistema assorbe una certa quantità di calore, esso immagazzina un'equivalente quantità di energia. Se il sistema esegue un lavoro o emette calore, perde un'equivalente quantità di energia.

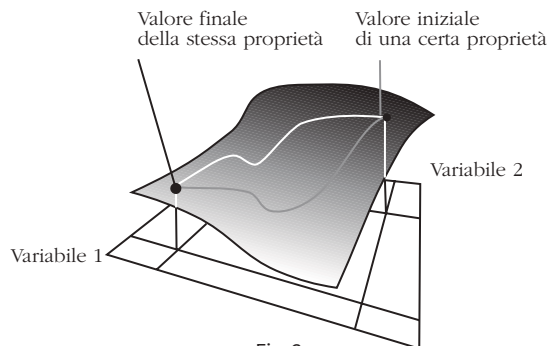


Fig. 3

3 Trasformazione termodinamica, trasformazione inversa e ciclo termico di un sistema

Una **trasformazione termodinamica** è il passaggio di un sistema termodinamico da uno stato iniziale a uno stato finale.

La figura 4 mostra un tipico esempio di sistema termodinamico costituito da una massa di gas contenuta in un pistone.

Gli stati di un sistema possono essere:

- **di equilibrio**, quando la pressione, il volume e la temperatura del sistema sono ben determinati;
- **dinamici**, quando la pressione o il volume o la temperatura del sistema non sono ben determinati.

Nell'equazione di stato di un gas perfetto, $pV = nRT$, essendo n ed R determinati, possiamo affermare che bastano due grandezze termodinamiche indipendenti per definire lo stato di equilibrio di un gas.

Possiamo anche identificare lo stato di un sistema utilizzando una coppia di assi cartesiani.



Fig. 4

a) Trasformazione termodinamica

Una **trasformazione termodinamica** indica il passaggio di un sistema termodinamico da uno stato iniziale ad uno finale.

Esistono due tipi di trasformazioni possibili:

- le trasformazioni **dinamiche** o **irreversibili**;
- le trasformazioni **quasistatiche** o **reversibili**.

• Trasformazioni dinamiche

Le **trasformazioni dinamiche** sono costituite sia da stati di equilibrio che da stati dinamici non rappresentabili e quindi esse stesse **non rappresentabili** e sono definite graficamente soltanto dal loro stato iniziale e finale (Fig. 5).

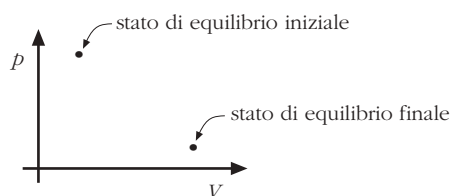


Fig. 5

• Trasformazioni quasistatiche

Le **trasformazioni quasistatiche** sono trasformazioni ideali che non esistono in natura.

Una trasformazione reale può essere approssimata ad una trasformazione quasistatica se avviene molto lentamente in modo che tutti gli infiniti stati che essa assume possano essere considerati di equilibrio.

Le trasformazioni quasistatiche, essendo costituite da un insieme infinito di stati di equilibrio, sono rappresentate da una curva che unisce tutti gli infiniti stati di equilibrio del sistema.

b) Trasformazione inversa

Si definisce **inversa** di una trasformazione data una trasformazione che unisce lo stato finale del sistema con quello iniziale passando per gli stessi stati intermedi.

Per quanto detto in precedenza soltanto le trasformazioni quasistatiche ammettono la trasformazione inversa.

c) Ciclo termico

Un **ciclo termico** è una trasformazione in cui lo stato iniziale e quello finale coincidono (Fig. 6).

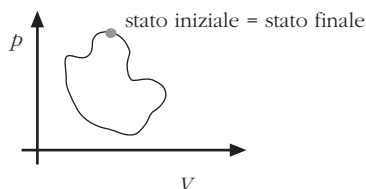


Fig. 6

4 Il lavoro in una trasformazione termodinamica

Studiamo ora il lavoro compiuto nelle varie trasformazioni.

Per fare ciò deve essere possibile tracciare la trasformazione su un piano pV e cioè la trasformazione deve essere quasistatica.

In realtà una trasformazione quasistatica non esiste ma possiamo approssimarla ad una trasformazione ideale che avvenga molto lentamente.

Consideriamo il cilindro di figura 4 e supponiamo che il gas nel cilindro si espanda molto lentamente in modo da consentire al sistema di cambiare stato passando per una serie infinita di stati di equilibrio (Fig. 7).



Fig. 7

a) Lavoro di una trasformazione isobara

Una trasformazione si dice **isobara** quando avviene a pressione costante p_0 .

Nella figura 8 è stata fatta un'istantanea di uno degli infiniti stati di equilibrio del sistema.

Essendo il sistema in equilibrio, la forza esercitata dal gas verso l'esterno è uguale in modulo alla forza esercitata dal pistone sul gas cioè:

$$|F_{\text{est}}| = |F_{\text{gas}}|$$

Dalla definizione di pressione si ha che:

$$|F_{\text{est}}| = |F_{\text{gas}}| = p_0 S$$

dove p_0 è la pressione (costante) e S è la sezione del pistone.

Dall'equazione di stato dei gas perfetti che in questo caso diventa:

$$p_0 V = nRT$$

si ha che ad una variazione ΔT di temperatura corrisponde una variazione ΔV di volume pari a:

$$\Delta V = \frac{nR}{p_0} \Delta T > 0$$

In seguito all'aumento ΔV del volume, il pistone subisce uno spostamento pari a:

$$\Delta y = \frac{\Delta V}{S}$$

Il gas compie quindi un lavoro positivo pari a:

$$\Delta L_{\text{gas}} = p_0 S \Delta y = p_0 \Delta V$$

Il lavoro è positivo in quanto la forza esercitata dal gas e lo spostamento compiuto dal pistone hanno lo stesso verso, mentre la forza esterna compie un lavoro negativo:

$$\Delta L_{\text{est}} = p_0 S \Delta y = p_0 \Delta V$$

Il lavoro è negativo in quanto la forza esercitata dall'esterno e lo spostamento compiuto dal pistone hanno verso opposto.

Indicando con V_i e con V_f il volume iniziale e finale del gas si ha che il lavoro compiuto dalla forza esterna è dato da:

$$L_{\text{est}} = -p_0 (V_f - V_i)$$



Fig. 8

e in un piano pV è rappresentato dall'area sottesa dalla curva:

$$p(t) = p_0$$

tra gli estremi V_i e V_f .

In figura 9 è rappresentato il lavoro fatto dalla forza esterna nella dilatazione e nella compressione del gas:

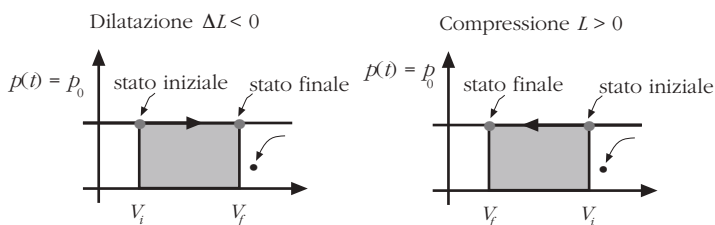


Fig. 9

In figura 10 sono state tracciate le trasformazioni isoterme dello stato iniziale e finale:

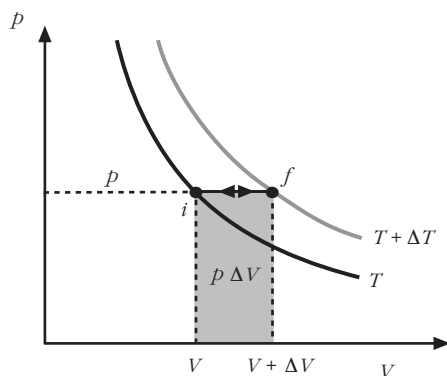


Fig. 10

b) Lavoro di una trasformazione isocora

Una trasformazione si dice **isocora** quando avviene a volume costante V_0 .

In questo caso, essendo il volume costante, $\Delta x = 0$ e quindi $\Delta L = 0$, le trasformazioni considerate sono supposte reversibili.

Si può dimostrare che il lavoro svolto da una trasformazione termodinamica dipende dal modo con cui si passa dallo stato iniziale allo stato finale e non è quindi una funzione di stato. Inoltre, il lavoro compiuto da una trasformazione reversibile è maggiore del lavoro compiuto da una trasformazione irreversibile.

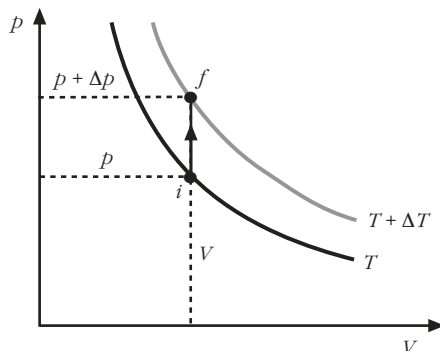


Fig. 11

5 Il primo principio della termodinamica

Si è osservato sperimentalmente che, indicando con Q il calore assorbito dal sistema e con L il lavoro compiuto dal sistema la quantità $Q - L$ si mantiene **costante** indipendentemente dal cammino effettuato. Essa dipende unicamente dallo stato iniziale e finale del sistema.

La differenza $Q - L$ è l'**energia interna** di un sistema (quindi è una funzione di stato).

Il **primo principio della termodinamica** afferma che la variazione di energia interna di un sistema termodinamico in una trasformazione è indipendente dal cammino svolto ed è pari a:

$$\Delta U = Q - L$$

L'energia interna di un sistema, essendo una funzione di stato, deve essere esprimibile mediante le variabili termodinamiche p, V, T cioè deve essere:

$$U = U(p, V, T)$$

Dall'equazione di stato dei gas perfetti non possiamo però ricavare alcuna informazione su U che deve essere calcolata caso per caso.

6 L'energia interna di un gas perfetto

Consideriamo la figura 12 in cui due recipienti A e B , comunicanti fra loro, sono posti all'interno di un termostato R .

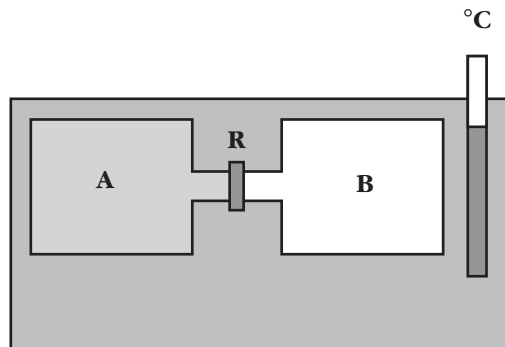


Fig. 12

Il recipiente A è riempito di un gas alla pressione p mentre nel recipiente B si è praticato il vuoto. Il sistema è in equilibrio termico ad una certa temperatura T .

Se si apre il rubinetto, il gas nel recipiente A si espande bruscamente nel recipiente B fino a raggiungere uno stato di equilibrio. Il termometro inserito nel termostato serve a controllare che effettivamente non ci siano stati scambi di calore tra il liquido e il calorimetro.

Poiché la trasformazione è avvenuta adiabaticamente, cioè senza scambi di calore, si ha che $\Delta Q = 0$.

Inoltre, poiché il gas ha esercitato una pressione su pareti immobili, si ha che $\Delta V = 0$ e quindi il gas non ha compiuto alcun lavoro:

$$L = p\Delta V = 0$$

Alla fine dell'espansione il gas occuperà un volume $V_f > V$ e avrà una pressione $p_f < p$. Per il primo principio della termodinamica $\Delta U = Q - L$ ed essendo $L = 0$ e $Q = 0$, si ha che:

$$\Delta U = 0$$

e quindi:

$$U_f = U_i$$

L'energia interna del sistema in una **trasformazione adiabatica** è una funzione della sola temperatura in quanto la temperatura è l'unica grandezza (delle tre: p, V, T) a rimanere costante.

Per determinare la variazione della temperatura possiamo utilizzare la trasformazione per noi più comoda. Essendo l'energia interna una funzione di stato, la sua variazione è indipendente dalla trasformazione e quindi il risultato trovato sarà valido per una qualsiasi trasformazione (Fig. 13).

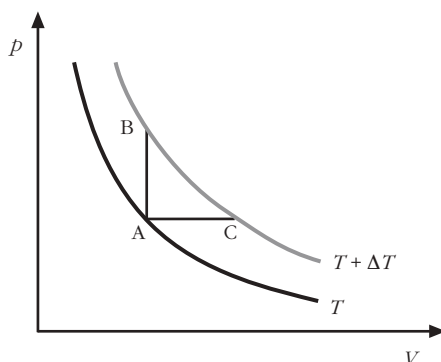


Fig. 13

a) Trasformazione a volume costante

La trasformazione forse più comoda è quella a **volume costante** (A – B) in quanto, in questo caso, $L = 0$ e quindi:

$$Q = \Delta U$$

Ricordando la relazione che esiste tra il calore assorbito o ceduto da un sistema e la sua variazione di temperatura:

$$Q = m \cdot c \cdot \Delta T$$

dove m è la massa del sistema e c il suo calore specifico, si ha che:

$$\Delta U = m \cdot c_v \cdot \Delta T \quad (6.1)$$

nella (6.1), essendo la trasformazione avvenuta a volume costante, abbiamo utilizzato c_v e cioè il calore specifico a volume costante.

Se n è il numero di moli e M è la massa di una grammomolecola, si ha che la massa del gas sarà pari a:

$$m = nM$$

• Variazione di calore

Se introduciamo il calore molare a volume costante ($C_v = Mc_v$) possiamo scrivere la (6.1) come:

$$\Delta U = n \cdot C_v \cdot \Delta T \quad (6.2)$$

che esprime il legame tra energia interna e temperatura.

b) Trasformazione a pressione costante

In una trasformazione a **pressione costante** (A – C):

$$Q = nC_p \cdot \Delta T$$

e inoltre:

$$L = p \cdot \Delta V$$

e quindi, applicando il primo principio della termodinamica:

$$\Delta U = nC_p \cdot \Delta T - p\Delta V \quad (6.3)$$

In un gas perfetto l'energia interna dipende soltanto dalla temperatura e, poiché la variazione di temperatura è la stessa, anche la variazione di energia interna deve essere la medesima. Uguagliando la (6.2) e la (6.3), si ha che:

$$nC_V \cdot \Delta T = nC_p \cdot \Delta T - p\Delta V \quad (6.4)$$

Consideriamo l'equazione dei gas perfetti $pV = nRT$, nel caso di una trasformazione a pressione costante si ha che:

$$p\Delta V = nR\Delta T \quad (6.5)$$

Sostituendo la (6.5) nella (6.4) si ha che:

$$nC_V \cdot \Delta T = nC_p \cdot \Delta T - p\Delta V = nC_p \cdot \Delta T - nR\Delta T$$

da cui segue che:

$$C_V = C_p - R$$

la quale è detta **relazione di Meyer**.

Da essa possiamo notare che il calore specifico molare di un gas perfetto a pressione costante è sempre maggiore del calore specifico a volume costante di una quantità pari alla costante universale dei gas R .

c) Trasformazione ciclica

In una **trasformazione ciclica** lo stato iniziale coincide con quello finale (Fig. 6) e quindi l'energia interna, essendo una funzione di stato, resta costante, ossia:

$$\Delta U = 0$$

da cui segue che:

$$Q = L$$

Da notare che il lavoro netto compiuto dalla trasformazione è pari all'area racchiusa dalla curva che rappresenta la trasformazione nel diagramma pV .

d) Trasformazione isocora

In una **trasformazione isocora** il volume si mantiene costante e quindi il lavoro è nullo:

$$L = p\Delta V = 0$$

Da ciò segue che:

$$\Delta U = Q - L = Q$$

e) Trasformazione adiabatica

In una **trasformazione adiabatica** non si ha uno scambio di calore con l'esterno, quindi $Q = 0$.

Applicando il primo principio a una trasformazione adiabatica si ha:

$$\Delta U = -L$$

Da ciò segue che:

- se il gas si espande adiabaticamente, il suo lavoro è positivo e quindi si ha una diminuzione dell'energia interna del sistema;

- se il gas viene compresso adiabaticamente, il suo lavoro è negativo e si ha un aumento dell'energia interna del sistema.

Ma:

- $\Delta U = nC_v \Delta T$ (energia interna di un gas perfetto)
- $L = p\Delta V$ (lavoro in una trasformazione)

e quindi in una trasformazione adiabatica:

$$p\Delta V = nC_v \Delta T \quad (6.6)$$

Dalla (6.6) possiamo notare che il lavoro positivo viene eseguito a spese dell'energia interna e della temperatura.

f) Processo di ebollizione

Supponiamo che una massa m di gas vaporizzi a pressione costante p . Determiniamo la sua variazione di energia interna.

Se indichiamo con V_l il volume dell'acqua allo stato liquido e con V_v il volume dell'acqua allo stato di vapore, il lavoro di espansione è dato da:

$$L = p(V_v - V_l)$$

Il calore che deve essere fornito per la completa vaporazione dell'acqua è:

$$Q = mL_v$$

dove L_v è il calore latente di vaporizzazione dell'acqua.

Quindi, per il primo principio della termodinamica:

$$\Delta U = Q - L = mL_v - p(V_v - V_l)$$

Il secondo principio della termodinamica

1 Gli enunciati di Kelvin-Planck e di Clausius

Il **secondo principio della termodinamica** può essere espresso con uno dei seguenti enunciati: quello di Kelvin-Planck e quello di Clausius.

a) Enunciato di Kelvin-Planck

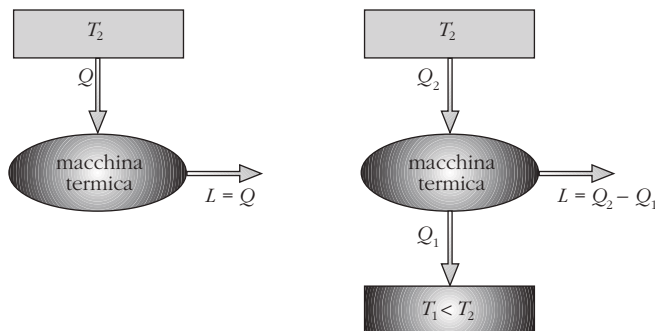
È impossibile realizzare una trasformazione ciclica il cui unico risultato sia quello di convertire in lavoro tutto il calore assorbito da una sola sorgente.

Una **macchina termica** è un dispositivo in grado di far compiere a una sostanza una trasformazione ciclica e cioè una trasformazione alla fine della quale la sostanza ritorna allo stato iniziale.

L'enunciato di Kelvin può anche essere formulato nel seguente modo:

È impossibile realizzare una macchina termica ciclica che abbia come unico effetto la totale trasformazione in lavoro del calore assorbito da un'unica sorgente (Fig. 1.a).

Per funzionare ciclicamente la macchina deve assorbire calore da una sorgente a temperatura maggiore e cederne una parte ad una sorgente a temperatura minore. La differenza fra l'energia assorbita e quella ceduta è uguale al lavoro compiuto dalla macchina (Fig. 1.b).



Figg. 1.a e 1.b

Per questo motivo è quindi impossibile che le navi o gli aerei possano convertire interamente in lavoro il calore assorbito dall'aria o dal mare.

b) Enunciato di Clausius

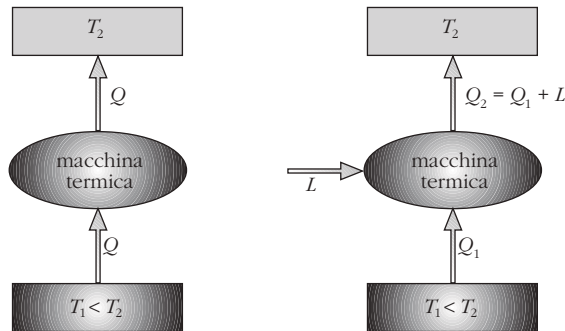
È impossibile realizzare una trasformazione il cui unico risultato sia quello di trasferire calore da un corpo a un altro avente temperatura uguale o maggiore a quella del primo.

L'enunciato di Clausius non vieta il passaggio di calore da un corpo più freddo a un corpo più caldo, vieta che ciò accada spontaneamente e che sia l'unico risultato della trasformazione (altrimenti non si spiegherebbero i frigoriferi).

L'enunciato di Clausius può anche essere formulato nel seguente modo:

È impossibile realizzare una macchina termica ciclica che abbia come unico effetto il passaggio di una certa quantità di calore da un corpo a temperatura più bassa verso un corpo a temperatura più alta (Fig. 2.a).

Il passaggio di calore da un corpo più freddo verso un corpo più caldo può avvenire soltanto a spese di un lavoro fornito dall'ambiente esterno: ad esempio, il frigorifero assorbe energia elettrica (Fig. 2.b).



Figg. 2.a e 2.b

In altre parole, l'enunciato di Kelvin-Planck afferma che non esistono macchine termiche perfette (ideali) mentre l'enunciato di Clausius afferma che non esistono macchine frigorifere perfette (ideali).

c) Equivalenza fra l'enunciato di Kelvin-Planck e l'enunciato di Clausius

Possiamo, per esempio, dimostrare che:

Se l'enunciato di Kelvin-Planck è falso anche l'enunciato di Clausius è falso.

DIMOSTRAZIONE

Supponiamo che l'enunciato di Kelvin-Planck sia falso e che possa esistere una macchina termica in grado di trasformare interamente il calore assorbito da una sorgente termica in lavoro.

Consideriamo un dispositivo formato da questa macchina termica e da una macchina frigorifera reale.

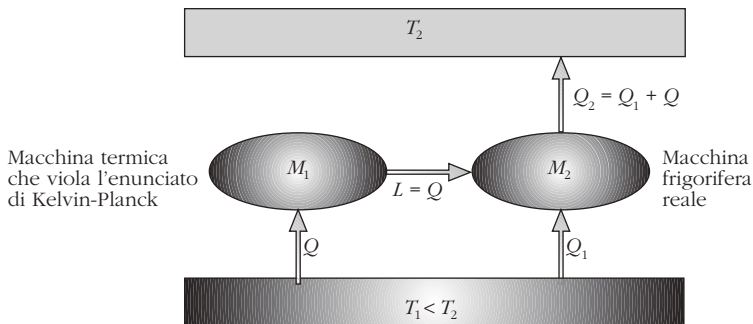


Fig. 3

Il lavoro prodotto dalla macchina termica viene totalmente scambiato all'interno di questo dispositivo e non rientra nello scambio energetico con l'ambiente esterno.

La macchina termica trasforma interamente in lavoro L il calore Q assorbito dalla sorgente termica.

La macchina frigorifera, grazie al lavoro esterno fatto dalla macchina termica, assorbe il calore Q_1 dalla sorgente più fredda e cede il calore Q_2 alla sorgente più calda.

Per il principio di Clausius il calore ceduto deve essere pari alla somma del calore assorbito e del lavoro fatto dall'esterno quindi, in questo caso:

$$Q_2 = Q_1 + L$$

Siccome la macchina frigorifera lavora ciclicamente la sua variazione di energia interna è nulla e quindi $\Delta U = 0$, da cui segue che:

$$L + Q_1 = Q_2$$

ma $L = Q$ e quindi:

$$Q + Q_1 = Q_2 = Q + Q_1$$

Pertanto, il dispositivo costituito dalle due macchine è in realtà un frigorifero che assorbe una quantità di calore pari a:

$$Q_2 = Q + Q_1$$

dalla sorgente a temperatura più fredda e cede la quantità di calore pari a:

$$Q_2 = Q + Q_1$$

alla sorgente a temperatura più alta senza che venga fatto del lavoro esterno.

Questo viola l'enunciato di Clausius.

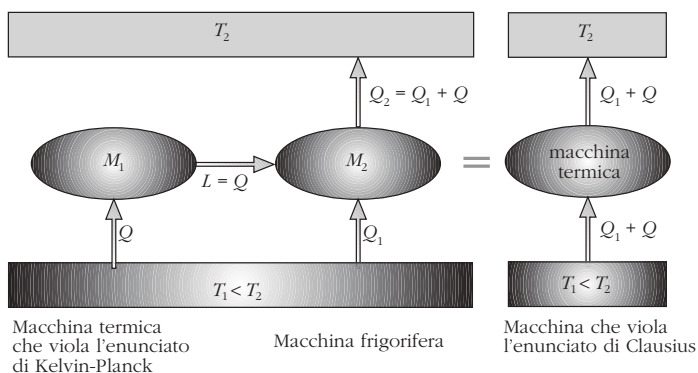


Fig. 4

Ciò vuol dire che se non si possono costruire macchine termiche ideali V_2 , non si possono neanche costruire frigoriferi ideali. Analogamente, si può dimostrare che se l'enunciato di Clausius è falso anche l'enunciato di Kelvin-Planck è falso.

d) Rendimento di una macchina termica

Il **rendimento** η di una macchina termica è dato dal rapporto tra il lavoro prodotto e il calore assorbito:

$$\eta = \frac{L}{Q}$$

Esso è dato, cioè, dal rapporto fra ciò che si ottiene da una macchina (il lavoro meccanico) e ciò che si è speso per ottenerlo (il calore).

Se la macchina termica fosse ideale, tutto il calore assorbito si trasformerebbe in lavoro e quindi il rendimento sarebbe unitario.

In una macchina termica reale, invece:

$$\eta = \frac{L}{Q} = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1}$$

2 Il teorema di Carnot

Il fisico francese Sadi Carnot nel 1824 definì un ciclo di lavoro detto **ciclo di Carnot** e dimostrò che una macchina che lavora con questo ciclo (detta **macchina di Carnot**) ha il massimo rendimento possibile in quanto dipende esclusivamente dalla temperatura delle sorgenti termiche con le quali essa scambia calore.

In altre parole, vale il seguente **teorema**, detto di **Carnot**.

Nessuna macchina termica che lavori tra due sorgenti di calore può avere un rendimento superiore a quello di una macchina di Carnot che lavori tra le stesse sorgenti:

$$\eta_{\text{Carnot}} < \eta_{\text{Reale}}$$

a) Ciclo di Carnot

Il ciclo di Carnot (Fig. 5) è un ciclo reversibile, composto da:

- un'espansione isoterma;
- un'espansione adiabatrica;
- una compressione isoterma;
- una compressione adiabatrica.

La macchina di Carnot è costituita da un cilindro chiuso con un pistone avente le pareti isolate adiabaticamente e contenente del gas perfetto che può scambiare calore soltanto con il fondo del pistone:

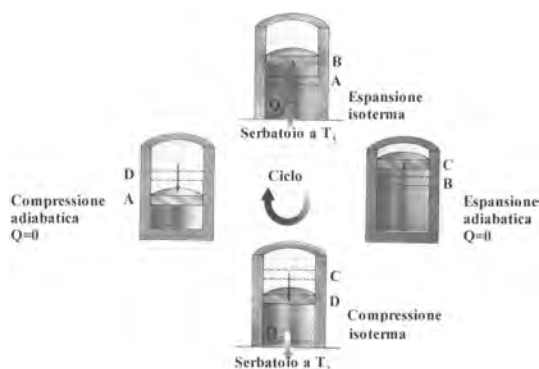


Fig. 5

• Espansione isoterma

Il pistone nel cilindro è inizialmente in posizione A.

Il cilindro viene posto a contatto con la sorgente più calda e il gas all'interno si espande fino a raggiungere la posizione B.

In questa trasformazione viene assorbito del calore e viene prodotto del lavoro.

La variazione di energia interna è nulla e quindi:

$$L_{A \rightarrow B} = Q_1$$

dove Q_1 è il calore assorbito dalla caldaia e:

$$L_{A \rightarrow B} = nRT_1 \ln \frac{V_B}{V_A}$$

- **Espansione adiabatica**

Il cilindro viene allontanato dalla sorgente e viene isolato termicamente. Il gas si espande adiabaticamente fino a raggiungere la temperatura del refrigerante in posizione C.

In questa trasformazione il lavoro è uguale alla variazione di energia interna:

$$L_{B \rightarrow C} = -\Delta U = -nC_V(T_C - T_B) = -nC_V(T_2 - T_1)$$

- **Compressione isoterma**

Il cilindro viene posto a contatto con il refrigerante e il pistone viene abbassato alla posizione D. Il gas subisce quindi una compressione isoterma.

In questa trasformazione viene ceduto del calore e viene assorbito del lavoro:

$$L_{C \rightarrow D} = nRT_2 \ln \frac{V_D}{V_C}$$

- **Compressione adiabatica**

Il cilindro viene allontanato dal refrigerante e isolato termicamente. Il gas viene compresso fino a raggiungere la temperatura della caldaia riportandosi nella posizione di partenza.

In questa trasformazione il lavoro è uguale alla variazione di energia interna:

$$L_{D \rightarrow A} = -\Delta U = -nC_V(T_D - T_A) = -nC_V(T_A - T_B) = -nC_V(T_1 - T_2) = -L_{B \rightarrow C}$$

Il lavoro totale in questo ciclo è dato da:

$$L = L_{A \rightarrow B} + L_{B \rightarrow C} + L_{C \rightarrow D} + L_{D \rightarrow A} = L_{A \rightarrow B} + L_{B \rightarrow C} + L_{C \rightarrow D} - L_{B \rightarrow C} = L_{A \rightarrow B} + L_{C \rightarrow D}$$

ed è quindi dovuto soltanto alle trasformazioni isoterme.

Esso è dato dall'area racchiusa nel cammino ABCD del diagramma pV (Fig. 6).

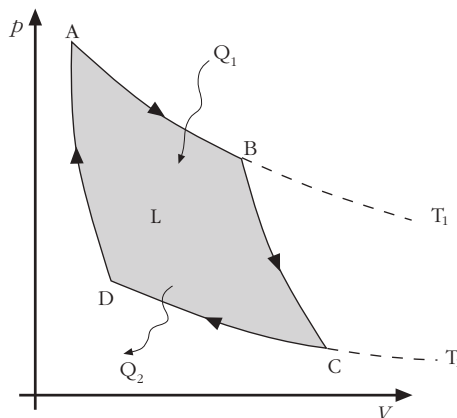


Fig. 6

Si può dimostrare che il rendimento di una macchina di Carnot è pari a:

$$\eta_{\text{Carnot}} = 1 - \frac{T_2}{T_1}$$

Inoltre, dalla relazione:

$$\eta_{\text{Carnot}} = \frac{L}{Q_1} = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1} = 1 - \frac{T_2}{T_1}$$

si ottiene che:

$$\frac{Q_2}{T_2} = \frac{Q_1}{T_1}$$

e cioè il rapporto dei calori scambiati è uguale al rapporto tra le temperature delle due sorgenti tra cui lavora la macchina di Carnot.

Possiamo, quindi, affermare che il rendimento di una macchina termica non può mai essere unitario (anche se il primo principio lo consentirebbe), ma può essere al massimo uguale al rendimento di una macchina di Carnot che lavora tra le stesse temperature della macchina termica considerata:

$$\eta_{\text{max}} = \eta_c = 1 - \frac{T_2}{T_1}$$

Questo risultato ci porta a fare alcune considerazioni.

Supponiamo che una macchina termica lavori tra due sorgenti: la caldaia ha una temperatura di 500°C (773K) e il refrigerante di 20°C (293K).

Il rendimento di Carnot è pari a:

$$\eta_{\text{Carnot}} = 1 - \frac{T_2}{T_1} = 1 - \frac{293}{773} = 0,62$$

Questo vuol dire che soltanto il 62% del calore iniziale è stato convertito in lavoro mentre il rimanente 38% è stato ceduto all'ambiente.

b) Efficienza di una macchina termica

L'**efficienza di una macchina** indica quanto il suo comportamento si avvicini a quello di una macchina di Carnot.

Essa è definita infatti dal rapporto:

$$\varepsilon = \frac{\eta}{\eta_{\text{Carnot}}}$$

3 L'entropia

Abbiamo visto con l'enunciato di Clausius che la propagazione del calore è unidirezionale.

Questo significa che la trasformazione termodinamica associata a una propagazione di calore ha una direzionalità che non può essere invertita ed è, quindi, una **trasformazione irreversibile**.

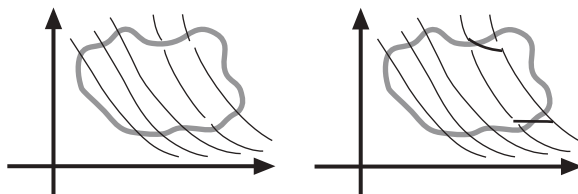
Tutti i fenomeni naturali hanno una propria direzionalità e sono, dunque, irreversibili; i fenomeni reversibili sono irreali.

Clausius cercò di dare una misura del grado di irreversibilità di una trasformazione, considerando il rapporto $\frac{Q}{T}$ in cui compaiono sia il calore scambiato durante la trasformazione sia la

temperatura associata a cui avviene tale trasformazione. Questa grandezza, inizialmente definita *valore equivalente di una trasformazione*, è chiamata **entropia**.

L'entropia è quindi quella grandezza fisica che fornisce un criterio per la misura dell'irreversibilità delle trasformazioni termodinamiche.

Consideriamo un ciclo reversibile qualunque e ricopriamolo con una rete di adiabatiche (Fig. 7.a).



Figg. 7.a e 7.b

Sostituiamo i tratti del ciclo con tratti di isoterme (Fig. 7.b).

Possiamo quindi considerare il ciclo reversibile come una successione di cicli di Carnot (Fig. 8).

Per ogni ciclo vale la relazione:

$$\frac{Q_2}{T_2} = \frac{Q_1}{T_1}$$

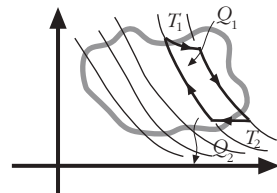


Fig. 8

e, tenendo conto del segno dei calori (positivo quello entrante e negativo quello uscente):

$$\frac{Q_2}{T_2} + \frac{Q_1}{T_1} = 0$$

Questo vuol dire che per tutta la rete si ha che:

$$\sum_i \frac{\Delta Q_i}{T} = 0 \quad (3.1)$$

Se infittiamo la rete di adiabatiche la (3.1) diventa:

$$\oint \frac{dQ}{T} = 0 \quad (3.2)$$

e quindi l'entropia definita da Clausius:

$$dS = \frac{dQ}{T} \quad (3.3)$$

è una funzione di stato.

La (3.1) e la (3.2) costituiscono l'enunciato del **teorema di Clausius** per la macchina termica di Carnot, la cui forma più generalizzata è la seguente:

In una macchina termica qualsiasi, costituita da una serie infinita di sorgenti, indicando con dQ la quantità di calore che la sostanza termodinamica scambia con la sorgente a temperatura T , si ha che:

$$\oint \frac{dQ}{T} \leq 0$$

dove l'uguaglianza vale soltanto nel caso in cui la macchina è reversibile.

a) Entropia dell'Universo

Integrando la (3.3) si ottiene che:

$$S_2 - S_1 = \int_1^2 \frac{dQ}{T}$$

da cui si ricava che l'entropia è definita a meno di una costante che viene abitualmente posta uguale a zero.

La dimensione dell'entropia è:

$$[S] = \frac{[Q]}{[T]}$$

e quindi la sua unità di misura è il $J \cdot K^{-1}$

Consideriamo una trasformazione che avviene senza scambio di calore con l'ambiente esterno e cioè una trasformazione adiabatica. In essa si ha che:

$$\Delta Q = 0$$

da cui segue che, essendo $dS = \frac{dQ}{T}$, anche $\Delta S = 0$ e dunque:

una trasformazione adiabatica è anche isoentropica.

Il fatto che le trasformazioni adiabatiche siano isoentropiche non significa che le trasformazioni isoentropiche siano per forza adiabatiche. Nella figura sono mostrate due trasformazioni isoentropiche: un'adiabatica reversibile e una trasformazione irreversibile.

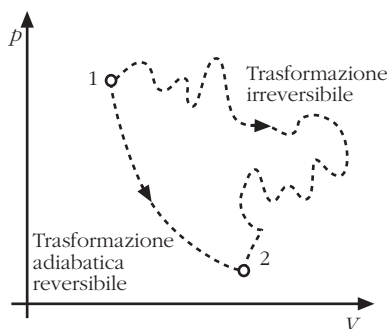


Fig. 9

Anche la trasformazione irreversibile non genera alcuna variazione di entropia nel sistema, ma questo non vuol dire che non faccia cambiare l'**entropia dell'Universo**.

Consideriamo un ciclo composto da una trasformazione reversibile e da una trasformazione irreversibile. Per il teorema di Clausius:

$$\oint \frac{dQ}{T} \leq 0$$

da cui segue che:

$$\int_1^2 \left(\frac{dQ}{T} \right)_{\text{IRR}} + \int_2^1 \left(\frac{dQ}{T} \right)_{\text{REV}} < 0$$

che si può scrivere anche come:

$$\int_1^2 \left(\frac{dQ}{T} \right)_{\text{IRR}} < \int_1^2 \left(\frac{dQ}{T} \right)_{\text{REV}}$$

Ricordando che per definizione:

$$\int_1^2 \left(\frac{dQ}{T} \right)_{\text{REV}} = S_2 - S_1$$

si ha che:

$$\int_1^2 \left(\frac{dQ}{T} \right)_{\text{IRR}} < S_2 - S_1$$

Se il sistema è isolato non scambia calore con l'esterno, ossia $dQ = 0$; da ciò segue che $S_2 - S_1 > 0$ e quindi:

$$S_2 > S_1$$

Pertanto, un sistema isolato, che compie una trasformazione irreversibile, evolve sempre verso stati termodinamici che implicano un aumento della sua entropia.

Poiché l'Universo è un sistema isolato, possiamo affermare che:

ogni trasformazione che avviene in natura comporta un aumento complessivo dell'entropia nell'Universo stesso.

Quest'affermazione, generalizzata a tutti i sistemi e a tutte le trasformazioni, costituisce l'enunciato più generale del **secondo principio della termodinamica**.

In altre parole, possiamo scrivere che:

$$\Delta S_U > 0$$

1 La struttura atomica

L'**atomo** è formato da protoni, neutroni ed elettroni oltre a un certo numero di particelle subatomiche. Il **protone** e il **neutrone** sono localizzati nel nucleo dell'atomo: il protone ha carica positiva mentre il neutrone non ha carica. Gli **elettroni** occupano delle regioni poste all'esterno del nucleo dette anche **gusci elettronici** ciascuna caratterizzata da un determinato livello energetico. I gusci più esterni hanno, rispetto a quelli più interni, un'energia maggiore e una stabilità minore. Gli elettroni che occupano un guscio più esterno possono «saltare» su un guscio più interno e questo salto è accompagnato dall'emissione di una certa quantità di energia. Analogamente, se a un elettrone viene fornita una certa quantità di energia esso può saltare su un guscio più esterno; se poi l'energia fornita è superiore ad un certo valore l'elettrone riesce a staccarsi dall'atomo.

Tutti gli atomi sono caratterizzati da un numero atomico e da un simbolo chimico.

Il **numero atomico** è indicato con la lettera Z e rappresenta il numero di protoni presenti nell'atomo. Per esempio per l'idrogeno $Z = 1$; per il carbonio $Z = 6$ etc.

Il **simbolo chimico** è, invece, composto da una o due lettere (tre lettere sono state assegnate agli elementi più recenti).

Gli elementi e i loro simboli chimici sono riportati nelle cosiddette **tavole periodiche**.

Normalmente un atomo è neutro e possiede tanti protoni quanti elettroni.

- Se l'atomo perde uno più elettroni assume una carica positiva, è detto **catione** e viene scritto con il simbolo $+$.
- Se l'atomo acquista uno o più elettroni assume una carica negativa, è detto **anione** e viene scritto con il simbolo $-$.

Per esempio, se l'atomo di carbonio (C) acquista un elettrone esso verrà scritto come C^+ ; se, invece, perde due elettroni verrà scritto come C^{2-} .

2 La quantità di carica

Siccome gli oggetti sono composti da atomi, essi posseggono una certa quantità di carica, la cui unità di misura è il Coulomb, che dipende dalla carica posseduta dai singoli atomi.

Essendo la carica dell'elettrone pari a $-1,6 \times 10^{-19}$ Coulomb e la carica del protone pari a $1,6 \times 10^{-19}$ Coulomb, si ha che:

- se una sostanza è elettricamente neutra, la sua carica è nulla;
- se il numero di protoni (N_p) supera il numero di elettroni (N_e), la sua carica (Q) sarà data dalla differenza fra il numero di protoni e quello degli elettroni moltiplicata per la carica del protone:

$$Q = (N_p - N_e)1,6 \times 10^{-19}C$$

- se il numero di elettroni supera il numero di protoni, la sua carica sarà data dalla differenza fra il numero di elettroni e quello dei protoni moltiplicata per la carica dell'elettrone:

$$Q = (N_e - N_p)(-1,6 \times 10^{-19})C$$

3 Conduttori, isolanti e dielettrici

Gli elettroni che occupano il livello più esterno sono detti **elettroni di valenza** e sono quelli che determinano il comportamento elettrico della sostanza. Infatti, gli atomi tendono ad avere l'ultimo livello elettronico occupato con il massimo numero di elettroni, per cui gli atomi che hanno l'ultimo livello quasi completato tendono ad attirare elettroni, mentre gli atomi che hanno pochi elettroni di valenza tendono a sbarazzarsene per avere come ultimo il livello sottostante pieno.

Dalla diversa configurazione dello strato di valenza nasce la differenza di comportamento dei materiali, i quali possono essere:

- conduttori;
- isolanti;
- dielettrici.

a) Conduttori

I **materiali conduttori** hanno il livello di valenza occupato da uno o due elettroni e quindi tendono a sbarazzarsene rendendo l'elettrone libero di muoversi all'interno del reticolo cristallino. Vedremo, nel seguito del volume, che, applicando una minima differenza di potenziale ai capi di un materiale conduttore, gli elettroni sono attratti dal potenziale maggiore generando così una corrente all'interno del materiale.

b) Isolanti

I **materiali isolanti** hanno l'ultimo livello energetico quasi del tutto occupato e quindi sono molto restii a cedere elettroni. Gli elettroni sono, in questo caso, saldamente legati al nucleo e quindi, anche applicando ai capi del materiale una differenza di potenziale, non saranno liberi di muoversi al suo interno.

c) Dielettrici

I **materiali dielettrici** spesso vengono confusi con gli isolanti in quanto in particolari condizioni non conducono corrente, ma questo comportamento non è dovuto alla conformazione degli atomi del materiale considerato ma alle sue molecole le quali si polarizzano quando viene applicata una differenza di potenziale.

Questo tipo di materiale è utilizzato nei condensatori per ridurre la tensione fra le sue armature.

L'acqua è un esempio tipico di dielettrico. A nessuno verrebbe in mente di buttare un phon acceso nella vasca in cui sta facendo il bagno e questo perché l'acqua conduce l'elettricità, se però l'acqua venisse versata tra le armature di un condensatore precedentemente vuoto la tensione fra le due armature diminuirebbe di circa 80 volte.

4 La legge di Coulomb

Un detto afferma che gli opposti si attraggono e i simili si respingono. Mai legge popolare fu più verificata in fisica, infatti:

- oggetti carichi dello stesso tipo si respingono;
- oggetti carichi di tipo opposto si attraggono.

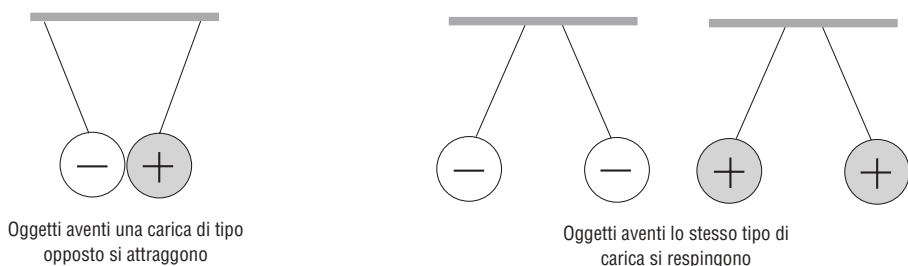


Fig. 1

Questo fenomeno è stato studiato dal fisico francese **Coulomb** che definì la **legge** che da lui prende nome:

La forza esercitata da una carica puntiforme su un'altra è:

- diretta lungo la retta che congiunge le due cariche;
- repulsiva se le cariche hanno lo stesso segno e attrattiva se le cariche hanno segno opposto;
- inversamente proporzionale al quadrato della distanza d che separa le cariche e direttamente proporzionale al prodotto delle cariche:

$$F = k \frac{|q_1| |q_2|}{d^2}$$

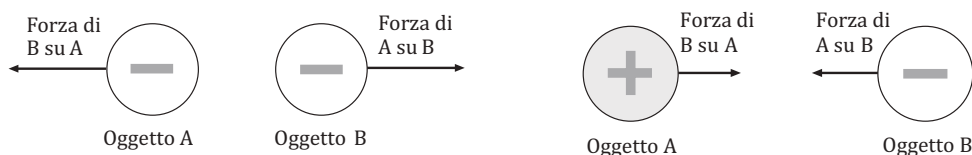
Questa forza, come tutte le forze, è misurata in newton.

La costante k è stata determinata sperimentalmente ed è pari a:

$$k = 8,98755 \times 10^9 \frac{\text{Nm}^2}{\text{C}^2} \cong 9 \times 10^9 \frac{\text{Nm}^2}{\text{C}^2}$$

Se i due oggetti sono elettrizzati con lo stesso tipo di carica essi si respingono e i versi delle forze elettriche sono schematizzati nella figura 2.a.

Se i due oggetti sono elettrizzati con cariche opposte essi si attraggono e i versi delle forze sono segnati nella figura 2.b.



Figg. 2.a e 2.b

È importante sottolineare che i due corpi non devono essere a contatto per risentire della forza di Coulomb. La forza di Coulomb è una forza a distanza.

5 Confronto tra forza elettrica e forza gravitazionale

La forza gravitazionale:

$$F_{\text{grav}} = G \frac{m_1 m_2}{d^2}$$

e la forza elettrostatica:

$$F_{\text{elet}} = k \frac{q_1 q_2}{d^2}$$

sono molto simili fra di loro. Infatti, entrambe:

- sono forze a distanza;
- sono inversamente proporzionali al quadrato della distanza fra i corpi;
- sono proporzionali al prodotto di ciò che causa la forza e quindi:
 - dalla carica dei due corpi nel caso della forza elettrica;
 - dalla massa dei due corpi nel caso della forza gravitazionale.

Dal confronto del valore della costante G con quello della costante k si deduce che la forza esercitata tra due corpi di massa unitaria posti alla distanza di $1m$ è notevolmente inferiore a quella esercitata fra due corpi di carica unitaria e posti alla stessa distanza.

6 Le forze elettriche in un sistema di cariche

Finora abbiamo studiato l'effetto di una carica elettrica su un'altra carica elettrica. Se invece di una carica avessimo un sistema di cariche dovremmo applicare il **principio della sovrapposizione degli effetti**.

Nella figura 3, la forza elettrica di cui risente la carica q dovuta alle cariche q_1, q_2, q_3 e q_4 è pari alla somma vettoriale delle forze elettriche dovute a ogni singola carica cioè:

$$\vec{F} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4 = k \frac{q_1 q}{r_1^2} \vec{u}_1 + k \frac{q_2 q}{r_2^2} \vec{u}_2 + k \frac{q_3 q}{r_3^2} \vec{u}_3 + k \frac{q_4 q}{r_4^2} \vec{u}_4 = \sum_{i=1}^4 \vec{F}_i$$

dove $\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4$ sono i versori delle rispettive forze.

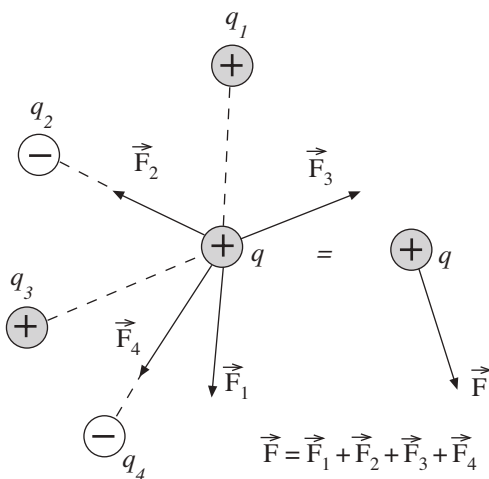


Fig. 3

7 Il campo elettrico

Riprendiamo il secondo esempio del paragrafo precedente e ripercorriamo al contrario le formule utilizzate per trovare l'intensità della forza elettrica sulla carica q_3 dovuta a q_1 e q_2 :

$$\begin{aligned} F_3 &= \sqrt{F_{3x}^2 + F_{3y}^2} = \sqrt{(F_{13x} + F_{23x})^2 + (F_{13y} + F_{23y})^2} = \sqrt{(F_{13} \sqrt{2} + 0)^2 + (F_{13} \sqrt{2} - F_{23})^2} = \\ &= \sqrt{\left(k \frac{q_1 q_3}{d_{13}^2} \sqrt{2}\right)^2 + \left(k \frac{q_1 q_3}{d_{13}^2} \sqrt{2} - k \frac{q_2 q_3}{d_{23}^2}\right)^2} = q_3 \sqrt{\left(k \frac{q_1}{d_{13}^2} \sqrt{2}\right)^2 + \left(k \frac{q_1}{d_{13}^2} \sqrt{2} - k \frac{q_2}{d_{23}^2}\right)^2} \end{aligned}$$

Possiamo notare che la forza che si esercita sulla carica q_3 è proporzionale alla carica q_3 stessa:

$$F_3 \propto q_3$$

e, quindi, possiamo dedurre che la forza di Coulomb che agisce su una carica dovuta alle cariche circostanti:

- è una forza a distanza;
- è direttamente proporzionale alla carica;
- si annullerebbe se si eliminassero le cariche circostanti.

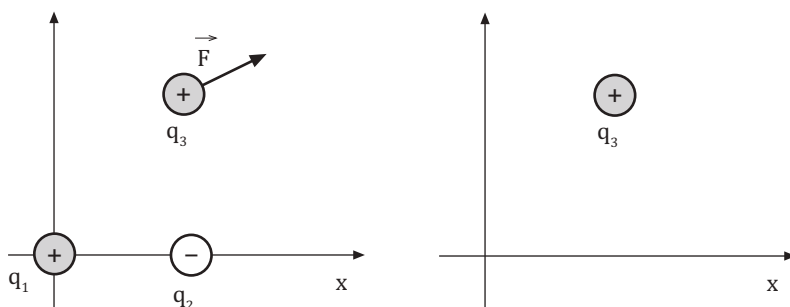


Fig. 4

Una forza con caratteristiche simili a queste è la forza gravitazionale, la quale è strettamente connessa con il campo gravitazionale; infatti, il campo gravitazionale è quella regione di spazio all'interno della quale un qualsiasi oggetto risente della forza gravitazionale.

Possiamo estendere il concetto di campo alle forze coulombiane definendo così un campo elettrico.

Un **campo elettrico** è una regione di spazio all'interno della quale gli oggetti carichi risentono di forze elettriche.

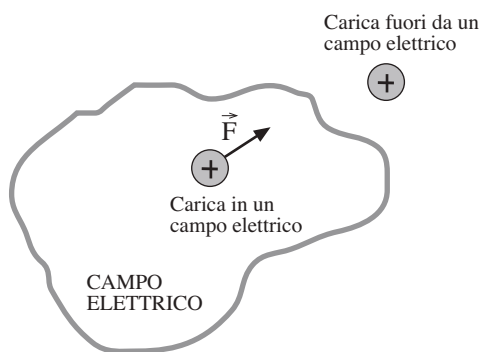


Fig. 5

La formula $F_3 \propto q_3$ può anche essere riscritta come:

$$F_3 = Eq_3$$

e, più in generale, tenendo conto della natura vettoriale della forza, come:

$$\vec{F} = q\vec{E}$$

dove $\vec{E}$ è il vettore intensità di campo elettrico o, semplicemente, **vettore campo elettrico**.

Esso è definito dal rapporto:

$$\vec{E} = \frac{\vec{F}}{q}$$

e si misura in N/C.

La carica q è anche detta **carica esploratrice** o **carica di prova** perché è su di essa che viene misurato il campo elettrico.

Spostando la carica elettrica all'interno del campo e misurando la forza elettrica che agisce su di essa, possiamo ricavare il vettore campo elettrico in ogni punto del campo.

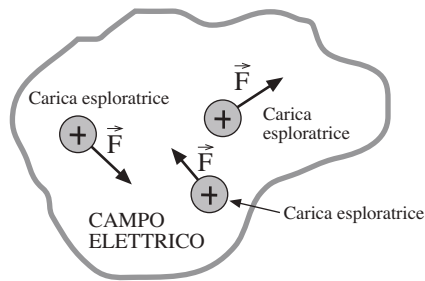


Fig. 6

Dividendo questo valore per la quantità di carica della carica esploratrice possiamo ricavare il valore del campo elettrico $\vec{E}$.

La carica esploratrice deve essere sufficientemente piccola da non modificare il campo prodotto dalle altre cariche.

a) Campo elettrico di una carica puntiforme

Se il campo elettrico è generato da una sola carica elettrica possiamo determinare la sua intensità utilizzando la legge di Coulomb:

$$F_{\text{elet}} = k \frac{q_1 q_2}{d^2}$$

dove possiamo supporre che q_1 sia la carica che genera il campo e q_2 sia la carica esploratrice:

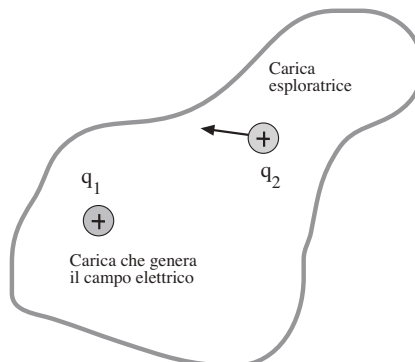


Fig. 7

Avremo infatti che:

$$E = \frac{k \frac{q_1 q_2}{d^2}}{q_2} = k \frac{q_1}{d^2}$$

Se il campo elettrico è invece generato da più cariche, per il principio di sovrapposizione, per determinare l'intensità possiamo calcolare i campi elettrici generati da ogni singola carica per poi sommarli vettorialmente.

b) Linee di forza

Un campo elettrico può anche essere visualizzato mediante le sue *linee di forza*.

Una **linea di forza** è una linea in ciascun punto della quale il vettore intensità di campo elettrico risulta tangente alla linea stessa.

Vediamo come si costruisce una linea di forza.

Consideriamo due cariche elettriche: una positiva e una negativa.

1. Poniamo una carica esploratrice in un punto P_1 del campo e tracciamo il vettore campo elettrico in quel punto.
2. Spostiamoci di un tratto infinitesimo lungo il vettore campo elettrico (punto P_2) e tracciamo il vettore campo elettrico in P_2 .
3. Spostiamoci di un tratto infinitesimo lungo il vettore campo elettrico in P_2 (punto P_3) e tracciamo il vettore campo elettrico in P_3 .
4. Continuiamo nello stesso modo in un punto P_4 .
5. In questo modo arriviamo a tracciare una linea che unisce le due cariche e cioè una linea di forza.

I punti sono stati, infatti, presi a distanze talmente piccole che i vettori campo elettrico dei punti della linea di forza sono tangenti alla linea stessa e questo rispecchia la definizione di linea di forza.

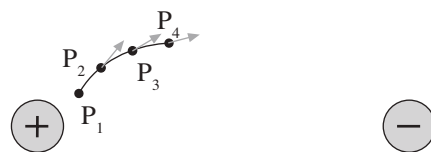


Fig. 8

Nella figura 9 sono mostrate le linee di forza di una carica positiva, di una carica negativa e di un dipolo elettrico come quello appena studiato.

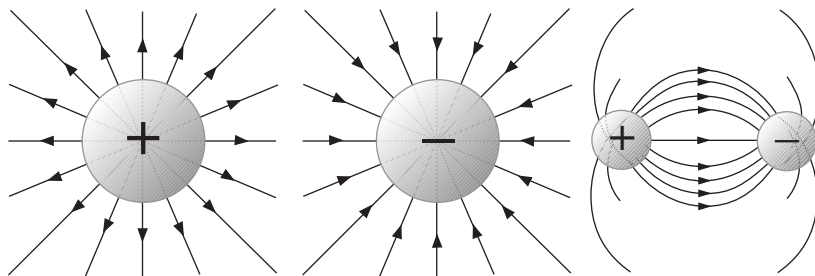


Fig. 9

• Regole per tracciare le linee di forza

Siccome il campo elettrico è una funzione continua dello spazio, in teoria si dovrebbero tracciare infinite linee di forza, una per ogni punto dello spazio occupato dal campo elettrico. Una cosa del genere, oltre a essere impossibile, non è di grande utilità. Sono state definite alcune regole su come tracciare le linee di forza di un campo elettrico, in modo che dalla loro visualizzazione sia possibile avere un'idea del campo elettrico stesso.

1. Le linee di forza escono dalle cariche elettriche positive ed entrano nelle cariche elettriche negative.

2. Il numero di linee che escono da una carica positiva o entrano in una carica negativa è direttamente proporzionale alla carica.
3. Le linee che escono o entrano in una carica puntiforme sono tracciate simmetricamente.
4. La densità delle linee, cioè il numero di linee riferito all'unità di area della superficie perpendicolare alle linee, è direttamente proporzionale all'intensità del campo.
5. Due linee non possono intersecarsi.

Se $\vec{E}$ è il campo elettrico su una superficie piana dS , sufficientemente piccola da poter ritenere che il campo elettrico sia costante su tutti i suoi punti, il numero di linee di forza che attraversano la superficie dS deve essere pari al prodotto EdS .

c) Campi elettrici generati da più cariche

Se il campo elettrico è generato da più cariche elettriche, per calcolarne il valore in un punto si utilizza il principio della sovrapposizione.

Si calcola il valore del vettore campo elettrico generato da ogni singola carica in quel punto e poi si sommano (vettorialmente) i valori ottenuti.

Nella figura 10.a sono mostrati i vettori campo elettrico generati dalle cariche q_1 e q_2 in alcuni punti dello spazio.

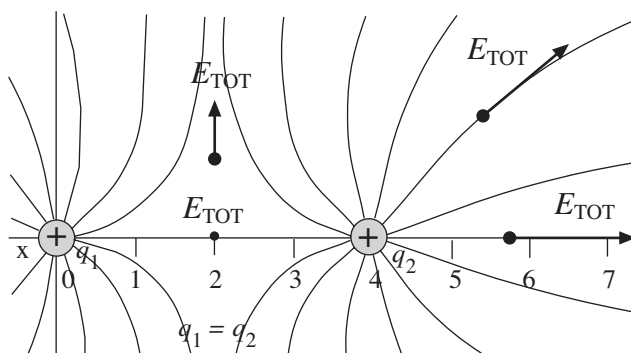


Fig. 10.a

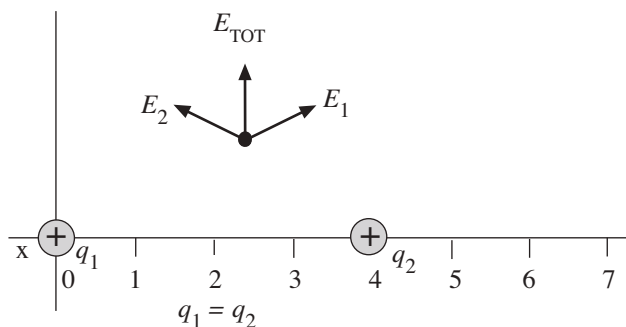


Fig. 10.b

Nella figura 10.b è mostrato come il valore del campo elettrico in un punto sia in realtà dato dalla somma vettoriale dei campi elettrici generati dalle singole cariche in quel punto.

8 Il flusso del campo elettrico

Data una superficie piana S attraversata da un campo elettrico E costante su tutta S , si definisce **flusso del campo elettrico E attraverso S** il prodotto scalare:

$$\Phi_E = \vec{E} \cdot \vec{A}$$

dove $\vec{A}$ è un vettore, detto **vettore superficie**, che caratterizza la superficie S stessa, avente:

- modulo uguale all'area della superficie;
- direzione perpendicolare alla superficie;
- verso arbitrario.

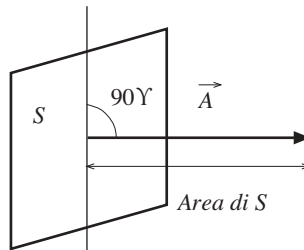


Fig. 11

Il flusso di un campo elettrico attraverso una superficie indica la quantità di campo che attraversa la superficie.

Il prodotto scalare fra due vettori è pari al prodotto di un vettore per la proiezione dell'altro vettore su di esso.

Dunque, se la **superficie è perpendicolare** al campo elettrico, il vettore $\vec{A}$ è parallelo al vettore $\vec{E}$ (Fig. 12.a) e il flusso è pari a:

$$\Phi_E = E \cdot A$$

Se, invece, la **superficie è parallela** al campo elettrico, il vettore $\vec{A}$ è perpendicolare al vettore $\vec{E}$ (Fig. 12.b), la proiezione di $\vec{A}$ sul vettore $\vec{E}$ è nulla e quindi:

$$\Phi_E = 0$$

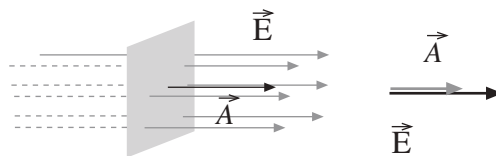


Fig. 12.a

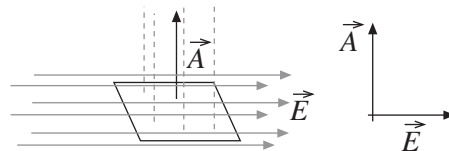


Fig. 12.b

Più in generale, indicando con ϑ l'angolo compreso fra il vettore campo elettrico e il vettore superficie, si ha che:

$$\Phi_E = \vec{E} \cdot \vec{A} = EA \cos \vartheta$$

L'angolo ϑ è l'angolo compreso fra il campo e il vettore area. Il coseno di questo angolo determina la componente del campo che attraversa perpendicolarmente la superficie.

Il flusso del campo elettrico è quindi definito dal prodotto scalare (o interno):

$$\Phi_E = \vec{E} \cdot \vec{A} = EA \cos \vartheta$$

Il flusso elettrico è misurato in Nm^2/C .

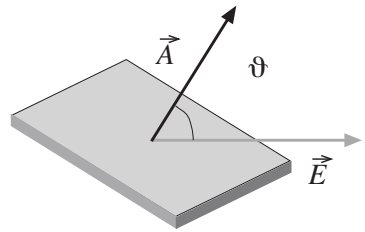


Fig. 13

9 Il teorema di Gauss

Finora siamo sempre partiti dalle cariche che generavano un campo elettrico per determinare il valore del campo stesso.

Il matematico tedesco Carl Friedrich Gauss fece il contrario: partendo dal valore del campo elettrico in una regione dello spazio analizzò le cariche elettriche che lo generavano sfruttando sia quelle che lui stesso chiamò **superfici gaussiane** sia un **teorema** che prende il suo nome.

Una **superficie gaussiana** è una superficie che avvolge completamente una certa regione di spazio, può assumere qualsiasi forma ed è utilizzata per analizzare i campi elettrici che la attraversano.

Il **teorema di Gauss** stabilisce una relazione tra il flusso elettrico che attraversa una superficie gaussiana e la quantità di cariche in essa racchiusa. Esso afferma che:

*Il **flusso elettrico** Φ_E che attraversa una superficie chiusa (gaussiana) è proporzionale alla quantità di cariche q_{incl} che essa racchiude:*

$$\Phi_E = \frac{1}{\epsilon} q_{incl}$$

dove ϵ è la costante di permittività elettrica.

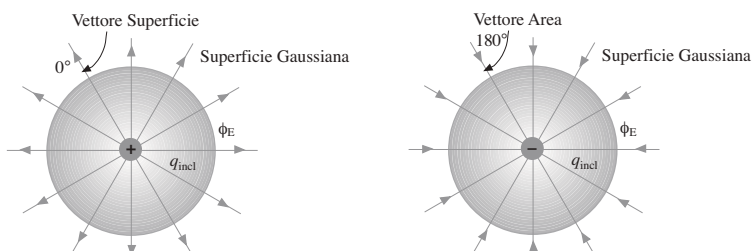
La costante di permittività elettrica nel vuoto è indicata con ϵ_0 e vale:

$$\epsilon_0 = 8,854\,18 \times 10^{-12} \frac{C^2}{Nm^2}$$

a) Flusso positivo e flusso negativo

Per convenzione, il vettore superficie è sempre diretto verso l'esterno di una superficie gaussiana.

Questo significa che se una carica positiva è posta nel centro di una superficie gaussiana sferica (Fig. 14.a), il vettore superficie ha, in ogni punto, la stessa direzione e lo stesso verso della linea di forza che attraversa la superficie gaussiana in quel punto.



Figg. 14.a e 14.b

Quindi, essendo i due vettori paralleli, il flusso attraverso l'intera superficie è dato da:

$$\Phi_E = EA$$

ed è positivo.

Se, invece, la superficie gaussiana avvolge una carica negativa (Fig. 14.b), il vettore superficie è in ogni punto della superficie gaussiana parallelo al vettore campo elettrico ma ha verso opposto:

$$\Phi_E = -EA$$

e il flusso è negativo.

La potenzialità del teorema di Gauss risiede nel fatto che non dipende dalla superficie di Gauss utilizzata infatti, per esempio, se si raddoppia la superficie gaussiana si dimezza l'intensità del campo elettrico ma il flusso totale che attraversa la superficie, essendo proporzionale al loro prodotto, rimane invariato.

Una conseguenza piuttosto sorprendente del teorema di Gauss è che le cariche poste all'esterno di una superficie gaussiana (Fig. 15) non alterano il flusso del campo elettrico attraverso la superficie. Questo in quanto **tutte** le linee di forza dei campi elettrici generati da cariche che si trovano all'esterno della sfera attraversano la superficie gaussiana e, quindi, il **flusso totale** di queste linee di forza attraverso la superficie gaussiana è **nullo**.

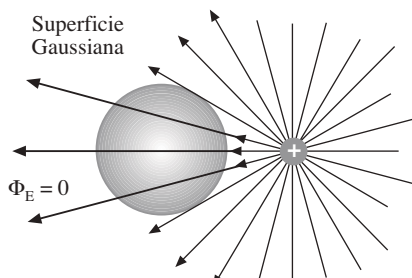


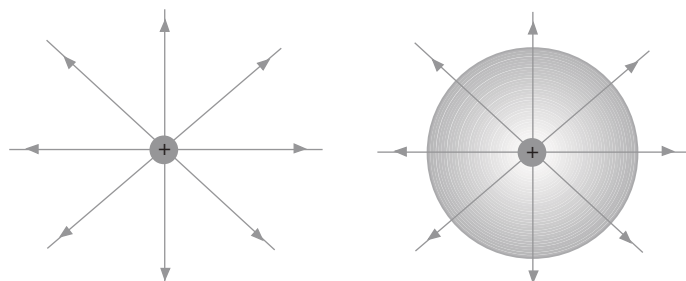
Fig. 15

b) Come scegliere la superficie gaussiana

Supponiamo di voler misurare il flusso di un campo elettrico generato da una singola carica in modo da poter applicare il teorema di Gauss.

Per iniziare possiamo tracciare le linee del campo generato dalla carica (Fig. 16.a).

Successivamente disegniamo una superficie gaussiana attorno a questa carica. Questa superficie può assumere una qualsiasi forma, noi consideriamo una superficie gaussiana sferica (Fig. 16.b).



Figg. 16.a e 16.b

In questo caso il vettore area è sempre perpendicolare alla sfera e quindi parallelo al campo elettrico. Quindi su tutta la superficie si ha che:

$$\vartheta = 0 \rightarrow \cos\vartheta = 1$$

Inoltre, siccome il campo elettrico ha lo stesso valore su tutta la superficie, il flusso del campo elettrico è dato da:

$$\Phi_E = EA = 4\pi r^2 E$$

10 La relazione tra il teorema di Gauss e la legge di Coulomb

Si può dimostrare che la legge di Coulomb deriva dalla legge di Gauss.

DIMOSTRAZIONE

La legge di Coulomb si utilizza per calcolare la forza elettrostatica fra due cariche q_1 e q_2 poste a una distanza r :

$$F = k \frac{q_1 q_2}{r^2}$$

Consideriamo la carica q_1 e disegniamo attorno a essa una superficie gaussiana sferica.

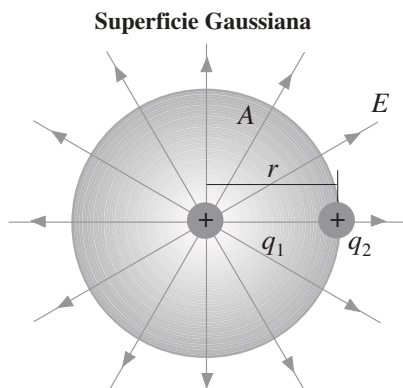


Fig. 17

Il flusso che attraversa la sfera di raggio r è pari a:

$$\cos\vartheta = 1 \Rightarrow \Phi_E = EA \cos\vartheta = EA$$

$$A = 4\pi r^2 = 4\pi(1,25)^2 = 19,6m^2$$

$$\Delta E = E(4\pi r^2)$$

A questo punto possiamo applicare la legge di Gauss per calcolare il campo elettrico sulla superficie della sfera:

$$q_1 = \epsilon_0 \Phi_E = 4\pi r^2 \epsilon_0 E$$

$$E = \frac{q_1}{4\pi r^2 \epsilon_0}$$

Infine, possiamo calcolare la forza che il campo elettrico esercita sulla carica q_2 :

$$F = q_2 E = q_2 \frac{q_1}{4\pi r^2 \epsilon_0} = \frac{1}{4\pi \epsilon_0} \frac{q_1 q_2}{r^2} = k \frac{q_1 q_2}{r^2}$$

dove $k = \frac{1}{4\pi \epsilon_0}$ è la costante di Coulomb.

Si può notare come la forza trovata è proprio la forza di Coulomb.

11 Le applicazioni del teorema di Gauss

a) Campo generato da una distribuzione sferica di cariche

Supponiamo di elettrizzare una superficie sferica conduttrice avente centro O e raggio R con una carica Q e di avere, quindi, una carica distribuita uniformemente su tutta la superficie sferica.

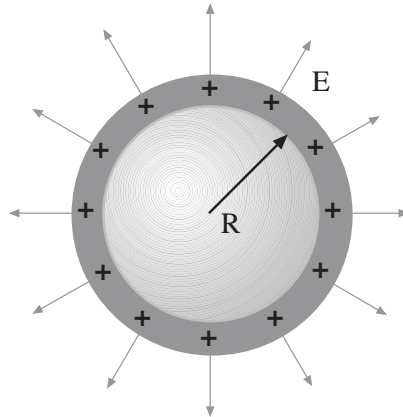


Fig. 18

Utilizzando il teorema di Gauss calcoliamo il campo elettrico:

- nei punti esterni alla sfera;
- nei punti sulla sfera;
- nei punti interni alla sfera.

• Campo elettrico nei punti esterni alla sfera

Consideriamo come superficie gaussiana la sfera di raggio r e centro O passante per il punto P .

Il campo elettrico su questa superficie è pari a:

$$E = \frac{\Phi}{S} = \frac{\Phi}{4\pi r^2}$$

Per il teorema di Gauss tale flusso è dato da:

$$\Phi = \frac{q_{\text{incl}}}{\epsilon_0} = \frac{Q}{\epsilon_0}$$

e quindi, sostituendo, si ottiene che:

$$E = \frac{\Phi}{4\pi r^2} = \frac{Q}{\epsilon_0 4\pi r^2} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \quad (11.1)$$

Da ciò si evince che il campo elettrico generato dalla carica Q , distribuita uniformemente sulla superficie della sfera, genera un campo elettrico identico a quello generato dalla stessa carica, ma concentrato nel centro della sfera.

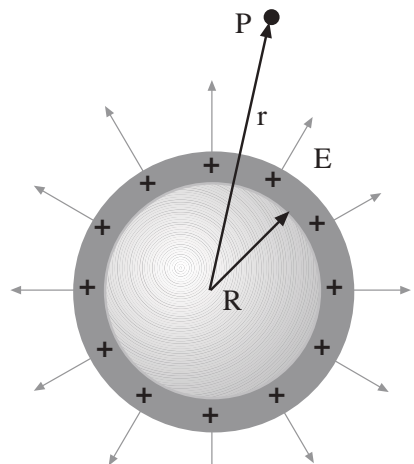


Fig. 19

- Campo elettrico nei punti sulla sfera

Nei punti molto prossimi alla superficie della sfera possiamo porre $r \approx R$ e la (11.1) diventa:

$$E = \frac{1}{4\pi\epsilon} \frac{Q}{R^2} = \frac{1}{\epsilon} \frac{Q}{4\pi R^2} = \frac{1}{\epsilon} \frac{Q}{S} = \frac{\sigma}{\epsilon} \quad (11.2)$$

dove $\sigma = \frac{Q}{S}$ è la **densità superficiale**.

La (11.2) è nota come **teorema di Coulomb**.

- Campo elettrico nei punti interni alla sfera

Nei punti interni alla sfera non esistono cariche elettriche ($Q = 0$) e quindi, per il teorema di Gauss, $\Phi = 0$.

Questo significa che il campo elettrico $E = \frac{\Phi}{S}$ è nullo all'interno della sfera.

b) Gabbia di Faraday

I risultati ottenuti per la sfera carica valgono per un qualsiasi conduttore. All'interno di un conduttore elettrizzato, il campo elettrico è nullo.

Su questo principio si basa la **gabbia di Faraday**.

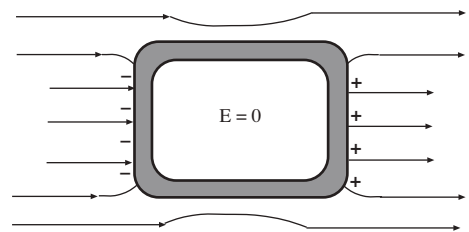


Fig. 20

c) Campo generato da una superficie piana su cui sono uniformemente distribuite cariche

Supponiamo di avere una distribuzione uniforme di cariche su una superficie piana e applichiamo il teorema di Gauss per calcolare il campo elettrico generato.

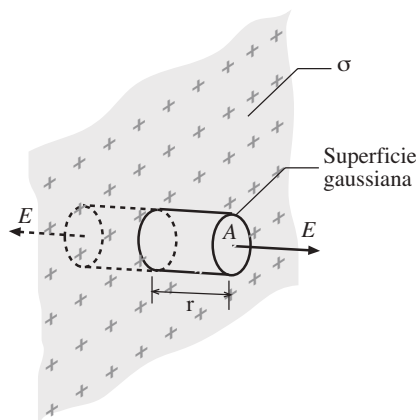
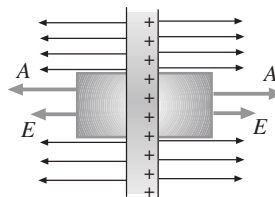


Fig. 21



Come superficie gaussiana possiamo considerare un cilindro di base A perpendicolare alla superficie piana.

Per il teorema di Gauss deve essere:

$$\Phi_E = \frac{1}{\epsilon} q_{\text{incl}} \quad (11.3)$$

dove q_{incl} è la carica posta all'interno del cilindro.

Il flusso uscente da ciascuna base del cilindro è pari a:

$$\Phi_b = EA$$

e quindi il flusso totale che attraversa entrambe le basi del cilindro è pari a:

$$\Phi = 2EA \quad (11.4)$$

Sostituendo la (11.3) nella (11.4) si ottiene:

$$2EA = \frac{q_{\text{incl}}}{\epsilon}$$

da cui segue che:

$$E = \frac{Q}{2A\epsilon} = \frac{1}{2\epsilon} \frac{Q}{A} = \frac{\sigma}{2\epsilon}$$

dove è stata introdotta la densità superficiale di carica $\sigma = \frac{Q}{A}$.

d) Campo generato da una doppia lastra carica

Consideriamo ora una doppia lastra caricata come nella figura 22.

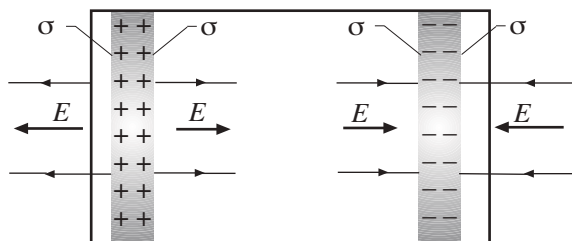


Fig. 22

In questo caso possiamo considerare come superficie gaussiana un cilindro avente la base perpendicolare alla lastra.

All'interno del cilindro la quantità di carica totale è nulla (in quanto le densità di carica sulle due piastre sono uguali e opposte) e quindi all'esterno del cilindro il campo elettrico è nullo.

Tra le due piastre, invece, si ha un campo elettrico $E = \frac{\sigma}{2\epsilon}$ diretto da sinistra verso destra, dovuto alle cariche positive, e un campo elettrico $E = \frac{\sigma}{2\epsilon}$ anch'esso diretto da sinistra verso destra, dovuto alle cariche negative.

Quindi, fra le due piastre il campo elettrico totale vale:

$$E = \frac{\sigma}{2\epsilon} + \frac{\sigma}{2\epsilon} = \frac{\sigma}{\epsilon}$$

ed è diretto dalla piastra positiva verso quella negativa (Fig. 23).

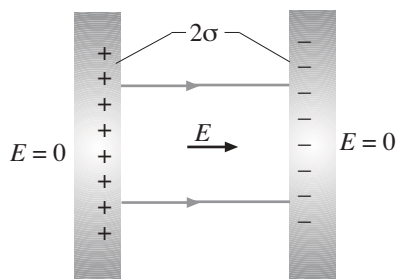


Fig. 23

1 Cos'è il potenziale elettrico?

È noto che alcune **forze**, come la forza gravitazionale, sono **conservative** e altre, come l'attrito, sono **forze non conservative**.

Se un corpo è soggetto:

- a sole forze conservative la sua energia meccanica si conserva;
- a forze conservative e a forze non conservative la sua energia meccanica non si conserva.

Per le forze conservative è possibile definire, in ogni punto dello spazio, una funzione scalare, detta **funzione energia potenziale**.

L'energia potenziale $U(A)$ di un corpo in un generico punto A è uguale al lavoro fatto dalle forze conservative del campo per spostare il corpo da un punto A a un altro punto preso come riferimento (indicato con O) cioè:

$$U(A) \equiv L(A \rightarrow O)$$

Nel caso della forza gravitazionale, prendendo come punto di riferimento la superficie terrestre, si ha che il lavoro fatto dalla forza gravitazionale per far cadere un corpo da un'altezza h è dato da:

$$L = F \cdot s = mgh$$

che è uguale all'energia potenziale posseduta dal corpo all'altezza h .

La variazione di energia potenziale di un corpo nel passare dal punto A al punto B è pari a:

$$U(A) - U(B) = -\Delta U$$

Il lavoro fatto dalle forze del campo per portare il corpo dal punto A al punto B è indipendente dal cammino percorso (è una funzione di stato) e quindi possiamo supporre che il corpo si sposti dal punto A al punto B passando per il punto identificato con lo O :

$$L(A \rightarrow B) = L(A \rightarrow O) + L(O \rightarrow B) = L(A \rightarrow O) - L(B \rightarrow O) = U(A) - U(B)$$

da cui segue che il lavoro fatto per spostare un corpo da un punto A a un punto B è uguale alla variazione di energia potenziale cambiata di segno:

$$L(A \rightarrow B) = -\Delta U$$

I campi elettrici sono campi conservativi e quindi, anche per loro, possiamo definire una funzione energia potenziale chiamata **energia potenziale elettrica**.

L'energia potenziale di una carica in un punto A è uguale al lavoro fatto dalle forze del campo per spostare la carica positiva dal punto A a un punto preso come riferimento a distanza infinita.

Il **potenziale elettrico** in un punto A di un campo elettrico è la grandezza scalare che misura il lavoro che la forza elettrica del campo deve compiere per spostare una carica positiva dal punto A fino a un punto situato a distanza infinita.

Il potenziale elettrico in un certo punto del campo è definito come l'energia potenziale riferita all'unità di carica, quindi:

il potenziale elettrico in un punto A di un campo elettrico è la grandezza scalare che misura il lavoro che la forza elettrica del campo deve compiere per spostare l'unità di carica positiva dal punto A fino ad un punto situato a distanza infinita.

Il potenziale elettrico è dunque uguale a:

$$V = -\frac{U}{q} = \frac{L}{q}$$

e si misura in volt (simbolo V):

$$1 \text{ volt} = \frac{1 \text{ joule}}{1 \text{ coulomb}}$$

2 La differenza di potenziale

La **differenza di potenziale** (d.d.p.) fra due punti A e B del campo elettrico è uguale a:

$$V_A - V_B = \frac{L_A}{q} - \frac{L_B}{q} = \frac{L_A - L_B}{q} = \frac{L(A \rightarrow B)}{q} = \frac{L_{AB}}{q} \quad (2.1)$$

Essa rappresenta il lavoro, riferito all'unità di carica positiva, che la forza elettrica esercitata dal campo deve compiere per spostare la carica da un punto all'altro del campo indipendentemente dal percorso effettuato.

Da questa definizione segue che due punti di un campo elettrico presentano una d.d.p. di 1 volt se la forza elettrica del campo compie un lavoro di 1 joule per trasportare una carica di 1 coulomb da un punto all'altro.

La (2.1) può anche essere scritta come:

$$L_{AB} = q(V_A - V_B) \quad (2.2)$$

a) Lavoro motore e lavoro resistente

Dalla (2.2) segue che il **lavoro** della forza elettrica può essere positivo (**motore**) o negativo (**resistente**) a seconda del segno della carica elettrica e del potenziale.

- **Carica elettrica positiva**

Se la carica elettrica è positiva, il lavoro del campo è motore (positivo) se la d.d.p. tra i punti A e B è positiva.

Questo vuol dire che le cariche positive immerse in un campo elettrico si sposteranno verso i punti a **potenziale minore**.

- **Carica elettrica negativa**

Se la carica elettrica è negativa, il lavoro del campo è resistente (negativo) se la d.d.p. fra i due punti A e B è negativa.

Questo significa che le cariche negative immerse in un campo elettrico si sposteranno verso i punti a **potenziale maggiore**.

Basterà, quindi, conoscere la distribuzione dei valori del potenziale nei diversi punti del campo elettrico per riuscire a prevedere, dato il segno di L, in che direzione si sposteranno le cariche immerse nel campo.

L'insieme dei punti del campo aventi lo stesso potenziale costituisce le cosiddette «**superfici equipotenziali**».

In figura sono state tracciate le linee di forza e, tratteggiate, le superfici equipotenziali.

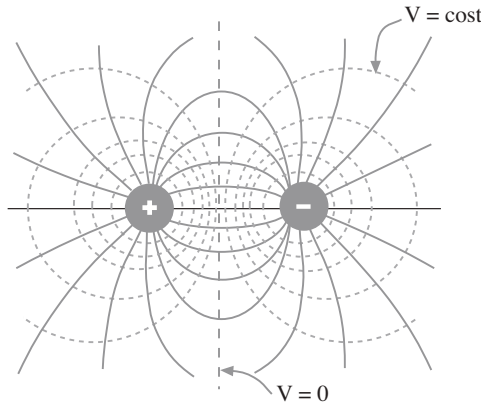


Fig. 1

In ogni punto del campo le linee di forza sono **perpendicolari** alle superfici equipotenziali.

DIMOSTRAZIONE

Il lavoro fatto dal campo elettrico E per spostare una carica su una superficie equipotenziale è nullo. Se A e B sono due punti della superficie equipotenziale si ha che:

$$L_{AB} = q(V_A - V_B) = 0$$

ma, essendo il lavoro definito dal prodotto scalare della forza esercitata dal campo sulla carica per lo spostamento della carica stessa, si ha che:

$$L_{AB} = \vec{F} \cdot \vec{s} = q\vec{E} \cdot \vec{s}$$

Siccome su una superficie equipotenziale il lavoro fatto dal campo elettrico è nullo, deve essere:

$$q\vec{E} \cdot \vec{s} = 0 \quad (2.3)$$

e quindi, essendo entrambi i fattori diversi da zero, il prodotto scalare (2.3) si annulla soltanto se i due vettori che lo compongono sono perpendicolari fra di loro e quindi soltanto se $\vec{E}$ è perpendicolare a $\vec{s}$. Poiché questo deve valere in ogni punto dello spazio, le linee di forza sono dappertutto perpendicolari alle superfici equipotenziali.

b) Elettronvolt

Nella (2.2) possiamo sostituire alla carica q il numero ne di cariche elementari che generano il campo, da cui segue che:

$$L_{AB} = q(V_A - V_B) = ne(V_A - V_B) \quad (2.4)$$

Ponendo nella (2.4) $n = 1$ e $V_A - V_B = 1$ volt, si ottiene il lavoro in joule riferito a una carica elementare.

*Questo lavoro può essere assunto come nuova unità di lavoro o di energia, detta **elettronvolt** (eV), e rappresenta il lavoro della forza elettrica per spostare un elettrone fra due punti aventi una differenza di potenziale (d.d.p.) di 1 volt.*

Siccome $e = 1,6 \times 10^{-19}$ coulomb:

$$1 \text{ eV} = 1,6 \times 10^{-19} \text{ C} \cdot V = 1,6 \times 10^{-19} \text{ J}$$

3 Il calcolo del potenziale elettrico

Calcoliamo ora il potenziale elettrico in alcuni casi particolari.

a) Campo elettrico radiale

Calcoliamo il lavoro fatto dalle forze del campo per spostare una carica da un punto A a un punto B.

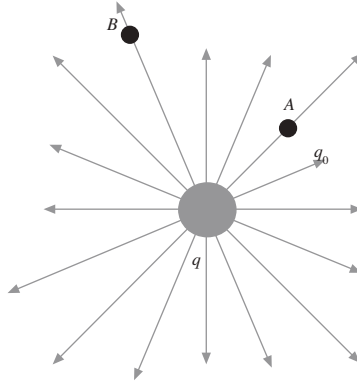


Fig. 2

Nella figura 3 è mostrata una generica traiettoria γ .

Se q_0 è la carica che deve essere trasportata dal punto A al punto B, si può dimostrare che la differenza di potenziale tra i due punti A e B è pari a:

$$V(B) - V(A) = \frac{U(B) - U(A)}{q_0} = \frac{q}{4\pi\epsilon_0 r_B} - \frac{q}{4\pi\epsilon_0 r_A} = \frac{q}{4\pi\epsilon_0} \left(\frac{1}{r_B} - \frac{1}{r_A} \right)$$

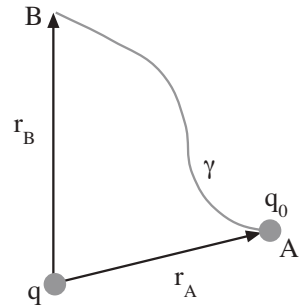


Fig. 3

DIMOSTRAZIONE

In un qualsiasi punto della traiettoria la forza del campo su q_0 è pari a:

$$\vec{F} = q_0 \vec{E} = q_0 E \vec{\mu} = \frac{q_0 q}{4\pi\epsilon_0 r^2} \vec{\mu}$$

dove $\vec{\mu}$ è il versore del vettore campo elettrico $\vec{E}$, e per uno spostamento infinitesimo $d\vec{s}$, il lavoro fatto dal campo E sarà pari a (Fig. 4):

$$L = \vec{F} \cdot d\vec{s} = \frac{1}{4\pi\epsilon_0} \frac{q_0 q}{r^2} \vec{\mu} \cdot d\vec{s} = \frac{1}{4\pi\epsilon_0} \frac{q_0 q}{r^2} dr$$

dove $dr = \vec{\mu} \cdot d\vec{s}$ è la proiezione del vettore $d\vec{s}$ lungo la direzione del vettore $\vec{E}$.

Il lavoro per spostare la carica q_0 dal punto A al punto B è dato da:

$$L(A \rightarrow B) = \int_A^B \vec{F} \cdot d\vec{s} = \int_{r_A}^{r_B} \frac{1}{4\pi\epsilon_0} \frac{q_0 q}{r^2} dr = \frac{q_0 q}{4\pi\epsilon_0} \int_{r_A}^{r_B} \frac{dr}{r^2} = \frac{q_0 q}{4\pi\epsilon_0} \left(-\frac{1}{r_B} + \frac{1}{r_A} \right) = - \left(\frac{q_0 q}{4\pi\epsilon_0 r_B} - \frac{q_0 q}{4\pi\epsilon_0 r_A} \right)$$

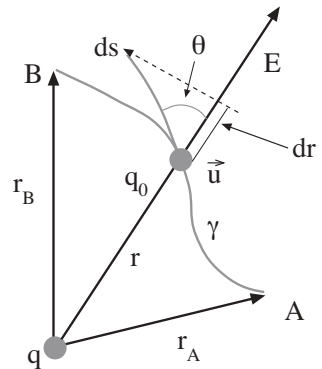


Fig. 4

Questo lavoro è uguale e opposto alla variazione di energia potenziale della carica $\Delta U = -L(A \rightarrow B)$ e quindi:

$$\Delta U = U(B) - U(A) = + \left(\frac{q_0 q}{4\pi\epsilon_0 r_B} - \frac{q_0 q}{4\pi\epsilon_0 r_A} \right)$$

Per trovare la differenza di potenziale fra i due punti A e B dobbiamo riferirci all'unità di carica, per cui essa è pari a:

$$V(B) - V(A) = \frac{U(B) - U(A)}{q_0} = \frac{q}{4\pi\epsilon_0 r_B} - \frac{q}{4\pi\epsilon_0 r_A} = \frac{q}{4\pi\epsilon_0} \left(\frac{1}{r_B} - \frac{1}{r_A} \right)$$

La differenza di potenziale si misura in volt (simbolo V):

$$1 \text{ volt} = \frac{1 \text{ joule}}{1 \text{ coulomb}}$$

Per definizione, il potenziale in un punto del campo è uguale al lavoro fatto dal campo per spostare l'unità di carica da quel punto all'infinito e quindi sarebbe anche uguale alla differenza di potenziale esistente tra il punto B considerato e il punto A ipoteticamente posto all'infinito.

Il potenziale nel punto B sarebbe quindi pari a:

$$V(B) = V(B) - V(\infty) = \frac{q}{4\pi\epsilon_0} \left(\frac{1}{r_B} - \frac{1}{\infty} \right) = \frac{q}{4\pi\epsilon_0} \frac{1}{r_B}$$

In generale, il potenziale in un qualsiasi punto di un campo radiale distante R dalla sorgente q del campo è pari a:

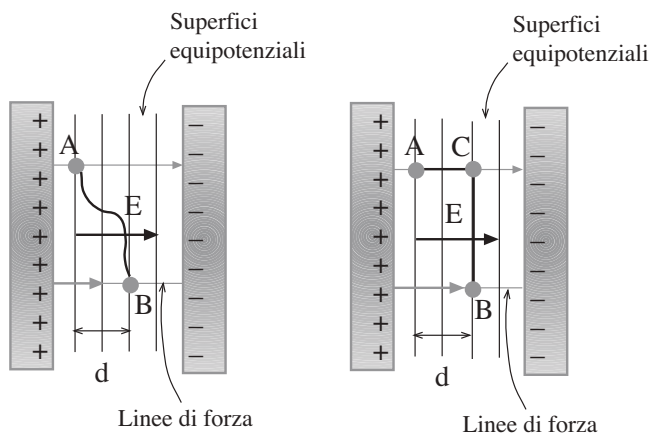
$$V(R) = \frac{q}{4\pi\epsilon_0} \frac{1}{R}$$

b) Campo elettrico uniforme

In figura 5.a sono mostrate due lastre piane cariche di segno opposto con distribuzione di carica uniforme.

Le linee di forza del campo elettrico sono quindi perpendicolari alle superfici e sono dirette verso la superficie negativa.

Le superfici equipotenziali, essendo costituite dagli infiniti punti che si trovano alla stessa distanza dalle superfici, sono infiniti piani paralleli alle superfici stesse.



Figg. 5.a e 5.b

Consideriamo due punti A e B che si trovano su due superfici equipotenziali diverse. Si può dimostrare che la differenza di potenziale fra il punto A e il punto B è pari a:

$$V(A) - V(B) = Ed = \frac{\sigma}{\epsilon} d$$

DIMOSTRAZIONE

Essendo il campo elettrico un campo conservativo, il lavoro fatto dal campo per portare la carica dal punto A al punto B è indipendente dal cammino percorso.

Si potrebbe pensare a un cammino come quello in figura 5.b passante per il punto C .

Il lavoro per portare la carica dal punto A al punto B è uguale a $L_{AB} = L_{AC} + L_{CB}$ e quindi:

$$V_A - V_B = \frac{L_{AB}}{q} = \frac{L_{AC} + L_{CB}}{q} = \frac{L_{AC}}{q} + \frac{L_{CB}}{q}$$

Essendo:

$$L_{AC} = \vec{F} \cdot \vec{AC} = q\vec{E} \cdot \vec{AC} = qEd$$

$$L_{AB} = \vec{F} \cdot \vec{CB} = q\vec{E} \cdot \vec{CB} = 0$$

e le superfici equipotenziali perpendicolari alle linee di forza, segue che:

$$V_A - V_B = \frac{qEd}{q} = Ed = \frac{\sigma}{\epsilon} d$$

c) Campo elettrico di N cariche puntiformi

Si può dimostrare che il potenziale in un punto P di un campo elettrico generato da N cariche puntiformi è pari a:

$$V(P) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \frac{q_i}{\vec{r}_i}$$

dove $\vec{r}_i$ è la distanza del punto P dalla carica q -esima.

DIMOSTRAZIONE

Consideriamo un campo generato da più cariche puntiformi (Fig. 6).

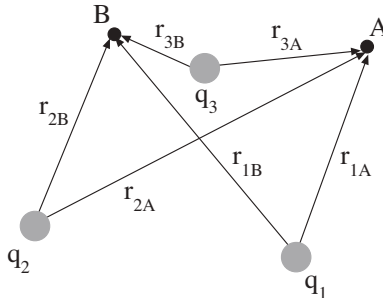


Fig. 6

Il lavoro fatto per portare una carica dal punto A al punto B è pari a:

$$L(A \rightarrow B) = \int_A^B \vec{F} \cdot d\vec{s} = q_0 \int_A^B \vec{E} \cdot d\vec{s} = q_0 \int_A^B \sum_i \vec{E}_i \cdot d\vec{s} = q_0 \sum_i \int_A^B \vec{E}_i \cdot d\vec{s}$$

Sostituendo il valore del campo elettrico:

$$L(A \rightarrow B) = q_0 \sum_i \int_A^B \vec{E}_i \cdot d\vec{s} = \frac{q_0}{4\pi\epsilon_0} \sum_i \int_A^B \frac{q_i}{\vec{r}_i^2} dr$$

e indicando con:

- $\vec{r}_{iA}$ la distanza della carica q_i da A
- $\vec{r}_{iB}$ la distanza della carica q_i da B

$$\Delta V = V(B) - V(A) = \frac{1}{4\pi\epsilon_0} \sum_i q_i \left(\frac{1}{\vec{r}_{iB}} - \frac{1}{\vec{r}_{iA}} \right)$$

Per calcolare il potenziale nel punto B possiamo immaginare di spostare il punto A all'infinito, ottenendo:

$$V(B) = V(B) - V(\infty) = \frac{1}{4\pi\epsilon_0} \sum_i q_i \left(\frac{1}{r_{iB}} - \frac{1}{\infty} \right) = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{r_{iB}}$$

Quindi, in un generico punto P del campo, il potenziale è dato da:

$$V(P) = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{r_i}$$

4 La capacità di un conduttore

Consideriamo un conduttore C isolato e non carico: il potenziale su C è nullo.

Se carichiamo il conduttore con una carica Q il suo potenziale non sarà più nullo ma assumerà un valore V_C proporzionale alla carica Q . Infatti, il potenziale in un qualsiasi punto del campo è uguale alla somma dei potenziali di tutte le cariche che generano il campo e quindi, raddoppiando o triplicando la quantità di carica presente su un conduttore, si raddoppia o si triplica anche il potenziale del conduttore.

La proporzionalità tra il potenziale di un conduttore e la carica presente sulla sua superficie si può esprimere in questo modo:

$$\frac{Q}{V_C} = C \quad (4.1)$$

dove C è una costante di proporzionalità caratteristica del conduttore detta **capacità elettrica del conduttore** che dipende dalla forma e dalle dimensioni geometriche del conduttore.

Il nome capacità deriva dal fatto che, per la (4.1), per un dato valore di potenziale V_C , la quantità di carica che un conduttore può immagazzinare è tanto maggiore quanto maggiore è la capacità C del conduttore stesso.

Dalla (4.1) segue che:

$$Q = C \cdot V_C$$

e $Q = C$ se $V_C = 1$.

La capacità elettrica di un conduttore è, quindi, la quantità di carica che si deve fornire al conduttore per far variare il suo potenziale di un volt.

La capacità si misura in coulomb/volt e la sua unità di misura è il *farad* (in onore di Faraday), indicata con il simbolo F :

$$1 \text{ farad} = \frac{1 \text{ coulomb}}{1 \text{ volt}}$$

Un conduttore che ha la capacità elettrica di 1 farad se caricato con una quantità di carica pari a 1 coulomb varia il suo potenziale di 1 volt.

Essendo il farad una capacità elettrica enorme si preferisce utilizzare i suoi sottomultipli:

- il microfarad $1\mu F = 10^{-6}F$;
- il nanofarad $1nF = 10^{-9}F$;
- il picofarad $1pF = 10^{-12}F$.

a) Capacità di un conduttore sferico

Calcoliamo ora la capacità di un conduttore sferico di raggio R isolato (Fig. 7).

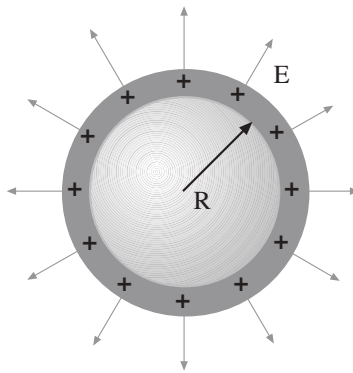


Fig. 7

Per ragioni di simmetria la densità di carica sulla sfera deve essere costante:

$$\sigma = \frac{Q}{S} = \frac{Q}{4\pi R^2}$$

Per il teorema di Coulomb, in prossimità della superficie del conduttore sferico, si ha che:

$$E = \frac{1}{4\pi\epsilon} \frac{Q}{R^2}$$

Abbiamo visto che il potenziale in un campo elettrico radiale è pari a:

$$V(R) = \frac{Q}{4\pi\epsilon_0} \frac{1}{R}$$

da cui segue che la capacità di un conduttore sferico di raggio R è uguale a:

$$C = \frac{Q}{V} = 4\pi\epsilon_0 R \quad (4.2)$$

Quindi la capacità di un conduttore sferico è proporzionale al suo raggio.

OSSERVAZIONE

Per convincersi dell'enorme capacità rappresentata da 1 farad, calcoliamo di quanto deve essere il raggio di un conduttore sferico per avere, appunto, la capacità di 1 farad.

Dalla (4.2) si ha che:

$$R = \frac{C}{4\pi\epsilon_0} = \frac{1}{4\pi(8,86 \times 10^{-12})} = 9 \times 10^9 m = 9000000 Km$$

Pertanto, una sfera carica, per avere la capacità di 1 farad, deve avere un raggio di 9 milioni di chilometri!

b) Capacità di un conduttore carico in presenza di un conduttore neutro

Finora abbiamo sempre calcolato il potenziale elettrico in conduttori isolati. Ma che cosa succede quando vicino a un conduttore carico mettiamo un conduttore neutro?

Si può dimostrare che, in presenza di un conduttore neutro, il potenziale di un conduttore carico diminuisce mentre la sua capacità aumenta.

In condizioni stazionarie tutti i punti sulla superficie di un conduttore si trovano allo stesso potenziale. Se avviciniamo un conduttore neutro a un conduttore carico esso viene attraversato dalle linee di forza del campo elettrico generato dal conduttore carico.

In questo modo i punti sulla superficie del conduttore neutro vengono a trovarsi su superfici equipotenziali diverse.

Facendo riferimento alla figura 8, i punti A e A' e i punti B e B' si trovano su due superfici equipotenziali diverse.

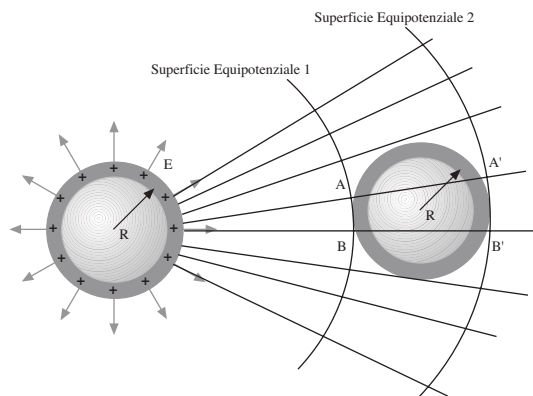


Fig. 8

Questo genera un moto di elettroni sul conduttore scarico dai punti a potenziale maggiore verso i punti a potenziale minore fino a riottenere una situazione di stazionarietà.

Si genera così una distribuzione di cariche come quella mostrata in figura 9.

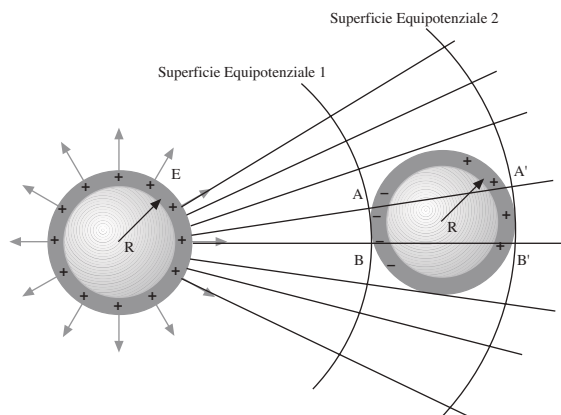


Fig. 9

In condizioni di stazionarietà il campo elettrico all'interno del conduttore scarico è nullo: se così non fosse il campo elettrico agirebbe sugli elettroni del conduttore accelerandoli e quindi andando contro l'ipotesi di stazionarietà.

La diversa distribuzione di cariche sul conduttore scarico genera un campo elettrico le cui linee di forza nei punti prossimi alla superficie sono perpendicolari a quest'ultima.

Le linee di forza del conduttore carico vengono, quindi, modificate dalla presenza del conduttore neutro (Fig. 10.a).

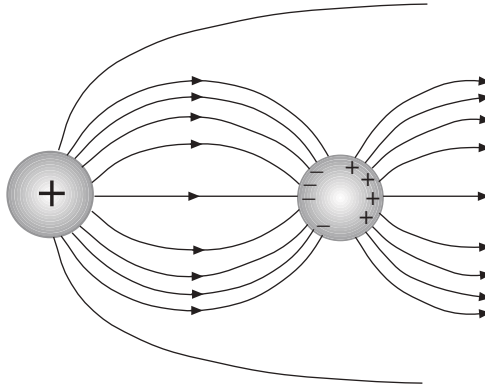


Fig. 10.a

Le linee di forza escono dal conduttore carico per entrare nel conduttore neutro e si flettono in modo da essere perpendicolari alla superficie del conduttore neutro.

Vediamo come variano il potenziale e la capacità di un conduttore carico in presenza di un conduttore neutro.

Supponiamo quindi di prendere una carica positiva q e portarla dal conduttore carico all'infinito.

La distanza fra q e le cariche negative indotte sul conduttore è minore della distanza fra q e le cariche positive (Fig. 10.b).

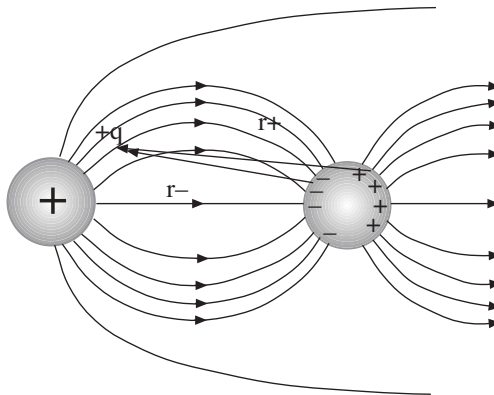


Fig. 10.b

La carica positiva verrà più attratta dalle cariche negative di quanto venga respinta dalle cariche positive e, quindi, la distribuzione di cariche indotta sul conduttore neutro aiuta il campo elettrico del conduttore carico a portare la carica all'infinito facendogli fare un lavoro minore.

Per definizione, il potenziale sulla superficie del conduttore carico è il lavoro fatto dalle forze del campo per trasportare una carica positiva dalla superficie del conduttore all'infinito.

Pertanto, il potenziale del conduttore carico diminuisce in presenza di un conduttore neutro.

La capacità del conduttore, invece, essendo inversamente proporzionale al potenziale, aumenta in presenza di un conduttore neutro.

5 I condensatori

Vediamo che cosa succede se un conduttore viene completamente circondato da un altro conduttore.

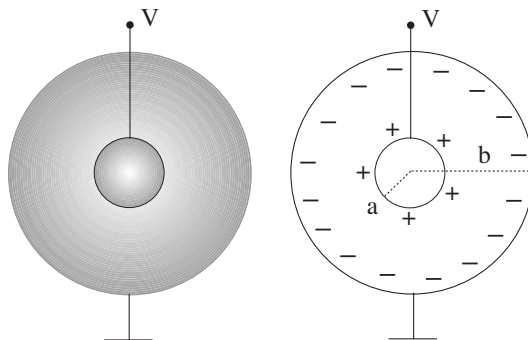


Fig. 11

Un sistema di questo tipo è detto **condensatore** e i due conduttori prendono il nome di **armature**.

Più in generale, un condensatore è costituito da due conduttori isolati, detti armature, caricati a potenziali differenti.

a) Capacità di un condensatore sferico

Un condensatore è costituito da due armature sferiche concentriche (Fig. 12).

La sua capacità può essere calcolata applicando il teorema di Gauss.

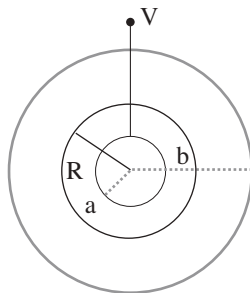


Fig. 12

Se tracciamo un'ipotetica superficie sferica di raggio R fra le due armature si ha che:

$$\Delta V = \frac{Q}{4\pi\epsilon_0} \left(\frac{1}{r_A} - \frac{1}{r_B} \right) = \frac{Q}{4\pi\epsilon_0} \left(\frac{r_B - r_A}{r_A r_B} \right)$$

La **capacità di un condensatore sferico** è, quindi, pari a:

$$C = \frac{Q}{\Delta V} = 4\pi\epsilon_0 \frac{r_A r_B}{r_B - r_A}$$

b) Capacità di un condensatore piano

Un condensatore piano è formato da due armature piane affacciate a una distanza molto piccola rispetto alle dimensioni trasversali delle armature stesse in modo da poter considerare il campo fra le armature uniforme.

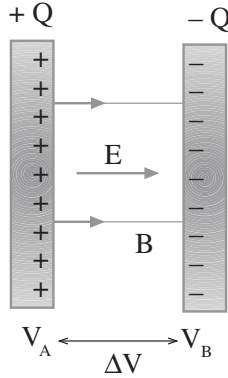


Fig. 13

Il campo elettrico fra le armature è pari a:

$$E = \frac{\sigma}{2\epsilon} + \frac{\sigma}{2\epsilon} = \frac{\sigma}{\epsilon}$$

e quindi la differenza di tensione fra le armature è pari a:

$$V_A - V_B = \frac{qEd}{q} = Ed = \frac{\sigma}{\epsilon}d$$

da cui segue che la **capacità di un condensatore piano** è uguale a:

$$C = \frac{Q}{\Delta V} = \frac{Q\epsilon}{\sigma d} = \frac{Q\epsilon}{\frac{Q}{S}d} = \epsilon \frac{S}{d}$$

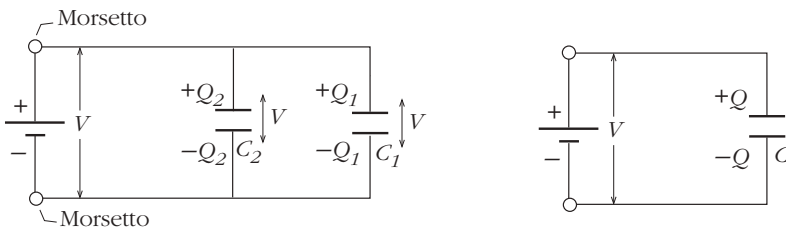
6 Condensatori in parallelo e condensatori in serie

Consideriamo due condensatori di capacità C_1 e C_2 .

All'interno di un circuito essi possono essere collegati in parallelo o in serie.

a) Condensatori in parallelo

Due o più condensatori sono collegati in **parallelo** se tra le loro armature è applicata la **stessa differenza di potenziale**.



Figg. 14.a e 14.b

Le armature illustrate nella figura 14.a sono sottoposte alla stessa tensione, per cui:

$$C_1 = \frac{Q_1}{V} \quad C_2 = \frac{Q_2}{V}$$

La carica totale immagazzinata da entrambi i condensatori è pari a:

$$Q = Q_1 + Q_2 = C_1 V + C_2 V = (C_1 + C_2) V$$

da cui segue che la capacità totale, detta anche **capacità equivalente** (Fig. 14.b) dei condensatori collegati in parallelo, è pari a:

$$C = \frac{Q}{V} = C_1 + C_2$$

Pertanto, due condensatori collegati in parallelo sono equivalenti a un unico condensatore di capacità pari alla somma delle capacità di ciascun condensatore.

In generale, la capacità equivalente di N condensatori collegati in parallelo è pari alla somma delle capacità dei singoli condensatori:

$$C = \sum_{i=1}^N C_i$$

b) Condensatori in serie

I condensatori sono detti in **serie** quando la **carica q** sulle loro armature è la **stessa per tutti i condensatori**.

Consideriamo i due condensatori inizialmente scarichi di figura 15 e colleghiamoli con un generatore di tensione.

In questo caso, la differenza di potenziale tra i due morsetti è uguale alla somma delle differenze di potenziale ai capi di ciascun condensatore:

$$V_A - V_B = (V_A - V_C) + (V_C - V_B) = \frac{Q}{C_1} + \frac{Q}{C_2} = Q \left(\frac{1}{C_1} + \frac{1}{C_2} \right)$$

da cui segue che:

$$\frac{V_A - V_B}{Q} = \frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$$

Quindi due condensatori collegati in serie sono equivalenti a un unico condensatore il cui reciproco della capacità è uguale alla somma dei reciproci delle capacità di ciascun condensatore.

In generale, il reciproco della capacità equivalente di N condensatori collegati in serie è pari alla somma dei reciproci delle capacità dei singoli condensatori:

$$\frac{1}{C} = \sum_{i=1}^N \frac{1}{C_i}$$

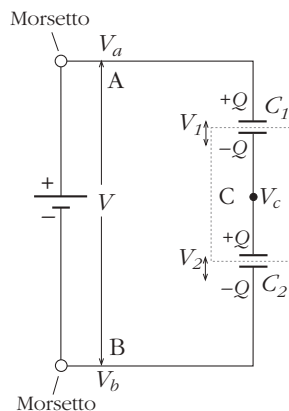


Fig. 15

7 L'energia immagazzinata in un condensatore

L'energia immagazzinata in un condensatore è pari a:

$$U = \frac{1}{2} CV^2$$

dove C è la capacità del condensatore, mentre V è la differenza di potenziale fra le sue armature.

DIMOSTRAZIONE

Consideriamo un condensatore formato da due conduttori di forma generica (Fig. 16). Un'armatura ha carica $+Q$ e potenziale V_1 , l'altra armatura ha carica $-Q$ e potenziale V_2 con $V_1 > V_2$.

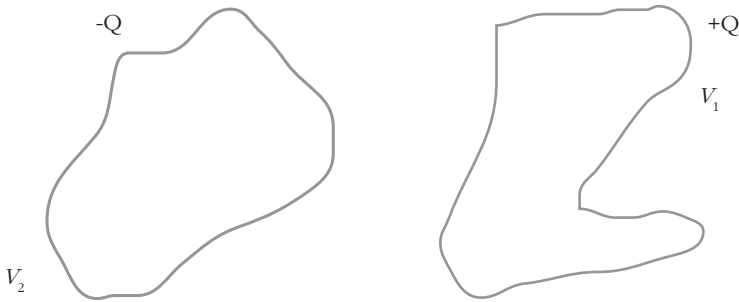


Fig. 16

Immaginiamo di trasferire una quantità di carica positiva dQ dall'armatura negativa verso l'armatura positiva.

Per fare ciò dobbiamo compiere un lavoro contro le forze del campo elettrico presente fra le due armature pari a:

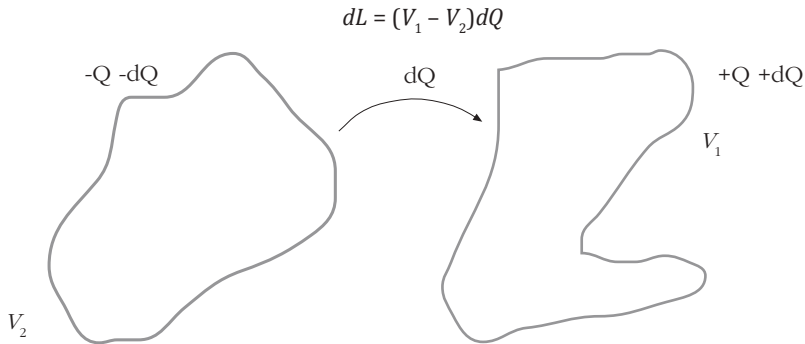


Fig. 17

Possiamo esprimere la differenza di potenziale fra le armature del condensatore mediante la capacità del condensatore:

$$V_1 - V_2 = \frac{Q}{C}$$

da cui segue che il lavoro fatto per spostare la quantità di carica dQ è:

$$dL = \frac{Q}{C} dQ$$

Questo lavoro, essendo fatto contro le forze del campo, fa aumentare l'energia potenziale del condensatore di una quantità:

$$dE = dL = \frac{Q}{C} dQ \quad (7.1)$$

Supponiamo che le armature di questo condensatore siano inizialmente scariche ($Q = 0$) e di volerle caricare fino ad avere sulle armature la quantità di carica Q . Per calcolare l'energia immagazzinata dal condensatore basta integrare la (7.1):

$$U = \int_0^Q \frac{Q}{C} dQ = \frac{1}{2} \frac{Q^2}{C} = \frac{1}{2} CV^2 \quad (7.2)$$

a) Condensatore piano: energia del campo elettrico

L'energia immagazzinata da un condensatore piano è pari a:

$$U = \frac{1}{2} \left(\epsilon \frac{S}{d} \right) (Ed)^2 = \frac{1}{2} \epsilon E^2 (Sd)$$

dove:

- S è la superficie delle sue armature;
- d è la distanza fra le armature;
- V è la differenza di potenziale fra le armature.

DIMOSTRAZIONE

Il campo elettrico E fra le armature è dato da $E = \frac{V}{d}$, da cui segue che:

$$V = Ed$$

Ricordando che la capacità di un condensatore piano è data da:

$$C = \epsilon \frac{S}{d}$$

e sostituendo entrambe queste formule nella (7.2), si ottiene che:

$$U = \frac{1}{2} \left(\epsilon \frac{S}{d} \right) (Ed)^2 = \frac{1}{2} \epsilon E^2 (Sd) \quad (7.3)$$

dove Sd è il volume di spazio racchiuso fra le armature del condensatore.

Possiamo quindi introdurre una nuova grandezza chiamata **densità di energia elettrica** η e definita da:

$$\eta = \frac{\text{Energia}}{\text{Volume}}$$

Dalla (7.3) si ottiene che per un condensatore piano:

$$\eta = \frac{\text{Energia}}{\text{Volume}} = \frac{\frac{1}{2} \epsilon_0 E^2 \cdot Sd}{Sd} = \frac{1}{2} \epsilon_0 E^2$$

e cioè l'energia riferita all'unità di volume di un campo elettrico è proporzionale al quadrato del campo elettrico.

Questo risultato può essere esteso a ogni tipo di condensatore.

1 Cos'è la corrente elettrica?

La **corrente elettrica** è definita come un flusso ordinato di elettroni che scorre all'interno di un conduttore in un certo intervallo di tempo.

Un conduttore viene attraversato da una corrente elettrica quando i suoi estremi si trovano a potenziali elettrici differenti.

Nella figura 1.a è mostrato che cosa accade a livello elettronico quando gli estremi di un conduttore vengono collegati ai morsetti di una pila.

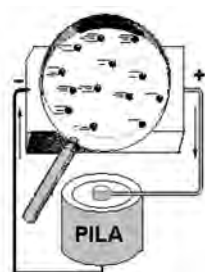


Fig. 1.a



Fig. 1.b

Gli elettroni, essendo cariche negative, vengono attratti, per la legge di Coulomb, dal polo positivo della batteria e il movimento avviene in maniera quasi immediata e a una velocità relativamente bassa.

In figura 1.b è mostrato ancora più dettagliatamente che cosa succede.

Il polo positivo della batteria attrae l'elettrone posto nel punto A.

Il punto A, avendo «perso» un elettrone, si carica positivamente e, per la legge di Coulomb, inizia ad attirare gli elettroni posti nel punto B.

Il punto B si carica a sua volta positivamente e inizia ad attirare gli elettroni posti nel punto C.

Il punto C inizia ad attirare gli elettroni dal punto D e così via.

Il moto di cariche, e quindi la corrente elettrica, all'interno del conduttore è proprio dovuto alla differenza di potenziale esistente tra i punti A e D, e, quindi, continuerà a scorrere fino a quando questi due punti, non verranno a trovarsi allo stesso potenziale.

a) Intensità di corrente elettrica

Dopo aver definito che cosa si intende per corrente elettrica è necessario quantificarla e cioè trovare un'unità di misura appropriata.

Per fare ciò dobbiamo, innanzi tutto, definire una grandezza derivata dalla corrente elettrica e cioè l'intensità di corrente elettrica.

L'**intensità di corrente elettrica** è definita come la quantità di carica che attraversa una sezione qualsiasi di un conduttore nell'unità di tempo.

Nella figura 2 è mostrata la sezione di un conduttore attraversata da cariche elettriche.

Se indichiamo con ΔQ la quantità di carica che attraversa la sezione S e con Δt l'intervallo di tempo considerato, l'intensità di corrente elettrica I è data dal rapporto:

$$I = \frac{\Delta Q}{\Delta t}$$

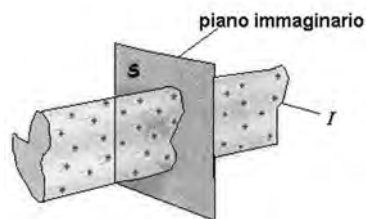


Fig. 2

Per definizione, il verso della corrente è opposto a quello degli elettroni. Quindi, siccome gli elettroni si muovono da punti a potenziale minore verso punti a potenziale maggiore, la corrente elettrica scorre da punti a potenziale maggiore verso punti a potenziale minore.

b) Ampere

L'intensità di corrente è una grandezza scalare; la sua unità di misura è l'**ampere**.

Una corrente ha l'intensità di 1 ampere quando in una sezione del conduttore scorre 1 coulomb in un secondo:

$$1A = \frac{1C}{1s}$$

Siccome 1 coulomb corrisponde alla quantità di carica trasportata da $6,25 \times 10^{18}$ elettroni, per determinare l'intensità di corrente si dovrebbe prima contare il numero di elettroni che attraversano la sezione di un conduttore in un certo intervallo di tempo e poi dividere questo numero per l'intervallo di tempo stesso.

Non essendo possibile contare gli elettroni non si può determinare l'intensità di corrente partendo da questa definizione.

Fortunatamente Ampere nel 1829 osservò sperimentalmente che due circuiti percorsi da corrente elettrica esercitano tra loro forze repulsive o attrattive a seconda se le correnti che percorrono i circuiti scorrono nello stesso verso o in verso opposti.

Egli scoprì che se due fili conduttori sono percorsi dalla stessa corrente I , l'intensità della forza esercitata tra essi è proporzionale al quadrato dell'intensità di corrente, è inversamente proporzionale alla loro distanza, ed è:

$$F = 2 \times 10^{-7} \frac{I^2}{d} l$$

da cui segue che un **ampere** è l'intensità di una corrente elettrica che, scorrendo in due conduttori rettilinei paralleli posti a una distanza di 1 m, provoca una forza di 2×10^{-7} N per metro di conduttore.

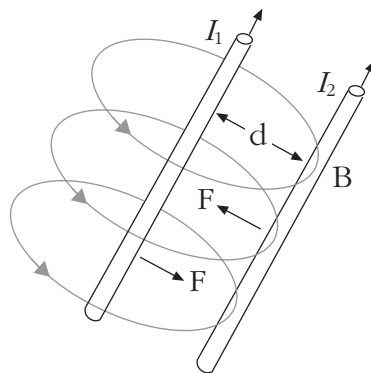


Fig. 3 - Forze esercitate fra due correnti

2 I generatori di forza elettromotrice

Abbiamo visto che in un conduttore scorre della corrente elettrica se ai suoi estremi viene applicata una differenza di potenziale.

Un dispositivo che fornisce questa differenza di potenziale è detto **generatore di tensione** o **generatore di forza elettromotrice** (f.e.m.).

Tra i generatori di tensione troviamo la pila, la dinamo della bicicletta, la batteria della macchina etc.

Tutti i generatori sono caratterizzati dall'avere due morsetti posti a due potenziali diversi:

- il morsetto posto a potenziale maggiore è anche detto **polo positivo** ed è caratterizzato dall'avere una carenza di elettroni;
- il morsetto posto a potenziale minore è anche detto **polo negativo** ed è caratterizzato dall'avere un eccesso di elettroni.

Se con un filo conduttore si connette il polo positivo con quello negativo, gli elettroni scorrono attraverso il filo dal polo negativo verso quello positivo. Questo passaggio di elettroni tenderà a ristabilire l'equilibrio di cariche dei due poli e quindi a livellarne il potenziale.

Nei circuiti elettrici che studieremo in seguito il generatore di tensione verrà rappresentato come in figura 4.

La linea verticale più lunga indica l'elettrodo a potenziale maggiore; la linea verticale più corta indica l'elettrodo a potenziale minore.

Una sorgente di f.e.m. compie del lavoro sulla carica che l'attraversa aumentando la sua energia potenziale.



Fig. 4

3 La caduta di tensione

Con il termine **caduta di tensione** o **caduta di potenziale** si indica la differenza di potenziale fra due punti qualsiasi di un conduttore attraversato da una corrente.

Una corrente scorre sempre da punti a potenziale maggiore verso punti a potenziale minore e, quindi, affinché ci sia della corrente in un circuito, ci dovranno sempre essere punti a potenziale diverso. Quando tutti i punti del conduttore hanno lo stesso potenziale non scorre più corrente nel circuito.

La tensione V tra due punti qualsiasi del conduttore è uguale al lavoro compiuto nell'unità di tempo per trasportare l'unità di carica tra i due punti; in simboli:

$$V = \frac{L}{Q} = \frac{L}{\Delta t} \frac{\Delta t}{Q} = P \frac{1}{I} = \frac{P}{I}$$

La sua unità di misura è il *volt*.

La potenza si misura in *watt* e la corrente in *ampere*, quindi:

1 volt è la differenza di potenziale elettrico fra due punti di un conduttore nel quale si deve dissipare 1 watt di potenza per far scorrere 1 ampere di corrente.

4 Resistenza elettrica e leggi di Ohm

Abbiamo visto che la corrente in un circuito è dovuta al flusso di elettroni che circolano al suo interno; questo moto è però ostacolato dalla presenza del reticolo cristallino del materiale in cui devono scorrere.

Nella prima metà del XIX secolo un professore tedesco, Georg Simon Ohm, scoprì che questa resistenza, chiamata **resistenza elettrica**, è data dal rapporto fra la tensione applicata ai capi del conduttore e la corrente che essa genera al suo interno:

$$R = \frac{V}{I}$$

La resistenza si misura in **Ohm**, in onore di chi l'ha scoperta, e viene indicata con la lettera greca Ω .

1 Ω è la resistenza elettrica fra due punti di un conduttore attraverso il quale, per una differenza di tensione di 1 volt, scorre una corrente di 1 ampere.

Quanto appena detto si traduce nella prima legge di Ohm.

a) Prima legge di Ohm

In un conduttore l'intensità di corrente è direttamente proporzionale alla tensione applicata ai suoi capi e inversamente proporzionale alla resistenza del conduttore.

È importante notare che la legge di Ohm vale sia per un conduttore che per un intero circuito e che la resistenza dipende dalla temperatura a cui si effettuano le misure.

Dopo aver trovato in che modo la resistenza elettrica dipendesse dalla tensione applicata e dalla corrente all'interno del conduttore, Ohm cercò anche di capire come la resistenza di un conduttore variasse con le sue dimensioni. Arrivò così a formulare un'altra legge che prende il nome di «seconda legge di Ohm».

b) Seconda legge di Ohm

In un conduttore la resistenza è direttamente proporzionale alla lunghezza del conduttore e inversamente proporzionale alla sua sezione.

Siccome la resistenza elettrica dipende dalla forma del conduttore, uno stesso materiale può presentare una resistenza diversa a seconda delle sue dimensioni. Per definire in modo univoco la resistenza di un materiale indipendentemente dalle sue dimensioni è stato necessario introdurre una nuova grandezza: la **resistività** ρ .

La resistività di un conduttore è la sua resistenza riferita all'unità di lunghezza e di sezione (Fig. 5).

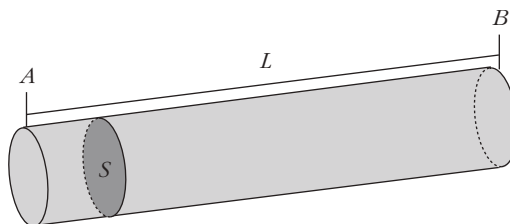


Fig. 5

Se indichiamo con:

- S la sezione del conduttore,
- L la lunghezza del conduttore,
- R la resistenza elettrica del conduttore,
- ρ la resistività del conduttore,

la seconda legge di Ohm afferma che:

$$R = \rho \frac{L}{S}$$

da cui segue che:

$$\rho = R \frac{S}{L}$$

Normalmente la resistività si misura in $\Omega \cdot m$.

L'inverso della resistività è la **conducibilità** σ :

$$\sigma = \frac{1}{\rho}$$

5 Resistività e campo elettrico

Consideriamo la figura 5 in cui è rappresentato un conduttore cilindrico di sezione S e lunghezza L percorso dalla corrente I .

Per la legge di Ohm:

$$V_A - V_B = \Delta V = RI$$

ma:

$$R = \rho \frac{L}{S}$$

e quindi:

$$\Delta V = \rho \frac{L}{S} I$$

la differenza di potenziale ai capi di un conduttore è pari al prodotto del campo elettrico per la lunghezza del conduttore:

$$\Delta V = E \cdot L$$

da cui segue che:

$$EL = \rho \frac{L}{S} I$$

e ancora, attraverso semplici passaggi:

$$\frac{E}{\rho} = \frac{I}{S} \quad (5.1)$$

Possiamo a questo punto introdurre la **densità di corrente elettrica** j definita come la corrente che scorre attraverso la sezione unitaria del conduttore:

$$j = \frac{I}{S}$$

che, sostituita nella (5.1) e dopo un semplice passaggio, dà:

$$\rho = \frac{E}{j}$$

Possiamo, quindi, definire un legame tra campo elettrico, resistività, conducibilità e densità di corrente:

$$j = \frac{1}{\rho} E = \sigma E$$

Questa relazione riassume le **due leggi di Ohm**.

6 I circuiti elettrici

Consideriamo il circuito nella figura 6; esso è formato da un generatore di tensione (la pila), un filo metallico, un interruttore e una resistenza rappresentata dalla lampadina.

Quando l'interruttore è:

- aperto, nel circuito non circola corrente e la lampadina è spenta;
- chiuso, nel circuito circola corrente e la lampadina si accende.

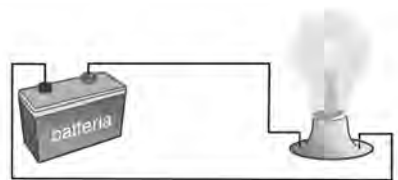


Fig. 6

All'interno di un circuito si possono individuare i rami, i nodi e le maglie, dove:

- per **ramo** si intende un collegamento in serie di più elementi in un circuito;
- per **nodo** si intende un punto in cui convergono almeno tre rami;
- per **maglia** si intende un insieme di due o più rami che formano un cammino chiuso.

Nella figura 7 è rappresentato un circuito in cui sono stati evidenziati i rami, i nodi e le maglie.

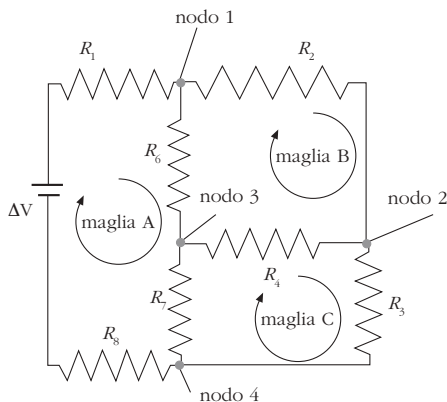


Fig. 7

a) I circuiti elettrici in corrente continua

I circuiti più semplici in corrente continua sono costituiti da:

- uno o più generatori di tensione indicati con il simbolo:



- alcuni fili metallici che servono a trasportare la corrente elettrica, indicati con il simbolo:



- una o più resistenze che indicano gli utilizzatori del circuito e cioè i componenti elettrici che hanno bisogno della corrente elettrica per funzionare. Il simbolo della generica resistenza è:



- uno o più interruttori che possono interrompere il passaggio di corrente all'interno del circuito, indicati con il simbolo:



7 Un'applicazione della legge di Ohm

Nella figura 8 è rappresentato un modello molto semplice di circuito elettrico composto da un generatore di tensione E e da una resistenza R .

Il generatore di tensione mantiene una differenza di tensione costante E tra i punti **a** e **b**, con il punto **a** a potenziale più alto.

Possiamo supporre che i fili che collegano il generatore alla resistenza non abbiano resistenza elettrica e quindi tutta la resistenza del circuito sia concentrata in R .

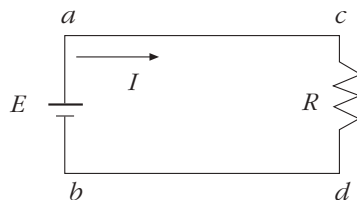


Fig. 8

Per la legge di Ohm, $V = RI$, dunque, non si hanno cadute di tensione lungo i fili e, quindi, la differenza di tensione fra i punti **a** e **b** è la stessa che fra i punti **c** e **d**, ed è pari a E . La corrente che scorre nella resistenza R è, pertanto, ricavabile dalla legge di Ohm, ed è pari a:

$$I = \frac{E}{R}$$

Inoltre, ha verso orario.

8 Generatori ideali e generatori reali

Il **generatore** separa al suo interno le cariche positive da quelle negative concentrando le prime sul suo polo positivo e le seconde sul suo polo negativo.

All'interno del generatore agiscono, quindi, forze che trasportano gli elettroni dal polo positivo verso il polo negativo. Il lavoro di queste forze, riferito all'unità di carica, è chiamato **forza elettromotrice** (f.e.m.) del generatore:

$$f = \frac{L}{q}$$

ed è misurata in volt.

In un **generatore ideale** la differenza di potenziale fra i suoi morsetti è indipendente dalla corrente nel circuito: la f.e.m. (f) è sempre uguale alla differenza di tensione fra i suoi morsetti (Fig. 9).

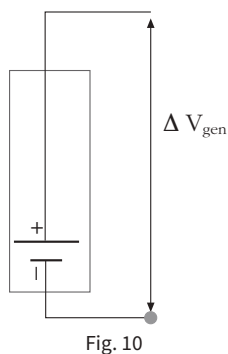


Fig. 10

In un **generatore reale**, invece, questo succede soltanto a circuito aperto quando non circola corrente (Fig. 10).

A circuito chiuso, invece, quando scorre corrente elettrica, la differenza di potenziale tra i morsetti del generatore è leggermente inferiore alla f.e.m. del generatore e questa differenza aumenta linearmente con l'intensità della corrente.

Possiamo quindi rappresentare un generatore di tensione reale come un generatore di tensione ideale con una sua resistenza che rappresenta la resistenza interna del generatore (Fig. 11).

Per la legge di Ohm, sulla resistenza R si ha una caduta di tensione pari a:

$$V = RI$$

E questa è anche la differenza di tensione fra i punti A e B:

$$V_A - V_B = RI$$

Dalla parte della batteria, invece, la differenza di tensione fra i punti A e B è pari alla forza elettromotrice del generatore diminuita della caduta di tensione sulla resistenza interna r :

$$V_A - V_B = f - rI$$

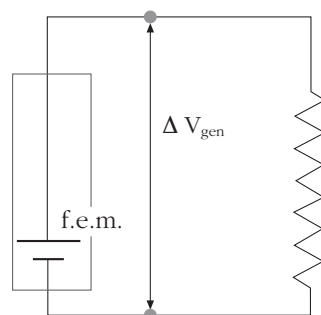


Fig. 9

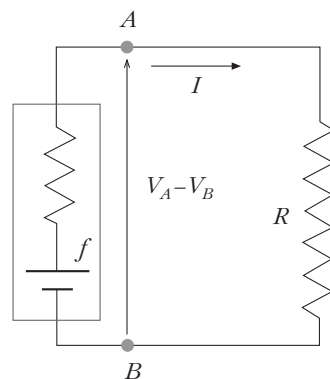


Fig. 11

Uguagliando membro a membro, si ottiene che:

$$RI = f - rI$$

da cui segue che la corrente che scorre nel circuito in cui sia presente un generatore reale è pari a:

$$I = \frac{f}{R + r}$$

mentre in un generatore ideale sarebbe pari a:

$$I = \frac{f}{R}$$

Dal confronto fra i due valori di I possiamo dedurre che la corrente nel circuito diminuisce all'aumentare della resistenza interna del generatore.

La resistenza R è detta **resistenza di carico**.

La resistenza interna del generatore assume un'importanza fondamentale nel caso in cui la resistenza di carico sia molto piccola. Infatti, in questo caso la corrente nel circuito è data da:

$$I = \frac{f}{R + r} \simeq \frac{f}{r}$$

e il fatto che r sia diversa da zero fa sì che la corrente nel circuito non sia infinita, evitando, in tal modo, il corto circuito.

9 I metodi di risoluzione dei circuiti elettrici

Esistono diversi modi con cui si possono risolvere i problemi relativi ai circuiti elettrici in corrente continua.

a) Risoluzione mediante i principi di Kirchhoff

Nel 1847 G.R. Kirchhoff, partendo dalla legge di Ohm, definì due principi con i quali è possibile risolvere molti problemi relativi ai circuiti elettrici in corrente continua:

- il **primo principio di Kirchhoff**, detto anche principio delle correnti o principio dei nodi;
- il **secondo principio di Kirchhoff**, detto anche principio delle tensioni o principio delle maglie.

• Primo principio di Kirchhoff

In un nodo la somma delle correnti entranti è uguale alla somma delle correnti uscenti.

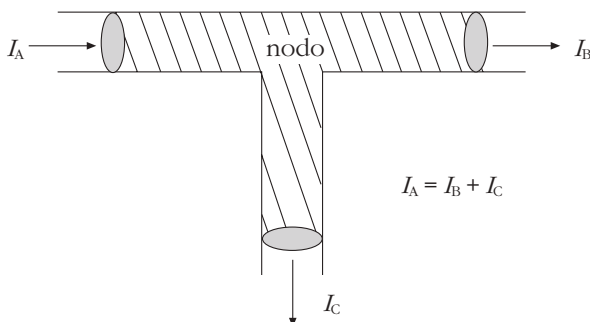


Fig. 12

Se consideriamo positive le correnti entranti e negative le correnti uscenti, il primo principio di Kirchhoff si può anche esprimere nel modo seguente: *in un nodo la somma algebrica delle correnti è nulla*.

Se indichiamo con $I_1, I_2, \dots, I_N$ le intensità delle N correnti nei rami di un nodo, per il primo principio si ha che:

$$\sum_{i=1}^N I_i = 0$$

Il primo principio di Kirchhoff deriva dalla **legge di conservazione della carica**.

Non essendoci accumulo di carica in un nodo, la quantità di carica che entra in un nodo è uguale alla quantità di carica che ne esce. Siccome, per definizione, la corrente elettrica è la quantità di carica che attraversa una qualsiasi sezione di un conduttore nell'unità di tempo, non essendoci accumulo di carica la corrente che entra in un nodo è uguale a quella che ne esce.

• Secondo principio di Kirchhoff

In una maglia la somma algebrica delle tensioni è uguale a zero:

$$\Sigma E + \Sigma RI = 0 \quad (9.1)$$

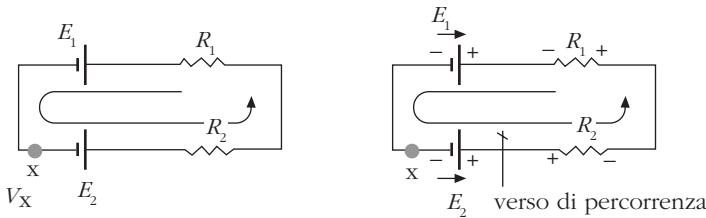
che equivale a dire che in una maglia la somma algebrica delle tensioni dei generatori è uguale a quella delle cadute di tensione sulle resistenze:

$$\Sigma E = \Sigma RI$$

Siccome nel circuito, a un generatore corrisponde in incremento di potenziale e al passaggio della corrente in una resistenza corrisponde una diminuzione di potenziale, possiamo anche dire che in ogni maglia la somma algebrica degli incrementi di potenziale è uguale a zero.

Il secondo principio di Kirchhoff deriva dalla **legge di conservazione dell'energia** in quanto afferma che *il lavoro fatto su una carica per farle percorrere un circuito chiuso deve essere uguale a zero*.

Se nella figura 13.a prendiamo un punto x avente potenziale V_x e poi percorriamo tutta la maglia in un verso qualsiasi, quando ritorniamo in x il potenziale in quel punto è ancora V_x . Se monitorassimo il potenziale di una carica del circuito partendo dal punto x vedremmo che il potenziale aumenta nel passaggio dall'armatura negativa verso l'armatura positiva di E_2 , diminuisce dopo il passaggio attraverso la resistenza R_2 , diminuisce ancora nel passaggio attraverso la resistenza R_1 e, infine, diminuisce ulteriormente nel passaggio dall'armatura positiva all'armatura negativa di E_1 .



Figg. 13.a e 13.b

Quando ritorniamo nel punto x dobbiamo ritrovare lo stesso potenziale e quindi possiamo scrivere la seguente equazione:

Aumento di potenziale in $E_1 =$ Diminuzione di potenziale in $R_1 +$
 Diminuzione di potenziale in $R_2 +$
 Diminuzione di potenziale in E .

In generale, una volta assegnato *in modo del tutto arbitrario* il verso di percorrenza di una corrente in un ramo o in una maglia del circuito, valgono le seguenti regole:

- se la corrente attraversa un generatore dal polo negativo verso il polo positivo, si ha un aumento di potenziale;
- se la corrente attraversa un generatore dal polo positivo verso il polo negativo, si ha una diminuzione di potenziale;
- se la corrente passa attraverso una resistenza, si ha una diminuzione di potenziale.

Questo significa che il punto in cui la corrente **entra** in una resistenza è maggiore del punto in cui la corrente **esce** dalla resistenza.

Visivamente parlando, una volta assegnato in modo arbitrario il verso di percorrenza della corrente nel circuito, segniamo con un **+** sia le *armature positive* dei generatori che i punti in cui le correnti *entrano* nei resistori, con un **-** sia le *armature negative* dei generatori che i punti in cui le correnti *escono* dai resistori.

Per comprendere meglio quanto detto consideriamo la figura 13.a in cui abbiamo preso come verso di percorrenza arbitrario della corrente il verso antiorario.

Segnamo con un **+** le armature positive dei generatori e i punti in cui la corrente entra nelle resistenze.

Segnamo con un **-** le armature negative dei generatori e i punti in cui la corrente esce dalle resistenze (Fig. 13.b).

Seguiamo, poi, il percorso della corrente.

Se passiamo da un **-** a un **+** la differenza di tensione è **positiva**.

Se passiamo da un **+** a un **-** la differenza di tensione è **negativa**.

Infine, applichiamo la (9.1).

Supponiamo di partire dal punto *x* della figura 13.a e percorriamo l'intera maglia secondo il verso segnato arbitrariamente.

Per prima cosa incontriamo il generatore E_2 e lo attraversiamo passando dal **-** al **+**: quindi la differenza di tensione è positiva.

Poi incontriamo la resistenza R_2 e la attraversiamo dal **+** al **-**: quindi la differenza di tensione è negativa.

Poi incontriamo la resistenza R_1 e la attraversiamo dal **+** al **-**: quindi la differenza di tensione è negativa.

Infine, incontriamo il generatore E_1 , lo attraversiamo dal **+** al **-**: quindi la differenza di tensione è negativa.

Dalla (9.1) otteniamo che:

$$\Sigma E + \Sigma RI = E_2 - E_1 - R_1 I - R_2 I = 0$$

da cui segue che:

$$I = \frac{E_2 - E_1}{R_1 + R_2}$$

Se la corrente I ricavata dai calcoli è positiva significa che il verso che abbiamo arbitrariamente assegnato è effettivamente il verso della corrente; se, invece, la corrente è negativa significa che ha verso opposto a quello ipotizzato.

• Applicazione dei principi di Kirchhoff

Per applicare i principi di Kirchhoff in un circuito dobbiamo procedere nel seguente modo:

1. Definire i rami del circuito e assegnare arbitrariamente una direzione di corrente per ogni ramo.
2. Definire le polarità (cioè i segni + e -) ai capi dei resistori del circuito e segnare anche le polarità di tutti i generatori del circuito.
3. Ricavare tutte le maglie e tutti i nodi del circuito.
4. Per quanto riguarda le maglie del circuito, partendo da un punto qualsiasi di ciascuna maglia, percorrere tutta la maglia tornando al punto di partenza e scrivere l'equazione della maglia tenendo conto delle polarità segnate.
5. Per quanto riguarda i nodi della rete, scrivere l'equazione dei nodi tenendo conto dei versi assegnati alle correnti.
6. Risolvere le equazioni.
7. In base ai risultati ottenuti ridefinire i versi delle correnti e le polarità.

OSSERVAZIONI

Quante maglie ha una rete?

Detto N il numero di nodi e R il numero di rami della rete, il numero di rami è:

$$R = M + N - 1$$

La risoluzione mediante i principi di Kirchhoff consiste nello scrivere un sistema di R equazioni in R incognite, dove le R incognite sono le correnti nei rami.

Le prime $N - 1$ equazioni derivano dal principio di Kirchhoff applicato agli $N - 1$ nodi.

Le rimanenti $R - N + 1$ equazioni derivano dal secondo principio di Kirchhoff applicato alle $R - N + 1$ maglie indipendenti.

Come individuare il numero di maglie indipendenti?

Partendo da una maglia, si individuano successivamente le altre in modo che ciascuna nuova maglia possieda almeno un ramo della rete che non fa parte delle precedenti.

Se N è il numero dei nodi, solo $N - 1$ sono indispensabili, cioè permettono di scrivere equazioni indipendenti per le correnti.

b) Risoluzione mediante il metodo di Maxwell

I circuiti possono anche essere studiati con il **metodo di Maxwell**.

In questo caso si assumono come incognite le correnti che circolano nelle maglie indipendenti.

La corrente che circola in un ramo comune a due maglie contigue sarà data dalla somma algebrica delle correnti.

10 Resistenze in serie e in parallelo

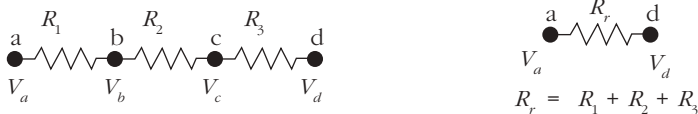
In un circuito, le resistenze possono essere collegate fra di loro in diversi modi; in particolare possono essere collegate in serie o in parallelo.

a) Collegamento in serie

Due o più resistenze sono **collegate in serie** quando sono attraversate dalla **stessa corrente**.

In figura 14.a, le resistenze R_1 , R_2 e R_3 sono collegate in serie e si può notare che:

- tutte le resistenze sono attraversate dalla stessa corrente;
- la tensione tra il punto a e il punto d si ripartisce sulle resistenze.



Figg. 14.a e 14.b

In un collegamento in serie, la **resistenza totale equivalente** è uguale alla somma delle singole resistenze.

Se $R_1, R_2, \dots, R_N$ sono N resistenze collegate in serie, l'effetto globale è uguale a quello di un'unica resistenza, chiamata *resistenza equivalente* R_r , avente come valore la somma delle singole resistenze:

$$R_r = \sum_{i=1}^N R_i$$

DIMOSTRAZIONE

Siccome ciascuna resistenza della figura 14.a è attraversata dalla stessa corrente si ha che:

$$I = \frac{V_b - V_a}{R_1} = \frac{V_c - V_b}{R_2} = \frac{V_d - V_c}{R_3}$$

La caduta di potenziale agli estremi delle resistenze può essere scritta come:

$$V_d - V_a = (V_d - V_c) + (V_c - V_b) + (V_b - V_a) = IR_1 + IR_2 + IR_3 = I(R_1 + R_2 + R_3)$$

Per la legge di Ohm applicata tra i punti d e a , la relazione fra tensione, corrente e resistenza fra questi due punti è pari a:

$$V_d - V_a = IR_r$$

e quindi, confrontando le due scritture, si ottiene:

$$R_r = R_1 + R_2 + R_3$$

La figura 14.a è equivalente alla figura 14.b.

b) Collegamento in parallelo

Due o più resistenze si dicono **collegate in parallelo** quando i capi delle resistenze sono disposti come in figura 15.

In due o più resistori collegati in parallelo:

- la differenza di potenziale agli estremi di ogni singolo resistore è la stessa;
- la corrente che arriva nel punto a si divide fra i vari resistori; quindi, se I è la corrente che scorre dal punto a al punto b , nel punto a la corrente si ripartisce fra i 3 resistori: I_1 attraverso R_1 , I_2 attraverso R_2 , I_3 attraverso R_3 .

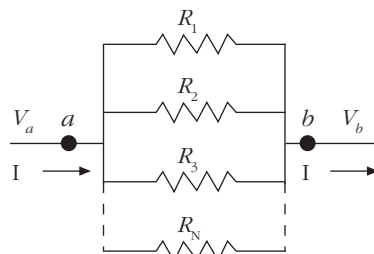


Fig. 15

In un collegamento in parallelo il reciproco (inverso) della **resistenza complessiva equivalente** R_p è uguale alla somma dei reciproci (inversi) delle singole resistenze.

Se $R_1, R_2, \dots, R_N$ sono N resistenze collegate in parallelo, l'effetto globale è uguale a quello di un'unica resistenza, chiamata *resistenza equivalente* R_r , il cui reciproco è pari alla somma dei reciproci delle singole resistenze:

$$\frac{1}{R_r} = \sum_{i=1}^N \frac{1}{R_i}$$

DIMOSTRAZIONE

Applicando la legge di Ohm a ogni singola resistenza, si ha che:

$$I_1 = \frac{V_{ab}}{R_1} \quad I_2 = \frac{V_{ab}}{R_2} \quad I_3 = \frac{V_{ab}}{R_3}$$

Per la legge dei nodi di Kirchhoff applicata al nodo a , si ha che se I è la corrente che arriva nel punto a :

$$I = I_1 + I_2 + I_3$$

e quindi:

$$I = \frac{V_{ab}}{R_1} + \frac{V_{ab}}{R_2} + \frac{V_{ab}}{R_3} = V_{ab} \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)$$

Per la legge di Ohm applicata tra i punti a e b , la relazione fra tensione, corrente e resistenza è pari a:

$$V_{ab} = V_b - V_a = IR_T$$

da cui segue che:

$$I = \frac{V_{ab}}{R_T}$$

Confrontando le due scritture segue che:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \quad (10.1)$$

Dalla (10.1) segue che la resistenza complessiva di due o più resistenze in parallelo è minore della più piccola resistenza che compone il parallelo.

11 I partitori di tensione e di corrente

a) Partitore di tensione

Può capitare di avere a disposizione un generatore con una tensione fissa V e aver bisogno di utilizzare, invece, una tensione diversa da quella fornita dal generatore.

In questo caso si utilizza un partitore di tensione il cui circuito è mostrato in figura 16.

Un **partitore di tensione**, come suggerisce il nome, serve a «partizionare» la tensione.

Nel circuito in figura 16, la tensione V è partizionata tra le due tensioni V_1 e V_2 .

Dalla figura possiamo dedurre che:

$$V = V_1 + V_2$$

Per la legge di Ohm, se I è la corrente che circola nelle resistenze si ha che:

$$V_1 = IR_1 \quad V_2 = IR_2$$

e quindi:

$$V = IR_1 + IR_2$$

da cui segue che:

$$I = \frac{V}{R_1 + R_2}$$

La tensione ai capi di ogni singola resistenza è pari a:

$$V_1 = \frac{R_1}{R_1 + R_2} V \quad V_2 = \frac{R_2}{R_1 + R_2} V$$

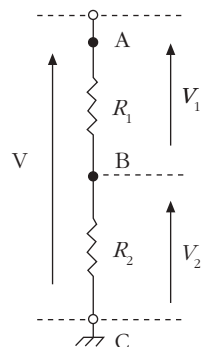


Fig. 16

da cui si deduce che, per calcolare la tensione ai capi di una resistenza collegata in serie ad altre resistenze e sottoposte a una tensione V , è sufficiente moltiplicare V per il rapporto tra la resistenza stessa e la resistenza totale del circuito.

b) Partitore di corrente

Un **partitore di corrente** serve, invece, a partizionare la corrente.

Consideriamo la figura 17 in cui i due resistori R_1 e R_2 sono posti in parallelo.

La corrente I_E si partiziona sulle due resistenze R_1 e R_2 , le quali, essendo collegate in parallelo, dovranno avere la stessa tensione ai loro capi.

Questo significa che la corrente si dividerà tra le due resistenze in modo che la tensione V prodotta ai loro capi sia la stessa.

Per la legge di Ohm deve essere:

$$V = R_1 I_1 = R_2 I_2$$

ma $I = I_1 + I_2$, quindi, per esempio:

$$I_2 = I - I_1$$

e, sostituendo, si ottiene:

$$R_1 I_1 = R_2 (I - I_1)$$

da cui segue che:

$$I_1 = \frac{R_2}{R_1 + R_2} I$$

Analogamente:

$$I_2 = \frac{R_1}{R_1 + R_2} I$$

Se al posto delle resistenze consideriamo le **conduttanze** G , cioè l'inverso delle resistenze:

$$G_1 = \frac{1}{R_1} \quad G_2 = \frac{1}{R_2}$$

e pertanto:

$$R_1 = \frac{1}{G_1} \quad R_2 = \frac{1}{G_2}$$

si ottiene:

$$I_1 = \frac{G_1}{G_1 + G_2} I \quad I_2 = \frac{G_2}{G_1 + G_2} I$$

Queste due formule servono per calcolare come la corrente si ripartisce fra due o più resistori in parallelo.

12 Lavoro e potenza elettrica

Consideriamo il circuito in figura.

Per la legge di Ohm la corrente nel circuito vale:

$$I = V/R.$$

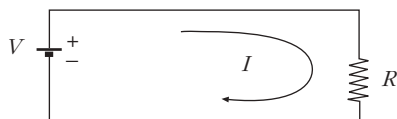


Fig. 18

La corrente è costituita dalle cariche libere che si muovono all'interno del circuito dal polo positivo verso il polo negativo del generatore, grazie all'azione del campo elettrico prodotto dal generatore stesso.

Il generatore compie, dunque, un lavoro nello spostare una carica Δq pari a:

$$L = F \cdot s = V \cdot \Delta q$$

dove V è la d.d.p. fra i morsetti del generatore.

Se dividiamo questo lavoro effettuato per il tempo impiegato a effettuarlo otteniamo la potenza P :

$$P = \frac{L}{\Delta t} = \frac{V \cdot \Delta q}{\Delta t} = V \frac{\Delta q}{\Delta t} = VI$$

Per la legge di Ohm $V = RI$ e quindi la potenza è anche uguale a:

$$P = VI = RI^2 = \frac{V^2}{R}$$

La potenza si misura in watt.

Ma dove finisce questa potenza fornita dal generatore?

a) Effetto Joule

La corrente in un conduttore è dovuta al movimento degli elettroni liberi che si muovono all'interno del reticolo cristallino.

Nel loro spostamento attraverso il reticolo, gli elettroni urtano contro gli ioni perdendo la loro energia cinetica, la quale si trasforma, per effetto dell'attrito, in energia termica. In questo modo un conduttore percorso da corrente si riscalda e questo fenomeno è detto **effetto Joule**.

L'energia fornita dal generatore viene dissipata sotto forma di calore.

b) Resistenza equivalente

La **resistenza equivalente** è la resistenza misurata tra due punti del circuito «guardando indietro» dentro al circuito. Se i nostri occhi vedono dei generatori di tensione o di corrente dobbiamo cortocircuitare i generatori di tensione e aprire i generatori di corrente.

1 Cos'è un campo magnetico?

Iniziamo a studiare il **campo magnetico** partendo da un'analisi qualitativa e quantitativa delle sue forze proprio come abbiamo fatto per il campo elettrico.

In questo caso lo studio risulta un po' più complicato in quanto, non esistendo cariche magnetiche, non possiamo definire un campo magnetico a partire dalle cariche.

Le sorgenti del campo magnetico possono infatti essere di vario tipo come è mostrato in figura 1:

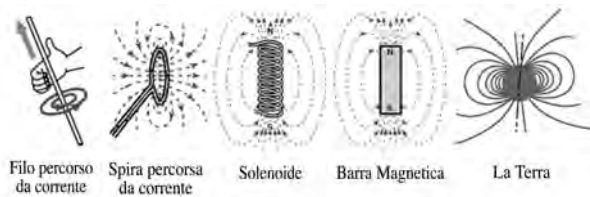


Fig. 1

Per studiare un campo magnetico dobbiamo esaminare la forza che esercita su altri magneti o su altre cariche elettriche senza conoscere le sorgenti del campo stesso.

Il campo magnetico è indicato dal vettore $\vec{B}$.

a) Campo magnetico terrestre

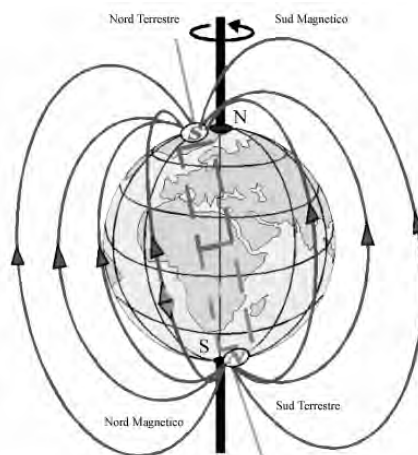


Fig. 2

Possiamo considerare la Terra come un enorme magnete circondato da un campo magnetico. Il Sud magnetico è quasi coincidente con il Polo Nord terrestre, analogamente il Nord magnetico è in prossimità del Polo Sud terrestre.

2 La forza di Lorentz

Un campo magnetico, a differenza di un campo elettrico, esercita una **forza su una particella carica soltanto** se essa è in movimento.

Questa forza è detta **forza di Lorentz** ed è definita dal prodotto:

$$\vec{F} = q\vec{v} \times \vec{B} \quad (2.1)$$

dove:

- q è la quantità di carica della particella;
- $\vec{v}$ è il vettore velocità della particella;
- $\vec{B}$ è il vettore che definisce il campo magnetico.

Dalla (2.1) possiamo dedurre che la forza di Lorentz è un vettore di modulo pari a:

$$F = qvB \sin\vartheta \quad (2.2)$$

(dove ϑ è l'angolo compreso fra i vettori $\vec{v}$ e $\vec{B}$) ed è quindi proporzionale alla quantità di carica della particella, alla velocità della particella, all'intensità del campo magnetico e al seno dell'angolo formato tra la direzione della velocità della particella e del campo magnetico.

La **direzio**ne della forza di Lorentz è perpendicolare al piano definito dai vettori $\vec{v}$ e $\vec{B}$, e il **verso** è determinato dalla regola della mano destra.

a) Regola della mano destra

Prendete la vostra mano destra (mi raccomando non la sinistra!!) e disponete pollice, indice e medio perpendicolarmente fra di loro.

Ogni dito rappresenta la direzione e il verso di un vettore.

L'indice e il medio sono i due vettori che compongono il prodotto vettoriale: l'indice è il primo vettore e il medio il secondo vettore.

Il pollice indica la direzione del vettore risultante.

Essendo la forza di Lorentz sempre perpendicolare al vettore velocità ed essendo il vettore velocità parallelo al vettore spostamento infinitesimale, si ha che la forza di Lorentz è, in ogni istante, perpendicolare al vettore spostamento.

Questo significa che il lavoro fatto dalla forza di Lorentz su una particella è sempre nullo. Infatti, il lavoro di una forza è definito dal prodotto scalare:

$$L = \vec{F} \cdot \vec{s} = Fs \cos\vartheta$$

e dunque, se $\vec{F}$ è perpendicolare a $\vec{s}$, $\vartheta = \frac{\pi}{2}$, da cui segue che:

$$\frac{\pi}{2} = Fs \cdot 0 = 0$$

Senza passare per le formule trigonometriche, si può anche dire che il lavoro è uguale al prodotto tra il vettore $\vec{F}$ e la proiezione di $\vec{s}$ su $\vec{F}$.

Siccome i due vettori sono perpendicolari, la proiezione di un vettore sull'altro è nulla e, quindi, il lavoro compiuto dalla forza di Lorentz è nullo.

Dalla (2.1) si possono dedurre le dimensioni del vettore campo magnetico. Infatti:

$$[B] = \left[\frac{F}{qv} \right]$$

Nel S.I. l'unità di misura del campo magnetico è il **tesla** (T).



Fig. 3

Il **gauss** (G) è un sottomultiplo del tesla:

$$1 \text{ tesla} = 10.000 \text{ gauss}$$

L'intensità del campo magnetico terrestre varia dai poli, dove è misurato in circa 0,75 gauss, all'equatore, dove scende a circa 0,35 gauss.

Analizziamo cosa succede quando il campo magnetico è perpendicolare o parallelo alla velocità della particella.

b) Campo magnetico perpendicolare alla velocità della particella

Supponiamo che $\vec{v}$ e $\vec{B}$ siano perpendicolari.

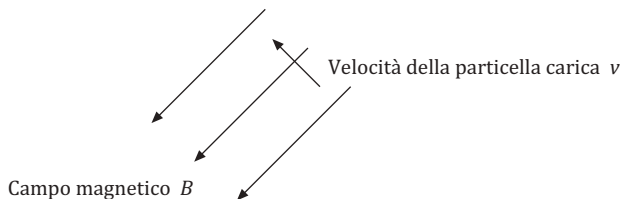


Fig. 4

In questo caso $\vartheta = \frac{\pi}{2}$ e quindi la particella risente di una forza F avente:

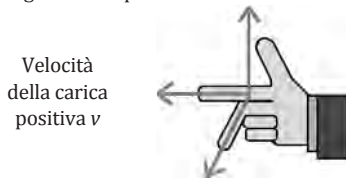
— **modulo:**

$$F = qvB \sin\left(\frac{\pi}{2}\right) = qvB \cdot 1 = qvB$$

— **direzione** perpendicolare al campo magnetico e alla velocità;

— **verso** definito dalla regola della mano destra (Fig. 5).

Forza agente sulla particella carica F



Campo magnetico B

Fig. 5

c) Campo magnetico parallelo alla velocità della particella

In questo caso, essendo il vettore $\vec{v}$ parallelo al vettore $\vec{B}$, si ha che $\vartheta = 0$ e quindi anche $\sin\vartheta = 0$, per cui la **forza di Lorentz è nulla**.

Di conseguenza, una particella che si muove in direzione parallela al campo magnetico non risente di alcuna forza di Lorentz.

3 Il moto di una particella con velocità perpendicolare alla direzione del campo

Nel paragrafo precedente abbiamo visto che:

1. la forza di Lorentz agisce su una particella carica in movimento **soltanto** se la velocità della particella non è parallela alla direzione del campo magnetico in cui è immersa;
2. il lavoro compiuto dalla forza di Lorentz è sempre nullo.

Tenendo presente queste considerazioni studiamo il moto di una particella la cui velocità è perpendicolare a quella del campo.

Dalla prima considerazione segue che la forza di Lorentz agisce sulla particella in quanto essa ha velocità perpendicolare a quella del campo.

Il lavoro compiuto dalla forza di Lorentz è uguale alla variazione dell'energia cinetica della particella. Quindi, sviluppando i calcoli, si ottiene che:

$$L = \Delta E_c = \Delta \left(\frac{1}{2} m v^2 \right) = \frac{1}{2} m (\Delta v)^2$$

Dalla seconda considerazione segue che, essendo $L = 0$:

$$\Delta v = 0$$

e pertanto la **velocità della particella** (in modulo) **non cambia**.

La forza di Lorentz farà variare la velocità della particella soltanto in direzione e non in modulo; è dunque una forza puramente centripeta, per cui la particella inizierà a percorrere orbite circolari.

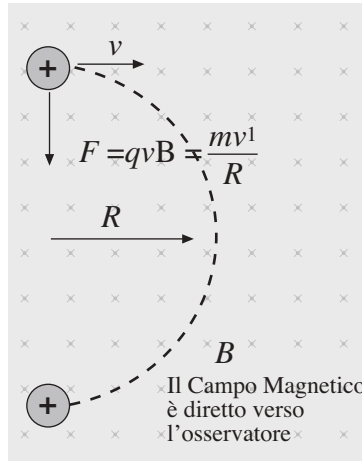


Fig. 6

La forza centripeta è legata alla velocità della particella e al raggio di curvatura della circonferenza percorsa dalla relazione:

$$F = \frac{mv^2}{R}$$

Essendo la forza di Lorentz una forza centripeta, possiamo scrivere:

$$F = qvB = \frac{mv^2}{R}$$

da cui segue che:

$$v = \frac{qRB}{m}$$

e che:

$$R = \frac{mv}{qB}$$

Il periodo della particella è pari a:

$$T = \frac{2\pi R}{v}$$

La frequenza della particella è:

$$f = \frac{qB}{2\pi m}$$

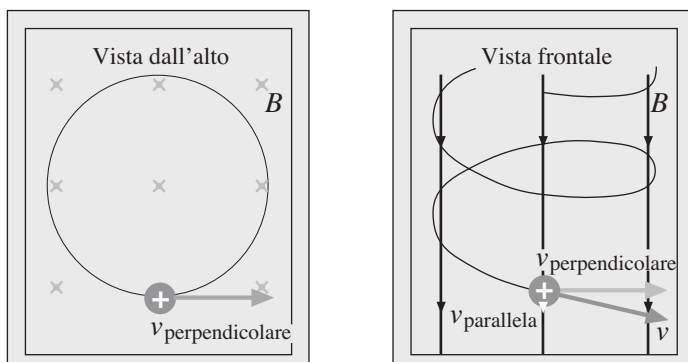
mentre la sua velocità angolare è:

$$\omega = \frac{qB}{m}$$

a) Moto di una particella con velocità obliqua rispetto al campo magnetico

Abbiamo visto che se una particella entra in un campo magnetico con velocità perpendicolare a quella del campo, essa inizia a percorrere delle traiettorie circolari.

Se la velocità della particella ha anche una componente parallela al campo magnetico, essa avrà una traiettoria elicoidale come quella mostrata in figura 7.b.



Figg. 7.a e 7.b

Per spiegare questo moto dobbiamo scomporre la velocità della particella nella componente parallela e in quella perpendicolare; dovremo studiare separatamente i due moti per poi ricomporli, trovando così le equazioni del moto.

1. Definizione della velocità del corpo all'istante iniziale e scomposizione lungo gli assi cartesiani:

$$\vec{v} = \vec{v}_{\text{parallela}} + \vec{v}_{\text{perpendicolare}}$$

2. Definizioni delle forze che agiscono sulla particella:

- sulla componente $\vec{v}_{\text{parallela}}$ non agisce alcuna forza;
- sulla componente $\vec{v}_{\text{perpendicolare}}$ agisce la forza di Lorentz.

3. Definizione del tipo di moto lungo ogni asse:

- lungo l'asse parallelo al campo magnetico, non agendo alcuna forza, la velocità della particella si mantiene costante;
- lungo l'asse perpendicolare al campo agisce la forza di Lorentz, la quale fa percorrere alla particella un **movimento circolare** su un piano **perpendicolare** al campo.

Dalla ricomposizione dei due moti si ottiene la traiettoria ellittica mostrata in figura.

b) Particella in moto in un campo elettromagnetico

Consideriamo una particella che si trova in una regione di spazio in cui oltre a un campo elettrico sia presente anche un campo magnetico.

Il **campo magnetico** esercita sulla particella una forza perpendicolare sia al campo magnetico che alla velocità della particella pari a:

$$\vec{F}_M = q\vec{v} \times \vec{B}$$

Il **campo elettrico**, invece, esercita sulla particella una forza pari a:

$$\vec{F}_E = q\vec{E}$$

Quindi, in presenza di un campo elettrico e di uno magnetico (elettromagnetico), la particella risentirà della forza:

$$\vec{F} = \vec{F}_E + \vec{F}_M = q\vec{E} + q\vec{v} \times \vec{B}$$

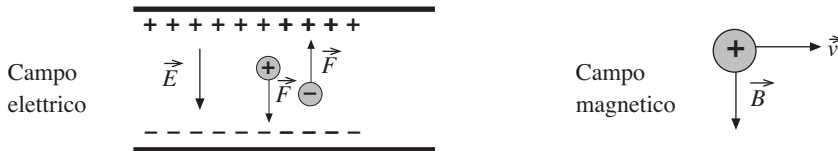


Fig. 8

4 L'effetto Hall

L'**effetto Hall**, dovuto all'azione della forza di Lorentz sulle cariche libere di un conduttore, permette di determinare il segno dei portatori di carica: prende il nome dal celebre fisico inglese Edwin Hall che lo scoprì nel 1879.

Se immergiamo in un campo magnetico una sottile striscia di materiale conduttore percorso da una corrente, si genera fra i suoi estremi una differenza di potenziale detta **tensione di Hall**.

In particolare, Hall studiò i portatori di carica presenti in una sottile lamina d'oro immersa in un campo magnetico e percorsa da una corrente perpendicolare al campo magnetico e dimostrò che la corrente all'interno della lamina era dovuta al movimento di cariche negative.

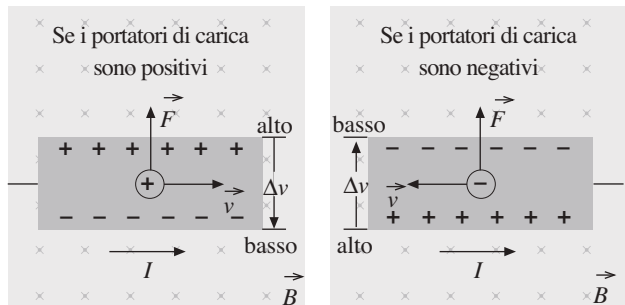
Successivamente, è stato scoperto l'elettrone e, quindi, tramite l'effetto Hall, è stato possibile dimostrare che la corrente elettrica è dovuta al moto degli elettroni.

Vediamo che cosa accade ai portatori di carica positivi (Fig. 9.a) e negativi (Fig. 9.b) all'interno della lamina.

Nel conduttore scorre una corrente continua I e le cariche elettriche si muovono, quindi, con velocità v .

Dalla figura, inoltre, è possibile evincere che i portatori di carica all'interno di un conduttore sono negativi.

Se i portatori di carica fossero positivi (Fig. 9.a), la loro velocità avrebbe lo stesso verso della corrente e la forza di Lorentz li spingerebbe verso il bordo superiore del conduttore.



Figg. 9.a e 9.b

In questo caso, il bordo superiore di un conduttore verrebbe a trovarsi a un potenziale elettrico più alto rispetto al bordo inferiore.

Se i portatori di carica fossero negativi (Fig. 9.b), la loro velocità avrebbe il verso opposto a quello della corrente e, per la regola della mano destra, la forza di Lorentz spingerebbe anch'essi verso il bordo superiore del conduttore.

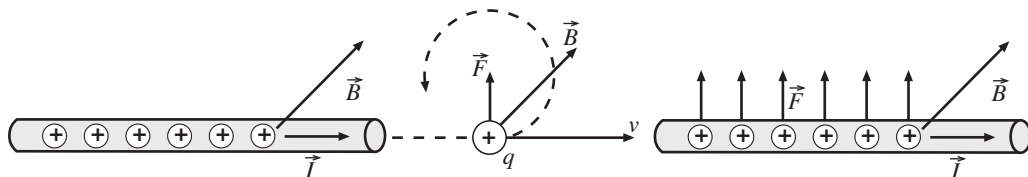
In questo caso, il bordo superiore di un conduttore verrebbe a trovarsi a un potenziale elettrico più basso rispetto al bordo inferiore.

In entrambi i casi ci sarebbe una differenza di potenziale fra la parte superiore e quella inferiore della lamina.

Hall osservò che il potenziale sulla superficie superiore era inferiore a quello sulla superficie inferiore e dedusse, quindi, che i portatori di carica erano gli elettroni.

5 La legge di Laplace

Esaminiamo l'azione di un **campo magnetico** $\vec{B}$ su un filo posto nel campo e percorso da corrente $\vec{I}$ (Fig. 10.a).



Figg. 10.a, 10.b e 10.c

Supponiamo che nell'unità di volume del filo (lunghezza = L e sezione = A) si trovino n cariche q che si muovono con velocità v .

Su ciascuna carica agisce una forza perpendicolare al campo magnetico e alla velocità della particella (Fig. 10.b).

Quindi, anche il filo stesso risentirà della forza di Lorentz (Fig. 10.c).

Essendo la corrente definita come la quantità di carica che attraversa una qualsiasi sezione di un conduttore nell'unità di tempo:

$$I = \frac{q}{\Delta t}$$

possiamo sostituire nella (2.2) $q = I\Delta t$ e dunque:

$$F = qvB \sin\vartheta = (I\Delta t)vB \sin\vartheta = I(v\Delta t)B \sin\vartheta = ILB \sin\vartheta \quad (5.1)$$

dove L è la lunghezza del filo.

In termini vettoriali, la (5.1) diventa:

$$\vec{F} = I\vec{L} \times \vec{B} = L\vec{I} \times \vec{B} \quad (5.2)$$

La (5.2) esprime la **legge di Laplace**.

La forza $\vec{F}$ che un campo magnetico $\vec{B}$ esercita su un tratto di filo rettilineo di lunghezza L e percorso da una corrente di intensità I è ortogonale sia a $\vec{I}$ che a $\vec{B}$; il suo modulo è pari a:

$$F = ILB \sin\vartheta$$

dove ϑ è l'angolo compreso fra i vettori $\vec{I}$ e $\vec{B}$, e il verso è quello dell'avanzamento di una vite destrorsa quando ruota da $\vec{I}$ verso $\vec{B}$.

La forza esercitata per unità di lunghezza è quindi pari a:

$$\frac{F}{L} = IB$$

a) Spira percorsa da corrente immersa in un campo magnetico

Consideriamo una spira percorsa da corrente (Fig. 11).

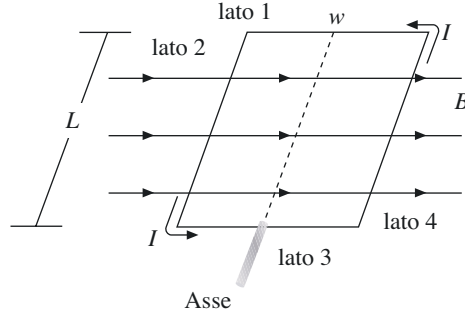


Fig. 11

Su ciascun lato della spira si esercita una forza pari a:

$$\vec{F} = L\vec{I} \times \vec{B}$$

Pertanto, la forza complessiva esercitata dal campo magnetico sulla spira sarà data dalla somma:

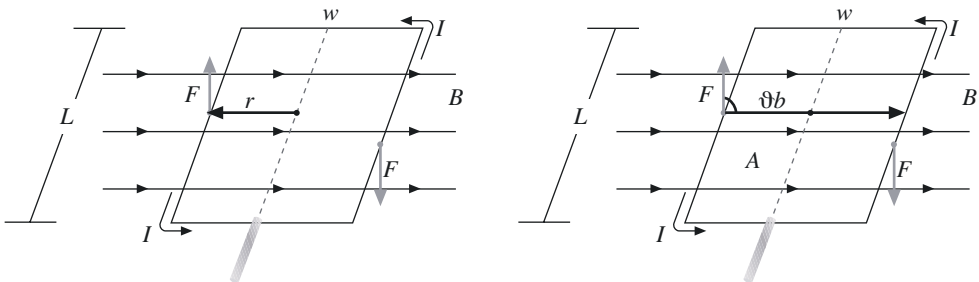
$$\vec{F} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4 \quad (5.3)$$

dove con $\vec{F}_1$, $\vec{F}_2$, $\vec{F}_3$ e $\vec{F}_4$ si intendono le forze sui rispettivi lati.

Sul lato 1 e sul lato 3 possiamo notare che i vettori $\vec{I}$ e $\vec{B}$ sono paralleli; quindi il loro prodotto vettoriale è nullo e la (5.3) diventa:

$$\vec{F} = \vec{F}_2 + \vec{F}_4$$

In figura 12 sono segnate le direzioni delle forze agenti sui lati 2 e 4 ricavate con la regola della mano destra.

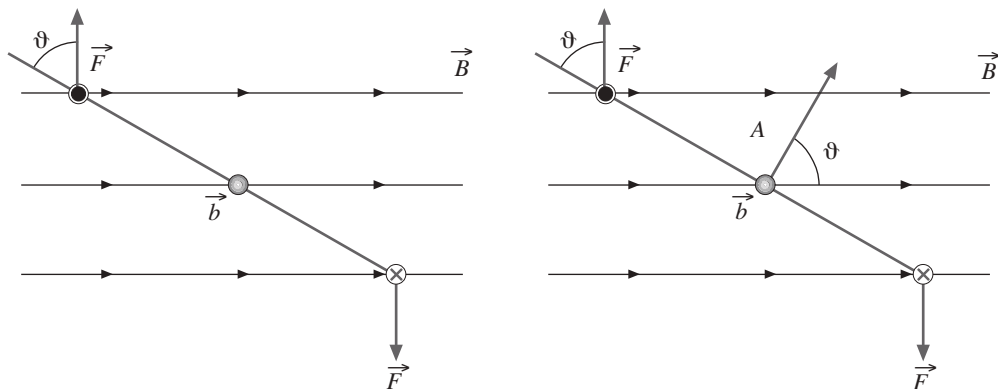


Figg. 12.a e 12.b

Entrambe le forze hanno intensità pari a LB , la stessa direzione, ma verso opposto. Formano, quindi, una coppia di forze che tende a far ruotare la spira.

Studiamo il momento di questa coppia di forze.

In figura 13 è mostrata la spira vista dall'alto:



Figg. 13.a e 13.b

La corrente sulla sinistra ha il verso uscente dal foglio mentre quella sulla destra ha il verso entrante nel foglio.

Se indichiamo con $\vec{b}$ il braccio della coppia e con ϑ l'angolo formato tra la forza e il braccio (Fig. 12.b), il momento di una coppia di forze è un vettore $\vec{M}$ definito dal prodotto vettoriale:

$$\vec{M} = \vec{b} \times \vec{F}$$

Dalla figura 13 possiamo notare che $\vec{b} = \vec{\omega}$, per cui:

$$\vec{M} = \vec{\omega} \times \vec{F}$$

ma $\vec{F} = L\vec{I} \times \vec{B}$

quindi:

$$\vec{M} = \vec{\omega} \times L\vec{I} \times \vec{B}$$

Il prodotto $\vec{\omega} \times L\vec{I}$ è un vettore perpendicolare sia al vettore $\vec{\omega}$ che al vettore $\vec{I}$ (e cioè perpendicolare alla spira stessa) avente modulo:

$$\omega L I = A I$$

dove A è l'area della spira.

Pertanto, se definiamo un vettore area (o vettore superficie) $\vec{A}$ avente:

- modulo uguale all'area della spira;
- direzione perpendicolare alla superficie della spira;
- verso pari a quello di avanzamento di una vite destrorsa che ruota secondo il verso della corrente nella spira;

il momento magnetico della coppia di forze è definito dalla formula:

$$\vec{M} = I\vec{A} \times \vec{B}$$

Se la spira fosse composta da N avvolgimenti, il momento magnetico della coppia di forze sarebbe:

$$\vec{M} = NI\vec{A} \times \vec{B} \quad (5.4)$$

b) Momento di dipolo magnetico

Possiamo riscrivere la (5.4) come:

$$\vec{M} = \vec{\mu} \times \vec{B}$$

dove il vettore:

$$\vec{\mu} = NI\vec{A} \quad (5.5)$$

ha:

- modulo pari a $NI\vec{A}$;
- direzione normale al piano della spira;
- verso di una vite destrorsa che ruota nel verso della corrente.

Il momento magnetico di dipolo $\vec{\mu}$ è mostrato nella figura 14.

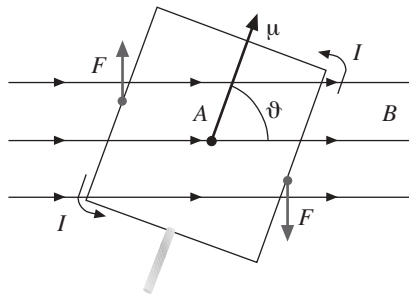


Fig. 14

Il momento magnetico della spira $\vec{M}$, avendo modulo pari a $M = \mu B \sin\vartheta$, è massimo quando il vettore $\vec{\mu}$ è perpendicolare al campo magnetico e cioè quando la spira è allineata al campo magnetico.

Il momento è nullo quando il vettore $\vec{A}$ è parallelo al campo magnetico e cioè quando la spira è perpendicolare al campo magnetico.

Questo significa che, quando il vettore $\vec{\mu}$ è orientato in direzione obliqua rispetto al campo magnetico, la spira risentirà di una forza che tenderà a far allineare $\vec{\mu}$ con il campo magnetico (e quindi a far allineare la spira perpendicolarmente al campo).

Possiamo quindi dedurre che il comportamento della spira è del tutto analogo a quello di un ago magnetico: quest'ultimo si dispone parallelamente al campo magnetico $\vec{B}$, mentre la spira orienta il suo momento magnetico $\vec{\mu}$ parallelamente a $\vec{B}$.

c) Momento di un magnete permanente

Consideriamo ora un magnete permanente di forma cilindrica e di dimensioni identiche a quelle del solenoide. Se immergiamo il solenoide in un campo magnetico, anch'esso, come la spira percorsa da corrente, risentirà di una coppia di forze che tenderà a farlo allineare perpendicolarmente al campo.

Possiamo definire, anche per il magnete, un momento magnetico di dipolo $\vec{\mu}_m$ e un momento magnetico della coppia di forze $\vec{M}_m$ tramite il quale è possibile rappresentare la coppia di forze che il campo magnetico esercita su di esso:

$$\vec{M}_m = \vec{\mu}_m \times \vec{B}$$

$\vec{\mu}_m$ dipende dalla natura del materiale ed è proporzionale al suo volume.

Se il magnete ha sezione A e lunghezza L , indicando con $\vec{m}_M = \frac{\vec{\mu}_M}{M}$ il momento magnetico per unità di volume (detto anche **magnetizzazione del magnete**), si ha che:

$$\vec{\mu}_M = V\vec{m}_M = AL\vec{m}_M \quad (5.6)$$

Un campo magnetico esercita la stessa coppia di forze su un magnete e su un solenoide se la magnetizzazione del magnete è uguale al momento di dipolo della spira, quindi se:

$$\mu_M = \mu$$

Dalla (5.6) e dalla (5.5) questo si verifica quando:

$$mLA = NIA$$

da cui si ricava che:

$$m = \frac{N}{L}I = nI$$

dove $n = \frac{N}{L}$ è la densità di spire e cioè il numero di spire per unità di lunghezza del solenoide.

Abbiamo quindi trovato una sorta di equivalenza fra un solenoide e un magnete permanente il quale viene espresso dal *principio di equivalenza di Ampere*.

• Principio di equivalenza di Ampere

Un magnete cilindrico è equivalente a un solenoide di pari sezione e lunghezza, percorso da una corrente di intensità I e avente un numero n di spire per metro tali che il prodotto di I per n risulti uguale al momento magnetico per unità di volume del magnete.

Si può anche dire che la magnetizzazione m di un magnete cilindrico è pari all'intensità di corrente per unità di lunghezza che circola nel solenoide equivalente.

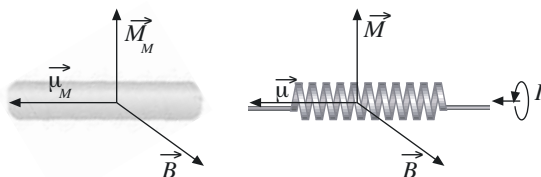


Fig. 15

d) Motore elettrico

Il **motore elettrico** a corrente continua si basa sullo stesso principio fisico che fa ruotare una spira in un campo magnetico.

Consideriamo la figura 16, in cui una spira percorsa da corrente viene immersa in un campo magnetico.

Abbiamo appena visto che essa è soggetta a una forza che tende a farla ruotare in modo che il suo momento magnetico si orienti nella direzione del campo magnetico e il piano della bobina sia perpendicolare al campo magnetico.

Il momento magnetico è definito da:

$$\vec{M} = \vec{r} \times \vec{F} = I\vec{A} \times \vec{B}$$

(dove A è l'area della spira) ed è un vettore perpendicolare al piano definito da r e da F , il cui verso è determinato dalla regola della mano destra.

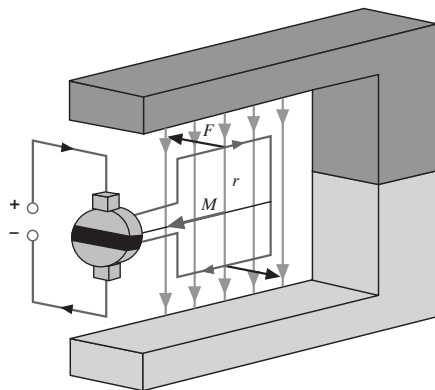


Fig. 16

Il valore di questo momento diminuisce man mano che la spira si dispone perpendicolarmente al campo e si annulla nel momento in cui la spira è esattamente perpendicolare al campo (Fig. 17).

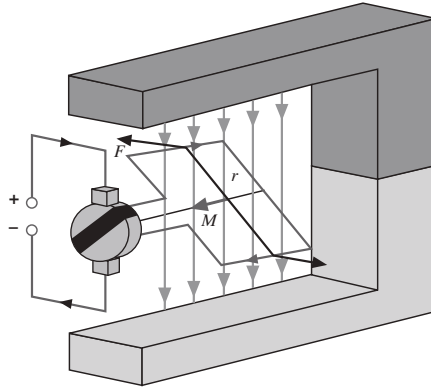
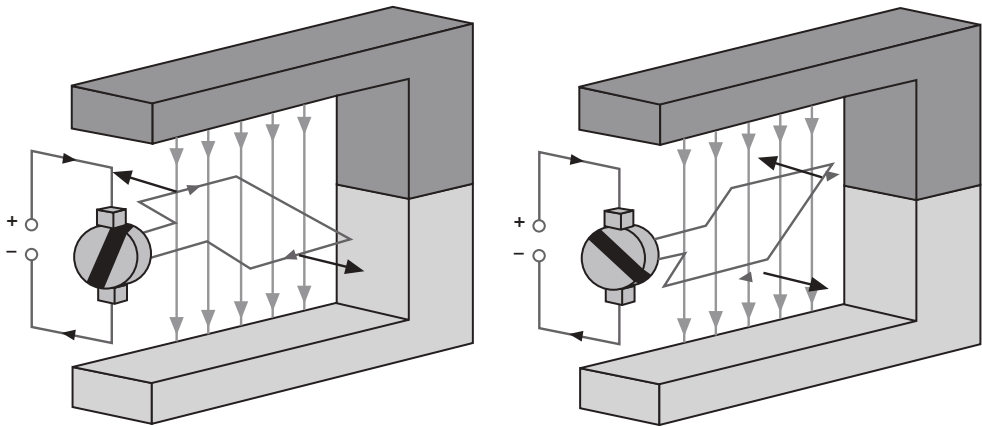


Fig. 17

La spira, però, una volta che arriva a disporsi perpendicolarmente al campo (Fig. 18.a), non si ferma immediatamente, ma supera, per inerzia, la sua posizione orizzontale, inclinandosi in modo opposto rispetto al campo (Fig. 18.b).



Figg. 18.a e 18.b

Si genera così un momento avente verso opposto, il quale tenderà a far ruotare la spira in verso antiorario in modo da farla disporre perpendicolarmente al campo. Anche in questo caso la spira non si ferma immediatamente una volta raggiunta la posizione ortogonale al campo, ma la supera leggermente.

Appena superata la posizione di equilibrio, si genera nuovamente un momento che tenderà a far ruotare la spira in verso orario per poi ruotare in verso antiorario. La spira inizia così ad oscillare.

Se si invertisse il verso della corrente nella spira, nell'istante in cui raggiunge la sua posizione di equilibrio, il momento magnetico non si invertirebbe quando la spira supera la posizione di equilibrio e, dunque, continuerebbe a ruotare nello stesso verso.

Alternando il verso della corrente in modo opportuno è, quindi, possibile far ruotare la spira in modo continuo. Su questo principio si basa il motore elettrico a corrente continua.

Nella figura 19 è illustrato il funzionamento di un motore elettrico.

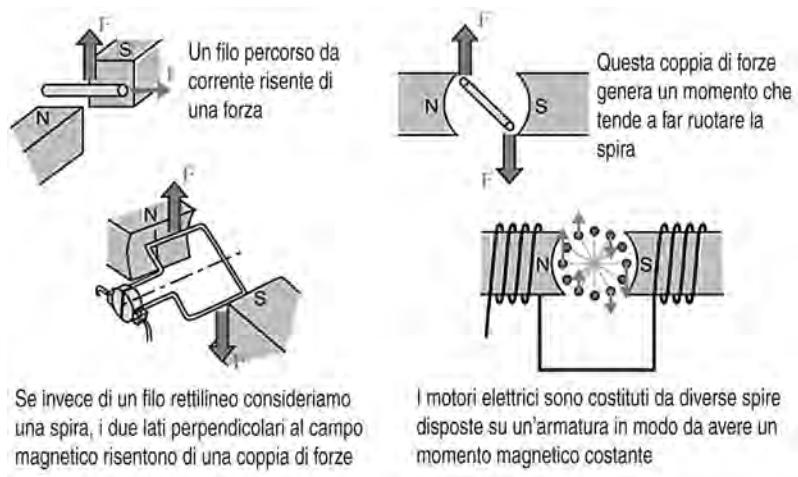


Fig. 19

Correnti elettriche e campi magnetici: l'elettromagnetismo

1 Correnti e campi magnetici

Abbiamo visto che il campo magnetico è in grado di esercitare una forza su un filo percorso da corrente definita dalla formula $F = BIL \sin\theta$. Da questa relazione segue che possiamo rilevare un campo magnetico in base alla forza che esso esercita su un filo percorso da corrente. In realtà, noi non possiamo rilevare il campo magnetico vero e proprio ma soltanto il suo effetto magnetico, il quale dipende dal materiale in cui il filo conduttore è immerso.

Per studiare il campo magnetico dobbiamo quindi definire due vettori:

- un vettore induzione magnetica $\vec{B}$, che rappresenta l'effetto magnetico del campo magnetico;
- un vettore campo magnetico $\vec{H}$, che rappresenta la causa tecnica del campo magnetico.

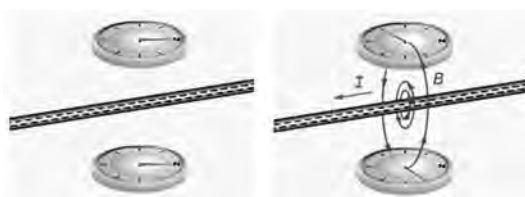
Il vettore $\vec{H}$ è quindi la causa del vettore $\vec{B}$ e il legame tra questi due vettori è dato dalla relazione:

$$\vec{B} = \mu \vec{H}$$

in cui μ è un parametro caratteristico del mezzo dal punto di vista magnetico, chiamato **permeabilità magnetica**. La permeabilità magnetica del vuoto è pari a: $\mu_0 = 4\pi \times 10^{-7}$.

2 Il campo magnetico generato da un filo percorso da corrente

Posizioniamo due bussole vicino a un filo elettrico (Fig. 1.a).



Figg. 1.a e 1.b

Se nel filo facciamo scorrere della corrente, gli aghi delle due bussole si posizionano perpendicolarmente al filo e possiamo, quindi, tracciare le linee di forza del campo magnetico come in figura 1.b.

Le linee di forza del campo magnetico sono pertanto concentriche e perpendicolari al filo; il loro verso è definito dalla *seconda regola della mano destra*.

a) Seconda regola della mano destra

Prendete la vostra mano destra e piegate le dita come se doveste fare l'autostop.

Se disponete il pollice nella direzione della corrente, le altre dita della mano indicheranno il verso delle linee del campo magnetico.

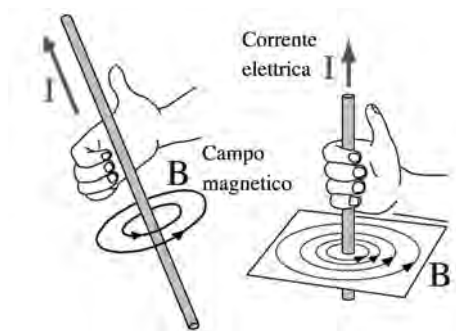


Fig. 2

b) Intensità del campo magnetico prodotto

Se posizionassimo le bussole dell'esperimento precedente a diverse distanze dal filo, noteremmo che la coppia di forze che orienta il loro ago diminuisce all'aumentare della distanza e aumenta all'aumentare della corrente.

Il vettore $\vec{B}$, che rappresenta il campo magnetico indotto, generato dalla corrente elettrica, ha in ogni punto del campo magnetico:

- direzione perpendicolare alla corrente e coincidente con la tangente alla linea di forza in quel punto;
- verso definito dalla seconda regola della mano destra;
- intensità proporzionale alla corrente e inversamente proporzionale alla distanza dal filo:

$$B = \frac{\mu I}{2\pi R} \quad (2.1)$$

dove:

- B è il valore del campo di induzione magnetica;
- I è la corrente che scorre nel filo;
- R è la distanza dal filo.

Questo significa che il vettore $\vec{H}$, il quale rappresenta il campo magnetico vero e proprio, ha modulo:

$$H = \frac{I}{2\pi R} \quad (2.2)$$

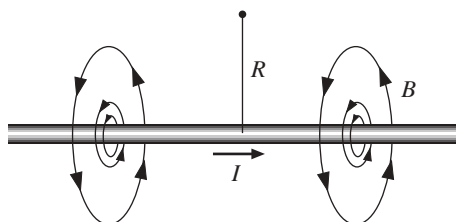


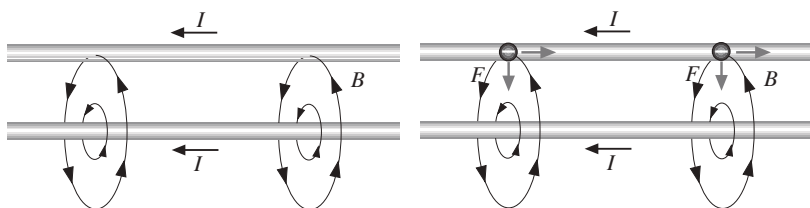
Fig. 3

Nella (2.1) si suppone che il filo abbia lunghezza infinita.

c) Forze tra due fili percorsi da corrente

Due fili percorsi da corrente si attraggono o si respingono a seconda della direzione delle correnti.

Analizziamo cosa accade quando due fili elettrici sono percorsi da corrente.



Figg. 4.a e 4.b

Il filo in basso in figura 4.a genera un campo magnetico il cui orientamento è definito dalla seconda regola della mano destra.

Questo campo magnetico interseca anche il filo superiore (Fig. 4.b) ed esercita una forza sugli elettroni, che si muovono al suo interno, avente direzione e verso definiti dalla regola della mano destra.

Quindi:

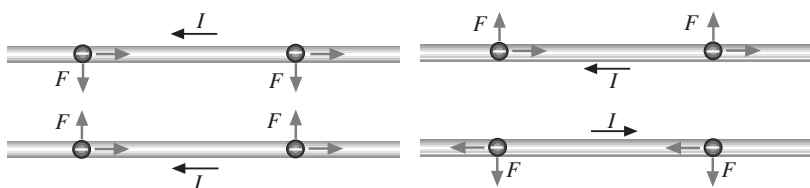
- se le correnti nei due fili hanno lo stesso verso, il filo superiore viene attratto da quello inferiore;
- se le correnti nei due fili hanno verso opposto, il filo superiore viene respinto da quello inferiore.

Se fossimo partiti dal filo superiore saremmo giunti alla conclusione che:

- se le correnti nei due fili hanno lo stesso verso, il filo inferiore viene attratto da quello superiore;
- se le correnti nei due fili hanno verso opposto, il filo inferiore viene respinto da quello superiore.

In definitiva:

- se le due correnti hanno lo stesso verso, le forze sono dirette come in figura 5.a e i due fili si attraggono;
- se le due correnti hanno verso opposto, le forze sono dirette come in figura 5.b e i due fili si respingono.



Figg. 5.a e 5.b

3 Il campo magnetico generato da due fili paralleli percorsi da corrente

Consideriamo due fili paralleli percorsi da corrente, posti ad una distanza R e percorsi dalle correnti I_1 e I_2 .

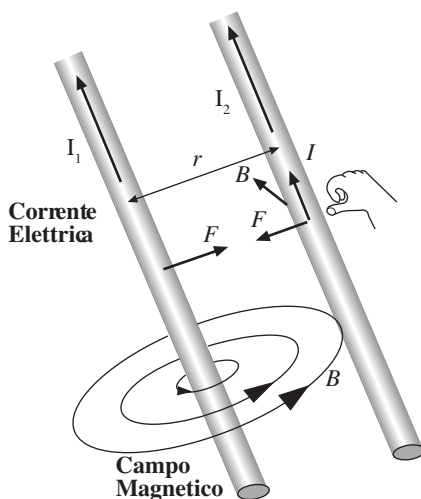


Fig. 6

Il campo generato dalla corrente I_1 nel primo filo, in prossimità del secondo filo ha intensità pari a:

$$B_1 = \frac{\mu_0 I_1}{2\pi R}$$

La direzione di questo campo, in accordo con la seconda regola della mano destra, è perpendicolare al piano definito dalle due correnti.

La forza esercitata dal campo magnetico è:

$$\vec{F} = L \vec{I} \times \vec{B}$$

Essendo la corrente perpendicolare al campo magnetico, si ha che:

$$F = L I B$$

Siccome conosciamo il valore del campo magnetico sul secondo filo, possiamo calcolare la forza che il campo magnetico generato dal primo filo esercita sul secondo:

$$\frac{F}{L} = I_2 B_1 = \frac{\mu_0 I_1 I_2}{2\pi R} \quad (3.1)$$

Essa è diretta verso il secondo filo.

Dalla (3.1) deriva la definizione dell'**unità di misura: l'ampere**.

Un ampere è definito come la corrente che, passando in due fili paralleli posti alla distanza di un metro, li fa attrarre con una forza di 2×10^{-7} newton per metro.

4 La legge di Biot-Savart

La **legge di Biot-Savart** afferma che il campo magnetico generato da una corrente costante è pari a:

$$\vec{B} = \frac{\mu_0}{4\pi} I \int \frac{d\vec{s} \times \vec{r}}{r^3} \quad (4.1)$$

dove:

- $\vec{B}$ = campo magnetico;
- μ_0 = permeabilità magnetica;

- I = corrente;
- $d\vec{s}$ = incremento di posizione;
- $\vec{r}$ = vettore posizione del generico punto P rispetto al tratto ds ;
- r = distanza fra P e ds .

Non fatevi impressionare dalla (4.1).

Siccome un integrale è una sommatoria infinita su elementi infinitesimi, possiamo scrivere la (4.1) in forma differenziale, cioè considerando il singolo elemento infinitesimo:

$$d\vec{B} = \frac{\mu_0}{4\pi} I \frac{d\vec{s} \times \vec{r}}{r^3} \quad (4.2)$$

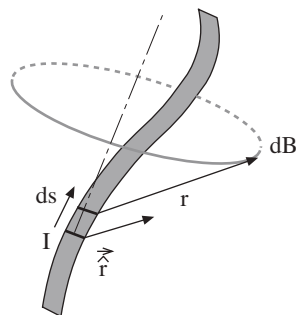


Fig. 7

$d\vec{B}$, essendo definito dal prodotto esterno fra $\vec{r}$ e $d\vec{s}$, è perpendicolare a entrambi questi vettori e ha direzione e verso definiti dalla regola della mano destra.

La (4.2) afferma semplicemente che un elemento infinitesimo del filo (ds) in cui scorre la corrente I genera un campo magnetico infinitesimo (dB).

Sommando gli infinitesimi campi magnetici generati dagli infinitesimi tratti di corrente su tutto il filo, si ottiene il campo magnetico B generato dalla corrente I .

Nella (4.2) alcune grandezze si trovano fuori dall'integrale, mentre altre si trovano nell'integrale.

Fuori dall'integrale troviamo i fattori che non dipendono dalla posizione sul filo; infatti:

- μ_0 e 4π sono due costanti;
- I si è supposta costante lungo tutto il filo.

All'interno dell'integrale troviamo fattori che dipendono dalla posizione sul filo; infatti:

- il vettore $d\vec{s}$ rappresenta un qualsiasi minuscolo segmento del filo;
- il vettore $\vec{r}$ va da $d\vec{s}$ al punto P in cui si vuole calcolare il campo magnetico;
- r indica la distanza fra $d\vec{s}$ e P .

Di seguito ricaveremo l'equazione per i campi magnetici partendo dalla legge di Biot-Savart.

a) Filo rettilineo percorso da corrente

Il campo magnetico generato da un filo percorso da corrente è espresso dalla (2.1). Si può dimostrare che questa equazione deriva dalla legge di Biot-Savart.

DIMOSTRAZIONE

Consideriamo la figura 8 e calcoliamo il valore del campo magnetico B nel punto P applicando la legge di Biot-Savart (4.1).

Il modulo di $\vec{B}$ è pari a:

$$B = \frac{\mu_0}{4\pi} I \int \frac{|d\vec{s} \times \vec{r}|}{r^3}$$

Il prodotto vettoriale $d\vec{s} \times \vec{r}$ genera un vettore perpendicolare ai vettori $d\vec{s}$ e $\vec{r}$, il cui modulo è pari a:

$$|d\vec{s} \times \vec{r}| = r \cdot ds \cdot \sin\vartheta$$

avendo indicato con ϑ l'angolo formato dai vettori $\vec{r}$ e $d\vec{s}$.

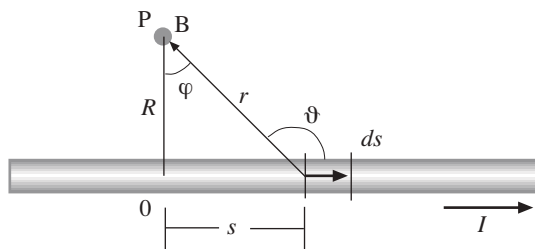


Fig. 8

La formula di Biot-Savart diventa:

$$B = \frac{\mu_0}{4\pi} I \int \frac{|d\vec{s} \times \vec{r}|}{r^3} = \frac{\mu_0}{4\pi} I \int \frac{r \sin\vartheta ds}{r^3} = \frac{\mu_0}{4\pi} I \int \frac{\sin\vartheta ds}{r^2} \quad (4.3)$$

Dalla figura possiamo notare che:

$$\vartheta = \varphi + 90$$

e quindi:

$$-\varphi = 90 - \vartheta$$

Dalla trigonometria sappiamo che:

$$\sin\vartheta = \cos(90 - \vartheta) = \cos(-\varphi) = \cos\varphi$$

Di conseguenza la (4.3) diventa:

$$B = \frac{\mu_0}{4\pi} I \int \frac{\cos\varphi ds}{r^2}$$

Siccome l'angolo φ varia al variare del tratto infinitesimo ds considerato, possiamo integrare su φ invece che su ds .

Per fare ciò dobbiamo prima esprimere tutte le variabili che compaiono nell'integrale, compreso ds , in termini di $d\varphi$.

Dalla figura 8 si ha:

$$s = R \tan\varphi$$

da cui, derivando, si ottiene:

$$ds = R \sec^2\varphi d\varphi$$

Sempre dalla figura si ha che:

$$\sec\varphi = \frac{r}{R}$$

per cui:

$$ds = R \frac{r^2}{R^2} d\varphi = \frac{r^2}{R} d\varphi$$

Quindi la formula di Biot-Savart diventa:

$$B = \frac{\mu_0}{4\pi} I \int \frac{\cos\varphi ds}{r^2} = \frac{\mu_0}{4\pi} I \int \frac{\cos\varphi}{r^2} \frac{r^2}{R} d\varphi = \frac{\mu_0}{4\pi R} I \int \cos\varphi d\varphi$$

I limiti di integrazione su ds sono $\pm\infty$:

$$\text{— per } ds = -\infty \Rightarrow \varphi = -\frac{\pi}{2}$$

$$\text{— per } ds = +\infty \Rightarrow \varphi = \frac{\pi}{2}$$

I limiti di integrazione diventano, dunque, $\pm\frac{\pi}{2}$, da cui segue che:

$$\begin{aligned} B &= \frac{\mu_0}{4\pi R} I \int_{-\pi/2}^{\pi/2} \cos\varphi d\varphi = \frac{\mu_0}{4\pi R} I [\sin\varphi]_{-\pi/2}^{\pi/2} = \frac{\mu_0}{4\pi R} I \left[\sin\left(\frac{\pi}{2}\right) - \sin\left(-\frac{\pi}{2}\right) \right] = \\ &= \frac{\mu_0}{4\pi R} I [1 - (-1)] = \frac{2\mu_0}{4\pi R} I = \frac{\mu_0}{2\pi R} I \end{aligned}$$

come volevasi dimostrare.

b) Campo magnetico nel punto centrale di un filo a forma di arco

Utilizziamo la legge di Biot-Savart per dimostrare che il campo magnetico nel punto centrale di un filo a forma di arco di lunghezza s e raggio R , percorso dalla corrente I , è pari a:

$$B = \frac{\mu_0 I}{4\pi R^2} s$$

Esprimendo il risultato in funzione dell'angolo φ sotteso sarà:

$$B = \frac{\mu_0 I \varphi}{4\pi R}$$

DIMOSTRAZIONE

Partiamo dalla formula di Biot-Savart (4.3).

È opportuno fare alcune considerazioni:

1. $\vec{r}$ è sempre perpendicolare a $d\vec{s}$ e dunque $\sin\theta = 1$ su tutto l'arco;
2. r è uguale al raggio R dell'arco.

La formula diventa quindi:

$$B = \frac{\mu_0}{4\pi} I \int \frac{1 \cdot ds}{R^2} = \frac{\mu_0}{4\pi R^2} I \int_0^s ds = \frac{\mu_0 I}{4\pi R^2} s$$

Ricordando che un arco sotteso da un angolo φ di raggio R ha lunghezza $R\varphi$, si ha:

$$s = R\varphi$$

e la formula di Biot-Savart diventa:

$$B = \frac{\mu_0 I \varphi}{4\pi R} \quad (4.4)$$

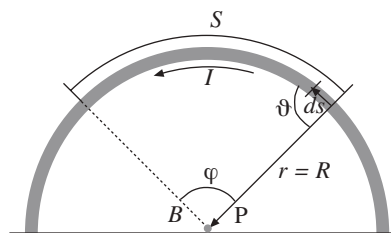


Fig. 9

c) Campo magnetico nel punto centrale di un cerchio percorso da corrente

Utilizziamo la legge di Biot-Savart per dimostrare che il campo magnetico nel punto centrale di un filo a forma di circonferenza di raggio R , percorso dalla corrente I , è pari a:

$$B = \frac{\mu_0 I}{2R}$$

DIMOSTRAZIONE

Riprendiamo la formula di Biot-Savart (4.4); essendo in questo caso $\varphi = 2\pi$, sostituendo otteniamo:

$$B = \frac{\mu_0 I}{2R}$$

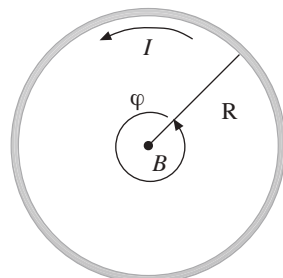


Fig. 10

5 La legge di Ampere

La **legge di Ampere** afferma che il campo magnetico nell'intorno di una corrente elettrica è proporzionale alla corrente elettrica stessa, proprio come il campo elettrico nell'intorno delle cariche elettriche è proporzionale alle cariche stesse.

Immaginiamo di tracciare un percorso chiuso che circonda la corrente elettrica, di suddividerlo in percorsi infinitesimali e di tracciare la componente del campo $B_{\parallel}$ parallela a essi (Fig. 11).

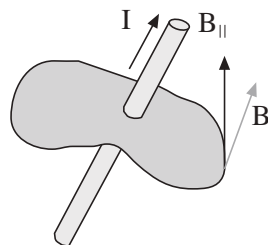


Fig. 11

Il **teorema di Ampere** afferma che:

Il campo magnetico parallelo a un segmento infinitesimo di un qualsiasi percorso chiuso, detto anche circuito amperiano, è proporzionale alla corrente elettrica che attraversa la superficie definita dal percorso chiuso:

$$\sum B_{\parallel} s = \mu_0 I \quad \sum B s \cos \vartheta = \mu_0 I \quad (5.1)$$

dove:

- B è il valore del campo sull'elemento infinitesimale del percorso chiuso;
- I è la corrente che passa nel circuito;
- $B_{\parallel}$ è la componente del campo parallela all'elemento infinitesimale del percorso chiuso;
- s è la lunghezza dell'elemento infinitesimale del percorso chiuso;
- ϑ è l'angolo fra il campo B e l'elemento infinitesimale del circuito.

Nella formula (5.1) è stato inserito il segno di sommatoria in quanto sia il campo magnetico che l'angolo fra il vettore campo magnetico e la direzione del circuito possono variare lungo il circuito.

Per sfruttare il teorema di Ampere è opportuno scegliere un circuito amperiano che faciliti al massimo i calcoli.

Per esempio, in figura 12 il circuito amperiano è stato tracciato concentrico al filo elettrico e perpendicolare a esso.

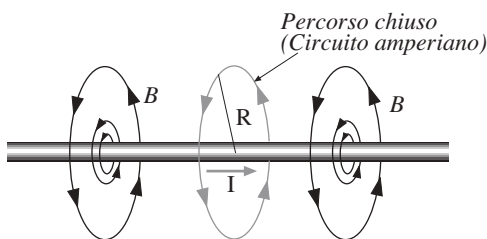


Fig. 12

In questo modo, il campo magnetico è costante su tutti i punti di questo circuito ed è parallelo, in ogni punto, al circuito stesso.

Essendo ϑ l'angolo formato fra il campo B e l'elemento infinitesimale del circuito, si ha che in questo caso è nullo su tutto il circuito amperiano e, quindi, il primo membro della (5.1) diventa:

$$\sum B_{\parallel} s = \sum B s \cos \vartheta = \sum B s \cos 0 = \sum B s$$

Ma, essendo B costante su tutto il circuito amperiano, abbiamo:

$$\sum B s = B \sum s = B \cdot 2\pi R = 2\pi R B$$

6 Le applicazioni della legge di Ampere

a) Campo magnetico in un conduttore cilindrico percorso da corrente

Il campo magnetico all'interno di un conduttore cilindrico percorso da una corrente uniformemente distribuita (Fig. 13) è pari a:

$$B = \frac{\mu_0 I r}{2\pi R^2}$$

dove:

- μ_0 = costante di permeabilità magnetica;
- B = campo magnetico;
- I = corrente;
- r = distanza dal centro del conduttore;
- R = raggio del conduttore.

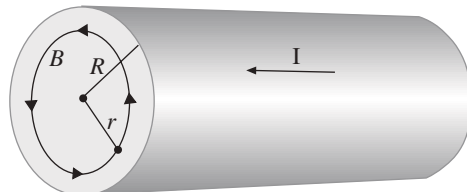
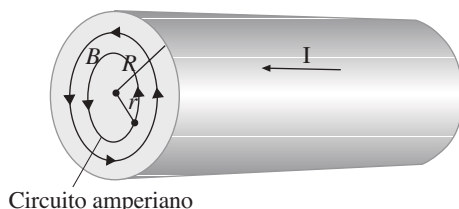


Fig. 13

DIMOSTRAZIONE

Per prima cosa si deve definire il circuito amperiano su cui andremo a integrare.

In questo caso, il circuito amperiano è identico a quello già utilizzato per ricavare il campo magnetico all'esterno di un filo rettilineo percorso da corrente, in quanto anche in questo caso sia il campo magnetico che l'angolo compreso tra il vettore campo magnetico e il cammino di integrazione è costante (Fig. 14).



Circuito amperiano

Fig. 14

Partiamo dalla legge di Ampere:

$$\sum \vec{B} \cdot d\vec{s} = \sum B \cos\vartheta ds = \mu_0 I_{\text{incl}} \quad (6.1)$$

dove I_{incl} è la corrente «inclusa» nel circuito amperiano.

Il vettore campo magnetico è sempre parallelo al circuito amperiano e, quindi, essendo sempre $\cos\vartheta = 1$ la (6.1) diventa:

$$\sum \vec{B} \cdot d\vec{s} = B \cos\vartheta s = Bs = \mu_0 I_{\text{incl}}$$

Da cui segue che:

$$B = \frac{\mu_0 I_{\text{incl}}}{s} \quad (6.2)$$

Essendo la densità di corrente costante all'interno del cilindro, il rapporto fra la corrente totale che scorre nel cilindro e quella che scorre all'interno del circuito amperiano è uguale al rapporto fra le aree delle rispettive sezioni, cioè:

$$\frac{I_{\text{incl}}}{I} = \frac{\pi r^2}{\pi R^2}$$

da cui segue che:

$$I_{\text{incl}} = \frac{r^2}{R^2} I$$

La lunghezza s del circuito amperiano è:

$$s = 2\pi r$$

Sostituendo i valori trovati nella (6.2) si ottiene:

$$B = \frac{\mu_0 I_{\text{incl}}}{s} = \frac{\mu_0}{2\pi r} \frac{r^2}{R^2} I = \frac{\mu_0 I r}{2\pi R^2}$$

come volevasi dimostrare.

Da questa equazione possiamo ricavare che nel centro del cilindro, che corrisponde a $r = 0$, non c'è campo magnetico.

b) Campo magnetico su un piano infinito

Il campo magnetico generato da una superficie piana percorsa da corrente ha modulo pari a:

$$B = \frac{\mu_0 K}{2}$$

dove:

- B = campo magnetico;
- μ_0 = costante di permeabilità magnetica;
- K = densità lineare di corrente e cioè la quantità di corrente che scorre per metro di lunghezza. Ha direzione parallela al piano e perpendicolare alla corrente. Inoltre, ha verso opposto al di sopra e al di sotto della superficie.

Nella figura 15 il campo magnetico è diretto verso sinistra al di sopra della superficie e verso destra al di sotto.

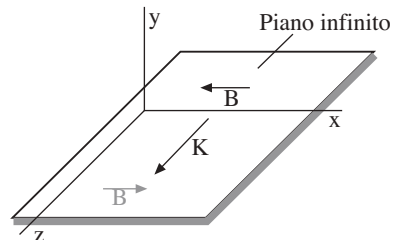


Fig. 15

DIMOSTRAZIONE

Iniziamo a dimostrare che il campo magnetico è diretto parallelamente al piano e perpendicolarmente alla corrente e che ha versi opposti al di sopra e al di sotto del piano.

Orientiamo la superficie percorsa da corrente parallelamente al piano xy .

Un piano infinito percorso da corrente può essere immaginato come un piano su cui scorrono infiniti fili paralleli disposti sullo stesso piano e percorsi dalla medesima corrente.

In figura 16 sono stati tracciati due ipotetici fili paralleli, diretti secondo l'asse z .

Ogni filo genera un campo magnetico rappresentato dai cerchi concentrici perpendicolari al filo.

Il campo magnetico risultante sarà dato dalla sovrapposizione di questi due campi magnetici.

Dalla figura possiamo notare che, nella sovrapposizione, le componenti verticali dei due campi si annullano, mentre le componenti orizzontali si sommano.

Al di sopra del piano, il campo magnetico ha componente orizzontale negativa; al di sotto del piano, ha, invece, componente orizzontale positiva.

Il campo magnetico risultante dalla sovrapposizione di queste due correnti è parallelo all'asse x ed è diretto verso sinistra al di sopra del piano e verso destra al di sotto del piano.

Dimostriamo ora che il valore del campo magnetico è pari a:

$$B = \frac{\mu_0 K}{2}$$

• Circuito amperiano

In questo caso, il circuito amperiano è un rettangolo parallelo al piano xy in modo che i tratti orizzontali del circuito sono paralleli al campo magnetico, mentre quelli verticali sono ad esso perpendicolari.

Partiamo dalla legge di Ampere:

$$\sum \vec{B} \cdot d\vec{s} = \sum B \cos \theta ds = \mu_0 I_{\text{incl}} = \mu_0 I$$

Suddividiamo il circuito amperiano nei quattro segmenti mostrati in figura 17.

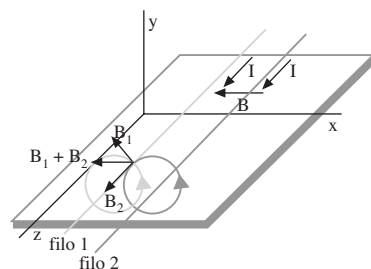


Fig. 16

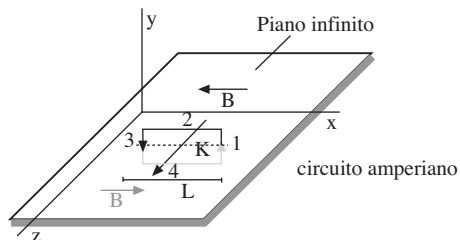


Fig. 17

Essendo il campo magnetico costante:

$$\sum \vec{B} \cdot d\vec{s} = \sum_{i=1}^4 B \cos \vartheta_i ds_i = B \cos \vartheta_1 ds_1 + B \cos \vartheta_2 ds_2 + B \cos \vartheta_3 ds_3 + B \cos \vartheta_4 ds_4 \quad (6.3)$$

I tratti verticali 1 e 3 sono perpendicolari al campo magnetico e quindi il prodotto scalare su questi tratti è nullo.

I tratti 2 e 4 sono invece paralleli al campo magnetico.

Quindi la (6.3) diventa:

$$\sum \vec{B} \cdot d\vec{s} = B ds_2 + B ds_4 = B(ds_2 + ds_4) = B(L + L) = 2BL$$

e, sostituendo nella legge di Ampere, si ottiene:

$$2BL = \mu_0 I$$

da cui si deduce che:

$$B = \frac{\mu_0 I}{2L} \quad (6.4)$$

Introducendo la densità lineare di corrente:

$$K = \frac{I}{L}$$

la (6.4) diventa:

$$B = \frac{\mu_0 K}{2}$$

come volevasi dimostrare.

c) Campo magnetico in un solenoide

Un **solenoide** è formato da un insieme di spire poste una accanto all'altra e aventi lo stesso raggio (Fig. 18).

Il campo magnetico all'interno di un solenoide è pari a:

$$B = \mu_0 I n$$

dove n rappresenta il numero di spire per unità di lunghezza.

All'esterno il campo magnetico è nullo:

$$B = 0$$

dove si suppone che il solenoide abbia lunghezza infinita.

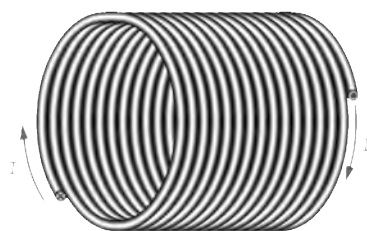


Fig. 18

DIMOSTRAZIONE

Nella figura 19 sono stati tracciati sia i campi magnetici delle singole spire del solenoide che il campo magnetico risultante.

Prima di derivare l'equazione del campo magnetico, cerchiamo di capirne la natura per poter definire un cammino amperiano opportuno.

Osserviamo la figura 19, in cui le spire sono state staccate per permettere di capire meglio che cosa succede in un solenoide.

Ogni spira del solenoide è percorsa dalla corrente I ; i punti da cui la corrente esce sono indicati con x e quelli in cui la corrente entra sono indicati con un pallino.

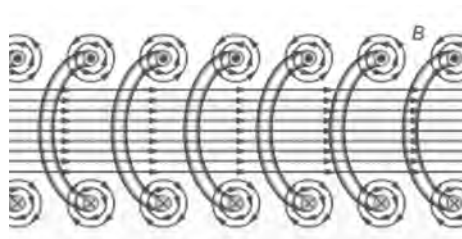


Fig. 19

La direzione del campo magnetico attorno a ciascuna spira è definita dalla seconda regola della mano destra. Sommando i campi magnetici generati da ciascuna spira, possiamo notare che:

- all'interno del solenoide, le componenti orizzontali del campo magnetico si sommano e quelle verticali si annullano;
- all'esterno del solenoide, le componenti del campo magnetico si annullano e quindi il campo magnetico è nullo.

• Circuito amperiano

Ricaviamo, ora, il valore del campo magnetico all'interno del solenoide tracciando un circuito amperiano nella figura 20.

Il circuito amperiano racchiude una porzione della parte superiore del solenoide ed è stato suddiviso in 6 porzioni.

Per la legge di Ampere:

$$\sum \vec{B} \cdot d\vec{s} = \sum B \cos \vartheta ds = \mu_0 I_{\text{incl}}$$

Avendo suddiviso il circuito amperiano in sei segmenti:

$$\begin{aligned} \sum \vec{B} \cdot d\vec{s} = \sum_{i=1}^6 B \cos \vartheta_i ds_i &= B \cos \vartheta_1 ds_1 + B \cos \vartheta_2 ds_2 + B \cos \vartheta_3 ds_3 + \\ &+ B \cos \vartheta_4 ds_4 + B \cos \vartheta_5 ds_5 + B \cos \vartheta_6 ds_6 \end{aligned} \quad (6.5)$$

I tratti 1 e 3 sono perpendicolari al campo magnetico, quindi:

$$B \cos \vartheta_1 ds_1 = 0 \quad B \cos \vartheta_3 ds_3 = 0$$

Il tratto 2 è parallelo al campo magnetico e dunque:

$$B \cos \vartheta_2 ds_2 = B ds_2 = BL$$

All'esterno del solenoide il campo magnetico è nullo e pertanto:

$$B \cos \vartheta_4 ds_4 = 0 \quad B \cos \vartheta_5 ds_5 = 0 \quad B \cos \vartheta_6 ds_6 = 0$$

Di conseguenza la (6.5) diventa:

$$\sum \vec{B} \cdot d\vec{s} = BL$$

e, sostituendo nella legge di Ampere, si ottiene:

$$BL = \mu_0 I_{\text{incl}}$$

da cui segue che:

$$B = \frac{\mu_0 I_{\text{incl}}}{L} \quad (6.6)$$

Possiamo sostituire la corrente I_{incl} racchiusa dal circuito con il prodotto tra la corrente I che scorre in ogni spira e il numero N di spire racchiuse all'interno del circuito, per cui la (6.6) diventa:

$$B = \frac{\mu_0 IN}{L}$$

Ma il rapporto N/L rappresenta la densità di spire del solenoide e cioè il numero n di spire per unità di lunghezza; il campo magnetico è quindi espresso dalla formula:

$$B = \mu_0 In$$

come volevasi dimostrare.

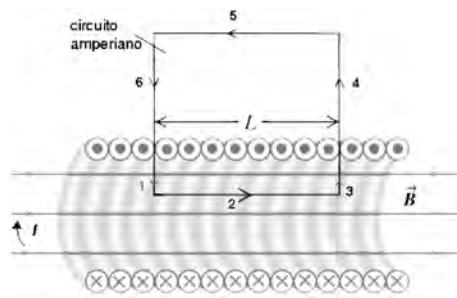


Fig. 20

d) Campo magnetico in un toroide

Un **toroide** è un solenoide curvato su se stesso in modo da assumere l'aspetto di una ciambella (Fig. 21).

All'**interno** del toroide, il campo magnetico a una distanza r dal centro è pari a:

$$B = \frac{\mu_0 IN}{2\pi r}$$

dove N è il numero di spire del toroide.

All'**esterno** del toroide il campo magnetico è nullo:

$$B = 0$$

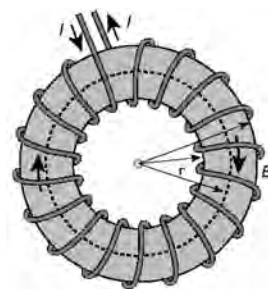


Fig. 21

DIMOSTRAZIONE

Il circuito amperiano che utilizzeremo per i nostri calcoli è un cerchio di raggio r , concentrico al toroide e interno a esso.

Per la legge di Ampere:

$$\sum \vec{B} \cdot d\vec{s} = \sum B \cos \theta ds = \mu_0 I_{\text{incl}} \quad (6.7)$$

Il campo magnetico è sempre parallelo al circuito amperiano e quindi la (6.7) diventa:

$$Bs = \mu_0 I_{\text{incl}} \quad (6.8)$$

Il circuito amperiano è una circonferenza di raggio r e dunque la sua lunghezza è pari a:

$$s = 2\pi r$$

La corrente racchiusa all'interno del circuito amperiano è pari alla corrente che circola in ogni spira per il numero di spire del toroide:

$$I_{\text{incl}} = IN$$

Sostituendo nella (6.8), si ottiene:

$$B2\pi r = \mu_0 IN$$

da cui segue che:

$$B = \frac{\mu_0 IN}{2\pi r}$$

come volevasi dimostrare.

7 Proprietà magnetiche della materia: il campo magnetico $\vec{H}$

All'inizio del capitolo abbiamo introdotto due diversi vettori: $\vec{B}$ e $\vec{H}$.

Abbiamo accennato al fatto che questi due vettori sono legati dalla relazione:

$$\vec{B} = \mu \vec{H}$$

Vediamo ora di esplicitare questa relazione.

Immaginiamo di introdurre all'interno di un solenoide un mezzo materiale omogeneo (Fig. 22).

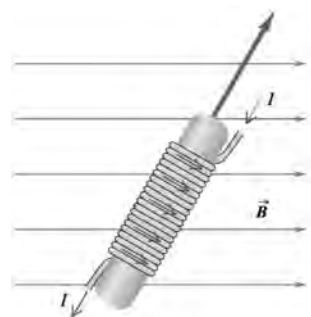


Fig. 22

Sperimentalmente si osserva che quando il solenoide è percorso da una corrente I il materiale posto al suo interno si comporta come un magnete permanente, le cui caratteristiche possono essere descritte dal vettore magnetizzazione definito nel capitolo precedente.

Per il principio di equivalenza di Ampere, un magnete avente vettore di magnetizzazione m e lunghezza L è equivalente a un solenoide percorso da una corrente di magnetizzazione I' :

$$m = \frac{N}{L} I'$$

Il campo magnetico presente all'interno del materiale omogeneo è dato dalla somma del campo magnetico generato dal solenoide esterno ($B = \frac{\mu_0 IN}{L}$) e del campo magnetico «fittizio» generato dal solenoide equivalente al magnete ($B = \frac{\mu_0 I' N}{L}$); esso sarà quindi pari a:

$$B = \mu_0 \left(\frac{N}{L} I + \frac{N}{L} I' \right) = \mu_0 (nI + m) \quad (7.1)$$

dove n è la densità di spire e cioè il numero di spire per unità di lunghezza.

Nella (7.1) i due termini della somma non sono indipendenti, in quanto il primo termine, che rappresenta il campo magnetico generato dalla corrente nella spira, è responsabile della magnetizzazione dovuta al secondo termine.

Per sottolineare ciò, il termine nI è detto **forza magnetomotrice** ed è indicato con la lettera H .

La (7.1) può quindi essere scritta come:

$$B = \mu_0 (H + m)$$

che in forma vettoriale diventa:

$$\vec{B} = \mu_0 (\vec{H} + \vec{m})$$

Il campo $\vec{B}$ è quindi composto dai campi $\vec{H}$ e $\vec{m}$.

Il campo $\vec{m}$ rappresenta il contributo al campo magnetico dovuto alle correnti vere.

Il campo $\vec{H}$ rappresenta il contributo al campo magnetico dovuto alle modificazioni indotte nel mezzo dalle correnti vere.

*$\vec{H}$ è il **campo magnetico** vero e proprio, mentre $\vec{B}$ è il **campo di induzione magnetica**.*

Nel vuoto $m = 0$ e quindi $\vec{B} = \mu_0 \vec{H}$, mentre in tutti gli altri casi, ponendo $\vec{m} = \chi \vec{H}$, si ottiene:

$$\vec{B} = \mu_0 (\vec{H} + \chi \vec{H}) = \mu_0 (1 + \chi) \vec{H} = \mu_0 \mu_r \vec{H}$$

La quantità $\mu_r = 1 + \chi$ è detta **permeabilità relativa al vuoto**; nel vuoto, infatti, $\mu_r = 1$.

a) Permeabilità magnetica

Dalla (7.1) possiamo ricavare che la **permeabilità magnetica** è una grandezza fisica che esprime l'attitudine di una sostanza a lasciarsi magnetizzare.

Le sostanze con $\mu_r < \mu_0$ sono dette **diamagnetiche**; le sostanze con $\mu_r > \mu_0$ sono chiamate, invece, **paramagnetiche**.

Esiste un gruppo di sostanze, dette **ferromagnetiche**, in cui μ_r è positivo.

8 La teoria microscopica del magnetismo

Per dare una spiegazione del tutto approssimativa sul perché le sostanze possono essere paramagnetiche, diamagnetiche o ferromagnetiche dobbiamo riprendere alcuni concetti sulla struttura atomica della materia.

Negli atomi le orbite elettroniche possono essere assimilate a delle piccole spire aventi un proprio momento magnetico.

Gli elettroni, oltre a ruotare attorno al nucleo, ruotano anche su sé stessi con una rotazione detta **spin** e, quindi, ogni materiale presenta un proprio momento magnetico dovuto alla somma dei momenti magnetici dei singoli elettroni.

A volte gli spin degli elettroni si annullano a vicenda, mentre in altri casi gli spin non si annullano e si creano all'interno del materiale dei domini magnetici, ovvero delle regioni in cui è presente un campo magnetico.

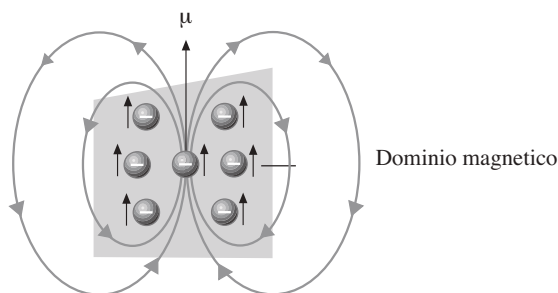


Fig. 23

Normalmente questi domini puntano in una qualsiasi direzione e pertanto il campo magnetico risultante è nullo.

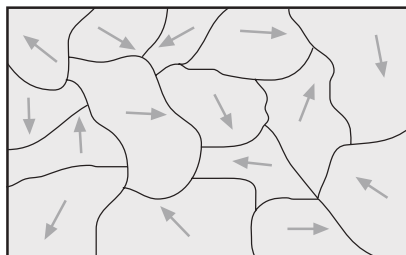


Fig. 24

Quando il materiale viene posto in un campo magnetico esterno, esercita un momento sui dipoli magnetici della materia e si magnetizza; il campo magnetico esterno si sovrappone a quello del materiale, facendo allineare i dipoli magnetici.

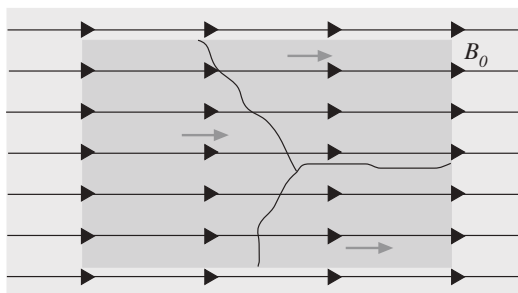


Fig. 25

Nei materiali ferromagnetici questi dipoli rimangono allineati anche dopo che la sorgente magnetica esterna è stata rimossa.

Un campo magnetico esterno è in grado anche di magnetizzare materiali che non sono ferromagnetici, detti paramagnetici o diamagnetici, ma questi materiali si smagnetizzano immediatamente una volta che la sorgente magnetica esterna è stata rimossa.

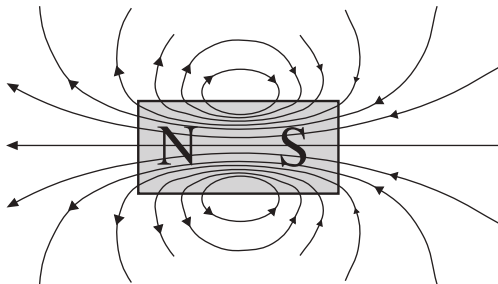


Fig. 26

a) Ciclo d'isteresi e punto di Curie

Se immergiamo un pezzo di ferro in un campo magnetico, esso si magnetizza; se poi rimuoviamo il campo magnetico, il pezzo di ferro continua ad essere magnetizzato. Questo effetto è chiamato **rimanenza**.

In figura 27 è mostrata una curva di magnetizzazione.

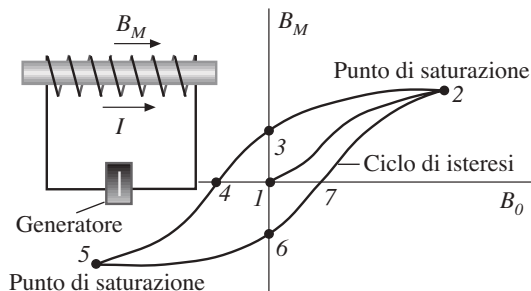


Fig. 27

Sull'asse orizzontale sono stati riportati i valori del campo magnetico esterno B_0 e sull'asse verticale sono stati riportati i valori del campo magnetico B_M in un materiale come il ferro.

A sinistra è anche mostrato il circuito che genera il campo magnetico.

La corrente nel solenoide genera un campo magnetico B_0 , il cui valore e la cui direzione dipendono dall'intensità e dal verso della corrente elettrica.

Questo campo magnetico esterno genererà anche un campo magnetico B_M all'interno del nucleo di ferro.

Il grafico mostrato in figura è il risultato di una serie di passaggi, durante i quali è stata fatta variare la corrente all'interno del circuito.

Vediamo questi passaggi in dettaglio.

1. L'intero processo ha inizio nel punto 1, in cui non si ha alcun campo magnetico sia all'esterno che all'interno del materiale.
2. Con un campo magnetico esterno si magnetizza il materiale, facendogli raggiungere il punto di saturazione 2.

3. Il campo magnetico esterno si riduce fino al valore 0 (punto 3). Per il fenomeno della *rimanenza* il materiale ferromagnetico continua ad essere magnetizzato e quindi ad avere un proprio campo magnetico sebbene sia minore che al punto 2.
4. Viene invertito il verso della corrente, con una conseguente inversione del campo magnetico esterno. In corrispondenza di un dato valore del campo magnetico esterno invertito, il campo magnetico nel materiale si annulla (punto 4). Da notare che per annullare il campo magnetico all'interno del materiale non basta annullare il campo magnetico esterno, ma occorre invertirlo.
5. Continuando ad aumentare il campo magnetico invertito, si raggiunge un altro punto di saturazione (punto 5) uguale ed opposto al punto 2.
6. Il campo magnetico esterno viene nuovamente portato a zero (punto 6), ma, per effetto della rimanenza, il materiale continua ad essere magnetizzato.
7. Nel punto 7, il campo magnetico esterno è abbastanza intenso da annullare il campo magnetico nel materiale ferromagnetico. Aumentando nuovamente il campo magnetico esterno, si raggiunge di nuovo il punto 2.

Il ciclo completo è detto **ciclo d'isteresi** e si chiama in questo modo dalla parola greca *hysteresis*, che significa «ritardo». La curva, infatti, nel secondo giro, raggiunge il punto 2 con un certo ritardo rispetto al primo giro e questo a causa del fenomeno della rimanenza.

La curva di isteresi varia a seconda del materiale considerato. Materiali ferromagnetici «dolci», come il ferro, hanno delle curve d'isteresi molto strette e sono più facilmente magnetizzabili e smagnetizzabili. Le curve d'isteresi dei materiali ferromagnetici «duri», invece, sono più allargate, in quanto essi mantengono il loro campo magnetico per valori più estesi del campo magnetico esterno.

La magnetizzabilità di un materiale dipende anche dalla temperatura esterna. Sebbene i dipoli magnetici tendano a restare allineati lungo le loro direzioni di magnetizzazione, l'agitazione termica dovuta alla temperatura in cui si trova il materiale può in qualche modo compromettere questo allineamento.

A temperatura ambiente non ci sono problemi, ma a temperature superiori i materiali possono smagnetizzarsi.

È interessante il fatto che il magnetismo non diminuisce gradatamente con la temperatura: l'intensità del campo magnetico all'interno dei materiali ferromagnetici rimane pressoché costante al variare della temperatura fino a quando, per un certo valore, diminuisce bruscamente.

Questo fenomeno può essere paragonato a una transizione di fase come quelle da solido a liquido: improvvisamente, infatti, il materiale passa da essere ferromagnetico a essere *non* ferromagnetico. La temperatura a cui avviene questa transizione è detta **punto di Curie**, in onore di Pietro Curie, che per primo studiò questo fenomeno.

Induzione magnetica ed elettromagnetismo

1 L'induzione elettromagnetica

Abbiamo visto nel capitolo precedente che i campi magnetici non sono entità indipendenti dai campi elettrici ma derivano dalla variazione di questi ultimi.

Nel 1831 Faraday scoprì che se si faceva variare il flusso del campo magnetico concatenato con una spira, in quest'ultima veniva indotta una forza elettromotrice (f.e.m.) e, quindi, una corrente; egli scoprì, inoltre, che questa forza elettromotrice era direttamente proporzionale alla velocità di variazione del campo magnetico concatenato con il circuito. In questo modo si dimostrò che neanche i campi elettrici erano indipendenti dai campi magnetici, in quanto la loro stessa esistenza era dovuta a una variazione di questi ultimi.

Si iniziò, dunque, a parlare non soltanto di elettricità e di magnetismo ma, soprattutto, di **elettromagnetismo**: una parola creata per sottolineare la fusione fra realtà elettrica e magnetica. Alla scoperta fatta da Faraday venne dato il nome di **induzione elettromagnetica**.

L'esistenza dell'induzione elettromagnetica si può facilmente dimostrare unendo gli estremi di una bobina a una lampadina e avvicinandovi un magnete (Fig. 1).

Quando il magnete è fermo rispetto alla bobina, la lampadina è spenta e, dunque, nella spira non circola corrente.

Se si avvicina o si allontana il magnete rispetto alla bobina, la lampadina si accende e, quindi, nella spira circola corrente che, non essendo stata indotta dal movimento del magnete verso la bobina, è detta **corrente elettrica indotta**.

Lo stesso fenomeno si poteva osservare spostando la spira rispetto al magnete.

Questo significa che in un circuito si genera una corrente indotta ogni volta che la sua posizione varia rispetto a un magnete, indipendentemente da come avviene questa variazione.

L'induzione elettromagnetica è regolata dalla **legge di Faraday-Neumann**, la quale afferma che se un campo magnetico varia nel tempo in un punto, in quel punto si crea un campo elettrico.

Per determinare questa legge è importante introdurre il concetto di flusso di un campo magnetico.

Abbiamo definito il flusso di un campo elettrico (Fig. 2.a) attraverso il prodotto scalare:

$$\Phi_E = \vec{E} \cdot \vec{A}$$

dove $\vec{A}$ è un vettore, detto **vettore superficie** o **vettore area**, avente:

- modulo uguale all'area della superficie;
- direzione perpendicolare alla superficie;
- verso arbitrario.

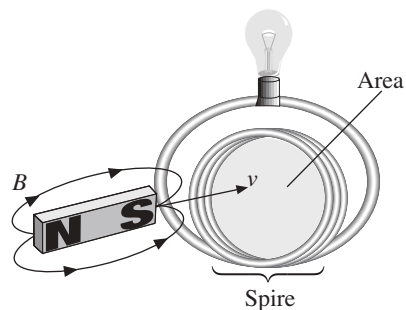
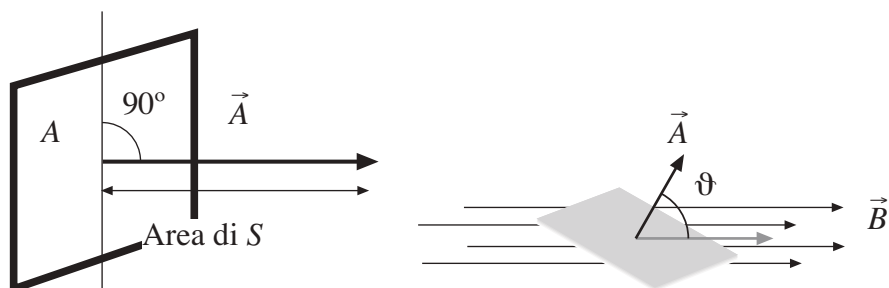


Fig. 1



Figg. 2.a e 2.b

Analogamente, il **flusso del campo magnetico** (Fig. 2.b) attraverso una superficie è dato dal prodotto scalare:

$$\Phi_B = \vec{B} \cdot \vec{A} = BA \cos \vartheta \quad (1.1)$$

dove ϑ è l'angolo formato tra il vettore campo magnetico e il vettore superficie.

Se al posto di una superficie avessimo considerato una spira coincidente con il bordo della superficie, il flusso magnetico concatenato sarebbe stato lo stesso.

In generale, se al posto di una spira avessimo avuto N spire, il flusso magnetico concatenato sarebbe stato pari a:

$$\Phi_B = N \vec{B} \cdot \vec{A} = NBA \cos \vartheta$$

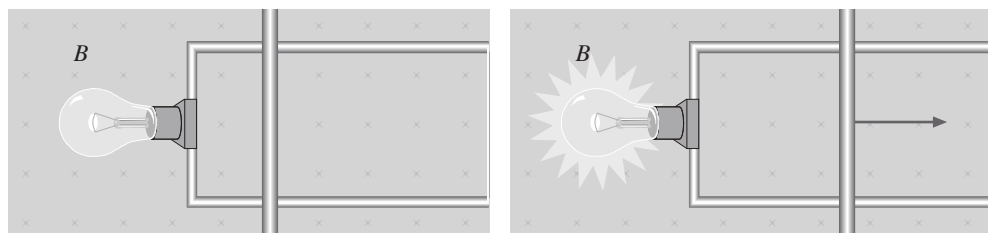
L'unità di misura del flusso del campo magnetico è il weber (Wb):

$$1T = 1Wb/m^2$$

Praticamente, il tesla (T) è la densità di flusso magnetico corrispondente a un flusso magnetico di 1 weber per metro quadrato.

2 La legge di Faraday-Neumann

Consideriamo la figura 3.a in cui un conduttore collegato con una lampadina è stato piegato a U e su di esso è libero di scorrere un conduttore rettilineo che chiude il circuito. Supponiamo, inoltre, che il circuito sia immerso in un campo magnetico B perpendicolare al piano individuato dal circuito stesso.



Figg. 3.a e 3.b

Facendo scorrere il conduttore rettilineo verso destra con velocità v , nel circuito si induce una corrente che fa accendere la lampadina (Fig. 3.b).

Per trovare il valore di questa corrente consideriamo il circuito in figura 4.

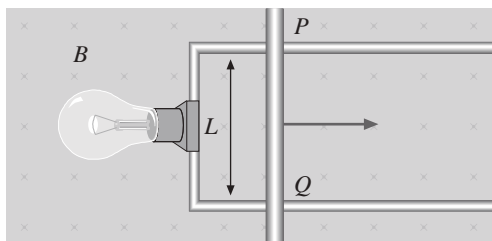


Fig. 4

Quando la sbarra PQ si sposta nel campo magnetico con velocità v , anche gli elettroni (e) al suo interno vengono spostati con la stessa velocità e, quindi, su ognuno di essi agisce la forza di Lorentz:

$$\vec{F}_L = -e\vec{v} \times \vec{B}$$

la cui direzione è determinata dalla regola della mano destra.

Essendo $\vec{v}$ perpendicolare a $\vec{B}$, la forza di Lorentz ha modulo:

$$F_L = -evB \quad (2.1)$$

ed è diretta lungo la direzione PQ verso Q .

Possiamo riscrivere la (2.1) come se essa fosse dovuta a un campo elettrico E_L :

$$F_L = -eE_L$$

da cui segue che il campo elettrico che agisce su un elettrone in movimento in un campo elettrico è pari a:

$$E_L = vB \quad (2.2)$$

Dalla (2.2) segue che E_L si annulla se $B = 0$ oppure se $v = 0$ e, dunque, esiste soltanto quando l'elettrone si muove in un campo magnetico. Se l'elettrone si dovesse fermare o se il campo magnetico venisse annullato, E_L si annullerebbe automaticamente. Esso è pertanto «indotto» dal movimento dell'elettrone nel campo, che per questo viene detto **campo elettrico indotto**.

Dalla (2.2), se la sbarra PQ è lunga L , tra P e Q si genera un **potenziale indotto**, detto anche **forza elettromotrice indotta**, pari a:

$$V = E_L L = BvL \quad (2.3)$$

OSSERVAZIONE

In questo caso abbiamo considerato una sbarra che si muove in direzione perpendicolare al campo magnetico. Se si fosse mossa in una direzione qualsiasi rispetto al campo magnetico, la forza esercitata da quest'ultimo sull'elettrone sarebbe stata pari a:

$$F_L = qvB \sin \vartheta$$

dove ϑ è l'angolo formato tra il vettore velocità della sbarra e il vettore campo magnetico.

La differenza di potenziale fra le sue estremità sarebbe stata pari a:

$$\Delta V = LvB \sin \vartheta$$

Vediamo ora come questa forza elettromotrice indotta possa, in realtà, essere collegata a una variazione di flusso all'interno del circuito.

Consideriamo la figura 5 in cui è indicata la superficie S (di area A) attraversata dal campo magnetico.

Il flusso concatenato al circuito è pari a:

$$\Phi = \vec{B} \cdot \vec{A} = BA$$

Quando la sbarretta PQ si sposta verso destra, a una variazione Δx della sbarretta PQ corrisponde una variazione dell'area A pari a:

$$\Delta A = L \Delta x$$

Siccome la sbarretta si muove verso destra con velocità v , si ha che:

$$\Delta x = v \Delta t$$

e quindi:

$$\Delta A = Lv \Delta t$$

A questa variazione di superficie corrisponde una variazione del flusso ad essa concatenato, pari a:

$$\Delta \Phi = B \Delta A = BLv \Delta t$$

Sostituendo in questa equazione la (2.3) si ha che:

$$\Delta \Phi = V \Delta t$$

da cui segue che:

$$V = \frac{\Delta \Phi}{\Delta t} \quad (2.4)$$

che esprime la legge di **Faraday-Neumann** e può essere così enunciata:

La forza elettromotrice indotta in un circuito è proporzionale alla velocità di variazione del flusso ad esso concatenato.

La (2.4) non tiene assolutamente conto di come debba avvenire la variazione del flusso magnetico concatenato ad un circuito.

Se al posto di una spira il circuito fosse stato composto di N spire la (2.4) sarebbe diventata:

$$V = N \frac{\Delta \Phi}{\Delta t}$$

3 La legge di Lenz

Per comprendere la direzione della f.e.m. creata in un circuito da una variazione di flusso concatenato con il circuito dobbiamo utilizzare la **terza regola della mano destra**.

Supponiamo che una spira venga percorsa da una corrente I , generando così un campo magnetico B (Fig. 6).

Se con le dita della mano «avvolgiamo» la corrente, il pollice indica la direzione del campo magnetico generato.

Per convenzione, se la direzione in cui scorre la corrente generata dalla f.e.m. indotta è tale da:

- aumentare il campo magnetico concatenato con il circuito, allora la f.e.m. è **positiva**;
- diminuire il campo magnetico concatenato con il circuito, allora la f.e.m. è **negativa**.

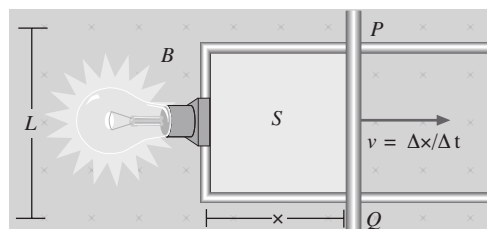


Fig. 5

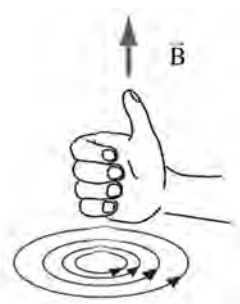


Fig. 6 - Spira conduttrice

Supponiamo di aumentare il campo magnetico concatenato con il circuito in figura 7.

La forza di Lorentz che agisce sulle cariche elettriche «spinge» gli elettroni verso l'estremità Q e, quindi, nel circuito si genera una corrente I che scorre in direzione opposta (da Q verso P).

Se applichiamo la terza regola della mano destra alla corrente, vediamo che essa genera un campo magnetico B indotto, avente direzione opposta a B e, dunque, tendente a opporsi all'aumento del flusso di B .

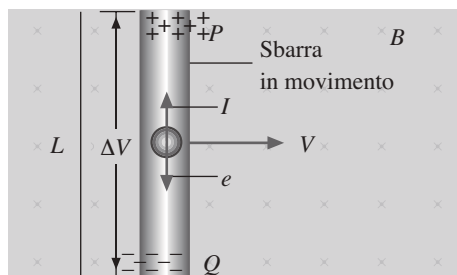


Fig. 7

Questo è un caso particolare di una legge (che ha una validità del tutto generale), nota come **legge di Lenz**, in onore del fisico tedesco Henrich Lenz, che per primo la formulò, la quale afferma che:

La direzione della forza elettromotrice indotta in un circuito è sempre tale da opporsi alla variazione di flusso che l'ha generata.

Possiamo riassumere la legge di Faraday e la legge di Lenz nella seguente formula:

$$V = - \frac{\Delta \Phi}{\Delta t} \quad (3.1)$$

dove il termine $\frac{\Delta \Phi}{\Delta t}$ deriva dalla legge di Faraday e il segno $-$ tiene conto della legge di Lenz.

La (3.1) vale se il flusso varia con velocità costante; se, invece, il flusso varia in modo casuale dobbiamo riscrivere il rapporto $\frac{\Delta \Phi}{\Delta t}$ nel limite di $\Delta t \rightarrow 0$ e, quindi, la forza elettromotrice indotta è pari all'opposto della derivata del flusso del campo magnetico rispetto al tempo:

$$V = - \frac{d\Phi}{dt}$$

Il circuito in figura 8 ha una resistenza R .

Dimostriamo che la potenza dissipata dal circuito, per effetto Joule dalla corrente indotta, è uguale alla potenza delle forze che spostano il conduttore.

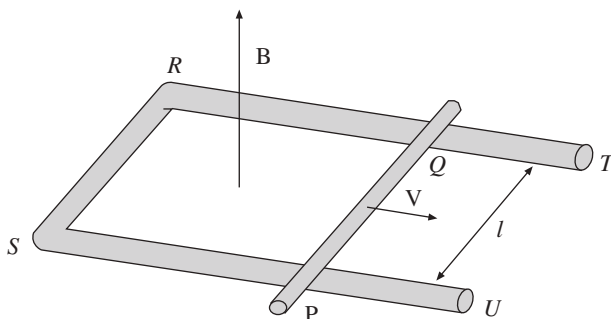


Fig. 8

DIMOSTRAZIONE

Dalla (2.3) sappiamo che la forza elettromotrice indotta nel circuito è pari a:

$$V = BvL$$

che genera nel circuito una corrente:

$$I = \frac{V}{R} = \frac{BvL}{R} \quad (3.2)$$

La potenza dissipata da questa corrente per effetto Joule è quindi pari a:

$$P = VI = \frac{(BvL)^2}{R}$$

Calcoliamo, ora, la potenza della forza meccanica che si esercita sul conduttore.

Poiché su una corrente immersa in un campo magnetico agisce una forza pari a $F = IBL \sin \vartheta$, dove ϑ è l'angolo formato tra la corrente e il campo magnetico, anche sulla corrente, dunque, il campo magnetico esercita una forza pari a:

$$F = IBL \quad (3.3)$$

Sostituendo la (3.2) nella (3.3) si ha che:

$$F = \frac{BvL}{R} BL = \frac{v}{R} (BL)^2$$

La potenza dissipata dalle forze meccaniche è pari al lavoro compiuto dalla forza F nell'unità di tempo:

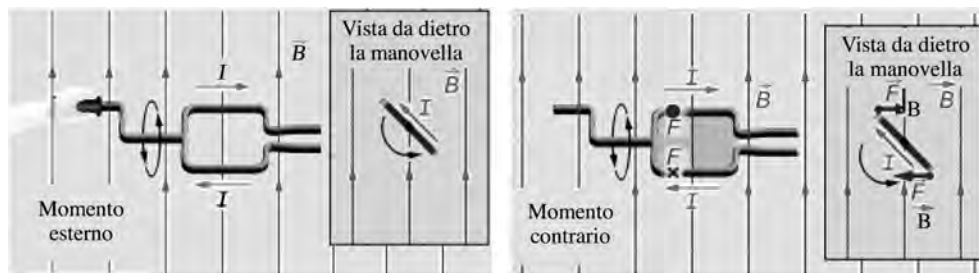
$$P = \frac{L}{\Delta t} = \frac{F \Delta s}{\Delta t} = Fv = v \frac{v}{R} (BL)^2 = \frac{(BvL)^2}{R}$$

che coincide con la potenza dissipata per effetto Joule.

4 I generatori di corrente alternata

Vediamo che cosa succede quando una variazione del flusso concatenato in un circuito è dovuta ad una rotazione di una spira in un campo magnetico.

Nella figura 9.a è stata riportata sia la vista laterale che quella frontale della spira rotante.



Figg. 9.a e 9.b

Mentre la spira ruota, sugli elettroni che scorrono al suo interno agisce una forza:

$$F_B = qvB \sin \vartheta$$

legata all'angolo formato fra il vettore velocità degli elettroni e il vettore campo magnetico (Fig. 9.b).

Siccome questo angolo varia nel tempo, varia anche la f.e.m. indotta in ogni segmento della spira; in particolare:

- è massima quando il vettore velocità degli elettroni, e quindi la corrente, è perpendicolare al campo (**spira verticale**);
- è nulla quando il vettore velocità degli elettroni è parallelo al campo (**spira orizzontale**).

Le forze indotte dal campo magnetico generano un momento opposto al momento della forza che l'ha generato.

La direzione della f.e.m. e della corrente cambiano durante la rotazione della spira; durante un ciclo completo, ciascun segmento ruota per metà nella direzione del campo e per metà in direzione opposta e quindi sia il verso della f.e.m. che quello della corrente indotta cambiano ogni mezzo giro, creando così nel circuito una corrente alternata.

La spira rotante in figura 11 può, pertanto, essere considerata un **generatore di corrente alternata**.

a) Forza elettromotrice indotta da un generatore

La forza elettromotrice indotta da un generatore, formato da N spire di area A che ruotano con velocità angolare ω in un campo magnetico uniforme B , è pari a:

$$V_{\text{TOT}} = NBA\omega \sin\omega t$$

DIMOSTRAZIONE

Supponiamo che le N spire che compongono il generatore abbiano lunghezza L e altezza h (Fig. 10).

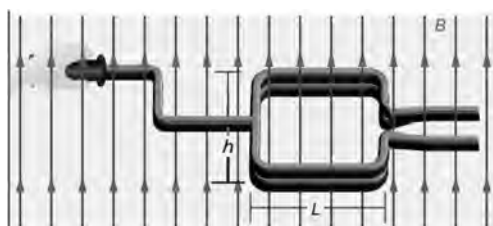
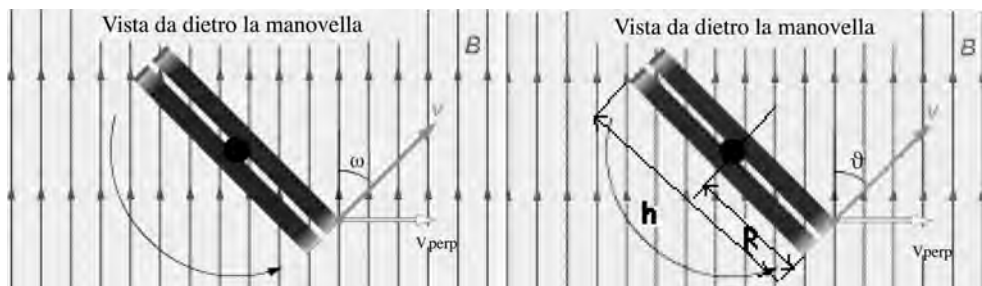


Fig. 10

Nelle figure 11.a e 11.b è stato tracciato il vettore velocità (v_{perp}) perpendicolare al campo magnetico.



Figg. 11.a e 11.b

Consideriamo la f.e.m. indotta in ciascun segmento.

I segmenti orizzontali si muovono in direzioni opposte: il segmento superiore, ruotando, si allontana dal foglio, mentre quello inferiore vi si avvicina.

Dalla direzione dei loro movimenti rispetto al campo possiamo dedurre che le f.e.m. indotte sui segmenti orizzontali hanno lo stesso verso e quindi si sommano.

I segmenti verticali (indicati con h) si muovono, invece, nello stesso verso e, quindi, le f.e.m. indotte hanno verso opposto: di conseguenza si annullano. I segmenti verticali non producono f.e.m.

La f.e.m. indotta in una spira che si muove di un angolo ϑ in un campo magnetico è pari a:

$$V = Lv_{\text{perp}}B = Lv\sin\vartheta B = BLv\sin\vartheta \quad (4.1)$$

dove v_{perp} è la componente della velocità normale al campo magnetico.

La velocità angolare è legata alla velocità tangenziale dalla formula:

$$v = \omega R$$

Dalla figura 11.b si ricava che:

$$R = \frac{1}{2}h$$

e quindi:

$$v = \frac{\omega h}{2} \quad (4.2)$$

inoltre:

$$\vartheta = \omega t \quad (4.3)$$

Pertanto, sostituendo la (4.2) e la (4.3) nella (4.1) avremo:

$$V = BLvsen\vartheta = BL \cdot \left(\frac{\omega h}{2}\right) sen\omega t$$

La f.e.m. prodotta dal generatore è pari a:

$$V_{TOT} = 2NV = 2NBL \cdot \left(\frac{\omega h}{2}\right) sen\omega t = NBA\omega sen\omega t$$

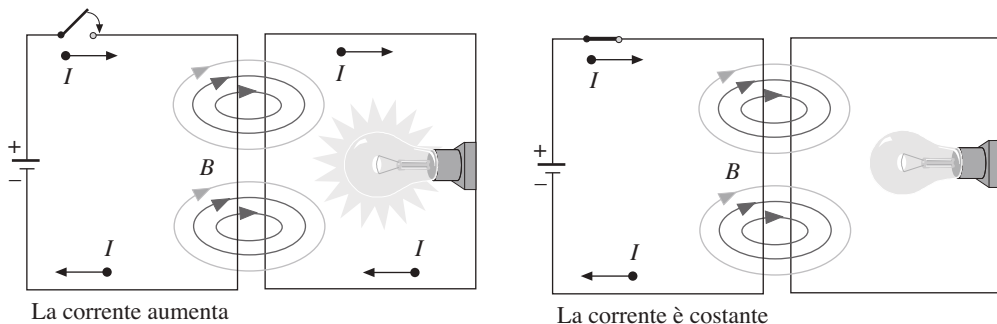
come volevasi dimostrare.

5 La mutua induzione

Con il termine **mutua induzione** si indica la f.e.m. indotta in un circuito in seguito alla variazione della corrente in un altro circuito ad esso accoppiato e cioè posto nelle immediate vicinanze.

Infatti, da quanto visto precedentemente, una corrente variabile genera in un circuito un campo magnetico variabile, il quale, se viene concatenato ad un secondo circuito, genera in esso una corrente, e quindi una f.e.m., variabile.

Consideriamo la figura 12.a, in cui il circuito di sinistra contiene una batteria e un interruttore.



Figg. 12.a e 12.b

1. Quando si chiude l'interruttore, una corrente inizia a scorrere nel circuito di sinistra fino a raggiungere il suo valore di regime.
2. All'aumentare della corrente aumenta anche il campo magnetico generato dalla corrente stessa.
3. L'aumento del campo magnetico fa aumentare il flusso magnetico, il quale risulta concatenato sia al circuito di destra che a quello di sinistra.
4. L'aumento del flusso magnetico nel circuito di destra genera una corrente che fa accendere la lampadina.
5. Quando la corrente nel circuito di sinistra raggiunge il suo valore di regime, e quindi smette di aumentare, non si ha più una variazione del flusso magnetico, per cui non si crea più una corrente indotta nel circuito di destra, e la lampadina si spegne (Fig. 12.b).

*Siccome una forza elettromotrice indotta non viene generata da un campo magnetico ma da una **variazione di campo magnetico**, se nel circuito di sinistra si fosse trovato un generatore di corrente alternata la lampadina non si sarebbe mai spenta.*

a) Trasformatori

Sul fenomeno della mutua induzione si basano alcuni dispositivi chiamati **trasformatori**.

Essi sono costituiti da un nucleo di ferro, su cui sono avvolte due bobine che fanno parte di due circuiti differenti.

In figura 13 è mostrato un trasformatore costituito da due avvolgimenti isolati su un nucleo di ferro.

Il nucleo di ferro fa sì che il campo magnetico venga trasmesso da un avvolgimento all'altro.

In un trasformatore, circa il 5% di energia viene dissipato per calore.

L'avvolgimento di **sinistra** è detto **avvolgimento primario** ed è collegato a un generatore di corrente alternata.

L'avvolgimento di **destra** è chiamato **avvolgimento secondario** e rappresenta una componente puramente resistiva, detta **carico**.

In figura 14 è mostrato il simbolo con cui vengono indicati i trasformatori.

In essa sono schematizzati l'avvolgimento primario e secondario, mentre le sbarrette verticali, disegnate fra i due avvolgimenti, simboleggiano il **nucleo ferromagnetico** sul quale si avvolgono i fili conduttori.

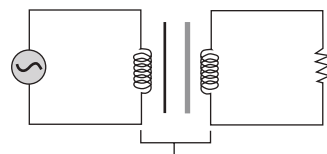


Fig. 14

La corrente variabile che circola nell'avvolgimento primario genera un campo magnetico variabile, il quale attraversa anche l'avvolgimento secondario, generando in esso una f.e.m. indotta e una corrente indotta.

Il trasformatore è così chiamato perché trasforma la f.e.m. del generatore, nel circuito primario, nella tensione ai capi dell'avvolgimento secondario.

Infatti, se indichiamo con ΔV_1 e ΔV_2 le tensioni ai capi dei due circuiti, e con N_1 e N_2 il numero delle loro spire, possiamo dimostrare che valgono le seguenti relazioni (Fig. 15):

$$\frac{\Delta V_1}{\Delta V_2} = \frac{N_1}{N_2}$$

$$\frac{I_2}{I_1} = \frac{N_1}{N_2}$$

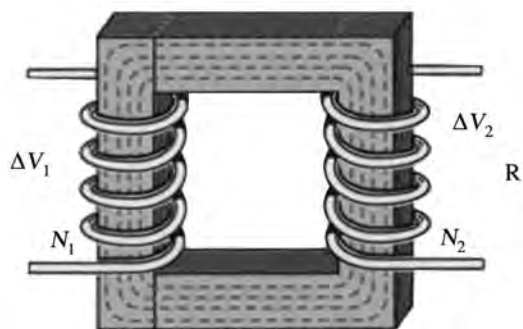


Fig. 15

DIMOSTRAZIONE

Per la legge di Faraday, la differenza di tensione ai capi di un avvolgimento di N spire è pari a:

$$\Delta V = -N \frac{\Delta \Phi_B}{\Delta t}$$

Dunque ai capi dell'avvolgimento primario si ha che:

$$\Delta V_1 = -N_1 \frac{\Delta \Phi_B}{\Delta t}$$

La variazione di flusso concatenata al circuito primario è anche concatenata al circuito secondario e, quindi, la tensione ai capi dell'avvolgimento secondario è pari a:

$$\Delta V_2 = -N_2 \frac{\Delta \Phi_B}{\Delta t}$$

Dividendo membro a membro si ottiene:

$$\frac{\Delta V_1}{\Delta V_2} = \frac{N_1}{N_2} \quad (5.1)$$

Se, per esempio, il rapporto fra il numero di spire fra il circuito secondario e quello primario è di 5:1, anche la tensione sul secondario sarà cinque volte maggiore di quella sul circuito primario.

Questo significa che i trasformatori possono «trasformare» una tensione, aumentandola o diminuendola a piacimento, a seconda del rapporto fra il numero di spire.

In fisica, però, governa sovrano il principio di conservazione dell'energia, per cui l'energia e, di conseguenza, la potenza sul circuito primario deve essere uguale a quella sul secondario.

La potenza è data dal prodotto $P = \Delta VI$ e, quindi, per il principio di conservazione dell'energia, deve essere:

$$\Delta V_1 I_1 = \Delta V_2 I_2$$

Pertanto:

$$\frac{\Delta V_1}{\Delta V_2} = \frac{I_2}{I_1}$$

Ma dalla (5.1) segue che:

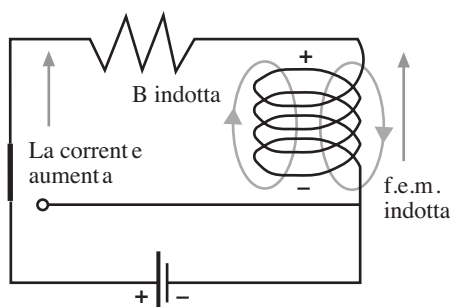
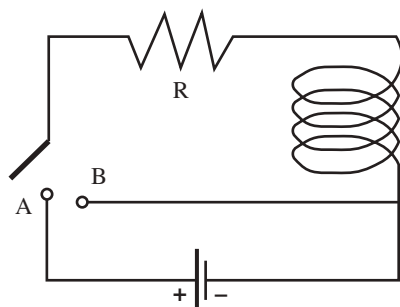
$$\frac{I_2}{I_1} = \frac{N_1}{N_2}$$

come volevasi dimostrare.

6 L'autoinduzione

Consideriamo il circuito in figura 16, nel quale è stato inserito un solenoide e un generatore di tensione continua.

Quando l'interruttore è aperto (Fig. 16.a), nel circuito non passa corrente.



Figg. 16.a e 16.b

Quando chiudiamo l'interruttore sul contatto A (Fig. 16.b), nel circuito inizia a scorrere corrente.

La corrente non raggiunge subito il suo valore di regime, dato dalla legge di Ohm, bensì presenta un andamento come quello mostrato in figura 17.

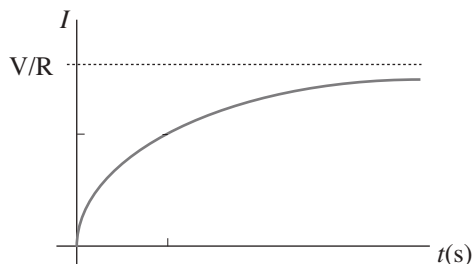


Fig. 17

Supponiamo che la corrente nel circuito di figura 16.b abbia raggiunto il suo valore di regime e mettiamo in corto circuito gli estremi del solenoide chiudendo l'interruttore sul contatto B ed escludendo in questo modo la batteria dal circuito (Fig. 18).

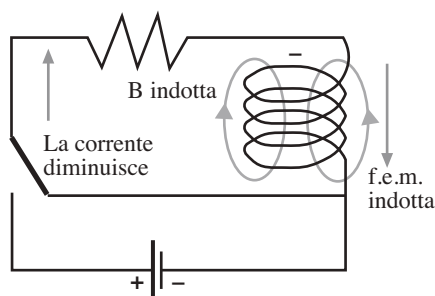


Fig. 18

La corrente nel solenoide non diminuisce istantaneamente ma segue l'andamento mostrato in figura 19.

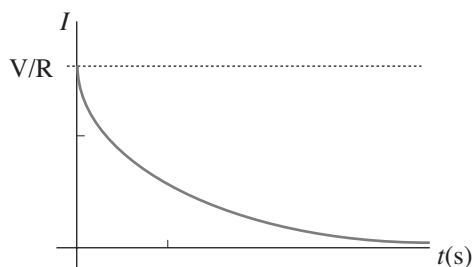


Fig. 19

Questo comportamento dipende dal fatto che la corrente, quando circola nel solenoide, genera un campo di induzione magnetica, il quale, a sua volta, genera una forza elettromotrice che si oppone alla variazione di campo magnetico che l'ha prodotto.

Questa forza è anche chiamata **forza controelettromotrice**.

Un fenomeno di questo tipo è detto di **autoinduzione** in quanto l'aumento di corrente induce automaticamente una corrente di verso contrario.

a) Induttanza

Sia il campo B che il flusso concatenato con un circuito sono in ogni punto proporzionali alla corrente I :

$$\Phi = LI \quad (6.1)$$

Il coefficiente di proporzionalità L è detto **coefficiente di autoinduzione** o **induttanza** del circuito.

Per la legge di Lenz, una variazione di flusso nel circuito genera una forza elettromotrice indotta pari a:

$$V = -\frac{d\Phi}{dt}$$

e quindi, sostituendo nella (6.1):

$$V = -L \frac{dI}{dt}$$

Dalla (6.1) si ricava che l'unità di misura di L è weber/ampere, **henry** (H).

1 henry rappresenta l'induttanza di un circuito che abbraccia il flusso di 1 weber quando è percorso dalla corrente di 1 ampere.

b) Calcolo dell'induttanza in un solenoide

L'**induttanza di un solenoide** (Fig. 20) è pari a:

$$L = n^2 s \mu_0 A$$

dove:

- n è la densità di spire;
- μ_0 è la permeabilità magnetica nel vuoto;
- s è la lunghezza del solenoide;
- A è la sezione del solenoide.

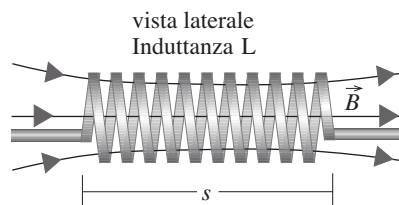


Fig. 20

DIMOSTRAZIONE

Calcoliamo, innanzi tutto, il flusso magnetico che attraversa una sola spira del solenoide. Per definizione, il flusso magnetico è pari a:

$$\Phi_B = BA$$

In un solenoide il campo magnetico è pari a:

$$B = \mu_0 I n$$

Quindi:

$$\Phi_B = (\mu_0 I n) A \quad (6.2)$$

Se s è la lunghezza del solenoide e N il numero di spire, la densità di spire è pari a:

$$n = \frac{N}{s}$$

da cui segue che:

$$N = ns \quad (6.3)$$

Riprendiamo la definizione dell'induttanza:

$$L = N \frac{\Phi_B}{I} \quad (6.4)$$

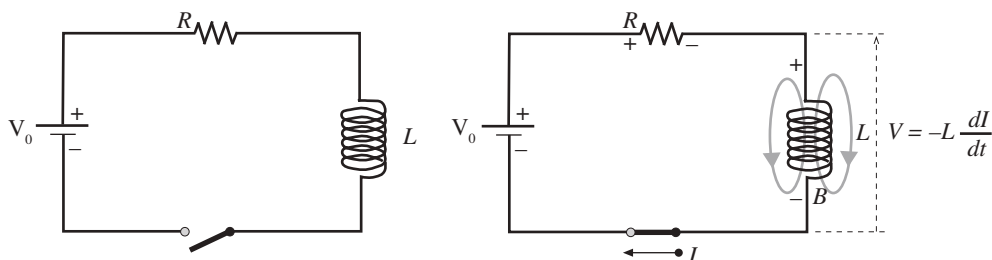
e sostituiamo nella (6.4) la (6.3) e la (6.2); si ottiene che:

$$L = N \frac{\Phi_B}{I} = ns \frac{(\mu_0 In)A}{I} = n^2 s \mu_0 A$$

come volevasi dimostrare.

7 I circuiti RL

Un circuito in cui è presente un resistore e un induttore in serie, come quello mostrato in figura 21.a, è detto **circuito RL**.



Figg. 21.a e 21.b

Siccome l'induttore, in questo circuito, si oppone a una qualsiasi variazione di corrente, essa non potrà aumentare o diminuire bruscamente.

In figura 21.b è mostrato il diagramma di un circuito RL.

Quando l'interruttore si chiude, la corrente inizia a fluire nel circuito e cresce fino a raggiungere il suo valore massimo. All'aumentare della corrente nel circuito, ai capi dell'induttore si genera una forza controelettromotrice pari a: $V = L \frac{dI}{dt}$.

Per analizzare questo circuito utilizziamo la **legge delle maglie di Kirchhoff**:

$$V_0 - IR - L \frac{dI}{dt} = 0 \quad (7.1)$$

Per risolvere la (7.1) possiamo separare le variabili:

$$L dI = -(IR - V_0) dt$$

da cui segue che:

$$\frac{dI}{(IR - V_0)} = -\frac{dt}{L}$$

Integriamo, ora, entrambi i membri:

$$\int \frac{dI}{IR - V_0} = -\int \frac{dt}{L}$$

ottenendo:

$$\frac{1}{R} \ln(IR - V_0) = -\frac{1}{L} t + \ln k$$

Ricordando che $a = \ln e^a$ e applicando le regole dei logaritmi, si ottiene:

$$\ln(IR - V_0)^{1/R} = \ln e^{-t/L} + \ln k$$

$$\ln(IR - V_0)^{1/R} = \ln k e^{-t/L}$$

Passando dai logaritmi ai numeri si ottiene:

$$(IR - V_0)^{V/R} = ke^{-t/L}$$

ossia:

$$IR - V_0 = ke^{-\frac{R}{L}t} \quad (7.2)$$

Nell'istante $t=0$ in cui il circuito viene chiuso, nello stesso non circola corrente ($I=0$) e quindi la (7.2) diventa:

$$-V_0 = ke^0 = k$$

Sostituendo questo valore nella (7.2) si ottiene:

$$IR - V_0 = -V_0 e^{-\frac{R}{L}t}$$

e pertanto:

$$I(t) = \frac{V_0}{R} (1 - e^{-\frac{R}{L}t}) \quad (7.3)$$

Possiamo riscrivere la (7.3) come:

$$I(t) = \frac{V_0}{R} (1 - e^{-t/\tau}) \quad (7.4)$$

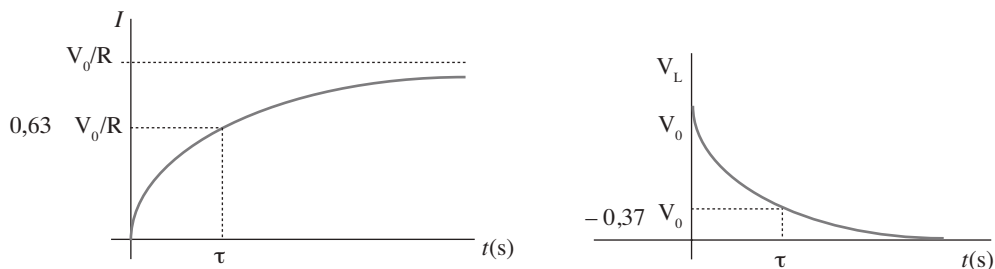
dove è stata introdotta la costante:

$$\tau = \frac{L}{R}$$

detta **costante di tempo**.

Essa indica in quanto tempo la corrente passa dal valore 0 al valore $(1 - \frac{1}{e})$ e si misura in secondi.

In figura 22.a è mostrato l'andamento della corrente nel circuito.



Figg. 22.a e 22.b

Ai capi dell'induttore la tensione vale:

$$V_L = V_0 - V_R = V_0 - RI$$

Dalla (7.4) si ottiene:

$$V_L = V_0 - V_0 (1 - e^{-t/\tau}) = V_0 (1 - 1 + e^{-t/\tau}) = V_0 e^{-t/\tau} \quad (7.5)$$

Dalla (7.4) e dalla (7.5) possiamo notare che, per $t=0$:

$$I(0) = 0 \quad V(0) = V_0$$

e quindi la tensione ai capi dell'induttore è uguale alla tensione ai capi del generatore: l'induttore si comporta, cioè, come un circuito aperto.

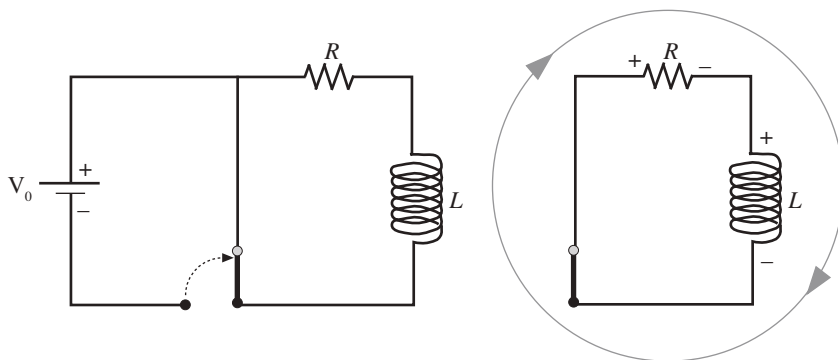
Inoltre, per $t = \infty$:

$$I(\infty) = \frac{V_0}{R}$$

$$V(\infty) = 0$$

L'induttore si comporta come un circuito chiuso e la corrente nel circuito è pari a $I_0 = \frac{V_0}{R}$.

Isoliamo ora il generatore dal circuito (Fig. 23.a)



Figg. 23.a e 23.b

Inizialmente nel circuito scorre una corrente $I_0 = \frac{V_0}{R}$, la quale inizierà poi a diminuire. Nell'induttore si creerà nuovamente una forza contro elettromotrice, che si opporrà alla diminuzione di corrente (Fig. 23.b)

Applicando la legge di Kirchhoff alla maglia del circuito, si ottiene:

$$IR + L \frac{dI}{dt} = 0$$

da cui segue che:

$$\frac{dI}{I} = -\frac{R}{L} dt$$

Integrando entrambi i membri:

$$\int \frac{dI}{I} = -\frac{R}{L} \int dt$$

si ottiene:

$$\ln(I) = -\frac{R}{L} t + K$$

da cui, passando dal logaritmo al numero, segue che:

$$I(t) = e^{\left(-\frac{R}{L}t + K\right)}$$

dove K è una costante di integrazione che dipende dalle condizioni iniziali del problema.

Sfruttando le proprietà delle potenze possiamo scrivere:

$$I(t) = e^{-\frac{R}{L}t} \cdot e^K$$

ed essendo e^K un numero, possiamo anche scrivere:

$$I(t) = K_0 e^{-\frac{R}{L}t} = K_0 e^{-\frac{t}{\tau}} \quad (7.6)$$

dove $\tau = \frac{L}{R}$ è la costante di tempo tipica del circuito RL.

Possiamo ricavare K_0 considerando che cosa succede all'istante $t = 0$.

Sostituendo $t = 0$ nella (7.6) si ottiene:

$$I(0) = I_0 = K_0 \quad (7.7)$$

e quindi K rappresenta la corrente che scorre nel circuito subito dopo la chiusura dell'interruttore.

Sostituendo la (7.7) nella (7.6) si ottiene:

$$I(t) = I_0 e^{-\frac{t}{\tau}} \quad (7.8)$$

All'istante iniziale, la corrente non risente della forza controelettromotrice «indotta» nell'induttore e, pertanto, l'induttore si comporta come un corto circuito. Cortocircuitando l'induttore nel circuito in figura 23.b si ottiene che la corrente I_0 è pari a:

$$I_0 = \frac{V_0}{R}$$

e quindi possiamo riscrivere la (7.8) come:

$$I(t) = \frac{V_0}{R} e^{-\frac{t}{\tau}}$$

La f.e.m. indotta varia, invece, come:

$$V(t) = V_0 - V_R(t) = V_0 - RI(t) = V_0 - V_0 e^{-\frac{t}{\tau}} = V_0 (1 - e^{-\frac{t}{\tau}})$$

1 Radiazione del corpo nero e ipotesi di Planck

All'inizio del 1900, i risultati ottenuti da un esperimento semplicissimo e del tutto insignificante, come la radiazione di un corpo nero, continuavano ad infastidire la comunità scientifica in quanto si discostavano in modo evidente dalle previsioni teoriche.

Un **corpo nero** ideale è un corpo in grado di assorbire completamente tutte le radiazioni che lo investono, il quale emette uno spettro di radiazione elettromagnetica che dipende esclusivamente dalla sua temperatura. La radiazione emessa è detta **Radiazione del corpo nero** o **Radiazione del corpo cavo**.

Per realizzare in pratica un corpo nero possiamo prendere un qualsiasi corpo cavo, per esempio una sfera, e inserirvi all'interno, tramite un piccolo foro, della radiazione elettromagnetica.

La radiazione viene quindi riflessa dalle pareti interne del corpo fino a quando non viene totalmente assorbita, causando così un aumento della sua temperatura. In queste condizioni è possibile calcolare teoricamente e misurare sperimentalmente la distribuzione spettrale del corpo nero.

In figura sono mostrate alcune funzioni di distribuzione spettrale corrispondenti a corpi neri a diverse temperature. Inoltre, sono riportati i valori della potenza P irradiata per unità di superficie dalla radiazione elettromagnetica in funzione della sua lunghezza d'onda λ .

Questi valori possono essere ricavati sperimentalmente. Quando tuttavia si confrontano i valori sperimentali con quelli previsti dalla teoria della radiazione del corpo nero ci si accorge che esistono notevoli discrepanze per le piccole lunghezze d'onda.

Secondo la teoria di Rayleigh-Jeans, la potenza irradiata è inversamente proporzionale alla quarta potenza della lunghezza d'onda:

$$P(\lambda, T) = \frac{8\pi k_B T}{\lambda^4}$$

dove k_B è la costante di Boltzmann, il cui valore è circa $1,38 \times 10^{-23} \text{ J/K}$, e T è la temperatura del corpo nero in gradi Kelvin.

Questo significa che, teoricamente, la potenza P della radiazione elettromagnetica dovrebbe crescere al decrescere della lunghezza d'onda, ossia:

$$\lim_{\lambda \rightarrow 0} P = \infty$$

Questa conclusione è in netto contrasto con quella ottenuta sperimentalmente in laboratorio per la quale:

$$\lim_{\lambda \rightarrow 0} P = 0$$

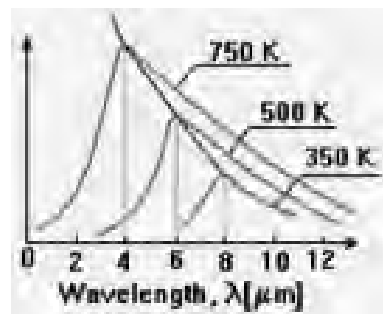


Fig. 1

Questa palese contraddizione tra risultati sperimentali e teorici per le piccole lunghezze d'onda viene indicato con il nome di **catastrofe ultravioletta**.

Per dare una soluzione a questa catastrofe Max Planck introdusse, durante una famosa conferenza tenuta il 14 dicembre 1900, una formula completamente nuova per la radiazione:

$$P(\nu, T) = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{(e^{h\nu/k_B T} - 1)}$$

Con questa formula egli fece automaticamente scattare la nascita della fisica quantistica e ne divenne immediatamente il padre.

La novità introdotta da Planck nella sua formula era che, al contrario di quanto affermato fino ad allora dalla Fisica classica, sia l'emissione che l'assorbimento della radiazione elettromagnetica all'interno di un corpo nero fossero **quantizzate**. Di conseguenza l'energia termica del corpo nero doveva essere assorbita o ceduta dalle pareti sempre in piccole porzioni, i cosiddetti quanti, che potevano essere considerati come dei piccoli pacchetti di energia.

Secondo l'ipotesi di Planck, l'energia di un singolo quanto dipende soltanto dalla frequenza ν e dalla costante $h = 6,626 \times 10^{-34}$ Js, detta Costante di Planck, ed è pari a:

$$E = h\nu$$

Da notare che Planck formulò questa ipotesi come un puro artificio di calcolo per far coincidere la funzione di distribuzione spettrale prevista dalla teoria con quella misurata sperimentalmente.

2 L'effetto fotoelettrico

L'**effetto fotoelettrico** venne scoperto da Hertz nel 1887 durante l'esperimento che egli fece per generare e rilevare onde elettromagnetiche. In quell'esperimento Hertz utilizzò uno spinterometro in un circuito accordato per generare onde e un altro circuito simile per rilevarle.

Egli notò che il passaggio di scintille nello spinterometro ricevitore del suo apparecchio era facilitato dalla luce proveniente dallo spinterometro generatore.

Nel 1900 il fisico Lenard scoprì che la luce incidente su una superficie metallica produce un fascio di elettroni, la cui energia è indipendente dall'intensità della luce incidente.

In figura è mostrato il dispositivo utilizzato.

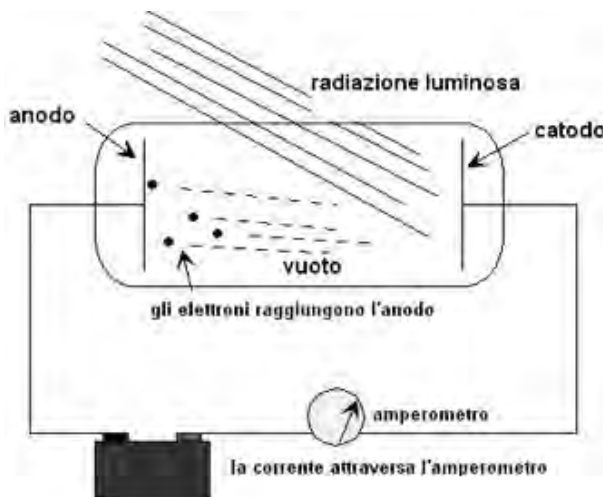


Fig. 2

Quando la luce incide sulla superficie metallica, dal catodo vengono emessi degli elettroni e, se alcuni di essi colpiscono l'anodo, si genera una corrente nel circuito esterno.

Il numero di elettroni che raggiungono l'anodo dipende dalla differenza di potenziale V tra anodo e catodo.

Se V è positivo gli elettroni sono attratti dall'anodo e per valori di V abbastanza grandi tutti gli elettroni estratti dal catodo raggiungono l'anodo e la corrente nel circuito esterno si stabilizza su un valore costante. Un ulteriore aumento di V non fa aumentare la corrente.

Lenard scoprì che la massima corrente nel circuito esterno dipendeva dall'intensità della luce incidente sul catodo e questo, tutto sommato, era un risultato atteso in quanto all'aumentare dell'intensità luminosa aumenta il numero degli elettroni emessi.

Se V è negativo, gli elettroni vengono respinti dall'anodo e soltanto quelli che hanno un'energia cinetica iniziale maggiore dell'energia potenziale $|eV|$ possono raggiungere l'anodo.

Se V è minore di un certo valore $-V_0$, nessun elettrone raggiunge l'anodo. Il potenziale V_0 , detto **potenziale d'arresto**, è legato all'energia cinetica massima degli elettroni emessi dalla relazione:

$$\left(\frac{1}{2}mv^2\right)_{\max} = eV_0$$

ed è quindi indipendente dall'intensità della luce incidente.

Nel 1905 Einstein dimostrò che questo risultato può essere spiegato se si suppone che l'energia è quantizzata in pacchetti chiamati **fotoni**, ognuno avente energia pari a $E = h\nu_0$, dove ν_0 è la frequenza e h è la costante di Planck. Egli, quindi, riuscì a dare una spiegazione all'effetto fotoelettrico utilizzando l'ipotesi di Planck della quantizzazione dell'energia!

Visto in questa ottica l'effetto fotoelettrico può essere spiegato nel seguente modo. I singoli elettroni che vengono estratti dal metallo illuminato ricevono una certa quantità di energia dai fotoni. Se si aumenta l'intensità della luce aumentano il numero di fotoni che colpiscono la superficie nell'unità di tempo ma non aumenta l'energia assorbita da ciascun elettrone.

Se indichiamo con Φ l'energia necessaria per estrarre un elettrone dal metallo, detta anche **lavoro di estrazione**, l'energia cinetica massima degli elettroni emessi è pari a:

$$\left(\frac{1}{2}mv^2\right) = eV_0 = h\nu - \Phi$$

Quest'ultima relazione è chiamata anche **equazione fotoelettrica di Einstein**.

3 I raggi X

I **raggi X** vennero scoperti da Röntgen nel 1895 mentre lavorava con un tubo a raggi catodici.

Egli osservò che i raggi che provenivano dal tubo erano capaci di attraversare i materiali opachi e di impressionare una pellicola fotografica.

Lo studioso non riuscì a deviare questi raggi con un campo magnetico e non osservò nemmeno spettri di interferenza come ci si sarebbe aspettato dalla loro ipotetica natura ondulatoria e, a causa di quest'incognita comportamentale, li chiamò *raggi X*.

Questo fatto non scoraggiò assolutamente il fisico tedesco, il quale analizzò a fondo le caratteristiche di questi raggi, arrivando alla conclusione che tutti i materiali erano più o meno trasparenti ad essi e che la loro trasparenza diminuiva con la densità.

La scoperta segnò la nascita della **radiografia** e Röntgen ricevette il primo premio Nobel della storia, nel 1901.

Se si analizzasse lo spettro emesso da un tipico tubo a raggi X, quando un bersaglio all'interno del tubo viene bombardato con elettroni, si noterebbe che lo spettro è costituito da una

serie di righe sottili, che prendono il nome di **spettro caratteristico**, sovrapposto a uno spettro continuo, detto **spettro di Bremsstrahlung**.

Lo spettro continuo è dovuto alla rapida decelerazione degli elettroni quando entrano nel bersaglio.

Se la differenza di potenziale nel tubo è V_0 , l'energia cinetica massima degli elettroni quando colpiscono il bersaglio è pari a eV_0 .

Normalmente, quando un elettrone rallenta vengono emessi parecchi fotoni; a volte, può succedere che venga emesso un solo fotone con energia pari a eV_0 .

La lunghezza d'onda di un fotone è inversamente proporzionale alla sua energia:

$$\lambda = \frac{hc}{E}$$

e, quindi, a una massima energia eV_0 corrisponde una minima lunghezza d'onda nello spettro di Bremsstrahlung, detta **lunghezza d'onda limite** λ_m :

$$\lambda_m = \frac{hc}{eV_0}$$

4 L'effetto Compton

Nel 1920 il fisico Arthur Compton analizzò la diffrazione di raggi X da parte di elettroni liberi.

In figura è schematizzato l'urto tra un fotone e un elettrone fermo.

Secondo la teoria classica, l'energia e la quantità di moto di un'onda elettromagnetica sono legate dalla relazione:

$$E = pc$$

Essendo $E = \frac{hc}{\lambda}$, la quantità di moto di un fotone è legata alla sua lunghezza d'onda dalla relazione:

$$p = \frac{h}{\lambda}$$

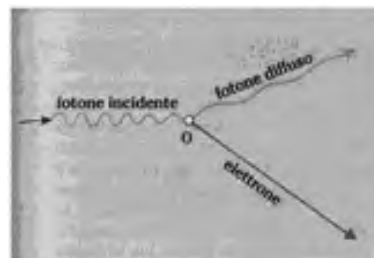


Fig. 3

Trattando la diffusione con le equazioni della meccanica relativistica e, applicando i principi di continuità dell'energia e della quantità di moto, Compton giunse alla conclusione che l'angolo di diffusione ϑ è legato alla variazione della lunghezza d'onda prima e dopo l'urto dalla relazione:

$$\lambda_2 - \lambda_1 = \frac{h}{mc} (1 - \cos \vartheta)$$

Possiamo quindi notare che la variazione di lunghezza d'onda è indipendente dalla lunghezza d'onda iniziale.

La grandezza $\frac{h}{mc}$ dipende soltanto dalla massa dell'elettrone ed è chiamata **lunghezza d'onda Compton**.

5 La lunghezza d'onda di De Broglie

L'introduzione della costante di Planck per spiegare sia la catastrofe ultravioletta che l'effetto fotoelettrico portò un radicale sconvolgimento nella comunità scientifica dell'inizio del 900.

Fino ad allora, infatti, nel pensiero dei fisici era radicato il concetto che la luce si comportasse come un'onda e che, in base a questo modello, fosse possibile spiegare tutti i fenomeni ottici.

D'altronde gli esperimenti sulla diffrazione della luce condotti da Young e Fresnel e le equazioni di Maxwell erano prove evidenti che la luce avesse una natura ondulatoria e che quindi si propagasse nello spazio come un'onda elettromagnetica di piccole lunghezze d'onda. Tale comportamento non era, però, in grado di interpretare l'effetto fotoelettrico.

Planck e Einstein dimostrarono, invece, che la luce fosse composta da quanti o fotoni e che quindi possedesse una natura corpuscolare.

Questa conclusione non poteva giustificare il fenomeno dell'interferenza e della diffrazione.

Il dualismo onda-corpuscolo si ripresentò anche per le particelle materiali e, nel 1924, De Broglie propose di concepirle come onde a cui era associata una lunghezza d'onda, detta **lunghezza d'onda di De Broglie**, pari a $\lambda = \frac{h}{p}$, dove h è la costante di Planck e p è la quantità di moto della particella.

I Fisici si trovarono, quindi, di nuovo di fronte ad una situazione che poteva sembrare contraddittoria, nel senso che si discostava completamente dall'esperienza quotidiana, in cui i concetti di onda e di corpuscolo sono legati a grandezze e caratteristiche completamente diverse.

Dalla meccanica classica possiamo derivare il concetto di particella come di un punto materiale con massa infinitesima identificabile con una posizione ben precisa nel tempo e nello spazio, mentre un'onda è una perturbazione che si ripete periodicamente nel tempo e nello spazio.

Le caratteristiche utilizzate per identificare un'onda (ampiezza, lunghezza d'onda e frequenza) sembrano perdere completamente significato se riferite a una particella.

Storicamente fu un'esperienza di diffrazione effettuata da Davidson e Germel nel 1927, con elettroni riflessi da un monocristallo di nichel, che dette un valido supporto sperimentale all'ipotesi di De Broglie. Tutt'ora la validità della relazione di De Broglie è verificata nei molti esperimenti eseguiti a livello di fisica atomica e con particelle elementari.

6 Il principio di indeterminazione di Heisenberg

Il **principio di indeterminazione di Heisenberg** costituisce uno dei perni della meccanica quantistica.

Esso stabilisce una relazione vincolante tra il grado di precisione Δx , con cui si può conoscere la posizione di un oggetto quantistico, e l'incertezza Δp della sua quantità di moto.

$$\Delta x \Delta p \geq \frac{h}{4\pi}$$

dove h è la costante di Planck.

Come possiamo notare, al secondo membro abbiamo una costante e quindi un valore che non cambia mai e che non può annullarsi: il diminuire di Δx porta ad un aumento di Δp e viceversa.

Il principio di indeterminazione limita a priori la conoscenza delle traiettorie esatte degli oggetti quantistici e questa limitazione non è dovuta a una imprecisione nelle misure ma a una proprietà della materia nel microcosmo. Un oggetto quantistico o non ha una posizione oppure non possiede una quantità di moto precisa.

7 L'equazione di Schrödinger

L'**equazione di Schrödinger** è un'equazione differenziale, la cui soluzione dà la «funzione d'onda» che descrive il comportamento di una particella microscopica.

Le scoperte di Planck, De Broglie e Heisenberg mostravano che esiste un'onda associata alla materia, con proprietà molto particolari. Nel caso dell'elettrone libero si sapeva com'era fatta. Ma negli altri casi? Per esempio se un elettrone è legato ad un atomo?

Ci voleva un'equazione che, risolta, desse la corretta funzione d'onda in grado di descrivere il comportamento dell'elettrone o di altre particelle nelle diverse situazioni.

Schrödinger risolse questo problema facendo il seguente ragionamento:

se gli elettroni si comportano come onde, forse anche per essi possiamo trovare una «funzione d'onda», $\Psi(x, t)$, analoga a quella valida per le onde elettromagnetiche e che descriva il comportamento delle onde di materia?

Egli considerò due diverse situazioni:

- le equazioni d'onda stazionarie;
- le equazioni d'onda dipendenti dal tempo.

Applicando a queste equazioni il risultato ottenuto da De Broglie, secondo il quale la lunghezza d'onda di una particella in presenza di potenziale è pari a:

$$\lambda = \frac{h}{\sqrt{2m(E-V)}}$$

ottenne due diverse equazioni:

- **l'equazione per le onde stazionarie**, utilizzabile nel caso in cui il potenziale del sistema sia costante:

$$-\frac{h^2}{8\pi^2m} \frac{d^2\psi(x)}{dx^2} + V\psi(x) = E\psi(x)$$

- **l'equazione per le onde dipendenti dal tempo**, utilizzabile negli altri casi:

$$-\frac{h^2}{8\pi^2m} \frac{\partial^2\psi(x,t)}{\partial x^2} + V\psi(x,t) = i\frac{h}{2\pi} \frac{\partial\psi(x,t)}{\partial t}$$

8 I modelli atomici

Nonostante numerose evidenze sperimentali avessero permesso di trovare l'esistenza delle particelle subatomiche, gli scienziati non erano ancora in grado di visualizzarne la disposizione all'interno dell'atomo e, a partire dalla fine del XIX secolo, iniziarono a definire dei modelli, i quali vennero via via perfezionati ogni volta che nuovi esperimenti conducevano a nuove interessanti conclusioni.

a) Teoria di Dalton

Dalton nel 1803, basandosi sull'interpretazione delle leggi fondamentali della chimica note a quei tempi (la legge della conservazione della massa, quella delle proporzioni definite e quella delle proporzioni multiple), trasse le seguenti conclusioni:

- la materia non è continua ma è composta da particelle, dette **atomi**, che non possono essere ulteriormente divisibili;
- gli atomi di un particolare elemento sono tutti uguali tra loro e hanno la stessa massa;
- gli atomi di elementi diversi hanno massa e proprietà differenti;
- le reazioni chimiche avvengono tra atomi interi e non tra frazioni di essi.

In una reazione chimica tra due o più elementi, gli atomi, pur conservando la propria identità, si combinano secondo rapporti definiti dando luogo a composti.

b) Modello di Thomson

Nel 1898 il fisico J.J. Thomson propose un modello in cui l'atomo, a differenza del modello compatto proposto da Dalton, fosse un aggregato di particelle più semplici. Alla luce dei pochi esperimenti in suo possesso, egli ipotizzò che l'atomo fosse costituito da una sfera omogenea carica di elettricità positiva, in cui gli elettroni erano distribuiti in modo del tutto casuale:

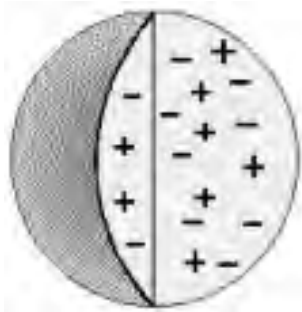


Fig. 4

c) Modello di Rutherford

Rutherford ipotizzò che la massa e la carica positiva fossero concentrate in una parte molto piccola dell'atomo, chiamata nucleo, e che gli elettroni si trovassero nella zona periferica a grande distanza dal nucleo.

Questa ipotesi nasceva dai risultati di un esperimento condotto da due allievi di Rutherford in cui una lamina sottilissima di metallo veniva bombardata con particelle alfa veloci. Uno schermo rivelatore indicava i punti di arrivo delle particelle, stabilendone così la traiettoria dopo il passaggio attraverso la lamina.

Comportamento delle particelle alfa quando attraversano la lamina metallica

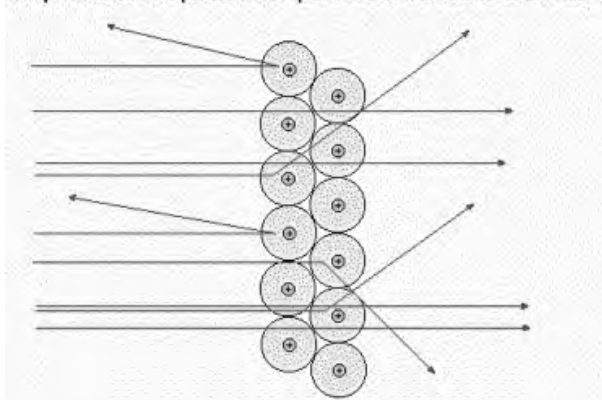


Fig. 5

Siccome la maggior parte delle particelle passarono attraverso la lamina senza subire deviazioni, mentre una minima parte veniva deviata o addirittura respinta, si arrivò alla seguente conclusione:

- le particelle che non venivano deviate erano quelle che passavano abbastanza distanti dai nuclei;
- le particelle che venivano deviate erano quelle che, passando in prossimità dei nuclei, risentivano della repulsione elettrica e il loro angolo di deviazione aumentava con la vicinanza al nucleo;
- le particelle che venivano respinte erano quelle che si dirigevano direttamente verso i nuclei.

Queste particelle erano molto poche in quanto il nucleo occupa una minima parte dello spazio e, quindi, la probabilità che una particella si dirigesse proprio contro un nucleo era bassa.

d) Modello di Bohr

Questo modello venne proposto da Bohr nel 1913 e risponde perfettamente a due esperimenti condotti da Balmer e da Rydberg verso la fine del 1800.

Nel 1884 un insegnante svizzero, Johann Balmer, aveva scoperto che alcune righe dello spettro dell'atomo di idrogeno potevano essere rappresentate dalla formula:

$$\lambda = 364,6 \text{ nm} \frac{n^2}{n^2 - 4}$$

dove n era un numero intero maggiore di 2.

Successivamente, i fisici Rydberg e Ritz trovarono una formula più generale degli spettri atomici, applicabile anche ad altri elementi:

$$\frac{1}{\lambda} = R \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$$

dove $n_2 > n_1$ e R è una costante uguale per tutte le serie spettrali, detta **costante di Rydberg**.

Dalle due formule possiamo notare che la formula di Balmer è un caso particolare della formula di Rydberg con $n_1 = 2$.

Il **modello di Bohr** si basa su alcune ipotesi fondamentali:

- nell'atomo gli elettroni ruotano attorno al nucleo su orbite circolari. Ogni orbita ha un raggio ben determinato;
- il momento angolare degli elettroni è quantizzato;
- finché un elettrone rimane nella sua orbita non emette e non assorbe energia.

Per passare da un'orbita con energia minore a un'orbita con energia maggiore (cioè da un'orbita più interna a una più esterna), l'elettrone deve ricevere dall'esterno una quantità di energia corrispondente alla differenza di energia fra le due orbite; se, invece, passa da un'orbita con energia maggiore a un'orbita con energia minore, l'elettrone emette una quantità di energia pari alla differenza di energia fra le due orbite. L'energia viene emessa o assorbita sotto forma di radiazione elettromagnetica. Esiste una relazione matematica fra i valori di energia delle orbite di partenza e di arrivo e la frequenza della radiazione:

$$E_1 - E_2 = h \nu$$

dove:

- E_1 è l'energia dell'orbita sulla quale si trovava l'elettrone all'inizio;
- E_2 è l'energia dell'orbita sulla quale si è portato l'elettrone;
- h è la costante di Planck;
- ν è la frequenza della radiazione emessa o assorbita.

Questo concetto, che equivale ad affermare la conservazione dell'energia mediante l'emissione di un fotone, è un'ipotesi chiave nella teoria di Bohr perché si discosta dalla teoria classica.

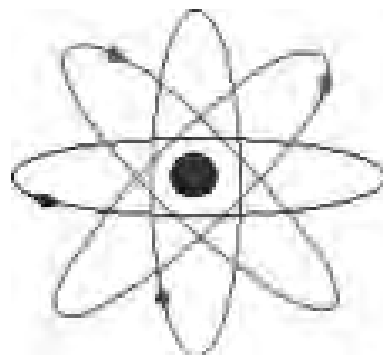


Fig. 6

L'ipotesi di Bohr sulla struttura dell'atomo spiega, quindi, perché gli spettri di emissione degli atomi sono spettri discontinui, ossia a righe: ogni riga corrisponde a un ben determinato valore di energia, che a sua volta corrisponde alla differenza di energia fra due orbite.

Essa spiega perfettamente i dati sperimentali ottenuti da Balmer e da Rydberg.

9 La teoria moderna

Il principio di indeterminazione di Heisenberg e la scoperta della doppia natura dell'elettrone da parte di De Broglie indicavano chiaramente una cosa: non era più possibile trattare l'elettrone come una particella classica.

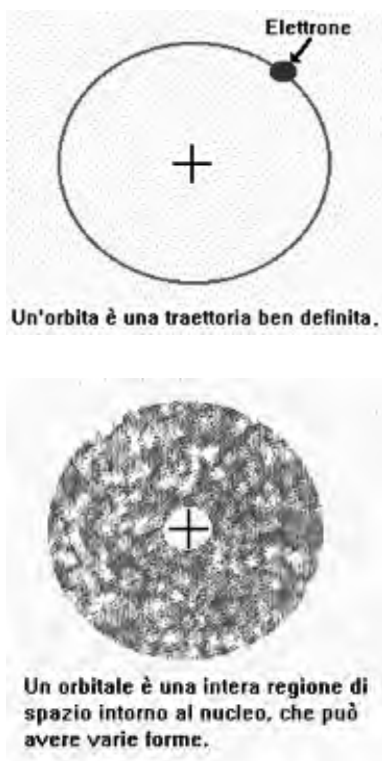


Fig. 7

Bohr, nel suo modello, aveva introdotto l'ipotesi della quantizzazione, ma per il resto aveva trattato l'elettrone come una particella classica, che si muove su orbite ben determinate, il cui raggio può essere calcolato in base a semplici considerazioni meccaniche sulle forze in gioco. Le nuove scoperte, però, imponevano un modo completamente diverso di affrontare il problema, che portò all'elaborazione di una nuova fisica, la **meccanica quantistica**.

Il termine **orbitali** indica le funzioni che si ottengono come soluzione dell'equazione di Schrödinger, che sono visualizzabili come regioni dello spazio intorno al nucleo, nelle quali è possibile trovare l'elettrone. Si può dire che gli orbitali hanno varie forme e si protendono lontano dal nucleo in modo diverso, in relazione ai numeri quantici che ne caratterizzano la funzione d'onda. Ogni funzione d'onda, o orbitale, descrive uno stato dell'atomo. Le diverse funzioni d'onda di un atomo si denotano indicando i valori dei tre numeri quantici: n, l, m ; a ogni terzetto di numeri quantici corrisponde un orbitale ben preciso.

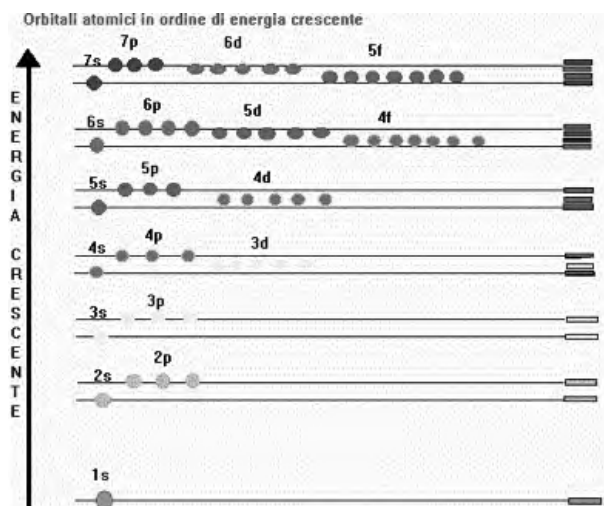


Fig. 8

Il numero quantico principale n può assumere valori maggiori o uguali a 1 e ha il ruolo più importante nel determinare l'energia di un dato orbitale.

Il numero quantico angolare l è legato al valore del momento angolare che l'elettrone ha nel suo moto intorno al nucleo. Esso stabilisce la forma degli orbitali e, insieme al numero n , contribuisce a determinarne l'energia.

Il numero quantico magnetico m definisce l'inclinazione del vettore momento angolare dell'elettrone e l'orientamento degli orbitali nello spazio.

1 La composizione dei nuclei atomici

L'atomo è la più piccola parte di un elemento chimico che conserva le proprietà dell'elemento stesso.

È formato da elettroni che si muovono attorno a un nucleo, composto (in uno schema molto semplificato) da protoni e neutroni. Nel nucleo è quasi del tutto concentrata la massa dell'atomo. Gli atomi sono particelle piccolissime e il loro diametro si misura in Angstrom, simbolo Å ($1\text{Å} = 10^{-10}\text{ m}$).

Gli elettroni esterni al nucleo sono numericamente uguali al numero di protoni.

Le sostanze semplici, o elementi, sono costituite da atomi della stessa specie; le sostanze composte, o composti, da atomi di specie diverse, ma sempre in rapporti numerici ben definiti e generalmente semplici.

Più atomi legati tra loro formano le **molecole**.

Il **protone**, corrispondente alla più piccola carica elettrica positiva esistente libera in natura, con il neutrone, rappresenta la particella fondamentale di ogni nucleo atomico.

La carica del protone è uguale (ma di segno contrario) a quella dell'elettrone, cioè $1,602 \cdot 10^{-19}\text{ C}$ (Coulomb); la sua massa a riposo è pari a $1,672 \cdot 10^{-27}\text{ Kg}$ ($= 938\text{ MeV}$).

Il **neutrone è una particella priva di carica elettrica**; ha una massa di $1,67 \cdot 10^{-27}\text{ Kg}$, uguale cioè a quella del protone e quindi corrispondente a $1u$.

L'elettrone ha una massa di $9,1 \cdot 10^{-31}\text{ Kg}$ e una carica elettrica di $-1,6 \cdot 10^{-19}\text{ C}$ (Coulomb). Tale particella, alla quale per convenzione è stato attribuito il valore di carica elettrica -1 , viene identificata con il simbolo e^- .

Numero atomico e numero di massa sono due grandezze necessarie per quantificare il numero di particelle atomiche presenti in un determinato isotopo.

a) Numero atomico

Il **numero atomico Z** corrisponde al **numero di protoni (p⁺) presenti nel nucleo di un atomo**. Atomi di uno stesso elemento hanno il medesimo numero di protoni.

In un **atomo neutro**, il numero delle cariche positive (protoni) contenute nel nucleo è uguale al numero di elettroni che vi ruotano intorno.

b) Numero di massa

Il **numero di massa A** corrisponde, invece, alla somma tra protoni e neutroni presenti nel nucleo di un atomo (protoni e neutroni sono indicati, complessivamente, con il termine di **nucleoni**).

Ogni atomo è caratterizzato dal **numero dei neutroni n°** del suo nucleo. Il numero di neutroni n° si calcola applicando la seguente formula:

$$n^{\circ} = A - Z$$

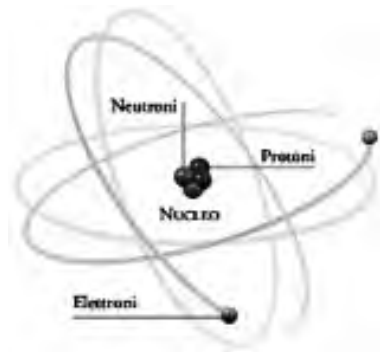


Fig. 1

in cui:

A = numero di massa

Z = numero atomico

Il *numero di massa* viene riportato in alto a sinistra del simbolo chimico dell'elemento, mentre il *numero atomico* viene evidenziato a sinistra, in basso (visto che il numero atomico Z è in corrispondenza biunivoca con il simbolo chimico dell'elemento può essere sottointeso e quindi omesso).

Per un generico atomo X , avente numero atomico Z e numero di massa A , scriveremo:

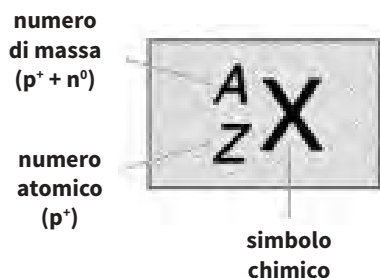


Fig. 2

2 Gli isotopi

Gli atomi di un elemento sono, dunque, caratterizzati dal numero atomico Z (indicato nella tavola periodica). Atomi che hanno lo stesso numero atomico (e quindi appartengono al medesimo elemento), ma differiscono per il numero di massa, prendono il nome di **isotopi**.

Questo termine deriva infatti da una parola greca che significa «stesso posto». Gli isotopi di un elemento chimico, infatti, hanno lo stesso numero atomico Z e quindi occupano identico posto nella tavola periodica. Inoltre, pur avendo masse diverse, non differiscono nel loro comportamento chimico, perché hanno lo stesso numero di protoni e quindi di elettroni.

a) Isotopi dell'idrogeno

L'idrogeno è una miscela di tre isotopi: *prozio*, *deuterio*, *trizio*. Essi hanno un protone ciascuno, ma differiscono per il numero di neutroni. Ilprozio non ha neutroni, il deuterio ne ha uno e il trizio ne ha due.



Ciascun elemento è costituito da una miscela di isotopi in percentuale sempre costante.

b) Fissione e fusione nucleare

La **fissione nucleare** consiste nella rottura di un atomo (o meglio del suo nucleo) in due parti, le cui masse sommate non raggiungono la massa originaria: la massa «mancante» si è trasformata in energia, secondo la relazione di Einstein $E = mc^2$.

Per innescare la fissione nucleare si deve disporre di materiale fissile, ovvero in grado di ottenerla. Due atomi adatti allo scopo sono l'uranio-235 e il plutonio-239. Poiché in natura l'uranio si presenta come la miscela di due isotopi (il 235 e il 238) e quello utile alla fissione (il 235, appunto) è il meno presente, occorre effettuare un'operazione chiamata «arricchimento», finalizzata all'aumento della percentuale di uranio. L'atomo (o meglio il suo nucleo) originario

viene bombardato da un neutrone e, in seguito all'impatto, si spezza in più parti, liberando neutroni che vanno a colpire altri atomi, proseguendo la reazione a catena e sprigionando energia.

La fissione è usata nella totalità delle centrali nucleari non sperimentali ed è anche il meccanismo sfruttato dalle due bombe atomiche che furono sganciate dagli USA sul Giappone alla fine della seconda guerra mondiale (per la precisione, a Hiroshima fu utilizzata una bomba all'uranio; a Nagasaki, invece, una al plutonio), con la differenza che in un ordigno bellico la reazione a catena è *incontrollata*, mentre in una centrale è (o, meglio, *dovrebbe* essere) un fenomeno *controllato*.

I prodotti di tale tipo di reazione sono atomi instabili, ovvero radioattivi, i quali cercano di raggiungere la stabilità e per farlo emettono radiazioni dannose per gli organismi biologici.

La **fusione nucleare** è, al contrario, il tipo di reazione che avviene in tutte le stelle. Due nuclei atomici leggeri (ad esempio, il deuterio e il trizio) si fondono e ne formano un terzo, la cui massa è minore della somma dei due di partenza. Anche questa volta la «massa mancante» si è tramutata in energia. Questo meccanismo è molto interessante poiché consentirebbe di produrre grandi quantità di energia in modo più sicuro ed ecologico delle attuali centrali nucleari (non ci sarebbero scorie radioattive), tuttavia è stato riprodotto dall'uomo solo a livello sperimentale, poiché ci sono alcune grosse difficoltà tecniche particolarmente difficili da superare. Ad esempio, le reazioni di fusione avvengono a temperature elevatissime, che rendono inadeguati i materiali a nostra disposizione e costringono all'uso di forti campi magnetici, i quali, per essere tenuti attivi, consumano grosse quantità di energia, anche maggiori di quella prodotta dalla fusione stessa. Per questo motivo è allo studio la cosiddetta «*fusione fredda*», ovvero una reazione di fusione a temperature più gestibili dall'uomo con le tecnologie a disposizione.

L'essere umano, comunque, è riuscito a realizzare bombe che sfruttano il meccanismo di fusione nucleare (le *bombe H* come la bomba Zar, il più potente ordigno atomico mai realizzato), le quali sono decisamente più devastanti di quelle che distrussero Hiroshima e Nagasaki.

c) Classificazione delle particelle

Attualmente possiamo classificare gli atomi in base alle loro proprietà chimiche.

Abbiamo visto che l'atomo è composto da un nucleo denso di carica positiva, costituito da protoni e da neutroni, e da una nube di elettroni di carica negativa.

I protoni e i neutroni sono a loro volta formati da particelle più elementari, chiamate **quark**. Per dare un'idea delle loro dimensioni possiamo immaginare che se i quark avessero le dimensioni di un capello, i protoni e i neutroni misurerebbero 1 centimetro e l'atomo sarebbe grande quanto 30 campi di calcio!

Ovviamente, in futuro, si potrebbe scoprire che i quark non sono le particelle più elementari, ma sono, a loro volta, composti da altre particelle ancora più piccole!

3 Il Modello Standard

Attualmente i fisici hanno scoperto circa 200 particelle (di cui la maggior parte *non fondamentali*) e hanno elaborato una teoria, denominata **Modello Standard**, che vuole descrivere tutta la materia e le forze dell'Universo (esclusa la gravità). La sua peculiarità sta nella capacità di «spiegare» l'enorme complessità di ciò che esiste attraverso un numero limitato di concetti fondamentali.

Il Modello Standard stabilisce che esistono 3 grandi categorie di particelle:

- i *leptoni*;
- gli *adroni*;
- le *particelle di campo*.

I **leptoni** non hanno una struttura interna e interagiscono soltanto tramite forze deboli elettromagnetiche; tra essi troviamo 3 particelle cariche (*elettrone*, *muone* e *tauone*) e tre particelle non cariche (i *neutrini* corrispondenti).

Gli **adroni** sono particelle composite: ciascuno di essi è formato da una combinazione di **quark**.

Attualmente esistono 6 tipi di quark, raggruppati in tre coppie: Up/Down, Charm/Strange e Top/Bottom.

A differenza degli elettroni che hanno carica -1 e dei protoni che hanno carica $+1$, i quark hanno *carica elettrica frazionaria*, di $2/3$ o $-1/3$, e sono inoltre dotati di un altro tipo di carica, chiamata *carica di colore*.

Gli adroni costituiti da 2 quark (e/o antiquark) sono detti *mesoni*, quelli costituiti da 3 quark, invece, *barioni*. Le particelle presenti nel nucleo sono dei barioni.

In particolare, un protone è costituito da 2 quark Up e da 1 Down, mentre un neutrone è composto da 2 quark Down e da 1 Up. L'unico barione stabile nel tempo è il protone, con una vita media superiore a 10^{33} anni. Il neutrone è stabile soltanto quando è confinato nel nucleo atomico, altrimenti ha una vita media di qualche centinaio di secondi, prima di decadere.

I mesoni sono tutte particelle instabili e decadono in altri mesoni o in altre famiglie di particelle.

Le **particelle di campo**, dette anche **bosoni**, sono costituenti essenziali che hanno il compito di trasferire il campo associato alle 4 interazioni fondamentali: *gravitazionale*, *elettromagnetica*, *forte* e *debole*.

Più in generale, tutta la materia dell'Universo è composta da particelle basilari, dette **fermioni**, le quali interagiscono tra loro attraverso le 4 forze fondamentali, scambiandosi «particelle di campo», appunto i *bosoni*.

a) Interazioni fondamentali

L'Universo esiste perché le particelle fondamentali interagiscono, ossia decadono, si «annichilano», reagiscono a forze legate alla presenza di altre particelle (per esempio nelle collisioni).

Per risalire alla vera natura delle forze è stato determinante un attento studio delle interazioni tra le particelle.

Si è capito che tali interazioni sono dovute ad uno scambio di **mediatori di forza**. Quelle che noi chiamiamo comunemente «forze» sono gli effetti dei mediatori di forza sulle particelle.

Esistono, come già accennato, quattro interazioni fondamentali:

- *gravitazionale*;
- *elettromagnetica*;
- *forte*;
- *debole*.

• Interazione gravitazionale

È probabilmente quella che ci è più familiare, ma non è compresa nel Modello Standard perché i suoi effetti sono minimi nei processi tra le particelle. Anche se non hanno ancora scoperto la particella mediatrice di forza per la gravità, i fisici ne hanno previsto l'esistenza e l'hanno già battezzata «**gravitone**».

• Interazione elettromagnetica

È quella che sperimentiamo di più, senza forse saperlo.

Per esempio, la forza di reazione del pavimento è in realtà dovuta alle interazioni elettromagnetiche nella materia, che si oppongono allo spostamento degli atomi dalla loro posizione di

equilibrio in un materiale. Anche l'attrito di un corpo che scivola su una superficie o l'attrazione di due cariche elettriche sono manifestazioni dell'interazione elettromagnetica.

La particella mediatrice di quest'ultima è il **fotone**. In base alla loro energia, i fotoni sono distinti in *raggi gamma*, *luce*, *microonde*, *onde radio* etc.

Se indaghiamo la materia più a fondo, a livello molecolare e atomico, scopriamo altri due tipi di interazioni che sono fondamentali per l'esistenza dell'Universo stesso, così come lo conosciamo oggi. Esse agiscono nell'infinitamente piccolo e nell'infinitamente grande, e sono l'*interazione forte* e l'*interazione debole*.

• Interazione forte

Gli atomi, di solito, hanno un numero uguale di protoni ed elettroni; quindi sono elettricamente neutri, dato che i protoni positivi bilanciano gli elettroni negativi. Ma che cosa tiene insieme gli atomi, per formare le molecole, se la maggior parte di essi non ha carica elettrica?

È stato stabilito che i quark hanno una carica, la quale non è di tipo elettromagnetico, chiamata **carica di colore**. Tra le particelle dotate di carica di colore l'interazione è molto intensa, tanto da meritarsi il nome di **interazione forte**. Dato che questa interazione tiene insieme i quark per formare gli adroni, la sua particella mediatrice è stata chiamata **gluone**, in quanto è particolarmente brava a «incollare» (in inglese «glue») i quark l'uno all'altro!

È importante precisare che ad avere carica di colore sono solo i quark e i gluoni. Gli adroni (tra cui i protoni e i neutroni) sono «neutri di colore», e così pure i leptoni.

Il problema resta aperto: cosa tiene insieme il nucleo se l'interazione forte agisce soltanto sui quark?

Non dimentichiamo che gli adroni sono composti da quark dotati di carica di colore: perciò i quark di un protone possono legarsi con i quark di un altro protone, anche se i protoni, di per sé, non hanno carica di colore. Questa viene chiamata **forza residua** ed è sufficientemente intensa da superare la repulsione elettromagnetica tra i protoni.

• Interazione debole

Ci sono 6 tipi di quark e 6 tipi di leptoni. Ma tutta la materia stabile dell'Universo è composta dai 2 tipi meno «massivi» di quark (*Up* e *Down*) e dal leptone carico meno massivo (elettrone): perché?

È tutta colpa dell'**interazione debole**, la quale è responsabile del fatto che tutti i quark e tutti i leptoni di massa maggiore decadono per produrre quark più leggeri ed elettroni. Una particella che «decade» sparisce e al suo posto appaiono due o più particelle. La somma delle masse delle particelle prodotte è sempre inferiore alla massa della particella di partenza. Questo è il motivo per cui la materia stabile che ci circonda contiene solo elettroni e i quark più leggeri.

Quando un quark o un leptone si «trasforma» (ad esempio un muone diventa un elettrone), si dice che cambia **sapore**. Tutti i mutamenti di sapore sono dovuti all'interazione debole.

Le particelle mediatrici dell'interazione debole sono i **bosoni W^+ , W^- , Z** . Le particelle W hanno carica elettrica, mentre Z è neutra.

b) Antimateria

Per ciascuna delle particelle di *materia* c'è una corrispondente particella di **antimateria**: in altre parole, per ogni *particella* c'è la corrispondente *antiparticella*.

Un'antiparticella è identica alla sua particella sotto qualunque aspetto, tranne che per la carica, che è opposta. Per esempio, il *protone* ha carica elettrica positiva, mentre l'*antiprotone* ha carica elettrica negativa; ma hanno identica massa, perciò sono soggetti alla gravità nella stessa maniera.

Quando una particella e la sua antiparticella si incontrano, si annichilano in energia pura. Questa energia può dar vita a particelle, prive di carica e mediatrici di forza, come fotoni, bosoni Z o gluoni.

Il simbolo di un'antiparticella è un trattino (barra) sopra il simbolo della corrispondente particella materiale. L'antiparticella di un protone è, come detto, l'**antiprotone**. L'antiparticella di un elettrone (e^-) è il **positrone** (e^+).

c) Decadimento radioattivo

Tra le particelle del nucleo, cioè tra i nucleoni, agisce una forza attrattiva molto grande, chiamata *nucleare forte*, che prevale nettamente sulle forze elettriche repulsive e consente al nucleo di non disintegrarsi.

Però, in Natura, esistono anche degli isotopi che non sono stabili e che quindi emettono spontaneamente delle particelle, trasformandosi nel nucleo di altri elementi. Questo fenomeno è denominato **decadimento radioattivo**, mentre il processo di radiazione prende il nome di **radioattività**.

La radioattività venne scoperta del tutto casualmente dal fisico francese Becquerel; qualche anno dopo, il neozelandese Rutherford dimostrò che la radiazione emessa poteva essere di tre tipi (*alfa*, *beta* o *gamma*), a seconda se la radiazione stessa:

- veniva respinta da un elettrodo carico positivamente (*raggi alfa*);
- veniva attratta da un elettrodo carico positivamente (*raggi beta*);
- proseguiva inalterata indipendentemente dalla carica dell'elettrodo (*raggi gamma*).

Egli dimostrò che:

- i *raggi α* sono costituiti da nuclei di elio, aventi carica $2+$ e numero di massa 4. Il decadimento α è tipico dei nuclei che hanno un numero elevato sia di protoni sia di neutroni, ossia con $Z > 83$ e $A > 220$. In questo decadimento, il numero atomico del nucleo di partenza diminuisce di due unità e il numero di massa di quattro unità;
- i *raggi β* sono fasci di elettroni veloci che non hanno massa rilevante e che portano una carica elettrica negativa. Il decadimento β è caratteristico dei nuclei troppo ricchi di neutroni rispetto ai protoni; tali nuclei devono aumentare il numero di protoni per stabilizzarsi, in modo da ripristinare l'equilibrio tra i nucleoni. Il numero atomico del nucleo che si forma è superiore di una unità rispetto al nucleo di partenza, mentre rimane inalterato il numero di massa;
- i *raggi γ* sono radiazioni elettromagnetiche, come la luce e i raggi X, ma di frequenza ed energia ancora maggiore. Le radiazioni γ consistono in una sorta di pacchetti di energia liberati dal nucleo dopo un'emissione α o β . In questo caso resta invariato sia il numero atomico sia il numero di massa.

Non esiste alcuna legge fisica che sia in grado di prevedere l'istante in cui stia per avvenire un decadimento radioattivo, però tutti i processi di decadimento hanno uno stesso andamento e possono essere schematizzati da una curva, denominata, appunto, *curva di decadimento*, lungo la quale si può osservare che a intervalli di tempo uguali corrisponde sempre il medesimo decadimento percentuale di sostanza radioattiva. Per sottolineare questo fenomeno, ogni isotopo è caratterizzato da un diverso *tempo di dimezzamento*, dove con questo termine si indica il tempo necessario per ridurre della metà la quantità di isotopo radioattivo.

III

Scienze biologiche

Sommario Sezione III

1 La biologia come scienza	12 Ereditarietà e ambiente
2 La chimica dei viventi	13 La classificazione dei viventi
3 I costituenti fondamentali della materia vivente	14 I cinque regni
4 La cellula come base della vita	15 I tessuti animali
5 Scambi di materia ed energia tra le cellule e al loro interno	16 Il movimento
6 La bioenergetica	17 Digestione e respirazione
7 La riproduzione cellulare	18 Circolazione sanguigna ed escrezione
8 Genetica mendeliana e genetica classica	19 I sistemi nervoso ed endocrino
9 La genetica molecolare	20 Riproduzione e sviluppo
10 La regolazione dell'espressione genica	21 Difesa dell'organismo, igiene e microbiologia
11 Le nuove frontiere della genetica	22 La biosfera

La biologia come scienza

1 La biologia: scienza della vita

La **Biologia** è, per definizione, la scienza che studia *la vita e gli organismi viventi*; ma cos'è la vita? Secondo il fisiologo francese *C. Bernard* (1813-1878): *la vita si può caratterizzare, ma non definire*.

Le caratteristiche e le attività fondamentali di tutti gli organismi viventi sono:

- struttura microscopica e composizione chimica;
- organizzazione;
- metabolismo;
- osmosi reattiva;
- sviluppo e ciclo vitale;
- riproduzione;
- ereditarietà;
- evoluzione.

In sintesi, corpi dotati di vita sono tutte quelle forme organizzate, capaci di reagire agli stimoli ambientali, riprodursi ed evolversi nel tempo in forme via via più complesse. La materia vivente è organizzata in **livelli di complessità** sempre maggiore, il primo dei quali è l'atomo, la più piccola particella di un elemento che ancora ne conserva le proprietà fisiche e chimiche.

Esaminiamo, dunque, i vari livelli: **atomo** → **molecola** (associazione di atomi uguali o diversi) → **organuli cellulari** (la più piccola struttura adibita a una specifica funzione) → **cellula** (la più piccola unità vivente formata da numerosi organuli) → **tessuto** (associazione di cellule simili che compiono una funzione specifica) → **organo** (associazione di tessuti diversi, organizzati in un'unità strutturale e funzionale) → **sistema di organi o apparato** (insieme di più organi che assolvono una specifica funzione del corpo) → **organismo** (insieme di apparati la cui interfunzionalità si concretizza in un essere vivente) → **specie** (organismi simili, interfecondi con prole fertile) → **popolazione** (gruppo di organismi della stessa specie che vivono nella medesima area) → **comunità** (l'insieme di più popolazioni che lavorano ed interagiscono nella stessa area) → **ecosistema** (l'insieme di più comunità e dell'ambiente fisico in cui vivono) → **biosfera** (parte della Terra abitata da organismi viventi, comprende viventi e non).

2 La componente biotica e abiotica della Terra

La Terra, l'unico pianeta del Sistema Solare ricoperto da grandi masse d'acqua liquida, è formato da una componente vivente ed una non vivente.

La **componente vivente o biotica** della Terra, definita **Biosfera**, è costituita dagli *organismi* microscopici fino a quelli più complessi, nonché da una sottile pellicola di circa 15 km di spessore, che comprende le *acque oceaniche e continentali*, la *parte di atmosfera a contatto con la superficie terrestre*, la *parte del sottosuolo dove giungono le radici e avvengono le interazioni tra mondo vivente e non*.

La **componente non vivente o abiotica** del nostro pianeta si può separare in tre sfere: la *litosfera*, l'*idrosfera* e l'*atmosfera*.

La **litosfera** o «sfera delle rocce» è la superficie solida che comprende i suoli e le rocce.

L'**idrosfera** o «sfera dell'acqua» comprende i mari, le acque continentali e l'umidità dell'aria.

L'**atmosfera** o «sfera dei gas» comprende la miscela dei gas che formano il sottile strato che avvolge la Terra.

Tra le diverse «sfere» c'è un continuo scambio di energia e materia, per cui non è possibile separare nettamente la componente vivente da quella non vivente, in quanto solo l'interazione di una con l'altra permette lo svilupparsi di quel meraviglioso fenomeno che è la **vita**.

3 Il metodo sperimentale

Sebbene le conoscenze scientifiche richiedano l'osservazione diretta di un fenomeno e la sperimentazione, per secoli la scienza è stata vincolata al *principio di autorità*, secondo il quale «una nozione è degna di fede solo se proveniente da un'autorità riconosciuta». Nelle scienze naturali, tale autorità era *Aristotele*, il quale abborriva la manualità, da lui considerata attività di esclusiva pertinenza degli schiavi.

Il primo a introdurre, sebbene ostacolato dal pensiero del tempo, un metodo basato sull'osservazione e sperimentazione fu *G. Galilei* nel XVII secolo. Il pensiero galileiano è riassumibile in una sua celebre frase: «Ove i fatti, ivi l'autorità superflua» (*Saggiatore*, 1623).

Il **metodo scientifico o sperimentale** consta di 4 passaggi:

- osservazione della realtà;
- formulazione delle ipotesi;
- preparazione e realizzazione dell'esperimento;
- relazione e sintesi dei risultati, con la formulazione di una legge.

Solo quando una lunga serie di esperimenti conferma l'ipotesi, è possibile definire una *Teoria o Legge di valore generale*, applicabile ogni qualvolta si verificherà un fenomeno simile. È comunque importante tener conto dei limiti del metodo scientifico, in quanto le *conclusioni scientifiche sono sempre da considerarsi provvisorie*, suscettibili di revisione qualora nuove osservazioni e nuovi esperimenti le contraddicano.

4 La dottrina della biogenesi

Il concetto di *generazione spontanea o abiogenesi* introdotto da *Aristotele*, secondo il quale la materia vivente si originerebbe spontaneamente dalla materia non vivente, fu per secoli considerato vero. Nel XVII secolo *F. Redi* (1626-1697) dimostrò che larve di insetti non si sviluppavano su pezzi di carne avariata posti in barattoli coperti, bensì solo su carne lasciata in barattoli aperti. Tale esperienza, seppure di una enorme semplicità, dimostrava inequivocabilmente che le larve di insetti non provenivano dalla carne, bensì dalle uova che le mosche depositavano su di essa.

Redi, purtroppo, non riuscì a portare a termine i suoi esperimenti, i quali furono continuati da *L. Spallanzani* (1729-1799) e conclusi brillantemente dal chimico e biologo francese *L. Pasteur* (1822-1895). Questi dimostrò inconfutabilmente la veridicità del **principio della biogenesi**, in base al quale *omne vivum ex vivo*: «ogni vivente proviene da ciò che già vive». Tale principio permise di abbandonare per sempre la teoria della generazione spontanea.

1 Gli atomi

Tutta la materia che ci circonda è costituita internamente da atomi diversamente combinati tra loro. Secondo le prime ipotesi «atomiste», l'atomo rappresentava la più piccola parte di materia che potesse essere isolata e, infatti, il termine «atomo» deriva dal greco *atomos*, che significa «indivisibile».

In realtà l'atomo stesso è costituito al suo interno da **particelle subatomiche** ancora più piccole e dovendone dare una definizione corretta si può affermare che **l'atomo è la più piccola parte di materia che mantiene le proprietà caratteristiche di un certo elemento chimico**.

Al suo interno l'atomo è organizzato in un **nucleo** centrale, costituito da **protoni** (particelle di carica elementare positiva) e **neutroni** (particelle prive di carica, aventi massa molto simile ai protoni). Attorno al nucleo si trovano gli **elettroni** (particelle di carica elementare negativa e di massa trascurabile rispetto a quella del protone), i quali, muovendosi attorno ad esso, generano una nube di carica elettrica negativa. Nell'atomo neutro il numero dei protoni del nucleo è uguale al numero degli elettroni e la carica totale è nulla.

Le particelle subatomiche (protoni, neutroni ed elettroni) sono uguali in tutti gli **elementi chimici**: ciò che rende differente un elemento dall'altro è il numero di protoni, neutroni ed elettroni che lo costituiscono.

Ogni elemento chimico è indicato con un **simbolo** (costituito da una o due lettere, di cui solo la prima maiuscola) ed è distinto da un **numero atomico** (indicato con Z), che rappresenta il numero dei protoni del nucleo, coincidente, nell'atomo «neutro», con il numero degli elettroni.

Tuttavia, scambiando modeste quantità di energia con l'esterno, l'atomo può perdere o acquistare elettroni, generando **ioni** che saranno, rispettivamente, di carica positiva o negativa. Ciò non cambia l'identità chimica dell'elemento, ma soltanto la sua carica totale: ad esempio, quando un atomo di sodio (Na) perde un elettrone, si genera uno ione sodio, indicato come Na^+ . Al contrario, il nucleo atomico non partecipa alle reazioni chimiche che avvengono ordinariamente e il numero dei protoni che lo costituiscono resta sempre immutato (fatta eccezione per le reazioni nucleari che coinvolgono quantità di energia enormi e non avvengono in condizioni ordinarie).

La maggior parte degli elementi chimici esiste in natura in più forme, le quali differiscono tra loro per il numero di neutroni che accompagnano i protoni nel nucleo: questi elementi mostrano tutti lo stesso comportamento chimico, ma al variare del numero dei neutroni possiedono masse diverse. Atomi dotati dello stesso numero atomico e di massa diversa sono detti **isotopi** e vengono distinti l'uno dall'altro attraverso il **numero di massa**, il quale esprime la somma del numero dei protoni e del numero dei neutroni del nucleo, e viene indicato come apice a sinistra del simbolo dell'elemento.

2 La struttura elettronica degli elementi

Lo studio delle proprietà chimiche e fisiche degli elementi ha permesso di comprendere la struttura interna degli atomi, essendo la reattività chimica di ogni elemento strettamente determinata dal tipo di **configurazione elettronica** che esso possiede. Con l'espressione «configurazione elettronica» si indicano il *numero*, la *disposizione* e l'*energia* degli elettroni che si trovano attorno al nucleo.

Ogni elettrone possiede un'energia totale dovuta in parte alla sua *energia cinetica*, ossia alla velocità con cui si muove, in parte alla sua *energia potenziale*, dipendente dalla posizione che occupa rispetto al nucleo atomico. Al pari di quanto accade ad un corpo nel campo gravitazionale, la cui energia potenziale è tanto maggiore quanto maggiore è la sua distanza dal centro della Terra, anche l'elettrone, immerso nel campo elettrico nucleare, possiede un'energia potenziale tanto maggiore quanto più grande è la sua distanza dal nucleo.

Secondo il modello attuale, gli elettroni possono assumere nell'atomo soltanto alcuni, ben definiti valori di energia, a cui corrispondono uno o più **orbitali atomici**, ossia porzioni di spazio di dimensioni e forma definite, all'interno delle quali si trova l'elettrone stesso.

Nell'atomo esistono quindi veri e propri **livelli energetici**, spesso rappresentati come gusci sferici centrati nel nucleo, l'energia dei quali, in accordo con l'andamento dell'energia potenziale, cresce man mano che ci si allontana dal nucleo. Ciascuno di questi livelli energetici può essere occupato da un determinato numero di elettroni: in particolare, secondo uno schema semplificato, il primo livello può essere occupato da due elettroni, il secondo ed il terzo da otto.

Per stabilire il tipo di configurazione elettronica posseduta da un atomo è sufficiente tener presente che gli elettroni tendono ad occupare, quando possibile, gli orbitali di energia più bassa; solo quando questi sono completi, occupano orbitali di energia superiore. Quindi la configurazione elettronica che ciascun elemento assume spontaneamente è quella a cui corrisponde la minima energia possibile e viene indicata come *configurazione dello stato fondamentale*.

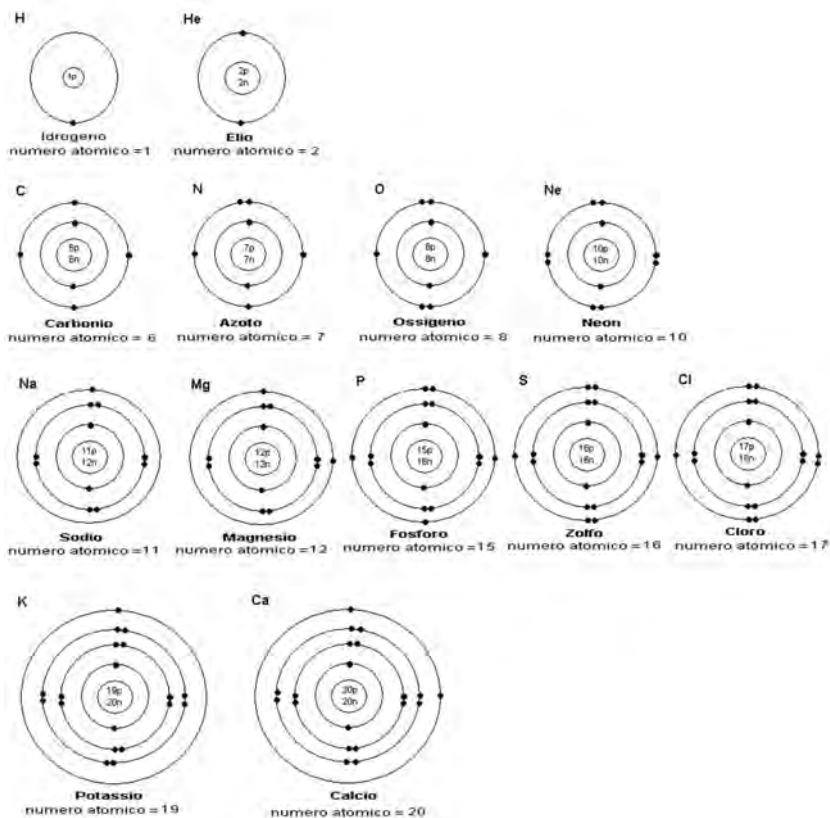


Fig. 1 - Configurazione elettronica di alcuni dei principali elementi chimici

Confrontando l'energia degli stati fondamentali di diversi elementi ci si è resi conto dell'esistenza di alcune configurazioni elettroniche di notevole stabilità, ovvero configurazioni a cui corrisponde un valore di energia totale particolarmente basso. Queste corrispondono a stati in cui si è raggiunta la completezza dell'ultimo livello energetico occupato dagli elettroni.

Solo pochi atomi possiedono questo tipo di configurazione: si tratta dei *gas nobili* come l'*elio* (He), il *neon* (Ne), l'*argon* (Ar). Tutti gli altri elementi tendono a raggiungere la completezza del livello esterno cedendo, acquistando o mettendo in comune elettroni con altri elementi, ossia formando **legami chimici**. Poiché il livello elettronico esterno, almeno per quanto concerne il secondo e il terzo livello energetico, si completa con otto elettroni, quanto detto si traduce nella nota **regola dell'ottetto**, secondo la quale tutti gli elementi tendono a completare l'ottetto esterno. Pertanto, gli atomi tendono a formare tanti legami chimici quanti sono gli elettroni necessari a raggiungere l'ottetto.

La formazione di un legame chimico tra due o più elementi dà origine a una **molecola**. Tutte le sostanze in cui gli atomi che costituiscono la molecola sono uguali, cioè appartengono allo stesso elemento chimico, vengono dette **elementi**; in caso contrario, quando gli atomi legati a formare la molecola appartengono ad elementi chimici diversi, la sostanza viene definita **composto**.

3 Il legame chimico

I tipi di legame chimico che possono instaurarsi tra gli atomi sono divisibili in due categorie principali: **legami elettrostatici**, derivanti dall'attrazione tra cariche elettriche di segno opposto generatesi negli atomi, e **legami covalenti**, derivanti dalla condivisione di elettroni tra due nuclei atomici.

Il **legame ionico** rientra nella famiglia dei *legami elettrostatici* e, come dice il nome stesso, si instaura tra *ioni*, cioè atomi che hanno acquisito una carica elettrica in seguito alla cessione o all'acquisto di uno o più elettroni. Esistono elementi per i quali il raggiungimento dell'ottetto può realizzarsi semplicemente perdendo o acquistando un elettrone. È questo il caso del sodio (Na) o del potassio (K), che, perdendo l'unico elettrone esterno, si trasformano rispettivamente in Na^+ e K^+ , le cui configurazioni elettroniche sono coincidenti rispettivamente con quella del Ne e dell'Ar. In modo analogo, l'atomo di fluoro (F) o quello di cloro (Cl) raggiungono la completezza del livello esterno semplicemente acquistando un elettrone: gli ioni F^- e Cl^- così generati possiedono, rispettivamente, la configurazione elettronica esterna del Ne e dell'Ar. È quindi plausibile aspettarsi la formazione di legami ionici tra elementi come il sodio e il potassio (metalli alcalini) ed elementi come fluoro o cloro (alogeni); il trasferimento di un elettrone dal metallo all'alogeno consente ad entrambi il raggiungimento dell'ottetto esterno e i due ioni formati restano legati per attrazione elettrostatica. In realtà il bilancio energetico del processo è tale che la formazione della singola molecola ionica è sfavorita, mentre si assiste alla formazione del cristallo ionico, in cui un gran numero di ioni sono impacchettati secondo una ben definita disposizione spaziale.

Il **legame covalente**, invece, deriva dalla messa in comune di elettroni tra due atomi dotati di livelli elettronici esterni incompleti. Ciascuno dei due atomi legati mette in comune con l'altro uno dei suoi elettroni; a tal fine i due atomi devono avvicinarsi fino a sovrapporre i propri orbitali, generando tra i nuclei una zona ad elevata densità di carica negativa, detta **zona di ricopertura**, che costituisce l'elemento legante tra i nuclei. Quando due atomi mettono in comune più di una coppia di elettroni si producono legami covalenti doppi come in O_2 ($\text{O}=\text{O}$), o tripli come in N_2 ($\text{N}\equiv\text{N}$). Il numero di legami covalenti che si stabiliscono tra due atomi viene indicato come **ordine di legame**.

Se i due atomi legati sono uguali, come accade in molecole quali H_2 , O_2 , N_2 , gli elettroni del legame covalente sono equamente condivisi tra i due nuclei e il legame è detto **covalente puro** o **omeopolare**.

Se invece i due atomi sono diversi, può generarsi un'asimmetria nella distribuzione della carica di legame a favore dell'elemento che esercita la maggiore forza di attrazione sugli elettroni di legame. L'addensamento della carica di legame su uno dei due atomi conferisce a questo una parziale carica negativa, mentre genera sull'altro una parziale carica positiva. Si forma così un **dipolo**, che attribuisce al legame covalente un parziale carattere elettrostatico, e il legame venutosi a formare viene chiamato **covalente misto o eteropolare**.

La capacità di attrarre a sé la carica degli elettroni di legame è una proprietà intrinseca dell'elemento, che si manifesta al momento della formazione di un legame covalente e viene indicata col nome di **elettronegatività**. Quanto più elevata è la differenza di elettronegatività tra due atomi legati, tanto più grande è la polarità del legame che si forma.

Oltre a quelli ionici e covalenti, che costituiscono i legami più forti esistenti in natura e legano gli atomi all'interno delle molecole (**intramolecolari**), esistono anche legami che si instaurano tra molecole (**intermolecolari**). Si tratta di legami più deboli, che vengono rotti molto facilmente, fornendo piccole quantità di energia, ma che rivestono un'enorme importanza in biologia. Alla base di queste interazioni ci sono le attrazioni tra i dipoli molecolari che portano le molecole ad attrarsi reciprocamente, in modo tale da avvicinare tra loro le parti dei dipoli dotate di carica opposta.

Il più importante di questa categoria di legami è il **legame idrogeno**, osservabile ogni volta che l'atomo di *idrogeno* è legato ad un atomo di elevata elettronegatività, come ad esempio l'*ossigeno*. Per le sue piccolissime dimensioni, l'atomo di idrogeno, se polarizzato, genera un campo elettrico molto elevato e tende ad attrarre a sé molecole polari, legandole con una discreta forza. L'energia del legame idrogeno è inferiore a quella di un legame ionico o covalente, ma superiore a quella di una comune interazione dipolare.

Ad ogni legame chimico è associata un'**energia di legame** che può essere definita come l'energia necessaria a sciogliere il legame stesso, portando i due atomi a distanza infinita tra loro, oppure, al contrario, come l'energia liberata in seguito alla formazione del legame. Confrontando i dati della tabella che segue, è evidente come l'entità dell'energia di legame dipenda dalla natura degli atomi legati e dall'ordine di legame: è ovvio, infatti, che i *legami multipli* abbiano energia superiore a quelli *semplici*. Si osservi come l'energia del legame semplice C - H sia decisamente superiore a quella dei legami C - O o C - C.

LEGAME	C = O	C = C	C - H	C - O	C - C
ENERGIA DI LEGAME (KCAL/MOL)	171	147	99	84	83

Tab. 1 - Energie di legame a confronto

4 Le trasformazioni chimiche

Viene definita **reazione chimica** una trasformazione che comporta un cambiamento di composizione della materia, lasciandone invariata la massa (*principio di conservazione della massa*). Infatti la trasformazione delle specie di partenza (**reagenti**) nei **prodotti di reazione** comporta la scissione di alcuni o di tutti i legami chimici presenti nelle molecole dei reagenti e la formazione di nuovi legami chimici nei prodotti, in cui gli atomi che costituivano i reagenti sono stati organizzati in nuove strutture. Se l'energia spesa per la demolizione dei reagenti è superiore a quella liberata dalla formazione dei legami nei prodotti, la reazione avviene con *assorbimento di energia dall'esterno* e viene detta **endoergonica**; viceversa, se l'energia liberata dalla formazione dei prodotti è superiore a quella spesa, la reazione *libera una certa quantità di energia all'esterno* e, quindi, viene definita **esoergonica**.

Dovrebbe essere chiaro, a questo punto, che all'interno delle molecole è custodita, sotto forma di legame chimico, una discreta quantità di energia, a cui gli esseri viventi possono attingere per soddisfare le proprie esigenze energetiche. Questo è quanto accade nella *respirazione cellulare*, nella quale, attraverso la combustione del glucosio, viene liberata energia (686 kcal/mol) utilizzabile dagli organismi. Allo stesso modo, ogni qualvolta bruciamo un *combustibile* per una delle nostre attività, sfruttiamo l'energia immagazzinata nei legami chimici, liberata in seguito alla reazione con l'ossigeno. Viceversa, tramite la sintesi di materia organica, l'energia può essere immagazzinata e conservata all'interno degli organismi viventi: questo è quanto accade nelle piante, che, mediante il *processo di fotosintesi*, assorbono l'energia solare e la utilizzano per sintetizzare molecole organiche a partire da sostanze inorganiche. Tale processo costituisce il mezzo attraverso il quale l'energia solare si trasforma in energia utilizzabile da tutti gli esseri viventi.

5 Le trasformazioni dell'energia: primo e secondo principio della termodinamica

L'**energia** può essere definita come l'*attitudine di un sistema a compiere lavoro*. Essa può manifestarsi in forme diverse, ossia come energia *termica, meccanica, chimica, luminosa, elettrica, magnetica* etc., ognuna delle quali possiede una differente attitudine ad essere trasformata in lavoro: in particolare, si è soliti distinguere le **energie di prima specie**, come quella *meccanica* o quella *elettrica*, convertibili in lavoro con rendimenti elevati, dall'*energia termica*, detta **energia di seconda specie**, che invece può essere convertita in lavoro con rendimenti molto bassi. Per questo motivo la trasformazione di energia di prima specie in energia di seconda specie viene indicata come **degradazione dell'energia**.

La disciplina che studia le trasformazioni dell'energia e le leggi che le governano è la **termodinamica**.

Il **primo principio della termodinamica** afferma:

«L'energia non può essere creata né distrutta, ma solo convertita da una forma all'altra».

In virtù di questo principio, quando un sistema isolato (tale, cioè, da non scambiare materia né energia con il suo ambiente) subisce una trasformazione, la sua energia totale resta invariata, mentre cambia la ripartizione dell'energia tra le diverse forme. Tuttavia, in natura si osserva che non tutte le trasformazioni dell'energia che rispettano il primo principio avvengono spontaneamente: in particolare, mentre tutte le energie di prima specie possono trasformarsi spontaneamente in calore, quest'ultimo non può convertirsi spontaneamente in un'altra forma di energia.

In pratica, in tutti i processi spontanei i sistemi tendono ad evolvere verso uno stato di maggiore «disordine» e l'energia, che nel corso di qualsiasi trasformazione non viene creata né distrutta, ma solo convertita da una forma all'altra, tende spontaneamente a portarsi in una forma più dispersa e degradata. Questo è quanto viene sancito dal *secondo principio della termodinamica*, che denuncia l'impossibilità di trasformare completamente il calore in lavoro.

La grandezza utilizzata per descrivere l'entità del «disordine» di un sistema è l'**entropia** ed il **secondo principio della termodinamica**, espresso in termini di entropia, può essere così enunciato:

«In un sistema isolato tutti i processi spontanei portano ad un aumento di entropia».

Se identifichiamo con l'*energia potenziale* la quantità di energia trasformabile in *lavoro*, una conseguenza del secondo principio della termodinamica è che i sistemi isolati tendono spontaneamente ad evolvere verso stati dotati di un minore contenuto di energia potenziale. Tutte le trasformazioni chimiche che passano da reagenti a più alto contenuto di energia a prodotti con

minore contenuto di energia decorrono spontaneamente e liberano energia all'esterno; viceversa, reazioni che portano da reagenti a basso contenuto di energia a prodotti ad alto contenuto energetico non sono spontanee e richiedono l'apporto di energia dall'esterno.

6 La velocità delle reazioni chimiche

Un altro aspetto da tenere in grande considerazione, a proposito delle reazioni chimiche, è quello della **velocità**. Infatti può accadere che reazioni favorite dal punto di vista termodinamico siano estremamente lente, cosa incompatibile con le esigenze degli esseri viventi, i quali necessitano in modo immediato della disponibilità di alcune sostanze.

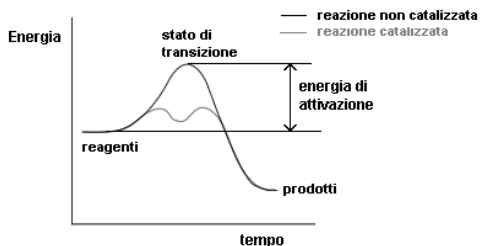


Fig. 2 - Reazione catalizzata e non catalizzata

transizione in cui alcuni legami sono in via di rottura, altri in via di formazione. Tale stato ha un contenuto di energia superiore rispetto ai reagenti: la differenza di energia esistente è detta **energia di attivazione**. Questa costituisce una sorta di barriera energetica che la reazione deve superare e rappresenta la minima energia che le molecole dei reagenti devono possedere per dare luogo alla reazione.

Esistono vari modi per aumentare la velocità di reazione, ad esempio l'innalzamento della temperatura: infatti, fornendo calore, quindi energia termica, aumenta la porzione di particelle dotate di energia sufficiente a superare la barriera rappresentata dallo stato di transizione. Anche la maggiore concentrazione delle specie reagenti può avere un'influenza sulla velocità, poiché, aumentando il numero totale di particelle, cresce la probabilità che queste si incontrino per dare luogo alla reazione.

Un metodo alternativo consiste nell'aggiungere al sistema di reazione una sostanza, detta **catalizzatore**, la quale non viene consumata né prodotta dalla reazione, ma possiede la capacità di modificarne il meccanismo, facendole attraversare uno o più stati di transizione di energia inferiore e determinando quindi un incremento della velocità. La figura 2 mostra il profilo energetico di una reazione chimica in presenza o in assenza di catalizzatore.

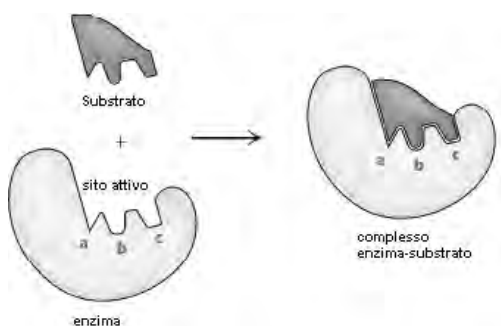


Fig. 3 - Meccanismo d'azione degli enzimi

Lo studio della velocità delle reazioni chimiche ha evidenziato come questa sia indipendente dallo stato iniziale e finale della reazione (ovvero dalla natura dei reagenti e dei prodotti), ma dipenda piuttosto dal meccanismo attraverso il quale la stessa avviene.

La trasformazione dei reagenti in prodotti si svolge passando attraverso uno **stato di**

Nei sistemi biologici, vincolati a condizioni di temperatura e di concentrazione delle sostanze variabili entro limiti molto ristretti, le reazioni avvengono in modo rapido grazie all'azione catalitica di sostanze specifiche, chiamate **enzimi**. Si tratta di molecole di natura proteica che possono assumere forme altamente specifiche a seconda della loro funzione. Il meccanismo d'azione degli enzimi è paragonabile al sistema chiave-serratura, schematizzato in figura 3.

L'enzima possiede un sito attivo la cui forma è esattamente complementare a quella del sub-

strato o dei substrati che devono reagire. Portando i reagenti a contatto tra loro ed indebolendo i legami interessati dalla reazione, l'enzima ne consente uno svolgimento rapido, in condizioni di temperatura compatibili con la vita delle cellule.

L'attività di molti enzimi è regolata da sostanze non proteiche, dette **cofattori**: lo ione magnesio, ad esempio, è il cofattore di reazioni enzimatiche che prevedono il trasferimento di un gruppo fosfato da una molecola a un'altra. Altri enzimi, invece, vengono attivati da molecole organiche, chiamate **coenzimi**, con cui si legano temporaneamente o in modo permanente. Molte vitamine sono coenzimi o parti di coenzimi, il che le rende elementi indispensabili al corretto svolgimento delle attività fisiologiche dell'organismo.

I costituenti fondamentali della materia vivente

1 La molecola d'acqua

a) Struttura e proprietà

L'acqua è estremamente diffusa sulla Terra, tanto da ricoprirne circa i tre quarti della superficie: non stupisce, quindi, che la vita sia cominciata nell'acqua e che questa costituisca tra il 50% e il 95% del peso di ogni organismo vivente.

La molecola d'acqua ha formula H_2O ed è costituita da due atomi di idrogeno uniti ad un atomo di ossigeno centrale per mezzo di legami covalenti semplici. Ciò è compatibile con le strutture elettroniche di questi elementi, poiché l'ossigeno ha sei elettroni sul livello esterno (v. Fig. 1, Cap. 2) e può formare due legami covalenti, mentre gli idrogeni completano il loro livello esterno con due elettroni e, dunque, possono formare un solo legame covalente. L'ossigeno è molto più elettronegativo dell'idrogeno, per cui il legame H-O è fortemente polare. Inoltre, poiché l'ossigeno è circondato da quattro coppie elettroniche, due di legame con gli idrogeni e due non condivise, i due legami H-O sono disposti a $104,45^\circ$ uno dall'altro. Tale geometria fa sì che i momenti di dipolo dei due legami si sommino vettorialmente, dando un momento di dipolo risultante elevato. La molecola d'acqua è, pertanto, una molecola altamente polare, capace di formare legami idrogeno: caratteristiche, queste, che le conferiscono proprietà del tutto particolari.

La **tensione superficiale** è una proprietà dei liquidi direttamente correlata all'entità delle interazioni esistenti tra le molecole del liquido medesimo (forze coesive). Nell'acqua queste forze sono abbastanza elevate — sia per la notevole polarità della molecola, sia per la capacità di formare legami idrogeno — e la loro presenza determina una certa resistenza della superficie liquida ad aumentare la propria estensione.

La **capillarità**, ossia la capacità dell'acqua di risalire all'interno di un capillare, è il risultato dell'azione combinata della coesione tra le molecole d'acqua (che tende a mantenerle unite) e dell'adesione delle molecole stesse alle pareti del capillare. Tale proprietà è sfruttata dalle piante per assorbire acqua dal terreno.

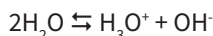
L'acqua è dotata anche di un'elevata **capacità termica** [1], la quale le conferisce una certa resistenza a cambiare la propria temperatura in seguito allo scambio di calore. Anche questa proprietà è correlata alla struttura molecolare, poiché i legami idrogeno, presenti tra molecole, costituiscono un vincolo all'aumento dell'energia cinetica media, a sua volta proporzionale alla temperatura. Si tratta di una proprietà estremamente importante per i viventi, perché contribuisce a mantenere stabile la loro temperatura corporea.

Infine, l'acqua ha un ottimo **potere solvente**, cioè possiede una buona capacità di sciogliere sostanze. Grazie alla sua polarità e alla possibilità di formare legami idrogeno, l'acqua scioglie facilmente le sostanze ioniche, solvatando efficacemente gli ioni sia positivi sia negativi, le specie polari (di cui talvolta può indurre la dissociazione in ioni) e le molecole in grado di formare legami idrogeno. Queste ottime capacità solventi fanno sì che in natura l'acqua non esista mai allo stato puro, ma porti sempre una certa concentrazione di soluti disciolti.

[1] Per *capacità termica* si intende la quantità di calore necessaria ad aumentare di un grado centigrado la temperatura di un corpo.

b) Autoprotolisi e pH

I legami covalenti fortemente polarizzati, come il legame -O-H, sono sensibili alle sollecitazioni esercitate da molecole polari circostanti che, attraendo l'idrogeno da un lato e l'ossigeno dall'altro, possono determinare la rottura del legame. Ciò è quanto accade a tutte le specie che si ionizzano in acqua, come gli acidi forti o deboli, e alle molecole dell'acqua stessa, le quali mostrano sempre un piccolo grado di dissociazione in ioni H_3O^+ e ioni OH^- . La reazione di *ionizzazione* o **autoprotolisi** dell'acqua può essere scritta come:



In tutte le soluzioni acquose il prodotto della concentrazione degli ioni H_3O^+ e OH^- è costante ed è pari a 10^{-14} (a 25°C):

$$K_w = [\text{H}_3\text{O}^+][\text{OH}^-] = 10^{-14}$$

In acqua pura le concentrazioni degli ioni H_3O^+ e OH^- sono molto piccole ed uguali tra loro (10^{-7} moli/litro).

L'acqua pura e tutte le soluzioni in cui le concentrazioni $[\text{H}_3\text{O}^+]$ sono uguali a $[\text{OH}^-]$ si dicono **neutre**. Le soluzioni in cui siano disciolte sostanze (acidi) che determinano un aumento della $[\text{H}_3\text{O}^+]$ si definiscono **acide**, mentre le soluzioni in cui siano disciolte sostanze (basi) che determinano un aumento della $[\text{OH}^-]$ si chiamano **basiche**. In soluzioni acide abbiamo $[\text{H}_3\text{O}^+] > 10^{-7}$; in soluzioni basiche, invece, $[\text{H}_3\text{O}^+] < 10^{-7}$. Spesso, trattandosi di numeri molto piccoli, rappresentati attraverso esponenziali negativi, le concentrazioni degli ioni H_3O^+ e OH^- vengono espresse utilizzando la **scala del pH**, definito come il logaritmo, cambiato di segno, della concentrazione degli ioni H_3O^+ di una soluzione:

$$\text{pH} = -\log [\text{H}_3\text{O}^+]$$

In base a quanto detto precedentemente, si ha:

$$\begin{aligned} \text{pH} < 7 &\rightarrow \text{soluzione acida} \\ \text{pH} = 7 &\rightarrow \text{soluzione neutra} \\ \text{pH} > 7 &\rightarrow \text{soluzione basica} \end{aligned}$$

2 Le molecole organiche presenti negli organismi viventi

Nonostante la loro straordinaria varietà, i composti organici che costituiscono gli esseri viventi derivano dalla combinazione di pochissimi elementi chimici: carbonio (C), idrogeno (H), azoto (N), ossigeno (O), fosforo (P), zolfo (S).

Tra questi, un ruolo centrale è svolto dall'atomo di carbonio che, grazie alla capacità di formare quattro legami, può generare catene più o meno lunghe, lineari, ramificate o cicliche, le quali formano lo scheletro di tutte le molecole organiche.

Le principali categorie di composti organici che troviamo negli esseri viventi sono:

- i **carboidrati**, composti costituiti da C, H ed O;
- i **lipidi**, costituiti anch'essi da C, H, O e, in alcuni casi, da N e P;
- le **proteine**, contenenti C, H, O, N e S;
- gli **acidi nucleici**, costituiti da C, H, O, N e P.

Tutte le *biomolecole* hanno in comune la caratteristica di essere molecole di grandi dimensioni, dette **polimeri**, ottenute dalla **reazione di condensazione** [2] di subunità più piccole, chiamate **monomeri**, dotate di strutture simili tra loro o addirittura identiche.

[2] In chimica organica è denominata «condensazione» la reazione che produce l'unione di due molecole per mezzo dell'eliminazione di una molecola d'acqua.

a) Carboidrati

I *carboidrati* sono composti organici che rispondono alla formula generale $C_n(H_2O)_n$. Gli *zuccheri semplici* o **monosaccaridi**, costituiti da molecole di piccole dimensioni, sono denominati, a seconda del valore di n , triosi, tetrosi, pentosi, esosi etc. Sotto il profilo chimico, i monosaccaridi sono polialcoli, in quanto ogni atomo di carbonio della catena è legato ad una funzione alcolica -OH, ad esclusione di uno, che può portare una funzione aldeidica -C(O)H (caso in cui si parla di aldotriosi, aldotetrosi, aldopentosi, aldosesi etc.) o una funzione chetonica -C(O)- (nel qual caso si parla, invece, di chetotriosi, chetotetrosi, chetopentosi etc.).

Tra questi i più comuni sono il *glucosio*, la principale fonte di energia delle nostre cellule, il *fruttosio*, zucchero presente nella frutta e nel miele, il *ribosio*, costituente degli acidi nucleici. I monosaccaridi hanno strutture lineari, anche se in soluzione acquosa tendono a portarsi in forma ciclica, generando anelli a cinque (furanosi) o a sei (piranosio) termini.

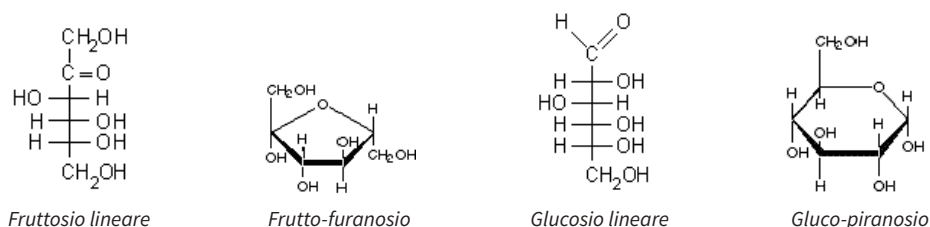


Fig. 1 - Strutture lineari e cicliche del fruttosio e del glucosio

Queste molecole semplici possono unirsi tra loro, eliminando molecole d'acqua e stabilendo legami a ponte di ossigeno (reazione di condensazione). Possono così originarsi molecole di piccole dimensioni come i **disaccaridi**, costituiti dalla condensazione di due molecole di monosaccaride, o molecole molto grandi come i **polisaccaridi**, generati dall'unione di un gran numero di monomeri.

Un *disaccaride* abbastanza diffuso è il *saccarosio* (il comune *zucchero da tavola*), risultato della condensazione di una molecola di glucosio e una di fruttosio, utilizzato dalle piante come zucchero di trasporto. Il disaccaride presente nel latte è il *lattosio*, risultato della condensazione di una molecola di glucosio e una di galattosio (zucchero analogo al glucosio, dal quale differisce solo per la posizione dei gruppi OH rispetto al piano dell'anello). Le molecole di monosaccaride possono essere liberate dalla molecola di disaccaride mediante **idrolisi**, reazione inversa rispetto alla *condensazione*, in cui viene riacquistata la molecola d'acqua eliminata e viene scisso il legame tra i due monosaccaridi.



Fig. 2 - Strutture cicliche del saccarosio e del lattosio

I **polisaccaridi**, infine, sono costituiti da lunghe catene di monosaccaridi. I più diffusi in natura sono l'*amido*, il *glicogeno* e la *cellulosa*, tutti e tre polimeri del glucosio, differenti l'uno dall'altro unicamente per la modalità di concatenamento dei monomeri, che tuttavia influenza notevolmente la geometria della molecola risultante e, di conseguenza, le sue proprietà e funzioni.

Quando si trova in forma ciclica il *glucosio* non ha una struttura perfettamente planare, ma assume la cosiddetta «conformazione a sedia», in cui i legami formati dagli atomi di carbonio possono essere distinti in *legami di tipo α* , più o meno perpendicolari al piano medio dell'anello, e *legami di tipo β* , giacenti sul piano medio dell'anello (Fig. 3).

L'*amido* è un polimero lineare i cui monomeri sono concatenati mediante legami di tipo α (1-4). Questo tipo di legame consente alla catena di ripiegarsi a spirale, formando i tipici granuli che costituiscono la riserva energetica nei vegetali.

Il *glicogeno* è caratterizzato, come l'amido, da legami di tipo α (1-4), ma possiede un gran numero di ramificazioni che si ripetono ogni 8-10 unità di glucosio. Questo polimero svolge la funzione di riserva energetica negli animali poiché la sua struttura, altamente ramificata e dotata di un elevato numero di terminali di catena, consente il rapido rilascio del glucosio nei momenti di notevole consumo energetico.

La *cellulosa*, infine, è un polimero lineare del glucosio, i cui monomeri sono concatenati tramite legami di tipo β (1-4). Questo tipo di legame consente, mediante la formazione di legami idrogeno, l'associazione delle catene di cellulosa in fasci detti *microfibrille*. La cellulosa, infatti, è un polimero strutturale, costituente fondamentale delle pareti delle cellule vegetali; i suoi legami non sono facilmente idrolizzabili e dunque essa non è utilizzabile come fonte energetica, se non da pochi microrganismi.

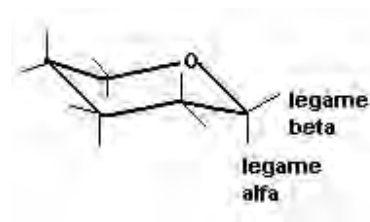


Fig. 3 - Legami di tipo α e β del glucosio

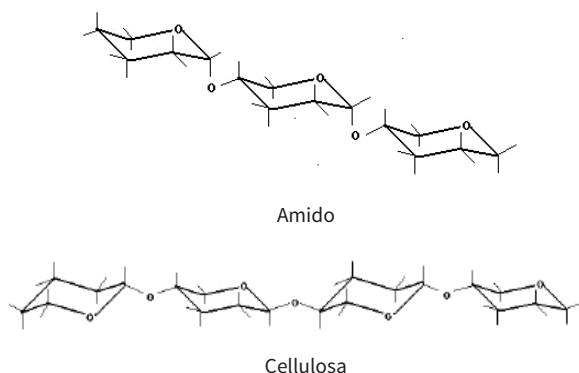


Fig. 4 - Legami dell'amido e della cellulosa

b) Lipidi

Conosciuti anche come **grassi**, i *lipidi* rappresentano un'ampia classe di composti organici, di natura prevalentemente idrocarburica, aventi in comune la caratteristica di essere apolari e quindi insolubili in acqua. Essendo costituiti essenzialmente da carbonio e idrogeno, possiedono legami covalenti di energia più alta rispetto ai carboidrati e, a parità di peso, immagazzinano una quantità di energia circa sei volte maggiore. Per questo i grassi rappresentano la principale forma di accumulo di energia nelle specie animali, per le quali l'aumento della massa corporea costituisce sicuramente un forte svantaggio.

Tra i lipidi figurano i **trigliceridi**, molecole derivanti dalla condensazione di una molecola di *glicerolo* con tre molecole di *acidi grassi*, ossia acidi organici caratterizzati da una lunga catena idrocarburica. Come illustrato in figura 5, il glicerolo possiede tre funzioni alcoliche (gruppi -OH)

che possono condensare con la funzione acida ($-\text{COOH}$) degli acidi grassi, generando una molecola di trigliceride.

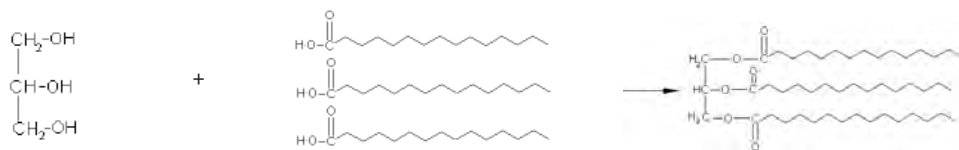


Fig. 5 - Costituzione di una molecola di trigliceride

Le proprietà fisiche dei trigliceridi sono fortemente influenzate dalle caratteristiche degli acidi grassi che li compongono, tra le quali sono di fondamentale importanza la *lunghezza della catena idrocarburica* e la *presenza di doppi legami (insaturazioni)*.

La *presenza di una o più insaturazioni* può provocare ripiegamenti nella catena idrocarburica, che impediscono uno stretto impacchettamento delle molecole. Al contrario, catene completamente sature (ossia contenenti tutti legami semplici) possono assumere assetti più compatti. Per questo motivo, a temperatura ambiente, i grassi **saturi** tendono ad essere solidi, mentre quelli **insaturi** tendono a rimanere liquidi.

Un altro fattore importante nel determinare lo stato fisico è la *lunghezza della catena idrocarburica*, poiché al crescere del numero di atomi di carbonio aumenta l'entità delle forze intermolecolari che favoriscono lo stato solido.

Ruoli strutturali di estrema importanza sono svolti dai **fosfolipidi**, composti simili ai trigliceridi in cui una delle molecole di acido grasso è sostituita da un *gruppo fosfato*, legato generalmente a un *gruppo polare R*. Essendo costituiti da una parte polare (*idrofila*) e una apolare (*idrofoba*), in soluzione acquosa i fosfolipidi tendono a disporsi sulla superficie dell'acqua, occupando l'interfase acqua-aria poiché le parti idrofile si trovano disciolte nell'acqua, mentre le catene idrofobe tendono ad uscirne. Se la concentrazione è sufficiente, si possono formare dei doppi strati secondo lo schema mostrato in figura 6, in cui tutte le «teste» idrofile sono disposte all'esterno, a contatto con l'acqua, mentre le «code» idrofobe sono confinate all'interno del doppio strato.

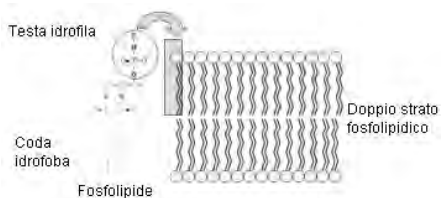


Fig. 6 - Formazione del doppio strato fosfolipidico in soluzione acquosa

Un comportamento analogo a quello dei fosfolipidi è proprio dei **glicolipidi**, molecole in cui uno dei carboni del glicerolo è legato ad una catena di natura carboidratica che costituisce una «testa polare» della molecola, mentre gli altri due carboni sono legati a molecole di acidi grassi. Anche questi composti svolgono un importante ruolo strutturale nei viventi, in qualità di *componenti delle membrane cellulari*.

Alla famiglia dei grassi appartengono anche le **cere**, sostanze idrofobe di composizione chimica complessa, che svolgono funzioni di protezione e impermeabilizzazione nelle piante e negli animali, e gli **steroidi**, che hanno in comune con gli altri grassi l'insolubilità in acqua, sebbene siano molto diversi strutturalmente.

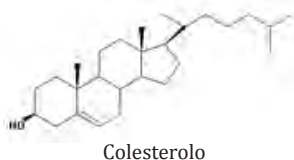


Fig. 7 - Struttura del colesterolo

In figura è mostrata la struttura del **colesterolo**, costituente delle membrane cellulari e precursore per la sintesi di molti ormoni, che tuttavia, se accumulato nel sangue, può provocare danni depositandosi nei vasi sanguigni.

c) Proteine

Le *proteine* sono le sostanze organiche più abbondanti negli esseri viventi, nei quali svolgono innumerevoli funzioni. Negli animali costituiscono oltre il 50% del peso secco, mentre nelle piante il contenuto proteico è percentualmente più basso a causa dell'elevato contenuto di cellulosa. Strutturalmente la proteina è un polimero risultante dal concatenamento di α -amminoacidi, molecole organiche contenenti una *funzione amminica* ($-\text{NH}_2$) e una *acida o carbossilica* ($-\text{COOH}$), legate allo stesso atomo di *carbonio*.

Gli **amminoacidi** che costituiscono le proteine naturali sono 20 e differiscono uno dall'altro unicamente per la natura del gruppo -R.

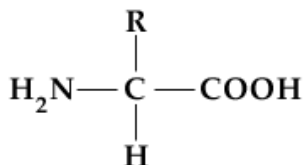


Fig. 8 - Struttura generale di un amminoacido

Nelle proteine gli amminoacidi si legano, come accade in tutte le biomolecole, attraverso una reazione di condensazione tra il gruppo amminico di un amminoacido e il gruppo carbossilico dell'amminoacido adiacente. Nella catena così generata uno dei due amminoacidi terminali mantiene inalterata la funzione amminica ed è indicato come *N-terminale*, mentre l'altro conserva intatta la funzione carbossilica e viene detto *C-terminale*. Il legame che si forma tra la funzione amminica e quella carbossilica genera la *funzione ammidica* $-\text{C}(\text{O})-\text{NH}-$ e si indica come **legame peptidico**. Di conseguenza, la catena che si forma è denominata *catena polipeptidica* (Fig. 9).

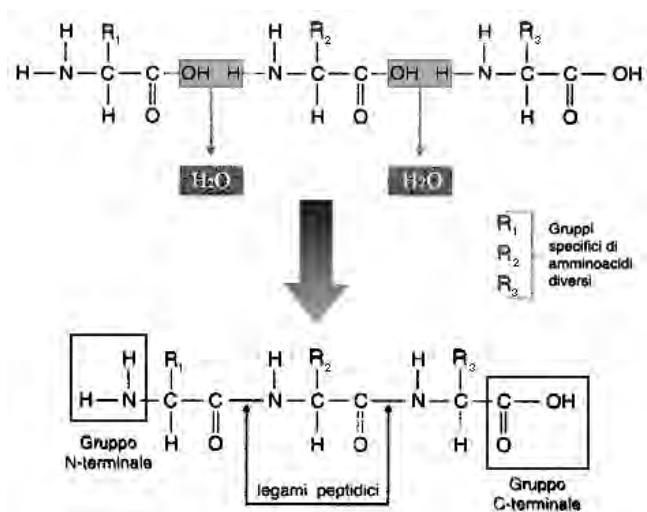


Fig. 9 - Formazione di una catena polipeptidica

Generalmente le proteine sono molecole molto grandi e arrivano a contenere diverse centinaia di amminoacidi. La varietà delle strutture che ne deriva è praticamente illimitata: può essere paragonata all'enorme quantità di parole che riusciamo a comporre con le lettere del nostro alfabeto. Inoltre nelle proteine sono riconoscibili diversi **ordini di struttura**.

• Struttura primaria

La *struttura primaria* è quella che influenza tutte le altre ed è rappresentata dalla sequenza degli amminoacidi di cui si compone la proteina. Occorre sottolineare che sotto il profilo strutturale non è sufficiente stabilire quanti amminoacidi di ciascun tipo siano contenuti nella molecola, ma bisogna conoscerne anche l'esatta sequenza. L'inversione di posizione di due soli amminoacidi su centinaia o la sostituzione di un amminoacido con un altro possono, infatti, mutare radicalmente le proprietà dell'intera proteina.

• Struttura secondaria

La *struttura secondaria* è rappresentata dal ripiegamento assunto dalla catena polipeptidica, la quale, essendo costituita da centinaia di amminoacidi, tende a ripiegarsi su sé stessa, assumendo conformazioni tipiche.

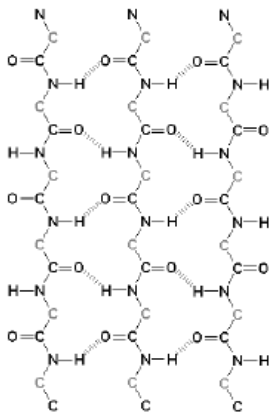


Fig. 11 - Struttura a foglio ripiegato

Una caratteristica struttura secondaria è quella **ad elica** (Fig. 10), stabilizzata dalla formazione di legami idrogeno tra il gruppo carbonilico di un legame peptidico e l'idrogeno ammidico di un secondo legame peptidico. I legami idrogeno sono piuttosto deboli e possono formarsi e rompersi con facilità: ciò conferisce a queste proteine proprietà elastiche, poiché si accorciano in seguito alla formazione dei legami idrogeno e si allungano in seguito alla loro rottura.

Altre proteine, invece, mostrano una struttura detta **a foglio ripiegato** (Fig. 11), in cui più catene polipeptidiche (o tratti della stessa catena che si ripiega, formando delle anse) si allineano in fasci paralleli uniti tra loro da legami idrogeno.

Esistono, infine, proteine che assumono una struttura **a cavo**, come il *collagene*, il costituente dei tendini, in cui più catene si avvolgono una sull'altra fino a generare strutture assai robuste, ulteriormente riunite in fibrille.

• Struttura terziaria

La *struttura terziaria* consiste in un ulteriore avvolgimento della struttura secondaria su sé stessa, in modo da generare una sorta di gomitolo. Per questo le proteine che possiedono una struttura terziaria sono dette **globulari** (Fig. 12), mentre quelle che mantengono per la maggior parte della loro lunghezza una struttura elicoidale, a foglio ripiegato o a cavo, sono dette **fibrose** (Fig. 13).

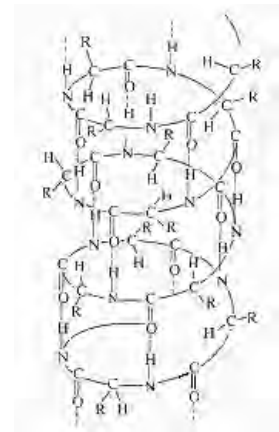


Fig. 10 - Struttura ad elica



Fig. 12 - Proteina globulare



Fig. 13 - Proteina fibrosa

La struttura terziaria è determinata solo raramente da legami forti, come nel caso della formazione di ponti disolfuro -S-S- tra molecole di cisteina, mentre generalmente è stabilizzata da interazioni deboli, come l'attrazione o la repulsione tra i gruppi R dei singoli amminoacidi, la formazione di legami idrogeno o interazioni idrofobiche, essendo l'acqua l'ambiente in cui le biomolecole sono disciolte più frequentemente.

• Struttura quaternaria

La *struttura quaternaria* riguarda le proteine formate da più catene polipeptidiche. Tra queste vi è l'**emoglobina**, la proteina trasportatrice di O₂ nel nostro sangue, costituita da quattro catene polipeptidiche, uguali a due a due, ognuna delle quali associata ad un gruppo non proteico, detto *gruppo prostetico*, il gruppo «eme».

Anche in questo caso le interazioni che mantengono la struttura quaternaria sono le stesse che stabilizzano la struttura terziaria e possono essere facilmente spezzate mediante il riscaldamento: per questo tutte le proteine sottoposte a riscaldamento «denaturano», ossia perdono la loro struttura secondaria, terziaria, quaternaria.

d) Acidi nucleici

Gli *acidi nucleici* sono sostanze di natura polimerica presenti nel nucleo cellulare. Costituiti da **DNA** o **acido deossiribonucleico** e da **RNA** o **acido ribonucleico**, sono dotati di strutture e funzioni diverse nell'ambito della cellula: il DNA è la molecola *depositaria del patrimonio genetico cellulare*, mentre l'RNA è un «trascritto» del DNA attraverso il quale viene inviato ed eseguito il messaggio genetico contenuto nel DNA.

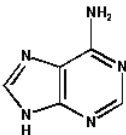
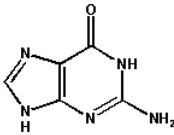
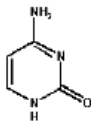
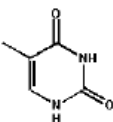
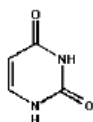
I monomeri costituenti degli acidi nucleici, più complessi di quelli incontrati nelle altre biomolecole, detti **nucleotidi**, sono formati a loro volta da una molecola di *zucchero*, una *base azotata* e una molecola di *acido fosforico*.

Nei nucleotidi che costituiscono l'RNA, lo zucchero è il *ribosio* (un aldopentoso), mentre in quelli che compongono il DNA è il *deossiribosio*, che differisce dal precedente per la perdita del gruppo -OH legato al carbonio 2 della catena. Nei nucleotidi entrambi gli zuccheri si trovano in forma ciclica.

Le *basi azotate* presenti negli acidi nucleici sono in tutto cinque: *adenina* (A), *citosina* (C) e *guanina* (G), contenute sia nel DNA che nell'RNA; *timina* (T), presente solo nel DNA; e *uracile* (U), presente solo nell'RNA.

L'*acido fosforico* è un acido inorganico poliprotico che condensa con una delle funzioni alcoliche dello zucchero, mantenendo le proprietà acide (da cui il nome di *acidi nucleici*).

Il nucleotide è il risultato della condensazione di una molecola di ribosio (o deossiribosio) con una base azotata in posizione 1 e con una molecola di acido fosforico in posizione 5.

BASI AZOTATE	
Purine	Pyrimidine
 Adenina  Guanina	 Citosina  Timina  Uracile

(Segue)

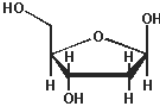
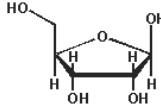
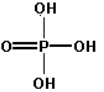
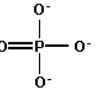
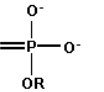
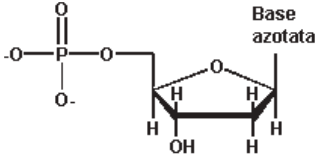
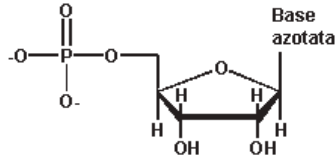
Zuccheri	Gruppo fosfato
 2-Deossiribosio  Ribosio	 Acido fosforico  Fosfato inorganico  Fosfato monoestere
Nucleotidi	
 5'-deossiribonucleotide  5'-ribonucleotide	

Fig. 14 - Struttura delle molecole che costituiscono i nucleotidi

La molecola dell'acido nucleico trae origine dalla condensazione di un gran numero di nucleotidi. Questa reazione avviene tra il gruppo fosfato di un nucleotide e il gruppo -OH ancora presente sul carbonio 3 dello zucchero del nucleotide adiacente. In tal modo, la catena dell'acido nucleico viene ad essere costituita dall'alternanza di una molecola di acido fosforico e una di zucchero, che a sua volta porta legata una molecola di base azotata. Le strutture risultanti sono mostrate nelle figg. 15 e 16.

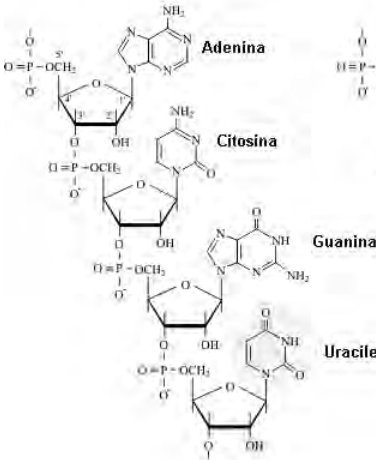


Fig. 15 - RNA

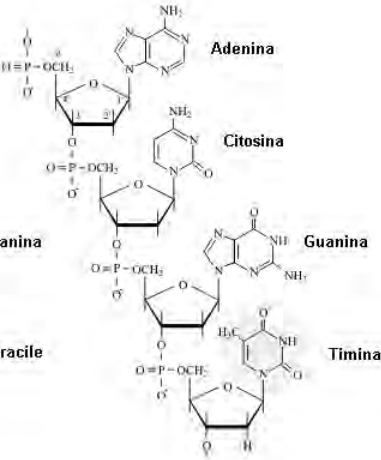


Fig. 16 - DNA

Oltre ad essere i costituenti fondamentali degli acidi nucleici, alcuni nucleotidi possiedono funzioni di grande importanza nella fisiologia cellulare, in quanto, modificati mediante l'aggiunta di due ulteriori gruppi fosfato a quello già presente, diventano efficienti trasportatori di energia. Uno di questi è l'**adenosina trifosfato** o **ATP**, la cui struttura è mostrata in figura 17.

I legami fra i tre gruppi fosfato sono abbastanza deboli e possono essere facilmente idrolizzati, liberando energia. In tal modo l'ATP, sintetizzata grazie a processi che producono energia (come, ad esempio, la respirazione cellulare), trasporta quest'ultima nei comparti cellulari in cui è richiesta. Qui si idrolizza e rilascia energia trasformandosi in **ADP (adenosindifosfato)**, che successivamente viene riconvertito in ATP, cosicché l'intero ciclo ricomincia.

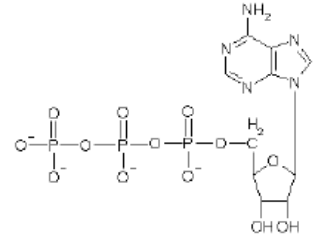


Fig. 17 - Struttura dell'ATP
(adenosina trifosfato)

La cellula come base della vita

1 La teoria cellulare

Uno dei nuclei fondanti della biologia moderna è la **teoria cellulare**, la cui formulazione ha richiesto una lunga serie di scoperte ed osservazioni. La parola *cellula* fu utilizzata per la prima volta da **Robert Hooke**, che nel 1665 osservò piccole cavità separate da pareti nel sughero e in altri tessuti vegetali, alle quali diede appunto il nome di «cellule».

Dovette poi trascorrere circa un secolo e mezzo dalla scoperta di Hooke perché si giungesse all'attuale teoria cellulare, che venne enunciata nel 1838 da **M.J. Schleiden** per le piante e nel 1839 da **Theodor Schwann** per gli animali. Nel 1850 fu poi **Max Schultze** a fornire la prima definizione di cellula quale «*particella microscopica elementare della sostanza vivente, costituita da una massa mucillaginosa contrattile ben delimitata, non miscibile con acqua, che contiene un corpicciattolo detto nucleo*».

Un passo decisivo fu infine compiuto da **Rudolf Virchow**, che nello stesso periodo osservò la *divisione cellulare* giungendo alla conclusione che le cellule possono originarsi solo da cellule preesistenti.

La formulazione attuale della *teoria cellulare* può essere riassunta nei seguenti punti:

- tutti gli esseri viventi sono costituiti da una o più cellule;
- tutte le reazioni chimiche alla base della vita hanno luogo nelle cellule;
- le cellule si originano da altre cellule;
- le cellule contengono le informazioni ereditarie degli organismi di cui sono parte e le trasmettono alle cellule figlie.

2 Le dimensioni cellulari: procarioti ed eucarioti

Caratteristica comune a tutte le cellule è quella di possedere una **membrana cellulare**, che permette di distinguere la cellula dall'ambiente esterno, e del **materiale genetico**, che dirige le attività della cellula e viene trasmesso alle cellule figlie.

L'organizzazione del materiale genetico all'interno della cellula consente, a sua volta, di distinguere *due tipologie cellulari*: le **cellule procariote**, che possiedono un'unica grande molecola di DNA circolare, associata a proteine, detta **cromosoma**, non isolata dal resto della cellula, ma semplicemente situata in una zona chiamata **nucleoide**; e le **cellule eucariote**, nelle quali, invece, il DNA è lineare e strettamente associato a proteine dette **istoni**, con cui forma un certo numero di cromosomi. In questo tipo di cellule i cromosomi sono racchiusi da una doppia membrana che li separa dal resto della cellula, individuando una zona nettamente distinguibile, denominata **nucleo**.

Tutto ciò che è contenuto all'interno della membrana cellulare, ad esclusione del nucleo e del nucleoide, è detto **citoplasma**; qui sono ospitati numerosi organuli preposti a funzioni specifiche. Nei procarioti la membrana cellulare è sempre circondata da una **parete cellulare** prodotta dalla cellula stessa, mentre fra le cellule eucariote solo quelle di piante e funghi sono provviste di parete, che è peraltro strutturalmente diversa da quella dei procarioti.

Un'ultima differenza rilevante tra procarioti ed eucarioti riguarda le dimensioni: le cellule procariote, rappresentate dagli attuali batteri, hanno un diametro dell'ordine di 2μ e sono circa dieci volte più piccole delle eucariote, i cui diametri variano tra i 10 e i 30μ . La maggiore grandezza

delle cellule eucariote comporta anche una maggiore complessità strutturale, legata soprattutto all'esigenza di avere una compartimentazione interna che consenta il rapido e ordinato svolgimento delle attività cellulari. Esiste, in ogni caso, un limite superiore alle dimensioni cellulari, dettato dalla necessità della cellula stessa di effettuare scambi con l'esterno: all'aumentare delle dimensioni, diminuisce il rapporto superficie/volume della cellula e ciò comporta una sempre minore capacità di scambio con l'esterno per unità di volume. Non a caso le cellule metabolicamente più attive, che necessitano di maggiori scambi con l'ambiente, tendono ad essere di piccole dimensioni.

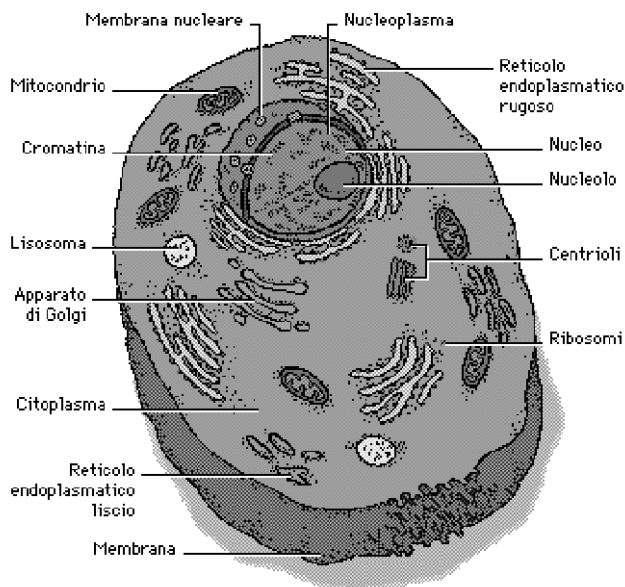


Fig. 1 - Struttura generale di una cellula eucariote animale

3 La membrana cellulare

Tutte le cellule sono separate dallo spazio esterno mediante una *membrana cellulare* dello spessore di 7-9nm, visibile al microscopio elettronico come una doppia linea continua. Tale membrana è costituita da un doppio strato di fosfolipidi, disposti in modo tale da porgere all'esterno e all'interno della cellula le teste idrofile, mentre all'interno del doppio strato si trovano le code idrofobe. Nelle cellule animali, all'interno del doppio strato, vi è anche una grande quantità di colesterolo, generalmente assente nelle cellule di piante e funghi.

La membrana contiene, altresì, numerose molecole proteiche, che si distinguono in **proteine integrali di membrana**, le quali attraversano il doppio strato sporgendo da entrambi i lati, e **proteine periferiche di membrana**, che invece aderiscono alle proteine integrali sul lato citoplasmatico.

Secondo il modello della *membrana a «mosaico fluido»*, illustrato in figura 2, le *proteine integrali* sono mantenute nelle loro posizioni dalle *proteine periferiche* e da filamenti proteici citoplasmatici; ma nonostante ciò possono subire spostamenti laterali all'interno del doppio strato, conferendo alla membrana una certa fluidità.

Le *proteine di membrana* svolgono un gran numero di funzioni: alcune di esse sono *enzimi*, altre agiscono da *proteine di trasporto* che consentono il passaggio di ioni e molecole attraverso la membrana, altre ancora sono *recettori* di ormoni o sostanze specifiche.

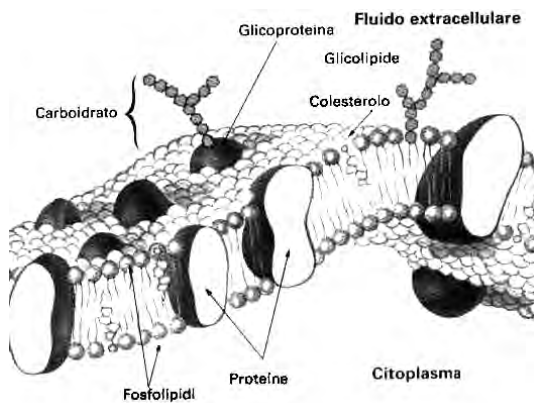


Fig. 2 - Modello della membrana a «mosaico fluido»

In molte cellule la superficie esterna della membrana è rivestita da uno strato carboidratico originato dalle catene di carboidrati associate ai lipidi di membrana (glicolipidi) e alle proteine integrali (glicoproteine). Questo strato gioca un ruolo fondamentale nell'adesione delle cellule e nel loro riconoscimento da parte di ormoni, anticorpi e virus.

4 Il nucleo

Nelle cellule eucariote è ben distinguibile un corpo sferico, il **nucleo**, separato dal *citoplasma* per mezzo di due membrane, ognuna costituita da un doppio strato fosfolipidico. Queste sono fuse insieme in numerosi punti per formare i **pori nucleari**, che consentono il passaggio di sostanze. All'interno del nucleo si trovano i **cromosomi**, ammassati in un unico groviglio, detto **cromatina**, ed è distinguibile una struttura, il **nucleolo**, in cui si producono i **ribosomi**.

Il nucleo cellulare è il vettore delle informazioni ereditarie che vengono trasferite dalla cellula madre alle cellule figlie, è il centro di controllo di tutte le attività cellulari, nonché il diretto responsabile della produzione di sostanze specifiche all'interno della cellula.

5 Il citoplasma

Grazie allo sviluppo dei microscopi elettronici è stato possibile individuare un gran numero di strutture contenute nel *citoplasma* cellulare. La parte fluida del citoplasma, il **citosol**, è costituita da una soluzione di ioni, piccole molecole e proteine presenti nella cellula in concentrazione maggiore rispetto all'esterno. Nel citosol si trovano anche i **ribosomi**, complessi di RNA e proteine, presso i quali avviene l'assemblaggio delle proteine, appunto, a partire dagli amminoacidi.

Nelle cellule eucariote il citoplasma è occupato per circa la metà da compartimenti circondati da membrane, gli **organuli**, piccole strutture in cui vengono svolte funzioni specifiche. Gli organuli sono ancorati al **citoscheletro**, una rete di filamenti proteici che mantiene l'ordine interno, conferisce alla cellula la sua forma e dirige il traffico molecolare.

6 Il reticolo endoplasmatico

Nelle cellule eucariote i *ribosomi* sono in parte dispersi nel *citosol*, in parte associati ad una struttura complessa, il **reticolo endoplasmatico**, costituito da una rete di sacchi e tubuli interconnessi tra loro, di cui si distinguono due diverse tipologie:

- il **reticolo endoplasmatico rugoso**, direttamente collegato allo strato esterno della membrana nucleare, che deve il suo aspetto all'associazione con i ribosomi;
- il **reticolo endoplasmatico liscio**, che costituisce la diretta continuazione del primo.

La quantità relativa dei due tipi di reticolo dipende dal tipo di attività che svolge la cellula. Di solito il reticolo rugoso, sede della sintesi delle proteine, è più abbondante di quello liscio, che invece risulta particolarmente abbondante nelle cellule specializzate nella sintesi e nel metabolismo dei lipidi.

La sintesi delle proteine cellulari, infatti, inizia sempre sui ribosomi presenti nel citosol e lì si completa per tutte le proteine che svolgono la loro funzione nel citosol stesso. Le proteine destinate a svolgere la loro azione lontano dal sito di produzione, dopo la formazione di una sequenza segnale, vengono trasportate, insieme al ribosoma su cui si stanno formando, sul reticolo endoplasmatico rugoso. La catena proteica che si forma passa nella cavità interna del reticolo endoplasmatico, dove viene associata a lipidi ed inviata verso l'*apparato di Golgi*; il ribosoma, invece, una volta completata la sintesi della catena proteica, si stacca dal reticolo endoplasmatico e torna nel citosol.

7 I mitocondri

In tutte le cellule eucariote sono presenti i *mitocondri*, strutture preposte alla trasformazione dell'energia chimica, contenuta nei composti organici, in energia direttamente utilizzabile dalla cellula per i suoi processi metabolici, cioè in ATP. Come mostrato nella figura 3, i mitocondri possiedono due membrane, la più interna delle quali costituisce il «banco di lavoro» su cui avvengono le reazioni mitocondriali e mostra numerosi ripiegamenti, in modo da estendere il più possibile il suo sviluppo superficiale.

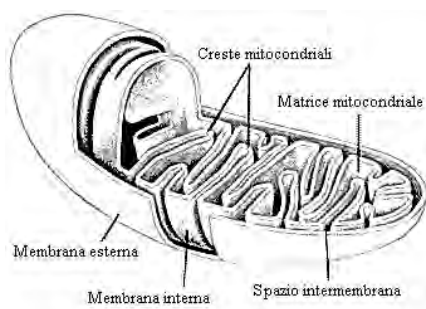


Fig. 3 - Struttura del mitocondrio

8 L'apparato di Golgi

Le cellule animali contengono da 10 a 20 strutture indicate come *apparato di Golgi*, mentre le cellule vegetali possono arrivare a contenerne diverse centinaia. Tale apparato si presenta come una serie di sacchi appiattiti, limitati da membrane, la cui funzione è quella di ricevere vescicole provenienti dal reticolo endoplasmatico e rielaborarne il contenuto. Le molecole rielaborate, chiuse in vescicole di trasporto, sono inviate a siti specifici nella cellula o sulla superficie cellulare, da cui verranno esportate.

Le principali funzioni dell'apparato di Golgi consistono nella sintesi dei lipidi e delle proteine destinate alla membrana cellulare o alle endomembrane, nonché nella rielaborazione e nell'immagazzinamento di sostanze destinate ad altre cellule, come ad esempio gli ormoni.

9 I lisosomi

I *lisosomi* sono particolari vescicole prodotte dall'*apparato di Golgi*, contenenti una notevole varietà di **enzimi idrolitici**, cioè in grado di demolire macromolecole mediante reazioni di «idrolisi». Questi enzimi, utilizzati dalla cellula per demolire molecole complesse e riutilizzarne i prodotti di idrolisi, sono lesivi nei confronti della cellula stessa ed è quindi indispensabile che rimangano confinati all'interno dei lisosomi, la cui rottura può provocare la *lisi cellulare*.

Una particolare funzione dei lisosomi si osserva nei globuli bianchi del corpo umano, in cui essi svolgono un'azione batteriolitica: un batterio catturato dalla cellula viene racchiuso in un vacuolo con il quale vanno a fondersi i lisosomi, i quali, liberando i propri enzimi idrolitici, ne determinano la distruzione.

10 Il citoscheletro

Il *citoscheletro* è una rete, interna al *citoplasma* cellulare e formata da proteine filamentose, che offre ancoraggio agli organuli, conferisce alla cellula la sua forma, dirige lo spostamento delle molecole al suo interno e gli consente di effettuare piccoli movimenti.

Le tre principali tipologie di filamenti proteici che costituiscono il citoscheletro sono mostrate in figura 4 e illustrate a seguire.

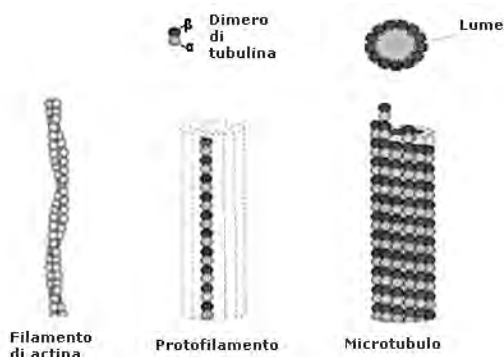


Fig. 4 - Filamenti proteici del citoscheletro

I **filamenti di actina**, sottili fibre proteiche costituite da molecole di actina, una proteina globulare, possono essere facilmente e velocemente assemblati e scomposti dalla cellula. Le fibre di actina sono ancorate alle proteine integrali di membrana e si estendono come cavi attraverso il citoplasma, formando una rete che sostiene la membrana cellulare.

I **microtubuli**, le strutture più grandi del citoscheletro, risultanti dall'assemblaggio di *dimeri* di due diverse proteine globulari, la *tubulina* α e la *tubulina* β , partono da un centro di organizzazione posto vicino al nucleo e si estendono verso la membrana cellulare. Questi costituiscono il sito di ancoraggio degli organuli e dirigono lo spostamento di vescicole all'interno della cellula.

I **filamenti intermedi**, così chiamati perché di dimensioni intermedie rispetto ai precedenti, sono costituiti da proteine fibrose e si irradiano a partire direttamente dal nucleo. Sono particolarmente numerosi in cellule sottoposte a stress meccanici, come quelle della pelle.

Oltre che come struttura di organizzazione e di sostegno, il citoscheletro assume importanza anche nel movimento cellulare, essendo una struttura dinamica che può essere modificata e spostata a seconda delle necessità. Piccoli movimenti cellulari, come le correnti attive nel citoplasma, i movimenti cromosomici e i cambiamenti di forma durante la divisione cellulare, sono determinati dall'assemblaggio di proteine contrattili, fra cui soprattutto i filamenti di actina.

11 Ciglia e flagelli

Alcune cellule possiedono vere e proprie strutture preposte al movimento, costituite da filamenti che si estendono dalla superficie cellulare, distinte tradizionalmente in **ciglia**, più corte e numerose, e **flagelli**, più lunghi e in numero minore.

Entrambe le tipologie derivano dall'assemblaggio di microtubuli secondo una disposizione caratteristica, indicata come **9 + 2**, caratterizzata da una coppia centrale attorno alla quale sono disposte altre nove coppie. Queste restano aderenti tra loro e a quella centrale per mezzo di molecole proteiche: si tratta di enzimi idrolitici dell'ATP che producono l'energia necessaria al movimento. Il sistema è ulteriormente stabilizzato da proteine disposte ad anello, che circondano le nove coppie.

Ciglia e flagelli prendono origine da ornanuli cellulari denominati **corpi basali**, a loro volta costituiti da microtubuli, dotati di una disposizione diversa rispetto a quella descritta. In molte cellule, infine, sono presenti i **centrioli**, corpi di struttura identica a quella dei corpi basali, che rappresentano i centri di organizzazione dei microtubuli.

12 La parete cellulare

Le cellule vegetali, come quelle di funghi e procarioti, sono dotate di una **parete esterna** alla membrana, costruita dalla cellula stessa. Nelle cellule vegetali, la cui struttura generale è mostrata in figura 5, la parete è rigida, costituita da cellulosa, e conferisce alla cellula la sua forma caratteristica. Le pareti di funghi e procarioti, invece, sono generalmente prive di cellulosa.

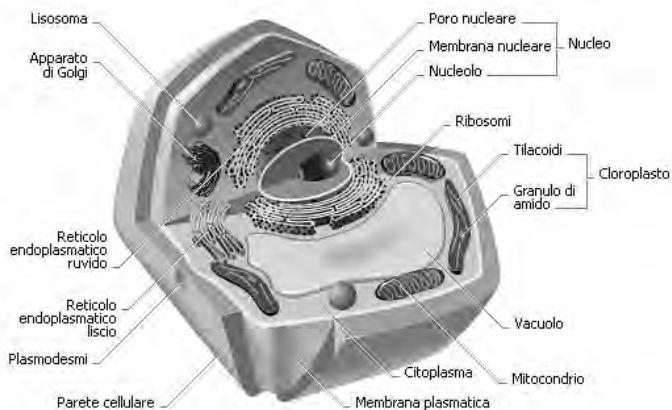


Fig. 5 - Struttura generale di una cellula vegetale

13 Vacuoli e vescicole

Nelle cellule vegetali sono presenti cavità, circondate da membrana, contenenti acqua e soluti, dette **vacuoli**. In genere le cellule giovani contengono vacuoli piccoli e numerosi, i quali, con la maturità, tendono a riunirsi in un unico «grande» vacuolo che *dà sostegno* alla cellula.

Le **vescicole** sono invece presenti in tutte le cellule eucariote. La loro struttura ricorda quella dei vacuoli, anche se le dimensioni sono notevolmente più piccole e la loro funzione è principalmente quella di trasportare le sostanze: le proteine sintetizzate nel *reticolo endoplasmatico*, ad esempio, vengono inviate all'*apparato di Golgi* in forma di vescicole.

14 I cloroplasti

I *cloroplasti* sono strutture tipiche delle *cellule vegetali*. Queste, a differenza di quelle *animali*, sono *cellule autotrofe fotosintetiche*, le quali utilizzano l'energia della luce solare per sintetizzare autonomamente i composti organici di cui necessitano, a partire da molecole inorganiche.

La trasformazione dell'energia solare e il suo immagazzinamento sotto forma di energia chimica avvengono grazie alla presenza di *pigmenti* in grado di assorbire la luce del Sole. Uno di questi è la **clorofilla**, che nella cellula si trova inserita in una membrana specializzata: il *tilacoide*. Come mostrato in figura 6, *pile di tilacoidi* costituiscono i *grani* verdi visibili all'interno dei cloroplasti.

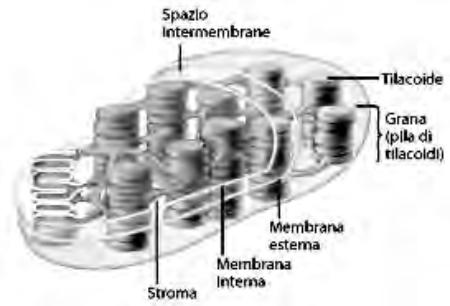


Fig. 6 - Struttura del cloroplasto

Scambi di materia ed energia tra le cellule e al loro interno

1 Gli scambi di materia

a) Diffusione e osmosi

La membrana cellulare ha una certa permeabilità, legata all'esistenza di piccole e temporanee aperture tra le molecole di fosfolipidi. Il movimento di molecole sufficientemente piccole attraverso la membrana è spesso governato dai processi naturali di *diffusione* e di *osmosi*.

La **diffusione** è un fenomeno spontaneo che garantisce l'omogeneità della composizione di un sistema. Ponendo una goccia di caffè in un bicchiere d'acqua, in breve tempo si assiste alla lenta e progressiva diffusione delle particelle colorate in tutto il volume, fino al raggiungimento di una colorazione omogenea. Questo fenomeno è dovuto al fatto che le particelle sono sottoposte a un movimento continuo e casuale e, in presenza di **gradienti di concentrazione** (cioè differenze di concentrazione tra una zona e l'altra della soluzione), il numero di quelle che passano dalla zona a maggiore concentrazione verso quella a minore concentrazione è statisticamente superiore rispetto a quelle che effettuano il passaggio contrario. La diffusione è un fenomeno lento, che tende a rallentare ulteriormente al diminuire della differenza di concentrazione tra le zone considerate. Le cellule, infatti, mantengono sempre gradienti di concentrazione elevati con l'esterno, in modo da rendere la diffusione un processo più efficiente. La lentezza degli scambi affidati al fenomeno della diffusione attraverso la membrana cellulare costituisce uno dei maggiori limiti alla dimensione delle cellule.

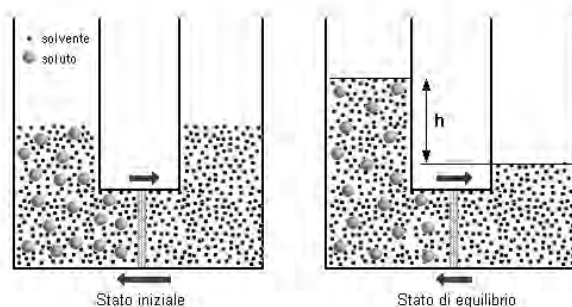


Fig. 1 - Apparato sperimentale per la rappresentazione del fenomeno dell'osmosi

L'**osmosi** è, a sua volta, un fenomeno indotto dall'esistenza di una differenza di concentrazione ai lati di una membrana semipermeabile. In figura 1 è mostrato un semplice apparato sperimentale costituito da un tubo ad U, le cui braccia, separate da una membrana semipermeabile, contengono da un lato la soluzione di un soluto di grandi dimensioni, incapace di attraversare la membrana, dall'altro il solvente puro. Le molecole di solvente si diffondono da un lato all'altro della membrana

e, in virtù della differenza di concentrazione, la diffusione risulta più veloce dal solvente verso la soluzione, determinando un passaggio netto di solvente nella soluzione. Ad equilibrio raggiunto, il dislivello h instauratosi tra le due braccia è tale da controbilanciare la spinta osmotica e le velocità di diffusione del solvente da un braccio all'altro si eguagliano. La pressione esercitata dalla colonna di soluzione di altezza h viene indicata come **pressione osmotica** della soluzione.

Va sottolineato che il valore della *pressione osmotica* di una soluzione dipende dal numero

delle particelle disciolte e non dalla loro natura chimica o dalla loro dimensione. Date due soluzioni, quella dotata di pressione osmotica maggiore si dice **ipertonica**, quella con pressione osmotica minore è chiamata **ipotonica**, mentre se le due soluzioni mostrano un uguale valore della pressione osmotica si definiscono **isotoniche**.

Il processo osmotico è di enorme importanza per gli esseri viventi e spesso può costituire un problema per tutti gli organismi che vivono in un ambiente non isotonico: infatti, essendo la membrana cellulare permeabile all'acqua, gli organismi che vivono in ambienti ipertonici rispetto ai loro fluidi corporei (come i pesci marini) tenderebbero a perdere acqua, mentre gli organismi che vivono in ambienti ipotonici tenderebbero ad assorbirla. Ovviamente questi organismi hanno dovuto sviluppare meccanismi di regolazione che consentono di mantenere la loro concentrazione interna ai giusti livelli.

b) Proteine di trasporto

Abbiamo visto come tra i componenti delle membrane cellulari ci sia una discreta quantità di *proteine*, le quali svolgono importanti funzioni nei *meccanismi di scambio* tra la cellula e l'ambiente esterno. Molti ioni o molecole che, per le loro dimensioni o polarità, non sarebbero in grado di attraversare la membrana cellulare possono entrare ed uscire dalla cellula grazie al trasporto operato dalle *proteine integrali di membrana*. Queste possiedono un'elevata specificità e spesso è proprio l'interazione con la molecola da trasportare ad indurre nella proteina un temporaneo cambiamento di conformazione, che consente il suo passaggio attraverso la membrana.

Se la proteina trasporta un unico tipo di soluto si parla di **uniporto**; se invece consente il passaggio di due differenti soluti si parla di **cotrasporto**, a sua volta distinto in **simporto**, se i due soluti transitano contemporaneamente e nella stessa direzione, o **antiporto**, se i due soluti transitano in direzioni opposte, contemporaneamente o in sequenza (Fig. 2).

Quando il trasporto di sostanze attraverso le proteine di membrana avviene secondo gradiente, ovvero nella direzione stabilita dalla differenza di concentrazione, si parla di **trasporto passivo** o «diffusione facilitata». Questo tipo di trasporto riguarda essenzialmente molecole non liposolubili, incapaci di attraversare lo strato idrofobico della membrana. Quindi, per potersi spostare da un lato all'altro sotto l'azione del gradiente di concentrazione, queste devono utilizzare il canale aperto dalle proteine di membrana, dette, per questo, **proteine canale**.

Se invece il trasporto avviene in senso inverso rispetto al gradiente di concentrazione si parla di **trasporto attivo**, così denominato perché si tratta di un processo che richiede dispendio di energia da parte della cellula.

Un esempio di trasporto attivo è quello realizzato dalla **pompa sodio-potassio** (Fig. 3), una proteina di trasporto attraverso la quale le cellule animali conservano gradienti di concentrazione degli ioni Na^+ e K^+ molto alti rispetto all'esterno, indispensabili per il mantenimento dell'equilibrio osmotico. La concentrazione del Na^+ è 14 volte più alta all'esterno della cellula rispetto all'interno, mentre quella del K^+ è da 10 a 30 volte più alta all'interno rispetto all'esterno. La proteina di trasporto possiede due conformazioni alternative: la prima ha una cavità rivolta verso l'interno della cellula adatta ad accogliere gli ioni Na^+ , mentre la seconda possiede una

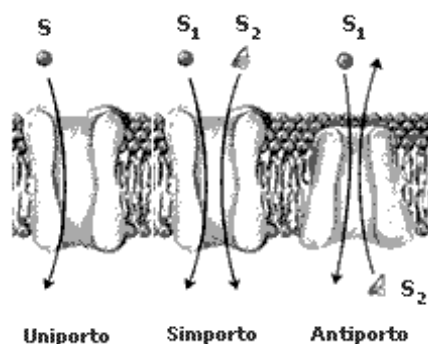


Fig. 2 - Tipologie di trasporto da parte delle proteine

cavità rivolta verso l'esterno in grado di accogliere gli ioni K^+ . Quando gli ioni sodio si legano alla proteina di trasporto, una molecola di ATP viene idrolizzata e lo ione fosfato, così liberato, si lega alla proteina stessa, provocandone il cambiamento di conformazione. In tal modo, gli ioni sodio vengono rilasciati all'esterno della cellula e la proteina è pronta ad accogliere gli ioni potassio. In seguito al legame degli ioni potassio, il gruppo fosfato viene liberato dalla proteina, la quale, riportandosi nella conformazione iniziale, libera gli ioni potassio all'interno della cellula. Ad ogni ciclo di pompaggio, con il consumo di una sola molecola di ATP, tre ioni Na^+ escono dalla cellula, mentre due ioni K^+ vi fanno ingresso. La pompa sodio-potassio può lavorare molto velocemente, arrivando fino a 100 cicli di pompaggio al secondo.

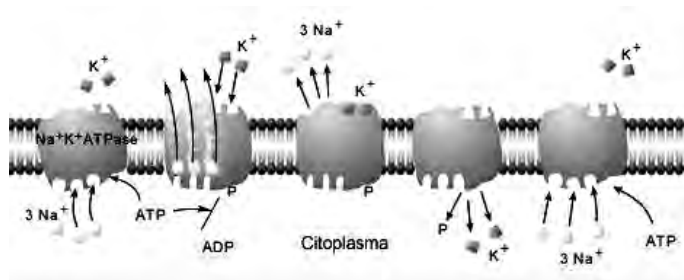


Fig. 3 - Pompa sodio-potassio

c) Endocitosi ed esocitosi

Diffusione, osmosi e trasporto tramite proteine consentono il passaggio attraverso la membrana di molecole di piccole dimensioni.

L'ingresso o l'uscita dalla cellula di macromolecole, particelle e microrganismi avviene per mezzo di *vescicole* che, fondendosi con la membrana cellulare, liberano il loro contenuto all'interno e all'esterno di essa. Molte sostanze prodotte dalla cellula vengono riversate all'esterno attraverso vescicole prodotte dagli *apparati di Golgi* e, analogamente, molecole o particelle di grandi dimensioni che dall'esterno entrano in contatto con la membrana cellulare, ne determinano un ripiegamento verso l'interno, con la conseguente formazione di una vescicola che trasporta le particelle all'interno della cellula (Fig. 4).

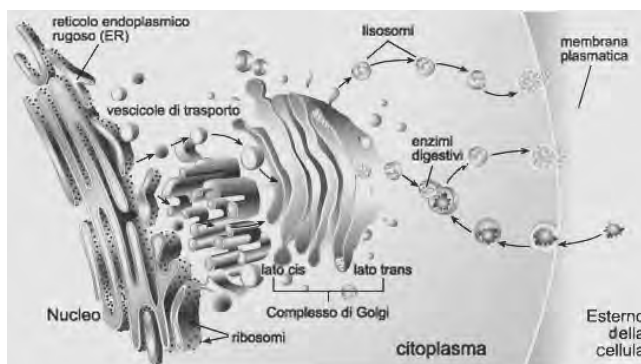


Fig. 4 - Formazione delle vescicole di trasporto

L'escrezione di sostanze attraverso vescicole prende il nome di **esocitosi**, mentre l'ingresso di sostanze o particelle viene detto **endocitosi**, a sua volta distinta in *fagocitosi*, se i prodotti

incamerati sono solidi o di dimensioni superiori ai 250nm, e *pinocitosi*, se le sostanze incamerate sono liquide.

In alcuni casi l'endocitosi può essere **mediata da recettori**: alcune proteine di membrana costituiscono siti di legame per sostanze specifiche, i **ligandi**, che devono fare ingresso nella cellula. Questi recettori si trovano in piccole depressioni della membrana, dette *fossette ricoperte*, contraddistinte dalla presenza di proteine denominate *clatrine*. Una volta formatosi, il complesso recettore-ligando dà origine a una vescicola ricoperta di clatrine e passa all'interno della cellula.

d) Giunzioni cellulari

Un organismo pluricellulare complesso è un sistema altamente organizzato, in cui gruppi di cellule che svolgono una funzione comune sono organizzati in **tessuti**; tessuti diversi sono riuniti insieme a formare **organi**; organi diversi sono coordinati fra loro per costituire **apparati**.

In queste complesse organizzazioni cellulari la comunicazione tra cellula e cellula avviene sotto forma di segnali chimici, cioè attraverso molecole che escono da una cellula per raggiungere un'altra. L'interazione con la cellula bersaglio può limitarsi al legame con un *recettore di membrana*, che innesci al suo interno un processo specifico, oppure può prevedere l'ingresso della molecola all'interno della cellula.

In alcuni tipi di tessuto le cellule sono strettamente a contatto tra loro ed evidenziano la presenza di vere e proprie **giunzioni cellulari**: nelle *cellule vegetali*, ad esempio, le pareti cellulari sono attraversate da canali, i **plasmodesmi**, che mettono in comunicazione diretta i citoplasmi di due cellule adiacenti. Nelle *cellule animali* esistono strutture analoghe, le **giunzioni comunicanti**, costituite da piccoli canali circondati da proteine che permettono il passaggio di molecole di piccole dimensioni.

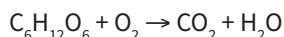
2 La produzione di energia

a) Reazioni di ossidoriduzione nei viventi

Molte delle reazioni chimiche che avvengono nelle cellule sono **reazioni di ossidoriduzione**, ossia reazioni che comportano il trasferimento di elettroni da un atomo o da una molecola ad un'altra: sono processi di ossidoriduzione, ad esempio, la **fotosintesi** e la **respirazione cellulare**.

Quando una specie chimica cede elettroni subisce un processo di **ossidazione**, mentre una specie che li acquista ne subisce uno di **riduzione**. Le caratteristiche fisiche dell'elettrone, in particolare l'elevata densità di carica elettrica, non gli consentono di trovarsi allo stato libero, sicché la sua liberazione da parte di una specie chimica è possibile solo in presenza di una seconda specie in grado di acquistarlo. Per questo motivo in natura si osservano unicamente processi di ossidoriduzione in cui l'ossidazione di una specie avviene contestualmente alla riduzione di un'altra.

In molte reazioni biochimiche il trasferimento di elettroni è accompagnato da quello di atomi di idrogeno e nelle corrispondenti equazioni chimiche si osserva il trasferimento di atomi di idrogeno dalla specie che si ossida a quella che si riduce. Consideriamo, a titolo di esempio, la *reazione di combustione del glucosio*, la quale, come tutte le combustioni, è un processo di ossidoriduzione in cui atomi di idrogeno, come detto, vengono trasferiti dalla specie che si ossida (il glucosio) a quella che si riduce (l'ossigeno):



In genere le combustioni sono reazioni fortemente *esoergoniche*, ovvero che liberano grandi quantità di energia, trasferendola all'ambiente sotto forma di calore. Se la combustione del

glucosio nella cellula avvenisse in questo modo, si avrebbe lo svantaggio di non poter utilizzare efficacemente l'energia liberata e, nello stesso tempo, il notevole incremento di temperatura comprometterebbe l'integrità cellulare.

Pertanto le cellule hanno sviluppato meccanismi per far sì che queste reazioni avvengano in modo controllato, attraverso **sequenze biochimiche** che consentono la liberazione di piccole quantità di energia in una serie di *step* di reazione successivi. Di norma ogni *step* della sequenza è catalizzato da un enzima specifico, in modo tale da avere, nel complesso, un processo sufficientemente veloce.

b) ATP: la «valuta energetica» della cellula

I carboidrati e il glucosio costituiscono le forme in cui l'energia può essere conservata, oppure trasferita da un organismo all'altro o tra cellula e cellula. L'energia chimica contenuta nelle molecole di glucosio può essere liberata attraverso una complessa serie di reazioni, che, nel loro insieme, vengono indicate come **respirazione cellulare**. Tuttavia, in diversi processi biochimici la cellula necessita di energia facilmente disponibile: in molte sequenze biochimiche, infatti, si possono verificare uno o più *step* di reazione *endoergonici*, cioè passaggi che hanno bisogno dell'apporto di energia dall'esterno. La cellula risolve questo problema mediante il **meccanismo delle reazioni accoppiate**: la reazione endoergonica si verifica contemporaneamente ad una esoergonica, la quale fornisce l'energia necessaria a far avvenire la prima. Nella maggior parte dei casi la molecola che fornisce l'energia per queste reazioni accoppiate è proprio l'**ATP (adenosina trifosfato)**, considerata per questo la «**valuta energetica**» della cellula.

Come già affermato in precedenza, l'ATP è un nucleotide a cui sono legati due ulteriori gruppi fosfato: liberando un gruppo fosfato per idrolisi, l'ATP si trasforma in **ADP (adenosina difosfato)** e libera 7 kilocalorie per mole; la liberazione di un secondo gruppo fosfato mette a disposizione altre 7 kilocalorie per mole e trasforma l'ADP in **AMP (adenosina monofosfato)**.

L'idrolisi dei gruppi fosfato è catalizzata da enzimi specifici, le *ATP-asi*, che molto spesso sono direttamente accoppiate con i sistemi di reazione ai quali devono fornire energia. Molte proteine che effettuano il trasporto attivo di sostanze attraverso la membrana sono direttamente associate con l'ATP-asi, che libera l'energia necessaria. Nel caso della pompa sodio-potassio, ad esempio, il gruppo fosfato non viene semplicemente rilasciato, ma trasferito ad un'altra molecola, secondo un processo indicato col nome di **fosforilazione**. In tal modo parte dell'energia contenuta nel legame della molecola di ATP viene conservata nel nuovo legame stabilito nel composto fosforilato, il quale, portato ad un contenuto di energia maggiore, può dare luogo alla reazione successiva.

1 L'ossidazione del glucosio

Per svolgere le proprie attività la cellula necessita di energia, la cui fonte principale è costituita dal glucosio. Il processo di liberazione dell'energia consiste nell'*ossidazione del glucosio* e nella produzione di molecole di ATP. Tale processo è svolto dalla cellula attraverso una serie di reazioni piuttosto complesse, che assicurano un elevato rendimento energetico.

La *reazione globale di combustione del glucosio* può essere scritta come:

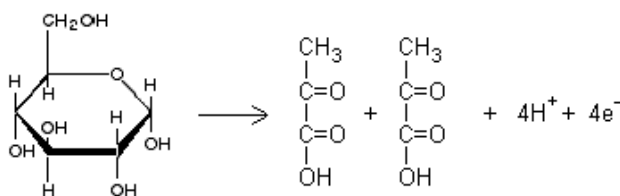


All'interno delle cellule questa reazione si svolge in due fasi, la prima delle quali, la **glicolisi**, ha luogo nel *citosol*, mentre la seconda, detta **respirazione cellulare**, avviene nei *mitocondri* e consta, a sua volta, di due fasi distinte: il *ciclo di Krebs* e la *catena di trasporto degli elettroni*.

In seguito all'ossidazione completa del glucosio si producono 38 molecole di ATP. Tuttavia, se le cellule si trovano in condizioni di carenza di ossigeno, come può accadere alle cellule muscolari nel corso di uno sforzo fisico, la respirazione non può avere luogo e la glicolisi viene seguita dalla **fermentazione**, un processo che ha un rendimento energetico molto basso, poiché produce solo due molecole di ATP a partire da una di glucosio, ma sufficiente a soddisfare le esigenze energetiche della cellula fino al ripristino di una corretta ossigenazione.

a) Glicolisi

Col termine *glicolisi*, che letteralmente significa «scissione del glucosio», si indica la prima fase dell'ossidazione del glucosio, consistente nella sua scissione in due molecole di *fosfogliceraldeide* (PGA), in seguito trasformate in due molecole di *acido piruvico*. Globalmente, la reazione può essere scritta come:



Durante questa reazione vengono rimossi dal glucosio quattro ioni H^+ insieme a quattro elettroni, con la liberazione di 143 kcal/mole, mentre le rimanenti 543 kcal/mole sono ancora contenute nelle molecole di acido piruvico (si ricordi che la combustione completa di una mole di glucosio libera 686 kcal/mole).

La glicolisi costituisce un esempio eloquente di reazione che si verifica attraverso una serie di stadi successivi, ognuno dei quali catalizzato da un enzima specifico e tale da mettere in gioco una piccola quantità di energia. Nello schema che segue sono indicate le diverse tappe della glicolisi: le prime quattro richiedono l'apporto di energia dall'esterno, fornita attraverso il trasferimento di gruppi fosfato da due molecole di ATP, mentre le ultime cinque producono energia, generando quattro molecole di ATP e due di NADH.

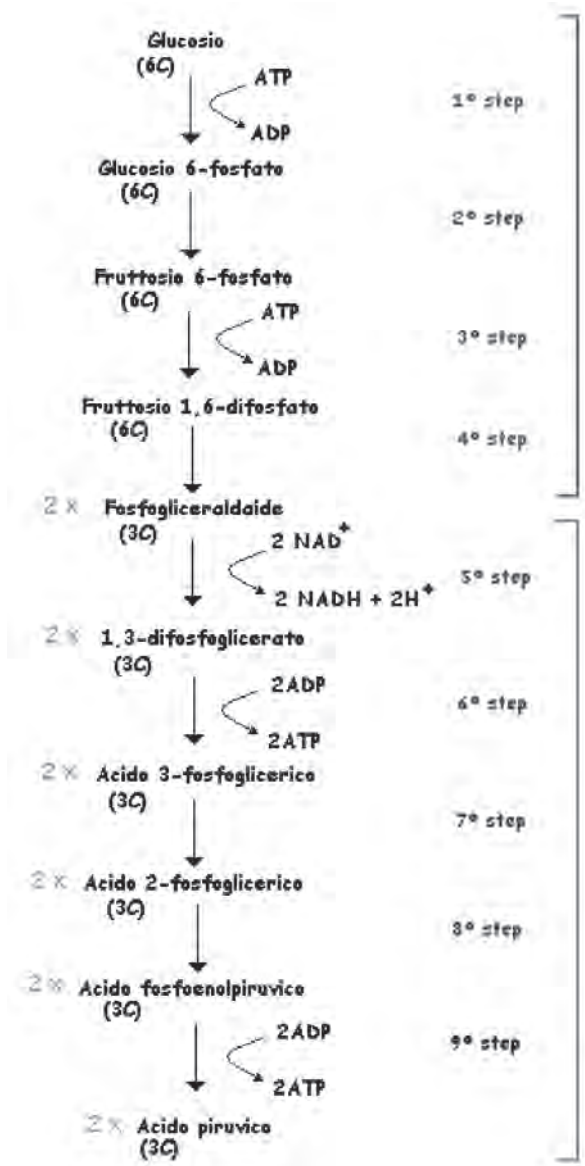
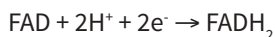
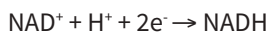


Fig. 1 - Le nove tappe della glicolisi

Sia nella glicolisi che nel ciclo di Krebs protoni ed elettroni vengono trasferiti a coenzimi che lavorano come trasportatori di elettroni: si tratta del nicotinammide-adenin-dinucleotide (NAD^+) e del flavin-adenin-dinucleotide (FAD), che acquistano energia riducendosi, rispettivamente, a NADH e FADH_2 secondo le seguenti semireazioni:



Il guadagno netto di energia nella glicolisi è, quindi, di due molecole di ATP e altrettante di NADH .

b) Respirazione cellulare

• Ciclo di Krebs

In presenza di una sufficiente quantità di ossigeno la glicolisi è seguita dalla completa demolizione del glucosio in anidride carbonica ed acqua, processo che nelle cellule eucariote ha luogo nei mitocondri. Questi organuli sono costituiti da due membrane, una esterna liscia e una interna dotata di numerosi piegamenti, detti *creste*. La membrana esterna è permeabile all'acqua e a molecole di piccole dimensioni, per cui la soluzione tra le membrane ha una composizione simile al citosol. La membrana interna, invece, ha permeabilità selettiva e limita il passaggio degli ioni H^+ , proprietà importantissima per la produzione di ATP.

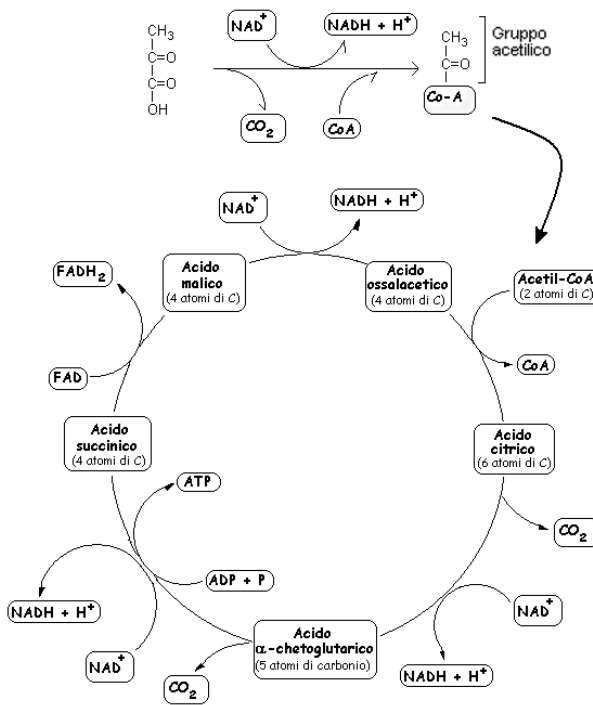


Fig. 2 - Schematizzazione delle reazioni che compongono il ciclo di Krebs

L'acido piruvico prodotto dalla glicolisi entra nella matrice dei mitocondri, dove subisce preliminarmente un'ossidazione e una decarbossilazione generando un *gruppo acetile* (CH_3CO). Questo si lega ad un composto detto **coenzima A**, generando l'addotto **acetil-coA**, che successivamente entra nel ciclo di Krebs. Naturalmente anche gruppi acetilici derivanti dalla demolizione di altre sostanze, come amminoacidi o grassi, possono, una volta legati al coenzima A, entrare nel ciclo di Krebs. Questo inizia con la combinazione del gruppo acetilico rilasciato dal coenzima A con l'acido ossalacetico, molecola contenente quattro atomi di carbonio. Si forma così un composto a sei atomi di carbonio che, nel corso del ciclo, perde due molecole di CO_2 rigenerando l'acido ossalacetico, da cui il ciclo ricomincia. Ogni ciclo consuma un gruppo acetilico e produce una molecola di ATP, tre di NADH e una di $FADH_2$. La demolizione di una molecola di glucosio richiede due cicli completi.

• Fosforilazione ossidativa

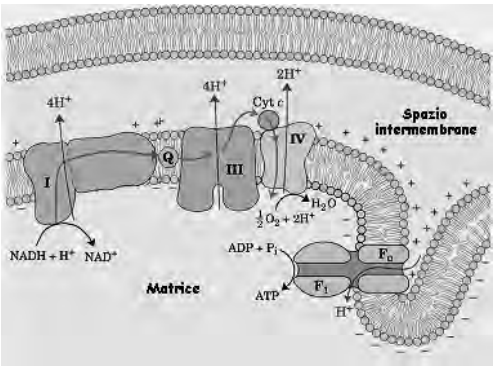


Fig. 3 - Processo di fosforilazione ossidativa

A questo punto gran parte dell'energia è ancora trattenuta nelle molecole di NADH e FADH₂ che hanno accettato elettroni dal glucosio. Questi, a loro volta, trasferiscono elettroni all'ossigeno attraverso una serie di molecole intermedie, dette *trasportatori di elettroni*, dotate di strutture assai simili tra loro, in modo tale che ad ogni passaggio si verifichi una piccola variazione di energia. La serie di reazioni che porta gli elettroni da NADH e FADH₂ all'ossigeno è detta **catena di trasporto degli elettroni** e l'energia da essa generata viene utilizzata per la sintesi di ATP a partire da ADP, secondo il processo denominato **fosforilazione ossidativa** (Fig. 3).

I componenti della catena di trasporto degli elettroni sono inseriti nella membrana interna dei mitocondri e risultano organizzati in tre complessi proteici.

L'energia liberata dalle reazioni di trasferimento degli elettroni viene utilizzata dai complessi proteici per pompare ioni H⁺ fuori dalla matrice mitocondriale, nello spazio tra le membrane. Poiché la membrana interna dei mitocondri è impermeabile agli ioni H⁺, le reazioni della catena di trasporto comportano l'instaurazione di un gradiente elettrochimico tra i due lati della membrana interna, dovuto alla diversa concentrazione di ioni H⁺ e, quindi, alla differenza di carica. Si stima che per ogni coppia di elettroni provenienti dal NADH vengano espulsi dieci ioni H⁺ e che per ogni coppia proveniente dal FADH₂ ne vengano espulsi sei.

Questo gradiente elettrochimico viene sfruttato dal complesso proteico **ATP-sintetasi** per produrre molecole di ATP: attraverso il canale interno del complesso gli ioni H⁺ tornano nella matrice e, spostandosi secondo il gradiente di concentrazione, liberano l'energia necessaria alla trasformazione di ADP in ATP. La sintesi di una molecola di ATP richiede il passaggio di tre protoni attraverso il complesso ATP-sintetasi. Il bilancio energetico totale della demolizione di una molecola di glucosio deve tener conto di tutte le molecole di ATP, NADH e FADH₂ prodotte a partire dalla glicolisi, ricordando che per ogni molecola di glucosio sono necessari due cicli di Krebs.

Processo		Citosol	Catena di trasporto degli elettroni	Molecole di ATP prodotte
Glicolisi		2 ATP		2 ATP
		2 NADH	6 ATP	6 ATP
Respirazione	Acido piruvico→ Acetil-CoA	2 × (1 NADH)	2 × (3 ATP)	6 ATP
	Ciclo di Krebs	2 × (1 ATP)		2 ATP
		2 × (3 NADH)	2 × (9 ATP)	18 ATP
		2 × (1 FADH ₂)	2 × (2 ATP)	4 ATP
Totale				38 ATP

Tab. 1

In ogni legame ADP-P sono immagazzinate 7 kcal/mol e, complessivamente, nelle 38 molecole di ATP prodotte si trovano 266 kcal/mol, cioè il 40% dell'energia liberata dalla combustione di una mole glucosio.

c) Fermentazione

In condizioni di carenza di ossigeno la demolizione del glucosio non può arrivare all'ossidazione completa, ma segue una via alternativa detta **fermentazione**, in cui l'*acido piruvico* prodotto dalla glicolisi viene ridotto ad *etanolo*, come accade durante la

fermentazione dell'uva, o ad *acido lattico*, come avviene nelle cellule muscolari sottoposte ad uno sforzo intenso. In questo caso, infatti, il rapido consumo di energia non è compensato da un altrettanto rapido rifornimento di ossigeno, cosicché l'ossidazione di NADH a NAD⁺ avviene a spese dell'acido piruvico. In tal modo viene rigenerato NAD⁺, disponibile per l'ossidazione di altre

molecole di glucosio. Il rendimento energetico della fermentazione è molto basso, poiché si ottengono solo due molecole di ATP da una di glucosio, ma sufficiente a soddisfare le esigenze energetiche immediate. Una volta ristabilita la normale ossigenazione, l'acido lattico accumulato nel muscolo viene riossidato ad acido piruvico e riutilizzato.

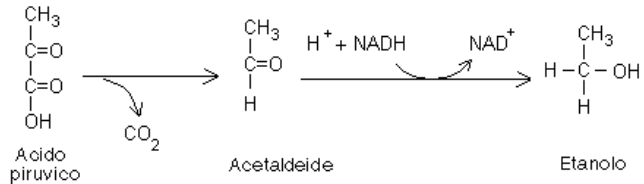


Fig. 4 - Fermentazione alcolica

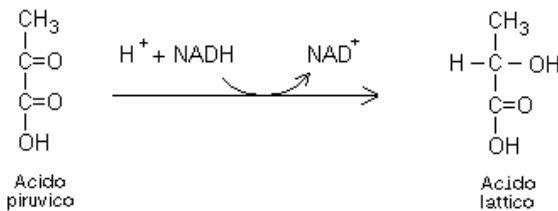


Fig. 5 - Fermentazione lattica

2 Il metabolismo cellulare

Sotto la denominazione di **metabolismo cellulare** si riunisce l'insieme dei processi chimici necessari alla vita. In particolare, si indica come **catabolismo** il complesso di reazioni di demolizione di composti preformati al fine di ricavarne energia o materia per la sintesi di nuove sostanze, mentre si usa il termine **anabolismo** per indicare l'insieme dei processi di sintesi delle sostanze necessarie alla vita.

A livello cellulare, la produzione di energia avviene all'interno di organuli specializzati, i mitocondri, attraverso una catena di reazioni che idealmente parte dal glucosio, ma nella quale, in realtà, si inseriscono i prodotti del catabolismo di tutte le sostanze di cui l'organismo si alimenta. Viceversa, i prodotti intermedi di questa stessa catena di reazioni possono essere utilizzati dalla cellula per la sintesi di nuove sostanze (Fig. 6).

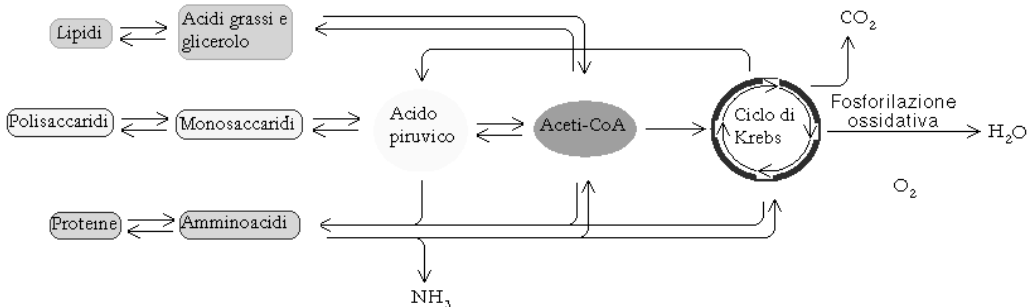


Fig. 6 - Schematizzazione del processo metabolico cellulare

3 Il processo di fotosintesi

Gli *organismi autotrofi* possiedono la capacità di sintetizzare composti organici a partire da composti inorganici operando la *fissazione del carbonio*. Nell'ambito di questa classe di organismi l'attività più rilevante si deve agli *organismi fotosintetici*, cioè quelli che svolgono questo processo sfruttando l'energia solare.

La radiazione solare che raggiunge il nostro pianeta è costituita da un insieme di **onde elettromagnetiche** [1] aventi **lunghezze d'onda** [2] diverse. La parte più cospicua di queste appartiene alla porzione visibile dello *spettro elettromagnetico*, ossia il campo di lunghezze d'onda che l'occhio umano percepisce come *luce*.

La porzione visibile dello spettro comprende le lunghezze d'onda dai 400 ai 700nm, corrispondenti a componenti di colore diverso variabili dal rosso, per le lunghezze d'onda maggiori, al violetto, per le lunghezze d'onda minori. L'insieme di queste radiazioni appare al nostro occhio come luce bianca, ma quando vengono separate una dall'altra e si propagano in direzioni diverse è possibile coglierne le diverse colorazioni: è quanto accade nell'arcobaleno. I **pigmenti** sono sostanze in grado di assorbire la luce solare: se l'assorbimento è totale questi appaiono neri; se assorbono solo alcune lunghezze d'onda, riflettendo le altre, appaiono del colore della luce riflessa.

Gli organismi fotosintetici possono utilizzare l'energia della luce solare per produrre sostanze organiche perché provvisti di appositi pigmenti: le **clorofille**. Il **processo fotosintetico**, benché assai complesso, può essere sintetizzato come il processo di formazione di molecole di glucosio e ossigeno a partire da anidride carbonica e acqua:



Tuttavia, ponendo anidride carbonica e acqua in presenza di un estratto di clorofilla la reazione precedente non avviene. Ciò significa che la capacità fotosintetica delle cellule vegetali non si deve tanto al possesso della clorofilla, quanto alla presenza di organuli specializzati: i **cloroplasti**. All'interno del cloroplasto sono visibili piccole macchie verdi, i *grani*, i quali, osservati al microscopio ottico, appaiono costituiti da pile di strutture membranose simili a sacchi appiattiti: i **tilacoidi**. Questi sono immersi in una soluzione densa, detta **stroma**, la cui composizione differisce dalla soluzione interna ai tilacoidi.

L'attività delle molecole di clorofilla è dovuta al loro inserimento nella membrana del tilacoide, dove, associate a proteine, formano complessi sistemi di sostanze, chiamati **fotosistemi**. Tale associazione permette, fra l'altro, di spostare la lunghezza d'onda del massimo assorbimento della clorofilla e di ottenere due tipi di fotosistemi, il *fotosistema I* o P_{700} e il *fotosistema II* o P_{680} , i cui massimi di assorbimento si collocano, rispettivamente, a 680 e 700nm. I fotosistemi consentono l'immediato trasferimento dell'energia solare assorbita a molecole reattive che portano avanti la reazione di fotosintesi.

a) Reazioni luce-dipendenti

La prima fase del processo di fotosintesi, o **fase luminosa** (Fig. 7), comprende reazioni che devono avvenire in presenza di luce. L'energia luminosa viene assorbita dal fotosistema II, in cui una molecola di clorofilla *a* (P_{680}) perde due elettroni, cedendoli a un accettore primario. La stessa molecola riacquista poi i due elettroni persi da una molecola d'acqua, che si ossida generando ossigeno molecolare e due protoni.

[1] L'onda elettromagnetica è costituita da un campo elettrico e da un campo magnetico oscillanti, con andamento sinusoidale su piani ortogonali tra loro.

[2] La lunghezza d'onda rappresenta la distanza tra due picchi successivi dell'onda sinusoidale.

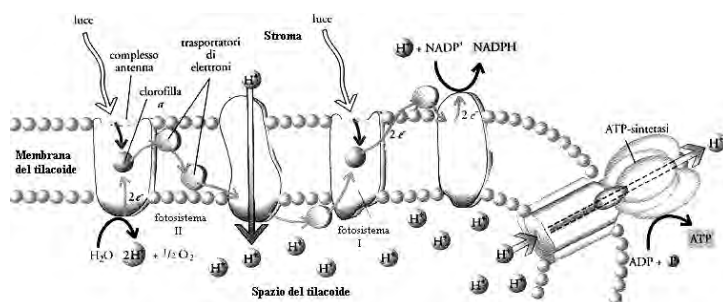
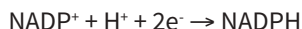


Fig. 7 - «Fase luminosa» del processo di fotosintesi

Attraverso una serie di trasportatori intermedi i due elettroni persi dal fotosistema II entrano nel fotosistema I. Qui, analogamente a quanto accade nel fotosistema II, in seguito all'assorbimento di energia luminosa, una molecola di clorofilla *a* (P_{700}) perde due elettroni, i quali, attraverso una seconda catena di trasportatori intermedi, vengono ceduti al NADP^+ , un accettore simile al NAD^+ , che si riduce secondo la semireazione:



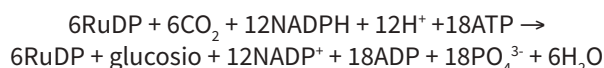
I due elettroni persi dalla clorofilla del fotosistema I vengono sostituiti con quelli provenienti dal fotosistema II. L'energia liberata dalla discesa degli elettroni lungo le catene di trasporto è utilizzata per trasferire protoni dallo stroma alla soluzione interna ai tilacoidi, generando un gradiente elettrochimico. Questo viene sfruttato dai complessi proteici ATP-sintetasi inseriti nella membrana del tilacoide, che utilizzano l'energia liberata dalla ridiscesa dei protoni secondo il gradiente di concentrazione per trasformare molecole di ADP in ATP. Il passaggio di due elettroni dal fotosistema II al fotosistema I determina la formazione di una molecola di ATP, secondo un processo denominato **fotofoosforilazione**.

b) Reazioni luce-indipendenti

Nella seconda fase della fotosintesi le molecole di NADPH e ATP prodotte dalle reazioni luce-dipendenti forniscono l'energia necessaria per ridurre le molecole di anidride carbonica, formando *zuccheri semplici*. Questa seconda fase del processo fotosintetico è detta **fase oscura**, in quanto può avvenire anche in assenza di luce, purché siano disponibili molecole di NADPH e ATP.

La serie di reazioni che porta alla riduzione dell'anidride carbonica avviene nello stroma dei cloroplasti ed è nota come **ciclo di Calvin** (Fig. 8). La prima tappa del ciclo consiste nell'aggiunta di una molecola di anidride carbonica a una molecola di zucchero a cinque atomi di carbonio, il *ribulosio difosfato* (RuDP), reazione catalizzata dall'enzima *RuDP-carbossilasi*, seguita dall'immediata scissione dell'intermedio a sei atomi di carbonio in due molecole di acido fosfoglicerico, contenente tre atomi di carbonio. Per questo motivo il ciclo di Calvin viene indicato anche come **via del C_3** .

Ogni giro completo determina la riduzione di una molecola di anidride carbonica e la rigenerazione del ribulosio difosfato. Occorrono sei giri del ciclo per formare due molecole di fosfogliceraleide, utilizzate come materiale di partenza per la sintesi di zuccheri, amminoacidi e grassi. La reazione complessiva può essere scritta come:



L'enzima *RuDP-carbossilasi* è in grado di catalizzare la reazione di carbossilazione del RuDP solo in presenza di una sufficiente concentrazione di anidride carbonica, mentre, se questa è

bassa rispetto alla concentrazione di ossigeno, lo stesso enzima catalizza la reazione tra ossigeno e RuDP innescando il processo di **fotorespirazione**, che abbassa il rendimento della fotosintesi. Questa condizione si verifica frequentemente per le piante che vivono in ambienti caldi e asciutti. Infatti, gli organismi fotosintetici terrestri assorbono l'anidride carbonica dall'aria, attraverso aperture dette *stomi*. Nei climi caldi e secchi, per limitare la perdita d'acqua, le piante chiudono gli stomi, provocando la diminuzione dell'anidride carbonica disponibile e l'accumulo dell'ossigeno prodotto dalla fotosintesi.

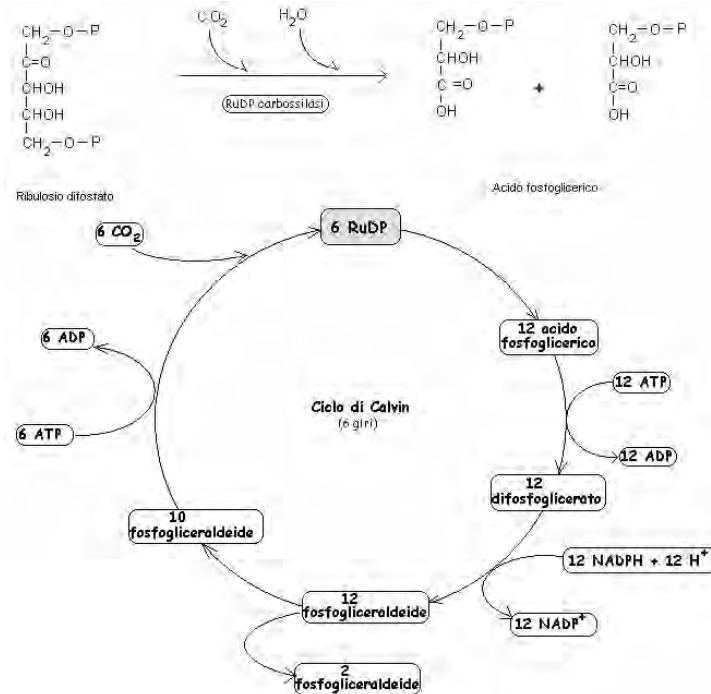


Fig. 8 - Schematizzazione del ciclo di Calvin

Queste condizioni rendono rilevante il processo di fotorespirazione, quindi alcune piante hanno sviluppato meccanismi alternativi per la cattura di anidride carbonica. Tra questi vi è la **via del C₄**, schematizzata in figura 9, che vede come prima tappa della fissazione del carbonio

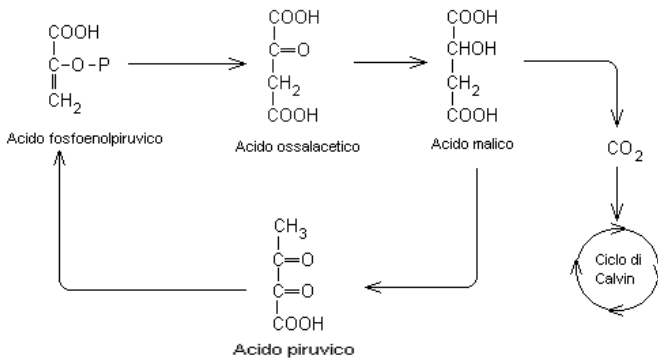


Fig. 9 - Schematizzazione della via del C₄

la reazione tra acido fosfoenolpiruvico (PEP) e anidride carbonica, con la formazione di una molecola a quattro atomi di carbonio, l'**acido ossalacetico**. La PEP-carbossilasi possiede una maggiore affinità per l'anidride carbonica rispetto alla RuDP-carbossilasi, quindi riesce a catalizzare efficacemente il processo di fotosintesi anche in condizioni di bassa concentrazione di CO₂. A sua volta l'acido ossa-

lacetico si trasforma in *acido malico* o in *acido aspartico*, che successivamente libera l'anidride carbonica destinata ad entrare nel ciclo di Calvin.

Un ulteriore meccanismo di cattura dell'anidride carbonica è quello utilizzato da piante, denominate **CAM** (*Crassulacean Acid Metabolism*), che aprono gli stomi solo di notte, quando la traspirazione è limitata dalla minore temperatura e dall'elevata umidità. Queste fissano l'anidride carbonica in composti acidi, per poi liberarla e utilizzarla durante il giorno allo scopo di alimentare il ciclo di Calvin, analogamente a quanto avviene nella via del C_4 .

1 Divisione cellulare e riproduzione asexuata

Abbiamo visto come le dimensioni della cellula siano limitate dalla necessità di avere una sufficiente superficie di scambio. Per questo la cellula, raggiunto un certo limite di accrescimento, si avvia alla divisione cellulare, processo che costituisce la diretta espressione di una importantissima proprietà del DNA, la sua **replicabilità**, cioè la capacità di originare da sé stesso una molecola identica.

La *divisione cellulare* determina la formazione, a partire da una cellula madre, di due cellule figlie simili tra loro e simili alla cellula che le ha originate. Al termine della divisione cellulare, infatti, le due cellule figlie ripartiscono tra loro le componenti citoplasmatiche della cellula madre e ciascuna di esse possiede un corredo completo di tutte le informazioni ereditarie contenute nel nucleo, o nel cromosoma per le cellule procariote.

Per gli *organismi unicellulari* la riproduzione cellulare comporta l'aumento del numero degli individui e costituisce, quindi, una via riproduttiva. Talvolta ciò avviene anche per organismi più complessi: gli anemoni di mare o le spugne, ad esempio, possono frammentarsi e generare nuovi individui, così come molte piante producono nuovi individui a partire da radici o stoloni. Questo tipo di riproduzione, che avviene attraverso la trasmissione dal genitore al figlio di una copia esatta del corredo cromosomico, è detta **riproduzione asexuata**.

Per gli *organismi pluricellulari* che si riproducono per via sessuata la riproduzione cellulare rappresenta da un lato la via attraverso la quale il nuovo individuo si accresce a partire da un'unica cellula, l'uovo fecondato, dall'altro il mezzo che consente il mantenimento dell'integrità dell'organismo adulto grazie alla riproduzione di cellule usurate o danneggiate.

a) Divisione nei procarioti

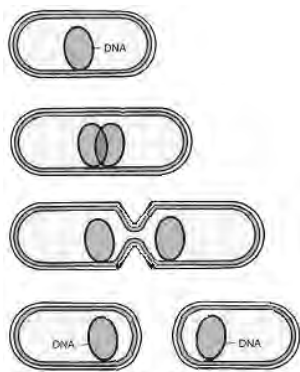


Fig. 1 - Divisione cellulare nei procarioti

Per le *cellule procariote*, la divisione cellulare è un processo piuttosto agevole e rapido grazie alla semplicità del materiale ereditario che esse contengono, costituito da un'unica, grande molecola di DNA circolare, immersa nel citoplasma cellulare. La divisione cellulare ha inizio con la replicazione del materiale ereditario da cui hanno origine due cromosomi, ciascuno dei quali resta unito ad un punto diverso della membrana cellulare. L'aumento delle dimensioni cellulari determina un progressivo allontanamento dei due cromosomi, fino alla loro completa separazione (Fig. 1).

A questo punto la membrana cellulare si inflette fino a generare una strozzatura e, risaldandosi a sé stessa, determina la formazione di due cellule separate. Nei procarioti la divisione cellulare è estremamente veloce, tanto che una cellula batterica può arrivare a dividersi ogni sei minuti.

b) Divisione negli eucarioti

La *cellula eucariote* contiene circa mille volte la quantità di DNA di una cellula procariote e ciò rende la divisione cellulare un processo più lungo e complesso. Le cellule eucariote vivono una regolare sequenza di accrescimento e divisione detta **ciclo cellulare**, composto da un'**interfase**, corrispondente al periodo di tempo compreso tra due divisioni cellulari successive e comprendente le fasi *G1*, *S* e *G2*, e dalla **divisione cellulare** vera e propria, a sua volta articolata in *mitosi* e *citodieresi*.

Il compimento di un ciclo può richiedere da poche ore a parecchi giorni, a seconda del tipo di cellula e delle condizioni di crescita. Infatti, prima di avviarsi alla divisione, la cellula deve duplicare il proprio DNA, produrre le proteine associate a questo nei cromosomi, generare una quantità di organuli sufficiente ad essere ripartita tra le cellule figlie: tutte attività che vengono svolte dalla cellula durante l'interfase.

Nel corso della **fase S**, o «fase di sintesi», viene duplicato il DNA e si producono le proteine associate. Nella **fase G2** vengono ultimati i preparativi per la divisione: nel nucleo cellulare i cromosomi si spiralizzano e diventano distinguibili; nel contempo, si comincia a preparare il *fuso*, la struttura che dirigerà la divisione cellulare.

La **fase G1** segue immediatamente la divisione cellulare e costituisce un momento di intensa attività della cellula, la quale deve accrescere notevolmente le proprie dimensioni, aumentando la quantità di citoplasma e organuli fino a poter avviare una nuova divisione cellulare. È in questa fase che si verifica la duplicazione di mitocondri e cloroplasti, che possiedono un cromosoma proprio e possono originarsi esclusivamente per duplicazione.

Come si osserva in figura 2, alcune cellule continuano ad accrescersi e a dividersi in modo continuo nel corso della vita, come le cellule staminali del midollo osseo umano, che si dividono continuamente per dare origine ai globuli rossi, mentre altre perdono questa capacità con la maturità, come accade, ad esempio, alle cellule nervose. La regolazione del ciclo cellulare e del ritmo di riproduzione delle cellule è fondamentale per conservare l'integrità degli organismi, che può essere compromessa in eguale misura da una riproduzione troppo lenta o da una troppo veloce.

Infine, la **mitosi** costituisce la serie di eventi mediante i quali ogni cellula figlia viene provvista di un corredo cromosomico completo e viene formalmente suddivisa in quattro fasi: *profase*, *metafase*, *anafase*, *telofase* (Fig. 3).

Nella **profase**, la più lunga tra le fasi del processo mitotico, i cromosomi sono già duplicati e spiralizzati ed appaiono costituiti da una coppia di *cromatidi*, ovvero due copie del cromosoma originario, unite tra loro in un punto

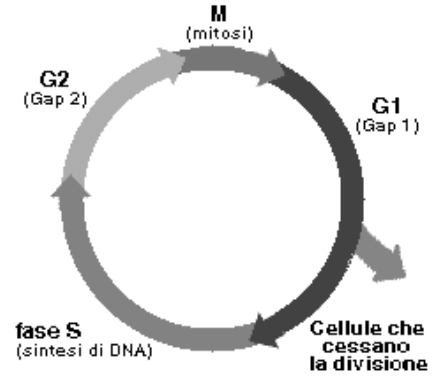


Fig. 2 - Divisione cellulare negli eucarioti

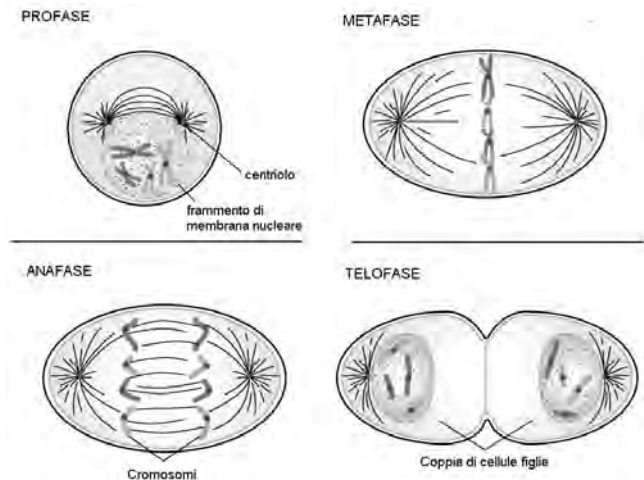


Fig. 3 - Fasi della mitosi

detto **centromero**. In prossimità di questo si trovano strutture proteiche chiamate **cinetocori**, che servono ad ancorare i cromosomi alle fibre del fuso. Durante la profase, il citoscheletro viene disassemblato e i microtubuli sono utilizzati per costruire le fibre del fuso: in questa fase, infatti, il citoplasma cellulare diventa più viscoso e la cellula assume un aspetto tondeggianti. Le coppie di centrioli, se presenti, migrano gradualmente ai poli della cellula, tra i quali si formano le **fibre polari**, che collegano i poli cellulari alla zona equatoriale. Dai centrioli si irradia un altro gruppo di fibre, denominate **aster**, il cui compito è quello di collegare ciascun polo del fuso con la membrana cellulare; infine, si formano le **fibre del cinetocore**. Al termine della profase anche la membrana nucleare viene riassorbita e i cromosomi sono immersi nel citoplasma cellulare.

Nella successiva **metafase** i cromosomi vengono progressivamente tirati dalle fibre del cinetocore e portati sul piano equatoriale della cellula.

Nell'**anafase** i cromatidi di ogni coppia si separano, divenendo cromosomi indipendenti, e vengono tirati dalle fibre del cinetocore verso i poli opposti della cellula.

Nella **telofase**, infine, il fuso comincia a dissolversi e intorno ai due corredi cromosomici completi iniziano a formarsi le membrane nucleari. I cromosomi si disperdono e il nucleo assume il tipico aspetto che possiede in una cellula non in divisione. Ricompaiono i nucleoli e a fianco di ciascun nucleo inizia la formazione di nuovi centrioli.

Generalmente la mitosi è seguita dalla **citodieresi**, ovvero dalla divisione della cellula in due parti pressappoco uguali. Nelle cellule animali la citodieresi ha inizio durante la telofase, quando sulla superficie cellulare compare un solco dovuto al restringimento della cellula lungo la circonferenza equatoriale. Responsabile di questa strozzatura è un cordone di filamenti di actina che, restringendosi progressivamente, determina la divisione della cellula. Nelle cellule vegetali, invece, sulla superficie equatoriale si raccoglie un insieme di vescicole prodotte dall'apparato di Golgi contenenti polisaccaridi, che, fondendosi tra loro, formano la piastra cellulare, delimitata da membrane su entrambi i lati. La fusione della piastra con la membrana cellulare provoca la definitiva divisione delle due cellule. Lo strato polisaccaridico che costituisce la piastra si impregna di pectina, trasformandosi nella lamella mediana sulla quale, successivamente, ciascuna delle due cellule depositerà cellulosa per formare la propria parete cellulare.

2 La riproduzione sessuata

a) Aploidia e diploidia

La maggior parte degli organismi pluricellulari, insieme a molti unicellulari, si riproduce per **via sessuata**, un tipo di riproduzione che richiede il contributo di due individui genitori e avviene attraverso due eventi fondamentali: la *meiosi* e la *fecondazione*. Studiando in dettaglio il contenuto del nucleo cellulare ci si è accorti che ogni organismo possiede un numero di cromosomi caratteristico della propria specie: nelle **cellule somatiche**, cioè quelle del corpo, tale numero è sempre pari e viene detto numero **diploide** o **2n**, mentre nelle cellule sessuali o **gameti** si riduce alla metà ed è denominato numero **aploide n**.

Le *cellule diploidi* hanno un *corredo cromosomico doppio*, poiché ciascun cromosoma possiede un omologo che gli somiglia per forma e dimensioni, e contiene lo stesso tipo di informazioni. Le *cellule sessuali*, invece, possiedono un *corredo cromosomico singolo*, in cui compare un solo cromosoma per ogni coppia di omologhi. Quando due cellule sessuali si fondono per generare lo **zigote**, ovvero la prima cellula di un nuovo individuo, il corredo cromosomico torna ad essere doppio e ciascuna coppia di omologhi è costituita da un cromosoma di origine paterna e uno di origine materna.

Il processo di formazione dei gameti prende il nome di **meiosi** e costituisce la via attraverso la quale il numero dei cromosomi viene ridotto da diploide ad aploide. Ciò consente di bilanciare gli effetti della fecondazione, che altrimenti provocherebbe il raddoppiamento del numero dei cromosomi.

Nei diversi organismi la meiosi e la fecondazione avvengono in momenti differenti del ciclo vitale, determinando una diversa alternanza delle forme aploide e diploide (Fig. 4).



Fig. 4 - Meiosi e fecondazione in organismi differenti

Negli animali la meiosi avviene immediatamente prima della fecondazione e l'individuo trascorre la vita nello stato diploide; protisti e funghi, invece, effettuano la meiosi subito dopo la fecondazione, quindi il ciclo vitale si compie nello stato aploide; nelle piante, infine, si verifica una vera e propria alternanza di individui aploidi (gametofiti) e diploidi (sporofiti), fenomeno indicato come «alternanza di generazioni».

b) Meiosi

La *meiosi* è un tipo particolare di divisione cellulare, in cui da un nucleo di partenza si generano, attraverso due divisioni nucleari successive, indicate come *meiosi I* e *meiosi II*, quattro nuclei, ognuno dei quali contiene un corredo cromosomico dimezzato, quindi un solo cromosoma per ogni coppia di omologhi.

Gli eventi essenziali della meiosi, schematizzata in figura 5, hanno luogo durante l'interfase e la profase della prima divisione meiotica.

- Nell'**interfase**, i cromosomi si duplicano, così che ognuno sia costituito da una coppia di cromatidi.
- Nella **profase I**, le coppie di cromosomi omologhi si uniscono e originano strette associazioni, dette **tetradi** perché formate da quattro cromatidi. All'interno della tetrade, i cromatidi dei due omologhi s'intrecciano in più punti e questo rende possibile lo scambio di segmenti di cromatidi tra cromosomi omologhi. Tale fenomeno, denominato **crossing-over**, costituisce un importante mezzo di ricombinazione del materiale genetico proveniente dai due genitori. Nel frattempo si formano le fibre del fuso e scompaiono il nucleolo e la membrana nucleare.
- Nella **metafase I**, le tetradi, legate alle fibre del fuso tramite i cinetocori, si dispongono sul piano equatoriale della cellula.

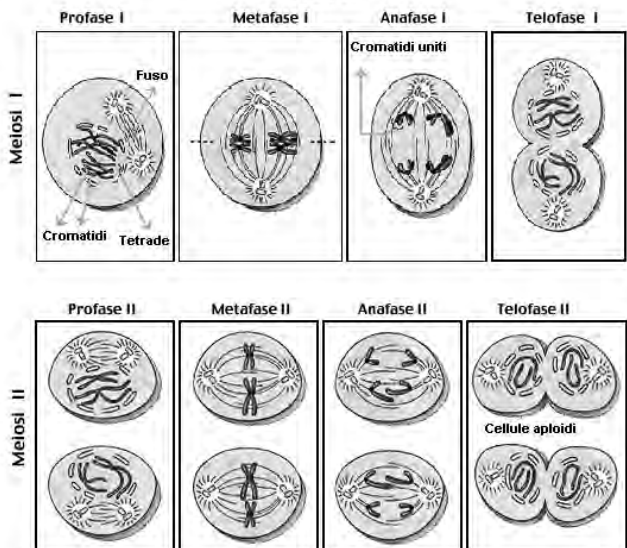


Fig. 5 - Fasi della meiosi

- Nell'**anafase I**, ciascuno dei cromosomi omologhi, ancora costituito da una coppia di cromatidi, viene tirato ad uno dei poli della cellula.
- Nella **telofase I**, ai due poli si sono formati due corredi cromosomici singoli, contenenti un solo cromosoma per ogni coppia di omologhi. A causa del *crossing-over*, tali cromosomi possono essere diversi da quelli della cellula d'origine. In alcuni casi, si assiste alla formazione delle membrane nucleari e alla citodieresi; in altri, la divisione prosegue direttamente attraverso la meiosi II, eventualmente preceduta da una breve interfase.
- Nella **profase II**, si crea un nuovo fuso e vengono riassorbite le membrane nucleari, nel caso si siano formate.
- Nella **metafase II**, i cromosomi, ancora costituiti da coppie di cromatidi, si dispongono sul piano equatoriale delle due cellule figlie.
- Nell'**anafase II**, i cromatidi si separano e vengono tirati ai poli opposti dalle fibre del fuso.
- Nella **telofase II**, infine, il fuso viene disassemblato e intorno ai due corredi cromosomici si formano le membrane nucleari. Segue poi la citodieresi, che porta alla formazione di quattro cellule aploidi a partire dall'unica cellula diploide iniziale.

La meiosi è dunque il risultato di due divisioni successive, la prima delle quali è *riduzionale*, perché determina la riduzione del numero dei cromosomi alla metà del numero iniziale, mentre la seconda è *equazionale*, poiché lascia invariato il numero di cromosomi, procedendo esattamente come una divisione mitotica.

Come evidenziato in figura 6, le quattro cellule aploidi generate dalla meiosi sono tutte diverse fra loro, in quanto il fenomeno del *crossing-over* determina un rimescolamento del materiale genetico tra coppie di cromosomi omologhi.

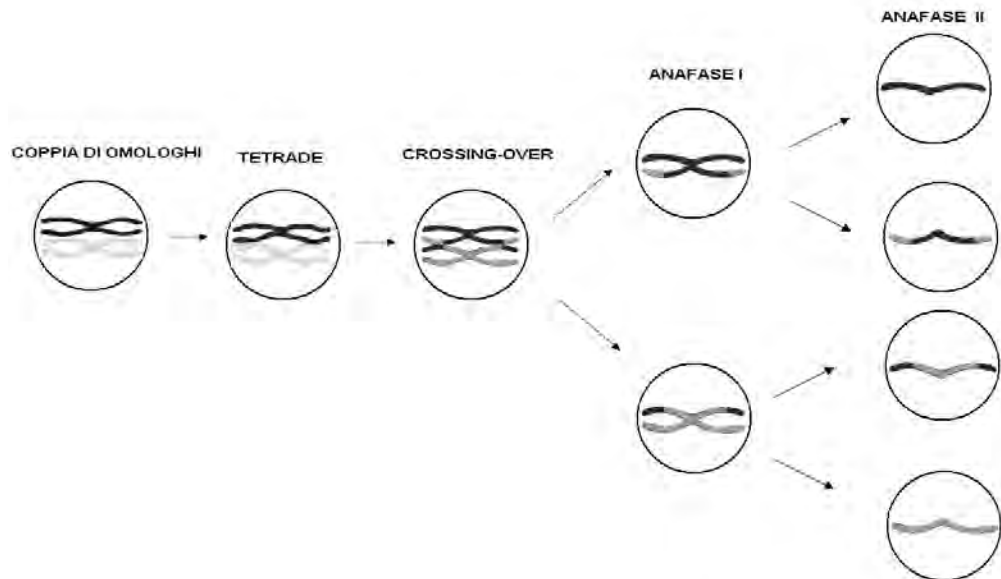


Fig. 6 - Cellule aploidi generate dalla meiosi

Genetica mendeliana e genetica classica

1 Le leggi di Mendel

Verso la metà dell'Ottocento **Gregor Johan Mendel**, monaco e biologo ceco-austriaco, avviò una serie di esperimenti sugli incroci di piante, utilizzando come «laboratorio» il giardino della sua abbazia. Per condurre questi esperimenti Mendel scelse le piante di pisello, che offrivano il vantaggio di essere economiche e facilmente reperibili, di crescere rapidamente, ma soprattutto di possedere una serie di caratteri ben distinguibili, che si tramandano inalterati da una generazione a quella successiva. In queste piante, infatti, le strutture riproduttive restano sempre racchiuse nei petali del fiore e quindi, normalmente, la pianta si autoimpollina.

Fra i caratteri mostrati dalle piante di pisello, Mendel scelse i sette riportati nella tabella seguente, poiché avevano la peculiarità di manifestarsi in due modalità antagoniste nettamente distinguibili, e selezionò **piante di linea pura** per tali caratteri, ossia piante che per molte generazioni avevano mostrato un determinato carattere sempre nella stessa modalità.

Carattere	Dominante	Recessivo
forma del seme	liscio	rugoso
colore del seme	giallo	verde
posizione del fiore	assiale	terminale
colore del fiore	porpora	bianco
forma del baccello	gonfio	sgonfio
colore del baccello	verde	giallo
altezza del fusto	alto	basso

Tab. 1 - I caratteri delle piante di pisello selezionati da Mendel

Su queste piante di linea pura Mendel eseguì incroci artificiali, asportando le antere di un fiore e cospargendone lo stigma con il polline di un fiore di un'altra varietà. Incrociando linee pure di piante dai semi gialli con linee pure di piante dai semi verdi (generazione parentale **P**), nella prima generazione filiale **F₁** Mendel ottenne unicamente individui dal seme giallo, mentre l'altro carattere sembrava scomparso. Tale comportamento, studiato su un gran numero di individui, si mostrò ripetitivo per ciascuno dei sette caratteri selezionati e permise allo studioso di enunciare la cosiddetta **legge della dominanza**, nota anche come *principio di uniformità degli ibridi di prima generazione*:

«Incrociando individui di linea pura che manifestano un certo carattere in modalità antagoniste, la prima generazione filiale mostra quel carattere soltanto in una delle due modalità, mentre l'altra sembra essere scomparsa».

Mendel chiamò «**dominanti**» le modalità manifestate nella prima generazione di incroci, «**recessive**» quelle che invece sembravano scomparire. Inoltre, lasciando che gli incroci di prima generazione si autoimpollinassero, nella seconda generazione filiale **F₂** si riscontrava la ricomparsa dei caratteri recessivi posseduti dalla generazione parentale P. Ripetendo questi

esperimenti per ciascuno dei sette caratteri separatamente, attraverso l'analisi statistica dei risultati ottenuti in un gran numero di prove, il biologo poté stabilire che i caratteri recessivi ricomparivano nella generazione filiale F_2 con una frequenza costante, pari al 25%: in pratica, un individuo su quattro mostrava il carattere recessivo. Tali risultati condussero Mendel a una brillante intuizione, enunciata nella **legge della segregazione**:

«I caratteri manifestati dal singolo individuo sono determinati da una coppia di fattori discreti, ereditati singolarmente, uno dalla madre e l'altro dal padre. Questi fattori si separano all'atto della formazione dei gameti e solo uno di essi viene trasferito al figlio».

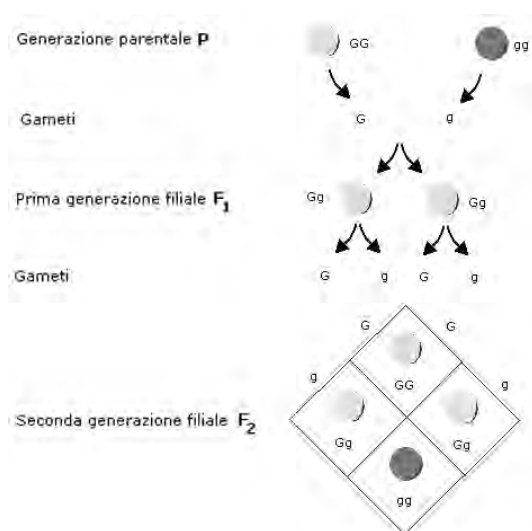


Fig. 1

Nella prima generazione filiale F_1 , tutti gli individui ereditano uno dei fattori dalla madre e l'altro dal padre, per poi manifestare, tra i due, il carattere dominante. Tuttavia, ciascuno di essi possiede, mascherato, anche il carattere recessivo, ragion per cui quando i due fattori si separano all'atto della formazione dei gameti, ricombinandosi a caso nella generazione successiva, originano individui in cui possono ricomparire i caratteri recessivi, come illustrato in figura 1. Quello riportato in figura è il cosiddetto **quadrato di Punnett**, che permette di visualizzare facilmente le possibili combinazioni di caratteri.

In seguito, furono denominati **geni** i fattori che determinano un carattere, **alleli** le diverse modalità con cui questi si possono

presentare. I due alleli che determinano un certo carattere sono indicati con la *stessa lettera: maiuscola per i dominanti, minuscola per i recessivi*.

Gli individui di linea pura incrociati da Mendel, che mostrano l'invariabilità di un carattere per molte generazioni, sono dotati evidentemente di due alleli uguali: le piante di linea pura con semi gialli possiedono la coppia di alleli GG, mentre le piante di linea pura con semi verdi possiedono la coppia di alleli gg. Simili individui sono detti **omozigoti**.

Gli ibridi di prima generazione, ereditando un allele da un genitore e uno dall'altro, possiedono alleli diversi (Gg) e sono detti **eterozigoti**. L'individuo eterozigote manifesta, dei due alleli, solo quello dominante e ciò determina l'uniformità degli ibridi di I generazione.

Esiste, quindi, una distinzione tra il **genotipo** e il **fenotipo** di un individuo: il *primo rappresenta il patrimonio genetico*, ossia il tipo di alleli che l'individuo possiede; il *secondo corrisponde alla parte del patrimonio genetico che viene effettivamente espressa e che determina le caratteristiche esteriori dell'individuo*. L'autoimpollinazione degli ibridi di prima generazione provoca la ricomparsa del carattere recessivo, perché i gameti formati dagli ibridi possiedono per metà l'allele G, per l'altra metà l'allele g. Nella generazione F_2 , gli omozigoti dominanti e gli eterozigoti evidenzieranno il carattere dominante «seme giallo», mentre gli omozigoti recessivi mostreranno il carattere recessivo «seme verde». La distribuzione statisticamente osservata è facilmente interpretabile in base all'analisi delle possibili combinazioni dei gameti.

La conferma all'ipotesi di Mendel è fornita dal test del **reincrocio** o **test-cross**. Per verificare che un carattere sia effettivamente determinato da una coppia di alleli che si separano nella formazione dei gameti, si può effettuare un reincrocio tra piante della generazione F_1 (eterozigoti) con piante di linea pura dai semi verdi (omozigoti recessivi). Questo test permette di stabilire il genotipo del

genitore con fenotipo dominante: infatti, se questo fosse un omozigote, tutti i figli avrebbero fenotipo dominante, mentre essendo un eterozigote i figli mostrano per il 50% il carattere dominante, per l'altro 50% il carattere recessivo, come si può facilmente verificare costruendo il *quadrato di Punnett* in figura 2.

Dopo aver studiato separatamente la trasmissione del singolo carattere, Mendel si dedicò a verificare quale rapporto esistesse tra la trasmissione di più caratteri contemporaneamente, selezionando linee pure di piante che differivano per più di un carattere.

Incrociando linee pure di piante dai semi gialli e lisci (GGRR) con linee pure di piante dai semi verdi e rugosi (ggrr), nella prima generazione filiale si riscontrava nuovamente l'uniformità degli ibridi, ottenendo tutte piante con semi gialli e lisci, cioè con fenotipo dominante. L'autoimpollinazione degli individui della prima generazione filiale produceva, per la maggior parte, individui con fenotipo dominante (315 individui con seme giallo e liscio), nonché una piccola frazione di individui con fenotipo recessivo (32 con seme verde e rugoso) e determinava, nello stesso tempo, la comparsa di due nuovi fenotipi, dovuti a due diversi riassortimenti dei caratteri: si ottenevano, infatti, sia piante con semi gialli e rugosi (101 individui), sia piante con semi verdi e lisci (108). Il rapporto esistente tra le diverse combinazioni di caratteri era 9:3:3:1, risultato coerente con quanto osservato nelle esperienze precedenti, a condizione di assumere che i due caratteri venissero ereditati dalla prole in modo del tutto indipendente. Ciò permise a Mendel di formulare la **legge del riassortimento indipendente**:

«Nella formazione dei gameti, le coppie di fattori che determinano caratteri diversi segregano (cioè si separano) indipendentemente le une dalle altre».

Infatti, esaminando le possibili combinazioni dei vari tipi di gameti che possono formarsi in seguito a una divisione indipendente e casuale dei fattori che determinano i due caratteri considerati, si ottiene la distribuzione mostrata in figura 3.

Si osservi come ogni parte del quadrato di Punnett rappresenti una combinazione avente probabilità identica a tutte le altre: le possibili combinazioni dei gameti sono 16; di queste 9 producono piante con semi gialli lisci, 3 producono piante con semi

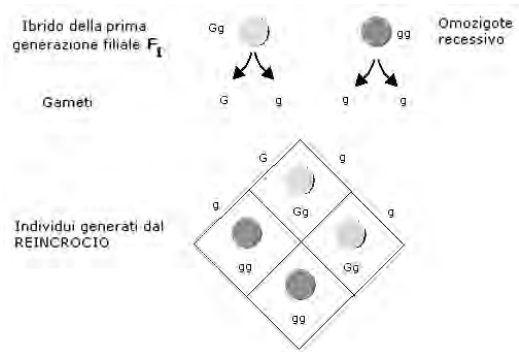


Fig. 2

Generazione parentale P

Gameti

Prima generazione filiale F_1

Gameti

Seconda generazione filiale F_2

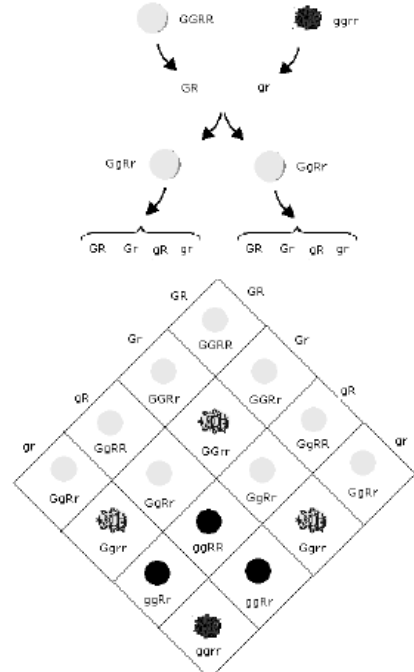


Fig. 3

gialli rugosi, 3 con semi verdi e lisci, mentre una sola combinazione produce piante con semi verdi rugosi. Quindi, i 9/16 (56,25%) degli individui ottenuti nella seconda generazione filiale devono avere semi gialli lisci, i 3/16 (18,75%) semi gialli rugosi, altri 3/16 (18,75%) semi verdi e lisci, e solo 1/16 (6,25%) semi verdi rugosi. Queste percentuali sono in buon accordo con quanto calcolato da Mendel nei suoi esperimenti.

Totale piante: 315 + 108 + 101 + 32 = 556			
% gialli-lisci	$\frac{315}{556} \cdot 100 = 56,65\%$	% verdi-lisci	$\frac{108}{556} \cdot 100 = 19,42\%$
% gialli-rugosi	$\frac{101}{556} \cdot 100 = 18,16\%$	% verdi-rugosi	$\frac{32}{556} \cdot 100 = 5,75\%$

Tab. 2

2

La riscoperta del lavoro di Mendel

Mendel, considerato a ragione il *padre della genetica*, completò e pubblicò il suo lavoro nel 1865. Purtroppo, però, i suoi studi vennero ignorati dalla comunità scientifica del tempo e solo agli inizi del Novecento, quando la microscopia permise di entrare più a fondo nel mondo della cellula e di arrivare ad osservare la divisione cellulare, i risultati da lui ottenuti furono riscoperti e rivalutati.

Nel 1902 **Walter S. Sutton**, studiando la divisione delle cellule spermatiche della cavalletta, notò la somiglianza dei cromosomi appaiati durante la divisione meiotica e riscontrò come fosse possibile individuare cromosomi aventi forme molto simili tra loro anche durante la metafase della mitosi. Tale osservazione chiarì immediatamente il significato di ciò che aveva affermato Mendel: i cromosomi dovevano essere i portatori dei geni e la coppia di alleli di cui egli parlava risiedeva su cromosomi omologhi, i quali si separano durante la meiosi e si ricombinano attraverso la fecondazione.

Quest’ipotesi spiegava in modo soddisfacente la «legge della segregazione», mentre incontrava qualche difficoltà con la «legge del riassortimento indipendente». Infatti ogni cromosoma porta informazioni per molti caratteri e la loro trasmissione alla generazione successiva si verifica nelle modalità osservate da Mendel solo se i geni esaminati sono situati su coppie distinte di cromosomi.

Tuttavia, non sempre la trasmissione dei caratteri è così semplice: i fenomeni di *crossing-over* osservabili durante la meiosi possono determinare un’alterazione della frequenza di alcune ricombinazioni di caratteri rispetto ad altre, al punto che diversi studiosi accusarono Mendel di avere in qualche modo «limato» i propri risultati. Inoltre, non sempre tra le diverse modalità di manifestazione di un carattere esiste un così ben definito rapporto di dominanza-recessività: in effetti, le caratteristiche che Mendel aveva scelto di studiare erano state accuratamente selezionate in modo da poter essere sempre riconosciute nettamente nel passaggio da una generazione all’altra, ma in realtà esistono casi più complessi.

3

Interazioni alleliche e geniche

Non sempre si osserva una netta relazione di dominanza di un allele sull’altro, ma esistono numerosi tipi di **interazioni alleliche** che si verificano quando entrambi gli alleli vengono espressi contemporaneamente negli eterozigoti, i quali, di conseguenza, possiedono un fenotipo proprio con caratteristiche comuni ad entrambi gli omozigoti o intermedie tra queste.

La **dominanza incompleta** rappresenta il caso in cui l'eterozigote mostra un fenotipo intermedio tra quello dei due genitori omozigoti. Un esempio è costituito dalle «bocche di leone»: incrociando piante dai fiori rossi con piante dai fiori bianchi si ottengono eterozigoti con fiori di colore rosa, in cui i due caratteri di partenza sembrano essersi mescolati. In realtà il fenotipo evidenziato dall'eterozigote è il frutto dell'espressione contemporanea dei due alleli e non di un loro effettivo rimescolamento, tanto che nella generazione ottenuta per autofecondazione degli eterozigoti questi tornano ad essere fenotipicamente ben distinti tra loro, come mostrato in figura 4.

Si parla, invece, di **codominanza** quando i due alleli risultano espressi contemporaneamente nell'eterozigote che possiede le caratteristiche di entrambi gli omozigoti, senza tuttavia arrivare ad avere un fenotipo intermedio.

Esistono casi di **allelia multipla**, cioè di geni a cui possono corrispondere più di due forme alleliche: ovviamente l'individuo diploide possiede solo una coppia di alleli, ma questi, nell'ambito della popolazione, esistono in più forme.

I **gruppi sanguigni** umani costituiscono un caso di codominanza e di allelia multipla. Infatti, gli alleli ad essi corrispondenti sono di **tre tipi (A, B, 0)** e determinano la presenza di particolari polisaccaridi sulla superficie dei globuli rossi e di specifici anticorpi nel sangue. Questi tre alleli implicano l'esistenza di **quattro gruppi sanguigni (A, B, AB, 0)**, poiché gli alleli A e B, oltre ad essere dominanti sull'allele 0, sono codominanti tra loro e vengono espressi contemporaneamente.

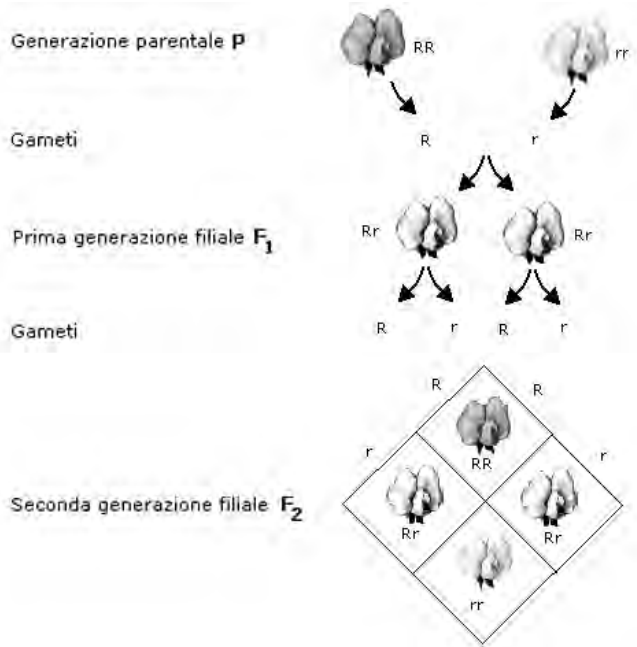


Fig. 4

I gruppi sanguigni umani			
Fenotipo	Genotipo	Polisaccaridi di superficie	Anticorpi nel plasma
0	00	—	anticorpo A anticorpo B
A	AA A0	polisaccaride A	anticorpo B
B	BB B0	polisaccaride B	anticorpo A
AB	AB	polisaccaride A polisaccaride B	—

Tab. 3

Oltre alle interazioni alleliche di cui abbiamo parlato, esistono anche **interazioni geniche**. Talvolta accade che l'espressione di un gene sia dipendente da quella di altri, fenomeno denominato **epistasia**. La **sordità** nell'uomo è il frutto dell'interazione di due geni e si verifica ogni

volta che l'individuo è omozigote recessivo per uno dei due: l'udito normale richiede la presenza di almeno un allele dominante per ognuno dei due geni.

Altri tipi di interazioni geniche si verificano perché diverse volte un carattere è determinato da più di un gene, così come capita spesso che uno stesso gene eserciti la propria influenza su più di un carattere.

Nel primo caso si parla di **ereditarietà poligenica**, tipica di caratteri, come la *statura* e il *peso*, determinati da numerosi geni contemporaneamente, per i quali non è osservabile una distinzione netta tra fenotipi diversi, ma piuttosto una variazione continua di quel carattere nell'ambito della popolazione.

Nel secondo caso si parla, invece, di **pleiotropia**, fenomeno sicuramente osservabile per geni che influenzano la produzione di sostanze dotate di una molteplicità di funzioni nell'organismo.

4 Interazioni con l'ambiente e mutazioni

A complicare ulteriormente la trasmissione dei caratteri ereditari interviene anche l'**interazione tra geni e ambiente**. L'espressione del genotipo di un individuo può essere fortemente influenzata dall'ambiente esterno: si pensi, ad esempio, ai conigli dell'Himalaya, che normalmente possiedono orecchie, zampe, naso e coda di colore nero, ma se allevati a temperature superiori ai 30°C hanno il mantello completamente bianco, a parità di genotipo. Analogamente, l'ambiente presente durante lo sviluppo dell'embrione (temperatura, pH, concentrazione di sostanze di varia natura) può influenzare notevolmente l'espressione genica.

Un'ulteriore, sorprendente scoperta fu fatta agli inizi del Novecento da **Hugo de Vries**, il quale, studiando l'ereditarietà dei caratteri sulla «rapunzia europea», riscontrò la comparsa improvvisa di caratteri nuovi, assenti nei genitori e negli antenati della pianta. Questo biologo olandese attribuì il fenomeno a improvvise trasformazioni dei geni, ereditabili dai discendenti, da lui definite **mutazioni**, e chiamò **mutanti** gli individui che recavano tali cambiamenti. Nella «rapunzia europea» de Vries individuò ben 2.000 mutazioni, delle quali, però, solo poche erano realmente tali, mentre la maggior parte era riconducibile alle ricombinazioni dei caratteri avvenute con la *meiosi*, i *crossing-over* e la *fecondazione*. La sua scoperta fu, comunque, di radicale importanza, poiché aprì uno spiraglio alla comprensione delle fonti di variabilità a partire dalle quali è possibile parlare di **evoluzione biologica**.

5 La teoria cromosomica dell'ereditarietà

La **trasmissione dei caratteri ereditari** è una realtà con cui inevitabilmente ci si confronta e il lavoro di Mendel aveva aperto la strada alla comprensione dei meccanismi con cui essa avviene. Tuttavia, per anni, i genetisti lavorarono sull'argomento senza che ci fosse la reale consapevolezza di «dove» risiedessero i geni e di «cosa» fossero concretamente.

Un'ipotesi sulla localizzazione fisica dei geni fu avanzata dal già citato citologo americano **W.S. Sutton**, le cui tesi, però, dovettero essere confermate da lavori successivi prima di essere prese seriamente in considerazione.

I primi cinquant'anni del Novecento furono segnati da un intenso lavoro nei campi della genetica e della citologia, che portò ad acquisire il bagaglio di conoscenze proprio della **genetica classica**.

a) Cromosomi sessuali

Un filone di ricerca che produsse prove a favore della *teoria cromosomica dell'ereditarietà* fu quello dedicato allo studio della *determinazione del sesso e delle caratteristiche ad esso legate*.

Attraverso l'osservazione di cellule al microscopio si scoprì che negli organismi diploidi tutti i cromosomi sono presenti in coppie di omologhi, gli **autosomi**, simili per forma e dimensione. Solo una coppia, quella dei **cromosomi sessuali**, evidenzia differenze tra individui maschi e femmine: negli uccelli il maschio ha i due cromosomi sessuali uguali (ZZ), mentre la femmina li ha diversi (ZW); **nei mammiferi, quindi anche nell'uomo, la femmina possiede i due cromosomi sessuali uguali (XX), mentre il maschio ha due cromosomi diversi (XY)**. Pertanto, nella specie umana i gameti femminili possiedono tutti il cromosoma X, mentre quelli maschili possiedono per il 50% il cromosoma X e per l'altro 50% il cromosoma Y. Il sesso del nuovo individuo è, quindi, determinato dal tipo di spermatozoo che feconda la cellula uovo. Il fatto che un carattere evidente come il sesso fosse legato a differenze cromosomiche fu una prova forte a sostegno della tesi secondo cui i geni risiedono effettivamente sui cromosomi.

b) Caratteri legati al sesso

Un gruppo molto attivo nelle ricerche in campo genetico fu quello guidato da **T.H. Morgan** presso la Columbia University, negli Stati Uniti. Morgan e i suoi collaboratori si dedicarono allo studio della *trasmissione dei caratteri ereditari* utilizzando il «moscerino della frutta» (*Drosophila melanogaster*), un organismo che si nutre di lieviti in fermentazione e cresce sulla frutta marcescente, dotato di occhi di colore rosso. Allevando colonie di moscerini, Morgan ottenne

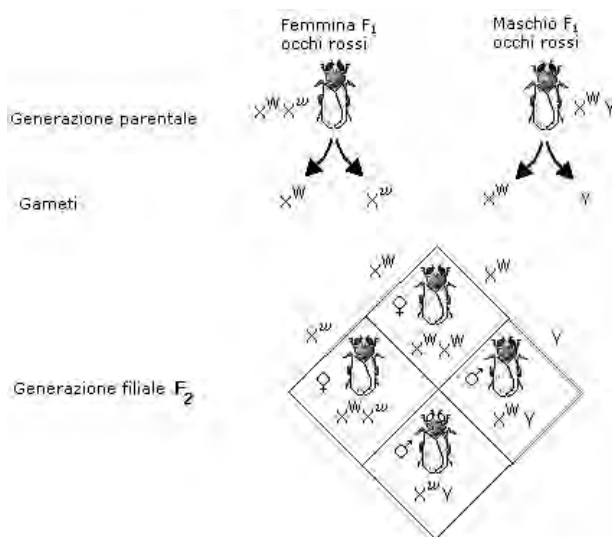


Fig. 6

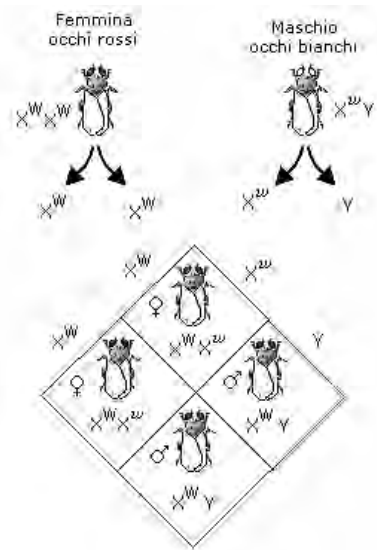


Fig. 5

un mutante, un moscerino con occhi bianchi di sesso maschile. Incrociando questo mutante con una femmina con occhi rossi, nella generazione F_1 nacquero tutti individui con occhi rossi, da cui fu possibile dedurre che il carattere «occhi bianchi» fosse recessivo (Fig. 5).

Procedendo poi all'incrocio di individui della F_1 (Fig. 6), Morgan ottenne, nella generazione F_2 , 2.459 femmine con occhi rossi, 1.011 maschi con occhi rossi, 782 maschi con occhi bianchi e nessuna femmina con occhi bianchi, risultati non conformi a quanto atteso secondo le leggi di Mendel, in cui era sorprendente l'assenza di femmine con occhi bianchi.

Infine, eseguendo un *test-cross* (Fig. 7), cioè incrociando il mutante maschio dagli occhi bianchi con una femmina della F_1 , si ottennero sia individui con occhi rossi, sia individui con occhi bianchi, anche se con una prevalenza di quelli con occhi rossi, equamente distribuiti tra maschi e femmine. In particolare: 129 femmine con occhi rossi, 132 maschi con occhi rossi, 88 femmine con occhi bianchi, 86 maschi con occhi bianchi.

Tali risultati indussero Morgan e i suoi collaboratori a concludere che il gene per il colore degli occhi è presente solo sul cromosoma X. L'allele «occhi bianchi» è recessivo, visto che nella F_1 non ci sono individui con occhi bianchi e quindi una femmina eterozigote ha occhi rossi: per questo motivo nella generazione F_2 non ci sono femmine con occhi bianchi. Al contrario un maschio possiede un solo allele per il colore degli occhi sul cromosoma X, cosicché se eredita l'allele «occhi bianchi» manifesta questo carattere. Ulteriori prove confermarono tale ipotesi, evidenziando anche una minore sopravvivenza dei moscerini con occhi bianchi rispetto a quelli con occhi rossi, segno che questo gene influenza anche altre caratteristiche.

Tali studi portarono quindi al concetto di **caratteri legati al sesso**, da un lato confermando che i geni sono collocati sui cromosomi, dall'altro avvalorando la tesi secondo cui ogni cromosoma porta informazioni riguardanti numerosi caratteri.

c) Gruppi di associazione e mappe cromosomiche

Si deve a **H.J. Muller**, uno dei collaboratori di Morgan, la scoperta che la frequenza delle mutazioni nella *Drosophila* potesse essere aumentata dai raggi X, dalla luce ultravioletta e da alcune sostanze chimiche. In tal modo fu più facile produrre individui portatori di mutazioni e studiare, su questi, gli incroci di moscerini che differivano per più di un carattere. Da questo tipo di studi emerse l'esistenza di **gruppi di associazione**, ossia gruppi di caratteri che tendono ad essere trasmessi insieme, poiché evidentemente risiedono sullo stesso cromosoma.

Inoltre si osservò che nella *Drosophila* e in altri organismi il numero dei gruppi di associazione è pari al numero di coppie cromosomiche. Tuttavia, anche per gli alleli appartenenti allo stesso gruppo di associazione potevano osservarsi piccole percentuali di **ricombinazione** dei caratteri. Nel caso della *Drosophila*, ad esempio, la maggior parte dei moscerini mostrava corpo di colore marrone e ali lunghe (caratteri dominanti). Incrociando individui omozigoti per questi caratteri con mutanti dal corpo di colore nero ed ali corte, in base alle ipotesi avanzate fino ad allora, se i due geni si fossero trovati su cromosomi diversi avrebbero dovuto segregare indipendentemente, producendo nuovi assortimenti di caratteri con i rapporti di combinazione 9:3:3:1 riscontrati da Mendel. Qualora i due geni si fossero trovati sullo stesso cromosoma, sarebbero stati trasmessi insieme ai gameti, producendo individui con assortimenti uguali a quelli dei genitori, nel rap-

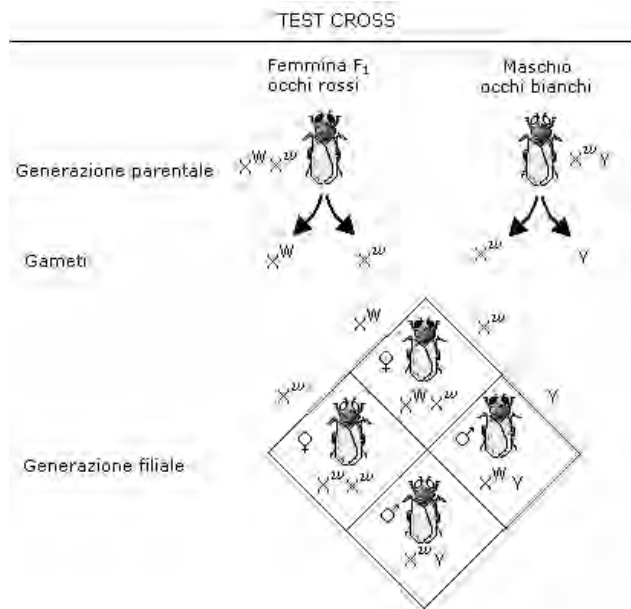


Fig. 7

porto di 3:1. I risultati ottenuti erano molto vicini a questo secondo comportamento, anche se si poteva osservare una piccola percentuale di individui con caratteri ricombinati, cioè moscerini con corpo marrone ed ali corte oppure con corpo nero ed ali lunghe.

Per spiegare questo fenomeno Morgan ipotizzò che potessero avvenire scambi di alleli anche tra cromosomi omologhi, cosa che effettivamente avviene con il *crossing-over* nella profase della meiosi I.

Uno scambio di materiale genetico così preciso tra cromosomi omologhi, che permetteva di sostituire un allele esattamente con il proprio omologo, evidenziava come ogni gene avesse nel cromosoma una ben definita posizione, detta *locus*. Inoltre, ogni coppia di geni appartenenti allo stesso cromosoma mostrava una propria frequenza di ricombinazione, che poteva essere correlata alla distanza tra i rispettivi *loci* sul cromosoma. Supponendo che i geni abbiano una sequenza lineare lungo il cromosoma, quelli che occupano *loci* molto vicini mostrano una frequenza di ricombinazione, tramite *crossing-over*, minore rispetto a quelli i cui *loci* si trovano a maggiore distanza tra loro.

In base a questa ipotesi si cominciarono a costruire le prime **mappe cromosomiche**, in cui la distanza tra due geni che presentano una frequenza di ricombinazione pari all'1% veniva assunta pari ad 1 unità.

1 La natura chimica dell'informazione ereditaria: DNA o proteine?

Le prime indagini sulla *natura chimica dei cromosomi*, ormai individuati come vettori dell'informazione ereditaria, sollevarono un problema ben preciso: analizzando il cromosoma eucariote si riscontrava la presenza di *acido deossiribonucleico* (DNA) e proteine in quantità pressappoco uguali. Il DNA, un polimero abbastanza semplice, formato da sole quattro basi azotate, sembrava essere inadeguato a conservare l'informazione ereditaria ed inizialmente molti ricercatori ritenevano che i geni fossero di natura proteica e che i cromosomi contenessero gli stampi utili per copiare proteine ed enzimi necessari alla vita della cellula. L'individuazione del DNA come depositario dell'informazione ereditaria richiese tempo e, come sempre, il contributo di molti ricercatori.

a) Esperimenti di Griffith

Esperimenti fondamentali in tal senso furono quelli di **Frederick Griffith**, un batteriologo inglese che, nel 1928, condusse una serie di studi sullo *pneumococco* con l'intento di realizzare un vaccino per una grave forma di polmonite batterica.

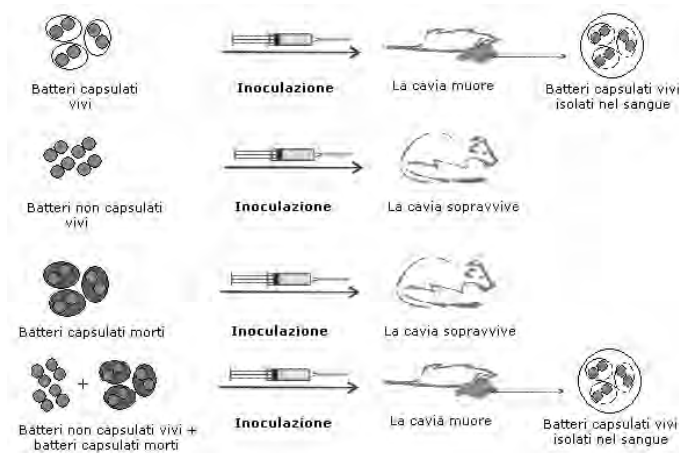


Fig. 1 - Schematizzazione degli esperimenti di Griffith

Gli pneumococchi esistono sia in forme virulente (ceppo S), caratterizzate dal possesso di una capsula polisaccaridica, sia in forme non virulente (ceppo R) prive di capsula. Questa caratteristica è geneticamente determinata, quindi ereditabile.

Gli esperimenti di Griffith sono riassunti nella figura 1: inoculando batteri capsulati vivi (S) nelle cavie, queste si ammalavano, morivano e nel loro sangue erano isolabili batteri capsulati vivi; inoculando batteri del ceppo R, non

virulento, o batteri del ceppo S uccisi col calore, le cavie non si ammalavano; mentre inoculando una miscela di batteri non virulenti vivi e batteri virulenti morti le cavie si ammalavano nuovamente e nel loro sangue potevano essere isolati batteri capsulati vivi. Evidentemente qualcosa, che Griffith chiamò **fattore trasformante**, era in grado di trasferirsi dai batteri morti a quelli vivi, conferendo loro una caratteristica ereditaria che prima non possedevano.

Nel 1943 **O.T. Avery** e i suoi collaboratori, dopo lunghi anni di lavoro, scoprirono che il fattore trasformante ipotizzato da Griffith era costituito da DNA. Tale scoperta permetteva di identificare il DNA come «depositario dell'informazione ereditaria», anche se furono necessarie ulteriori conferme perché questa idea venisse definitivamente accettata.

b) Esperimenti con batteriofagi

Negli anni Quaranta del Novecento ebbe inizio una serie di studi sui **batteriofagi** o **fagi**, un gruppo di virus in grado di parassitare i batteri, determinandone la *lisi*. I fagi, facilmente riconoscibili al microscopio grazie alla loro forma, simile ad uno spillo, possiedono una testa più o meno esagonale, seguita da un collare che termina in una coda sfrangiata in 5 o 6 fibre, utilizzate per fissarsi alla superficie dei batteri. All'analisi chimica questi virus risultavano costituiti unicamente da DNA e proteine.

Alfred Hershey e **Martha Chase** progettarono allora un esperimento attraverso il quale stabilire quale delle due sostanze costituisse i geni virali capaci di determinare la sintesi di nuovi virus all'interno della cellula ospite. Prepararono due campioni di virus, uno marcato con ^{32}P , l'altro con ^{35}S , con cui infettarono due colture separate di batteri. Una volta iniziato il *ciclo infettivo*, le cellule batteriche venivano frullate e centrifugate per separarle dal materiale extracellulare. L'analisi del materiale extracellulare ed intracellulare evidenziò che all'interno del batterio si trovava solo il ^{32}P , mentre lo ^{35}S rimaneva all'esterno della cellula. Considerando che le proteine contengono zolfo e non fosforo, mentre per gli acidi nucleici vale il contrario, questi esperimenti permisero di dedurre che è il DNA virale ad entrare nel batterio, provocando la produzione di nuovi virus, e che questo evidentemente contiene tutta l'informazione ereditaria. In pratica il fago, fissatosi alla superficie della cellula batterica, inocula al suo interno l'acido nucleico, lasciando fuori il capsido proteico.

c) Ulteriori conferme a favore del DNA

A sostegno di quanto scoperto da Griffith, Hersey e Chase giunsero poi i risultati di altre ricerche: **A.E. Mirsky**, dedicandosi allo studio delle *cellule somatiche* e *sessuali* di diversi organismi, evidenziò che generalmente le cellule somatiche di un organismo possiedono una quantità di DNA all'incirca doppia rispetto a quella presente nei gameti.

Erwin Chargaff, invece, dedicandosi all'analisi quantitativa delle basi azotate contenute nel DNA di diverse specie, evidenziò come queste siano presenti in eguali proporzioni in tutte le cellule di tutti gli individui di una stessa specie, mentre le proporzioni cambiano passando da una specie all'altra. Inoltre, la porzione di guanina risulta sempre uguale a quella di citosina e quella di adenina risulta sempre uguale a quella di timina: dato, questo, che successivamente si sarebbe rivelato di notevole importanza.

2 La struttura del DNA: modello di Watson e Crick

Nell'elaborare un *modello di struttura per il DNA*, **J.D. Watson** e **Francis Crick** partirono anzitutto da un riesame dei dati disponibili fino a quel momento: il DNA risultava essere una molecola lunga e filiforme, di alto peso molecolare, costituita da nucleotidi. L'analisi quantitativa di **Chargaff** aveva evidenziato l'uguaglianza tra le quantità di adenina e timina, e tra quelle di guanina e citosina. Inoltre, gli studi ai raggi X effettuati da **Rosalind Franklin** e **Maurice Wilkins** avevano evidenziato che il DNA possiede la struttura di una grande elica, analogamente a quanto riscontrato da **L.C. Pauling** per le molecole proteiche.

Inoltre, la molecola depositaria dell'informazione genetica doveva essere in grado di assumere una varietà di strutture sufficiente a codificare l'enorme quantità di informazioni che i viventi si tramandano, così come, nello stesso tempo, doveva essere dotata di un preciso meccanismo di duplicazione, tale da consentire la trasmissione dell'informazione da una cellula all'altra.

Fu così che Watson e Crick arrivarono ad ipotizzare per il DNA un **modello a doppia elica**, in cui le strutture portanti corrispondono a filamenti costituiti dall'alternanza di *unità di zucchero* e di un *gruppo fosfato*.

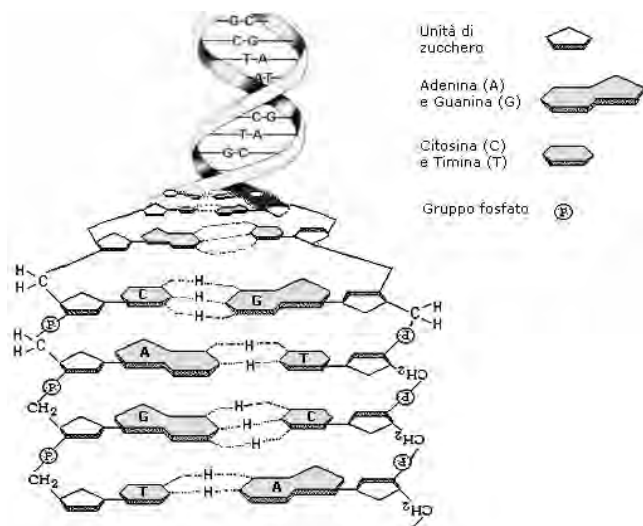


Fig. 2 - Modello di Watson e Crick

Tale distanza sarebbe stata superata se l'accoppiamento fosse avvenuto tra due purine e non sarebbe stata raggiunta dall'accoppiamento di due pirimidine.

Le coppie di basi *adenina-timina* e *guanina-citosina* risultavano complementari tra loro e ciò spiegava i dati quantitativi ottenuti da Chargaff. I due filamenti dell'elica sono quindi **complementari**, in quanto dalla struttura di uno è possibile dedurre quella dell'altro. Inoltre, poiché ogni gruppo fosfato è legato da un lato con il carbonio 5 di uno zucchero e dall'altro con il carbonio 3 di un altro zucchero, in ogni filamento si può distinguere un'estremità 5' e una 3'. I due filamenti della doppia elica si dicono **antiparalleli** perché corrono in senso inverso, uno dal 5' al 3' e l'altro dal 3' al 5'.

Una catena di DNA può arrivare a contenere migliaia di basi azotate e l'enorme varietà delle sequenze di basi possibili consente di avere quella varietà di strutture che si riteneva indispensabile per la molecola depositaria dell'informazione ereditaria.

3 La duplicazione del DNA

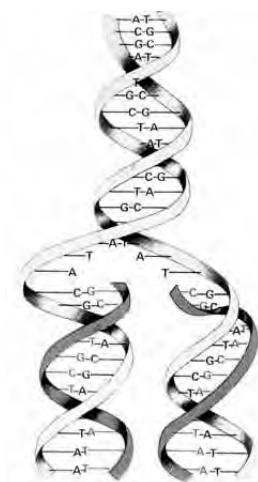


Fig. 3 - Duplicazione del DNA

Come mostrato in figura 2, alle unità di zucchero sono legate le basi azotate, che, nella doppia elica, si accoppiano come a formare i pioli di una scala a chiocciola. I due studiosi, costruendo un modello in ferro e stagno del DNA, si resero conto che gli accoppiamenti tra le basi azotate non potevano essere casuali, ma dovevano avvenire sempre tra adenina e timina (che si legano attraverso due legami idrogeno) e tra guanina e citosina (che stabiliscono tre legami idrogeno), in modo da mantenere una distanza costante tra i due filamenti (pari a 2nm), come evidenziato dagli studi della Franklin.

La struttura a doppia elica descritta nel paragrafo precedente rende conto anche di come la molecola di DNA possa produrre facilmente e precisamente una copia di sé stessa. La duplicazione del DNA rappresenta l'evento portante della duplicazione dei cromosomi, attraverso il quale da una molecola di DNA se ne producono due identiche a quella di partenza (Fig. 3).

Il processo è **semiconservativo** perché, al momento della duplicazione, le basi azotate si separano, determinando l'apertura della doppia elica ed ogni filamento funziona da stampo per la costruzione di un nuovo filamento. In tal modo, ciascuna delle due molecole di DNA prodotte possiede un filamento della molecola originaria e uno di nuova sintesi.

Sotto il profilo biochimico il meccanismo di duplicazione è molto complesso e chiama in causa un gran numero di enzimi. La duplicazione prende sempre avvio da una sequenza specifica di nucleotidi, detta **punto di origine della duplicazione**, in cui proteine di attivazione spezzano i legami idrogeno tra le basi azotate. In tal

modo si forma una **bolla di duplicazione** caratterizzata da due estremità aventi struttura ad Y, dette *forcelle di duplicazione*.

La duplicazione è detta **bidirezionale** perché la sintesi dei nuovi filamenti avviene a partire da entrambe le forcelle di duplicazione, determinando il progressivo allargamento della bolla. Nel cromosoma procariote vi è un unico punto d'inizio e la duplicazione si completa in seguito all'allargamento di un'unica bolla. Nei cromosomi degli eucarioti, invece, ci sono numerosi punti d'inizio e ogni bolla si ingrandisce progressivamente fino a fondersi con quelle adiacenti.

La sintesi del nuovo filamento ha luogo sotto il controllo dell'enzima DNA-polimerasi, che, muovendosi lungo il filamento stampo, aggiunge al nuovo filamento un nucleotide per volta. L'enzima è dotato di un sistema di **correzione di lettura** che gli consente di aggiungere nucleotidi al nuovo filamento solo se tutte le basi azotate precedenti sono correttamente accoppiate. In caso di errore la DNA-polimerasi inverte il proprio senso di marcia sul filamento stampo per correggere l'errore. Ciò assicura una precisa duplicazione della molecola, sottoposta all'ulteriore controllo di **enzimi riparatori** in grado di tagliare la doppia elica là dove ci sono errori e di sostituire i nucleotidi errati con quelli corretti.

4 Dal DNA alle proteine: la sintesi proteica

a) Un gene, una proteina

Dopo aver stabilito che il DNA costituisce la molecola depositaria dell'informazione ereditaria e una volta determinata la sua struttura, restava da capire in che modo l'informazione fosse codificata al suo interno e come potesse essere trasmessa. La stretta relazione tra il DNA di un organismo e le proteine che questo produce era stata messa in evidenza più volte.

Già nel 1908 un medico inglese, **Archibald Garrod**, aveva avanzato l'ipotesi che molte malattie di natura ereditaria fossero dovute a carenze enzimatiche, ipotesi implicante l'assunto che i *geni* influenzassero la produzione degli *enzimi*.

Nel 1941 **G.W. Beadle** ed **E.L. Tatum** iniziarono una serie di studi sulla muffa del pane, di cui produssero un certo numero di mutanti utilizzando i raggi x. Effettuando la *mappatura genetica* di alcuni mutanti, furono in grado di dimostrare che una certa mutazione determinava la perdita di funzione di uno specifico enzima; ne derivò l'ipotesi che ogni gene fosse responsabile della produzione di un enzima o, più in generale, di una *proteina*.

Anche **Pauling** si muoveva sulla stessa linea, sostenendo che l'anemia falciforme avesse origine da anomalie nella struttura dell'emoglobina. A supporto di quest'ipotesi analizzò, tramite **elettroforesi** [1], campioni di sangue di individui affetti da anemia falciforme (omozigoti recessivi), portatori sani (eterozigoti) ed individui sani (omozigoti per l'allele dominante), evidenziando come l'emoglobina di individui sani e malati avesse un comportamento elettroforetico diverso, quindi struttura differente. I portatori sani possiedono entrambi i tipi di emoglobina ed evidentemente in questi individui l'emoglobina normale è presente in quantità sufficiente affinché la malattia non si manifesti. Di lì a pochi anni fu dimostrato che la differenza tra le due molecole di emoglobina era dovuta alla sostituzione di 2 soli amminoacidi su 600.

Ulteriori prove a sostegno della relazione esistente tra DNA e proteine vennero anche dagli studi sui batteriofagi, dai quali emerse che, in seguito all'infezione virale, all'interno del batterio ha luogo la sintesi del DNA e delle proteine del rivestimento virale. L'insieme di questi dati rendeva più che attendibile l'ipotesi che fosse il DNA a portare l'informazione necessaria alla sintesi delle proteine, da cui la nascita dell'espressione: «**un gene, una proteina**».

[1] Tecnica analitica e separativa basata sulla diversa velocità di migrazione in un fluido di particelle elettricamente cariche, sotto l'azione di un campo elettrico applicato.

b) RNA messaggero e processo di trascrizione

Le ricerche sulle modalità di traduzione dell'informazione codificata nel DNA condussero alla scoperta di un altro acido nucleico, l'RNA o *acido ribonucleico*, particolarmente abbondante nelle cellule che producono grandi quantità di proteine, quindi ricche di *ribosomi*, di cui l'RNA è il costituente principale. Tale molecola sembrava essere un intermedio nel processo di sintesi delle proteine a partire dal DNA: infatti, quando un batteriofago infetta una cellula batterica prima che inizi la sintesi delle proteine virali, ha luogo la sintesi di RNA. Questo ha struttura simile al DNA, con la differenza di essere costituito da un unico filamento e di contenere nei suoi nucleotidi il ribosio invece del deossiribosio e la base azotata uracile al posto della timina. Nella cellula se ne distinguono diverse tipologie:

- RNA messaggero (m-RNA);
- RNA ribosomiale (r-RNA);
- RNA di trasporto (t-RNA).

L'**RNA messaggero** è la trascrizione di un tratto di DNA che viene copiata da uno dei due filamenti: analogamente a quanto accade durante la duplicazione, i nucleotidi, presenti nella cellula in forma di trifosfati, si accoppiano con basi azotate complementari della catena di DNA e l'enzima RNA-polimerasi ne catalizza la condensazione.

Nel **processo di trascrizione**, all'adenina del DNA corrisponde l'uracile nell'RNA, mentre alla timina del DNA corrisponde l'adenina nell'RNA (Fig. 4). Il filamento di m-RNA in formazione si accresce aggiungendo nucleotidi alla sua estremità 3', poiché l'RNA-polimerasi si muove lungo il filamento di DNA dall'estremità 3' alla 5'.

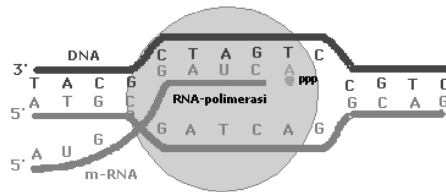


Fig. 4 - RNA messaggero

La sintesi delle molecole di m-RNA avviene a partire da specifiche sequenze di nucleotidi dette **promotori** (che costituiscono siti di legame per l'RNA-polimerasi) e termina in corrispondenza delle cosiddette **sequenze di terminazione**. Il DNA costituisce dunque la matrice originale, sempre disponibile nel nucleo della cellula, a partire dalla

quale vengono prodotte copie utilizzate direttamente per la sintesi delle proteine, che si svolge nel citoplasma cellulare.

c) Traduzione del codice genetico

Giunti a questo punto restava da comprendere come il DNA e l'RNA, costituiti dal concatenamento di quattro tipi di nucleotidi diversi, potessero codificare le informazioni necessarie alla sintesi di catene proteiche, costituite dal concatenamento di 20 tipi diversi di amminoacidi. Evidentemente doveva esistere una corrispondenza tra una sequenza di nucleotidi e i singoli amminoacidi: supponendo che ogni amminoacido fosse codificato da una coppia di nucleotidi, le quattro basi azotate avrebbero prodotto solo $4 \times 4 = 16$ combinazioni, insufficienti a codificare per i 20 amminoacidi. L'ipotesi più semplice, rivelatasi poi quella corretta, era che i nucleotidi necessari a codificare il singolo amminoacido fossero tre e, in tal caso, il numero di combinazioni sarebbe stato $4 \times 4 \times 4 = 64$, numero più che sufficiente a codificare per 20 amminoacidi.

Lavorando su questa ipotesi, **Marshall Nirenberg** e **Heinrich Matthaei** condussero i primi esperimenti che avrebbero portato a *decifrare il codice genetico*. Nirenberg aveva osservato che aggiungendo ad estratti cellulari di *Escherichia coli* tratti di RNA di varia origine e amminoacidi marcati radioattivamente si poteva riscontrare la formazione di proteine radioattive, segno che

gli estratti cellulari erano in grado di condurre la sintesi delle proteine anche a partire da un RNA estraneo alla cellula. I due scienziati sintetizzarono quindi molecole di RNA artificiale contenenti una sola base azotata, ad esempio l'uracile. La molecola ottenuta, che chiamarono *poli-U*, posta in presenza di estratti cellulari e di uno dei 20 amminoacidi, dava luogo alla formazione di *catene polipeptidiche* solo con l'aggiunta di fenilalanina. Da ciò si dedusse che la tripletta di basi azotate UUU corrispondeva all'amminoacido fenilalanina. Procedendo in questo modo e utilizzando molecole di RNA con sequenze di nucleotidi ripetute, fu possibile pervenire alla completa decifrazione del codice genetico (Fig. 5). Delle 64 triplette possibili, dette **codoni**, solo tre risultano non codificanti e vengono indicate come *triplette di stop*; le altre 61, invece, codificano per i 20 amminoacidi, a ciascuno dei quali corrispondono più triplette, spesso differenti fra loro per l'ultimo nucleotide: per tale motivo il codice genetico è detto **degenere**.

		Seconda base								
		U		C		A		G		
U		UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
		UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
		UUA	Leu	UCA	Ser	UAA	<u>STOP</u>	UGA	<u>STOP</u>	A
		UUG	Leu	UCG	Ser	UAG	<u>STOP</u>	UGG	Trp	G
C		CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
		CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
		CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
		CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A		AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
		AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
		AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
		AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G		GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
		GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
		GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
		GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G
										Terza base (estremità 3')
										U
										C
										A
										G

Fig. 5 - Decifrazione del codice genetico

Una scoperta importante quanto la decifrazione del codice genetico è stata la constatazione della sua **universalità**: il codice, infatti, risulta essere lo stesso per tutti i viventi, dai *procarioti* agli *eucarioti*, e ciò rende plausibile che tutti gli esseri viventi si sono evoluti a partire da un antenato comune.

Si è detto che, oltre all'*RNA messaggero*, nelle cellule esistono altri due tipi di RNA, diversi per struttura e funzione: l'*RNA ribosomiale* e l'*RNA di trasporto*, che vengono trascritti dal loro gene specifico, presente nel DNA. L'**RNA ribosomiale** è il costituente principale dei ribosomi, organuli preposti alla sintesi delle proteine. Ogni ribosoma è formato da due subunità: una più piccola, dotata del sito di legame per l'm-RNA, e una più grande, dotata di due siti di legame per gli RNA di trasporto. L'**RNA di trasporto** (Fig. 6) è la molecola che opera la traduzione del messaggio codificato nell'm-RNA, consentendo la formazione della catena polipeptidica corrispondente. Nella cellula esistono almeno 20 molecole diverse di t-RNA, tutte dotate della caratteristica forma a trifoglio: l'estremità 3',

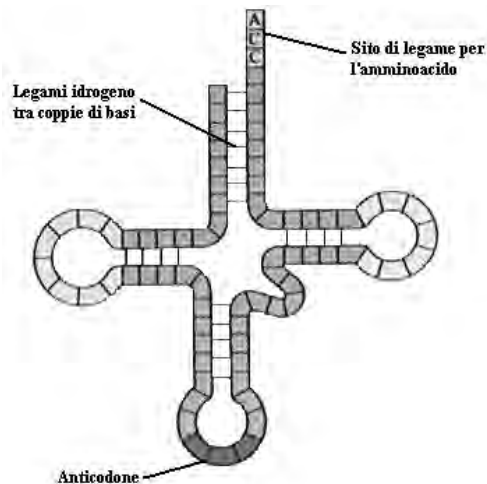


Fig. 6 - RNA di trasporto

caratterizzata dalla tripletta ACC, costituisce il sito di legame per l'amminoacido, mentre dalla parte opposta si trova una tripletta di basi, l'**anticodone**, attraverso la quale il t-RNA si lega al codone complementare posto sull'RNA messaggero.

Vi è poi una regione in cui si trova il sito di riconoscimento dell'enzima **amminoacil-tRNA-sintetasi**. Nella cellula sono contenuti almeno 20 enzimi amminoacil-tRNA-sintetasi diversi, dotati di un sito di legame per il t-RNA e di uno per l'amminoacido corrispondente. Questi enzimi assicurano che ogni molecola di t-RNA si leghi a uno specifico amminoacido. La **traduzione**, ossia il processo attraverso il quale dal filamento di m-RNA si arriva alla sintesi della proteina corrispondente, è schematicamente illustrata in figura 7. La prima fase del processo vede la formazione del cosiddetto **complesso d'inizio**, costituito dal legame tra l'unità più piccola del ribosoma, l'estremità 5' dell'm-RNA (di cui viene esposto il primo codone) e il primo t-RNA, che si appaia con questo. Una volta formato, il complesso si incastra nella subunità più grande del ribosoma, inserendo il primo t-RNA con il relativo amminoacido nel sito P (sito del peptide). Inizia così la fase di **allungamento**, in cui il secondo codone dell'm-RNA si colloca davanti al sito A (sito dell'amminoacile); qui si inserisce il secondo t-RNA con l'anticodone complementare, occupando il sito A.

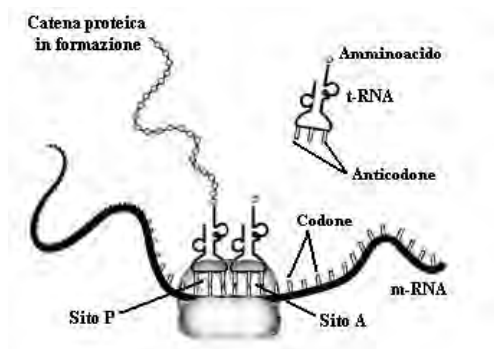


Fig. 7 - Processo di traduzione

Tra i due amminoacidi adiacenti sul ribosoma si forma il **legame peptidico**; il ribosoma si sposta lungo l'm-RNA, liberando il primo t-RNA, mentre il secondo t-RNA va ad occupare il sito P, liberando il sito A, davanti al quale viene a trovarsi il terzo codone della catena dell'm-RNA. Su di esso si inserisce il terzo t-RNA con l'anticodone complementare, cosicché si forma il legame peptidico tra il dipeptide legato al sito P e l'amminoacido legato al sito A. Il processo va avanti in questo modo fin quando, scorrendo lungo l'm-RNA, il ribosoma incontra una **sequenza di terminazione**, cioè un codone di stop. A questo punto la traduzione si interrompe, la catena polipeptidica viene liberata e le due subunità ribosomiali si separano.

Accade spesso che mentre un ribosoma procede lungo il filamento di m-RNA, un secondo ribosoma si leghi all'estremità iniziale, avviando a sua volta la traduzione: simili complessi, in cui uno stesso filamento di m-RNA viene letto contemporaneamente da più ribosomi, sono detti **polisomi** e permettono la sintesi di più molecole proteiche allo stesso tempo, a partire dalla medesima molecola di m-RNA.

5 Le mutazioni geniche

È ormai accertato che alcune **malattie genetiche** sono determinate dall'alterazione o dalla perdita di funzione di molecole proteiche. Negli individui affetti da anemia falciforme, ad esempio, l'anomalia si deve alla sostituzione, nella molecola di emoglobina, di una molecola di acido glutammico con una di valina. I codoni corrispondenti a questi due amminoacidi sono rispettivamente GAA o GAG per l'acido glutammico e GUA o GUG per la valina. Ciò significa che, in termini di DNA, la malattia è determinata dalla sostituzione di un solo nucleotide su circa 450 presenti nel gene dell'emoglobina. Simili anomalie possono verificarsi in seguito a **mutazioni geniche**, di cui aveva iniziato a parlare de Vries agli inizi del Novecento.

Si dicono **mutazioni puntuali** quelle che riguardano il singolo nucleotide. Il risultato di un tale tipo di mutazione può essere la sostituzione di un amminoacido con un altro, come nel

caso dell'anemia falciforme, con conseguente cambiamento morfologico e funzionale della proteina; in tal caso si parla di **mutazioni di senso**. Se invece la mutazione puntuale porta alla sostituzione del codone originario con un codone di stop, la proteina non viene sintetizzata e allora si parla di **mutazione non senso**. Oltre alla sostituzione di un nucleotide con un altro, possono verificarsi casi di delezione o di inserzione di nucleotidi in un gene che hanno come conseguenza uno **spostamento del sistema di lettura**. Cambiando il modo in cui i nucleotidi vengono raggruppati in triplette, la proteina prodotta può essere del tutto nuova e completamente priva di funzione.

Le mutazioni si originano in modo assolutamente casuale, indotte da fattori esterni come raggi x, raggi ultravioletti, sostanze radioattive etc., con frequenze variabili da 1/1.000 a 1/1.000.000 a seconda del gene. Quelle che hanno luogo nelle cellule somatiche si trasmettono alle cellule figlie prodotte per mitosi, mentre quelle che si verificano nei gameti vengono trasmesse alle generazioni successive. Gli alleli diversi di uno stesso gene si originano da mutazioni spontanee: l'allele dominante nell'eterozigote produce normalmente una quantità di proteina tale da compensare l'allele difettoso o recessivo, fenomeno indicato in genetica classica come **dominanza**. Solo in assenza dell'allele normale, non potendo essere prodotta la proteina normale, viene espresso il fenotipo recessivo. In alcuni casi, quando entrambi gli alleli codificano per proteine funzionali, si osservano fenomeni di **codominanza** o **dominanza incompleta**.

Un altro fenomeno osservato in genetica classica, spiegabile in modo efficace in termini di genetica molecolare, è l'*epistasia*, fenomeno per il quale alcuni caratteri di un individuo possono essere espressi solo in presenza di altri. Ciò è dovuto al fatto che normalmente le reazioni chimiche cellulari avvengono tramite sequenze biochimiche: se un individuo è omozigote recessivo per un gene che codifica per un enzima intermedio in una sequenza biochimica, venendo a mancare l'enzima, viene conseguentemente bloccata tutta la sequenza biochimica delle reazioni e tutti gli alleli normali, che codificano per gli enzimi successivi nella sequenza, vengono mascherati.

La regolazione dell'espressione genica

1 Cosa avviene nei procarioti

Nonostante che il codice genetico sia comune a tutti gli organismi, di qualunque grado di complessità, esistono enormi differenze riguardo ai meccanismi di regolazione dell'espressione genica. Il cromosoma procariote, in particolare, è costituito da un'unica, grande molecola circolare di DNA, che, srotolata, raggiunge dimensioni lineari ben superiori a quelle dell'intera cellula. Nel DNA vengono conservate tutte le informazioni utili per la produzione di proteine ed enzimi necessari alla vita e al metabolismo della cellula, e ogni segmento che codifichi per una proteina è detto **gene strutturale**. Di solito, nel cromosoma procariote, geni strutturali che codificano per proteine aventi funzioni correlate, come ad esempio gli enzimi di una stessa sequenza biochimica, si trovano raggruppati e spesso vengono trascritti in un unico filamento di m-RNA, in modo che le proteine di cui la cellula necessita contemporaneamente e in uguali quantità possano essere prodotte tutte insieme.

Tuttavia, la cellula produce enzimi e proteine codificati nel suo DNA solo in caso di necessità: ad esempio, *Escherichia coli* produce la β -galattosidasi, enzima necessario alla digestione del lattosio, solo in presenza di lattosio, appunto, condizione in cui arriva a contenerne fino a 3.000 molecole. Al contrario, in assenza di lattosio, ogni cellula contiene mediamente una molecola di β -galattosidasi. Evidentemente nella cellula esistono sistemi in grado di attivare o disattivare la trascrizione e la traduzione dei geni strutturali a seconda delle esigenze momentanee, ossia in rapporto al tipo di ambiente, alla presenza o assenza di alcuni nutrienti etc., evitando così sprechi di energia per produrre sostanze di cui non ha bisogno.

Nei procarioti la regolazione della sintesi proteica avviene essenzialmente a livello della trascrizione, secondo un meccanismo esplicato dal cosiddetto **modello dell'operone**, proposto da **François Jacob** e **Jacques Monod**. La trascrizione di un tratto di DNA inizia in seguito all'adesione della RNA-polimerasi a una sequenza specifica del DNA, detta **promotore**. Di qui, scorrendo lungo il DNA, la RNA-polimerasi copia un tratto che codifica per uno o più geni strutturali, fino ad incontrare una sequenza di terminazione. Secondo la teoria dell'operone, tra il promotore e i geni strutturali si trova una sequenza, detta **operatore**, alla quale può legarsi una proteina, a sua volta codificata da un gene **regolatore**, denominata **repressore**. Legandosi all'operatore, il repressore impedisce lo scorrimento dell'RNA-polimerasi lungo il DNA o addirittura ne evita il legame con il promotore, bloccando la trascrizione del gene strutturale. Il legame del repressore con l'operatore dipende dalla presenza di altre molecole, indicate come **corepressori** o **induttori**: i primi attivano il repressore, consentendogli di legarsi all'operatore; i secondi lo disattivano, impedendone il legame con l'operatore. Nell'*Escherichia coli*, ad esempio, la produzione di β -galattosidasi è sotto il controllo dell'operone *lac* (Fig. 1), regolato da un repressore che ha come induttore l'allolattosio, il primo prodotto del metabolismo del lattosio. In tal modo, quando *Escherichia coli* si trova in presenza di lattosio, le molecole di enzima presenti nella cellula ne avviano la digestione e l'allolattosio, prodotto come primo intermedio, si lega al repressore dell'operone *lac*, determinandone il distacco dall'operatore e avviando la trascrizione del gene strutturale della β -galattosidasi.

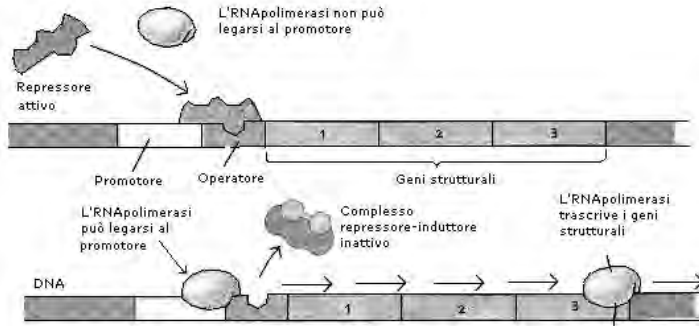


Fig. 1 - Attività dell'operone lac nell'*Escherichia coli*

2 Cosa avviene negli eucarioti

La quantità di DNA negli eucarioti è enormemente più grande che nei procarioti, anche se vi sono tantissimi segmenti ripetuti, molti dei quali apparentemente non utilizzati. Inoltre il DNA negli eucarioti mostra un'organizzazione molto più complessa, essendo strettamente associato a proteine che determinano la struttura del cromosoma, influenzando, di conseguenza, l'espressione genica.

L'associazione di DNA e proteine nel nucleo cellulare è indicata col nome di **cromatina** (per la sua sensibilità ai coloranti), di cui le proteine rappresentano oltre il 50% in peso. La maggior parte di queste è rappresentata da piccoli polipeptidi, detti **istoni**, che svolgono un ruolo fondamentale nel ripiegamento e nell'avvolgimento del DNA. La restante parte delle proteine associate al DNA varia molto da un organismo all'altro ed è costituita da enzimi, proteine di regolazione, molte delle quali ancora sconosciute. Si distinguono cinque tipi di istoni contenuti nel nucleo, quattro dei quali possiedono strutture assai simili anche in organismi sensibilmente diversi e sono presenti in quantità doppia rispetto agli istoni del quinto tipo. Tali proporzioni risultano facilmente comprensibili esaminando la struttura dell'unità fondamentale della cromatina, il **nucleosoma** (Fig. 2), costituito da una parte centrale, formata da otto molecole istoniche, due per ogni tipo, attorno alle quali è avvolto il DNA, e da una proteina istonica di tipo diverso rispetto a quelle interne, affiancata esternamente. I nucleosomi posti in sequenza costituiscono una sorta di «collana di perle», che, riavvolgendosi su sé stessa, forma una fibra del diametro di 30nm. Questa, a sua volta, si ripiega in **domini ad ansa** spiralizzati, i quali si condensano fino a costituire il cromosoma, il cui spessore arriva a 1.400nm.

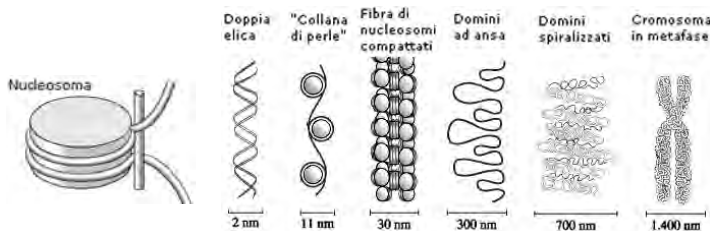


Fig. 2 - Componenti strutturali del nucleosoma

La cellula eucariote, contrariamente a quella procariote, esprime solo la piccola parte del suo DNA che viene permanentemente attivata nel corso del **differenziamento cellulare**. Questo processo consente a cellule dotate dello stesso patrimonio genetico, in quanto generate dal medesimo zigote, di differenziarsi, assumendo nell'organismo funzioni completamente diverse.

Numerose prove hanno dimostrato che l'espressione genica è strettamente collegata al grado di condensazione della cromatina. Colorando il nucleo cellulare, infatti, si possono distinguere zone in cui la cromatina è più densa, dotate di colorazione più accentuata, indicate come **eterocromatina**, e zone colorate più debolmente, caratterizzate da una maggiore dispersione della cromatina, denominate **euromatina**. Naturalmente, il processo di trascrizione e la conseguente espressione di un gene dipendono dal grado di accessibilità del gene stesso, per cui la parte di DNA che ogni cellula esprime è limitata alla porzione di euromatina.

Inoltre si è notato che, in seguito al differenziamento, il grado di condensazione di regioni diverse della cromatina varia da una cellula all'altra e il rapporto tra euromatina ed eterocromatina diminuisce, segno che ogni cellula differenziata esprime solo una piccola, specifica porzione del proprio DNA.

Infine, anche negli eucarioti sono presenti sistemi di regolazione basati su proteine come nei procarioti, sebbene dotati di un maggiore grado di complessità. Infatti, ogni gene è regolato solitamente da più proteine, alcune attivanti, altre disattivanti, i cui siti di legame sono spesso molto lontani dal gene, il che rende la loro individuazione estremamente difficoltosa.

3 Il DNA eucariote

Si è detto che la quantità di DNA negli eucarioti è nettamente maggiore rispetto a quella dei procarioti e molto di questo DNA sembra essere inutilizzato. Mentre i batteri esprimono praticamente la totalità del loro DNA, negli eucarioti meno del 10% del DNA codifica per le proteine, e nelle cellule umane questa percentuale potrebbe arrivare addirittura all'1%. Lo studio della struttura del DNA eucariote ha permesso di individuare sequenze di diversa tipologia e funzione.

Il **DNA a sequenza semplice** è formato da sequenze di 5-10 nucleotidi che si ripetono testacoda un gran numero di volte. Tali sequenze non sono codificanti, ma sembrano avere importanza per la struttura del cromosoma. Lunghi tratti di questo tipo si trovano nella zona del centromero o all'estremità dei cromosomi, dove formano una sorta di «cappuccio» di protezione che conferisce stabilità al cromosoma. Il DNA umano è costituito per il 20-30% da brevi sequenze ripetitive.

Il **DNA a sequenze ripetitive interposte** è, invece, un tipo di DNA codificante per proteine, costituito da tratti di 150-300 nucleotidi ripetuti più volte, raramente uno dopo l'altro, più spesso sparsi all'interno del DNA. Appartengono a questa classe i geni che codificano per gli istoni o per l'r-RNA, presenti nella cellula in numero di 50-500 copie. Altri esempi di sequenze ripetitive interposte sono le cosiddette *famiglie geniche*, costituite da sequenze non identiche ma molto simili tra loro, probabilmente originatesi da un gene ancestrale. Un esempio di questo tipo è dato dalla famiglia delle *globine*, in cui il gene ancestrale codificava probabilmente per una proteina simile alla mioglobina (proteina costituita da una sola catena polipeptidica con un solo gruppo *eme*), da cui, a seguito di mutazioni casuali, si sarebbe originata l'attuale famiglia genica.

Il **DNA a copia unica**, infine, corrispondente al 50-70% di tutto il DNA, rappresenta la classe alla quale appartengono tutti i geni strutturali, ossia quelli che codificano per proteine, anche se solo una piccola frazione di questo DNA (circa l'1%) viene effettivamente espressa dalla cellula.

Una sorprendente scoperta riguardo alla struttura dei geni codificanti concerne il fatto che le sequenze codificanti all'interno di un gene, dette **esoni**, non sono continue, bensì intervallate da sequenze non codificanti, denominate **introni**. Infatti, in molti casi la lunghezza del gene è maggiore di quella dell'm-RNA che viene tradotto in proteina nel citosol. La funzione degli introni non è stata ancora chiarita, ma si è osservato che spesso sono posti a separare esoni che codificano per parti di una proteina aventi funzioni distinte: ad esempio, i geni della famiglia delle globine sono formati da tre esoni e due introni, e l'esone centrale codifica per la parte della catena che è associata al gruppo *eme*.

4 I processi di trascrizione e di traduzione in procarioti ed eucarioti

Negli eucarioti il **processo di trascrizione** è simile a quello dei procarioti solo in linea generale: l'RNA-polimerasi si attacca alla sequenza iniziale (*promotore*) e, scorrendo lungo la molecola di DNA dall'estremità 3' a quella 5', sintetizza la molecola di m-RNA complementare. Al contrario di quanto accade nei procarioti, in cui più geni strutturali regolati dallo stesso operone sono trascritti e tradotti allo stesso tempo, nel cromosoma eucariote ogni gene strutturale viene regolato, trascritto e tradotto indipendentemente dagli altri. Inoltre, mentre nei procarioti la traduzione della molecola di m-RNA può avere inizio a partire dall'estremità 5' della molecola anche mentre all'estremità opposta si sta completando il processo di trascrizione, negli eucarioti trascrizione e traduzione costituiscono due processi del tutto separati sia spazialmente sia temporalmente.

Il processo di trascrizione (Fig. 3) ha luogo nel nucleo e produce una molecola di m-RNA in cui vengono copiati sia esoni sia introni. All'estremità 5' della molecola viene legata una sequenza **leader**, la cui funzione è quella di congiungere la molecola di m-RNA al ribosoma, mentre all'estremità 3' viene aggiunta una sequenza **trailer**, costituita da un filo di nucleotidi di adenina. Successivamente, prima che l'm-RNA venga esportato dal nucleo nel citoplasma, la molecola subisce una rielaborazione mediante un processo detto di **splicing**, attraverso il quale vengono eliminate le sequenze relative agli introni. Esistono numerosi esempi di molecole di m-RNA che, rielaborate attraverso *splicing* diversi, producono proteine differenti. In questi casi, in pratica, uno stesso gene strutturale viene utilizzato per produrre più di una proteina, semplicemente cambiando il sistema di lettura, cosicché quello che era un introne per una proteina può diventare un esone per l'altra, o viceversa. Tutto ciò conferma come nelle cellule eucariote il problema della regolazione e dell'espressione genica sia nettamente più complesso rispetto alle cellule procariote.

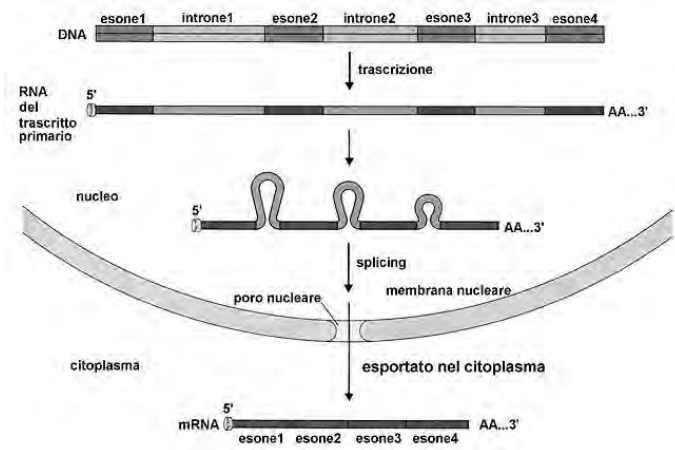


Fig. 3 - Processi di trascrizione e di splicing

5 Lo spostamento dei geni

Per molti anni i ricercatori sono stati convinti della sostanziale stabilità dei cromosomi, modificati unicamente in seguito a fenomeni di *crossing-over* o di mutazioni casuali. In realtà esistono molte altre cause che possono determinare modificazioni del DNA nei procarioti e negli eucarioti.

a) Plasmidi e coniugazione

In tutte le cellule batteriche, oltre al grande cromosoma batterico, esistono molecole di DNA più piccole, contenenti da 2 a 30 geni, dette **plasmidi**. Questi possono riprodursi contemporaneamente al cromosoma batterico, in modo che ogni cellula possieda una copia del plasmide, oppure a velocità maggiore o minore, casi in cui, rispettivamente, ciascuna cellula può posse-

derne più copie o addirittura nessuna. I plasmidi possono essere trasferiti da una cellula all'altra attraverso il contatto *cellula-cellula*: una volta creato un ponte citoplasmatico tra le cellule, un singolo filamento del DNA del plasmide viene trasferito dalla cellula donatrice a quella accettrice mediante una duplicazione definita «**a cerchio rotante**», e successivamente in entrambe viene ricostruito il filamento complementare (Fig. 4).

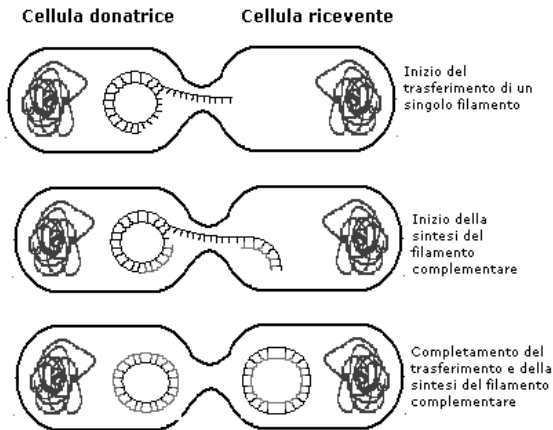


Fig. 4 - Duplicazione a cerchio rotante

Fra i diversi tipi di plasmidi individuati, uno di particolare interesse è quello della resistenza ai farmaci, detto **plasmide R**.

Si era osservato che una popolazione di batteri sensibile a un antibiotico poteva diventare resistente in meno di un'ora se posta in presenza di batteri resistenti. Successivamente si scoprì che il gene della resistenza, che in genere conferisce al batterio la capacità di produrre un enzima in grado di distruggere o rendere meno attivo l'antibiotico, risiede su un plasmide e ciò consente la rapida produzione di copie del gene e il suo trasferimento da una cellula batterica a un'altra. I geni della resistenza ai farmaci

possono anche passare dal plasmide al cromosoma batterico, oppure essere trasferiti a virus o batteri di specie diverse.

b) Virus come vettori d'informazione

I **virus**, costituiti da una molecola di DNA o di RNA racchiusa all'interno di un capsido proteico, non possiedono citoplasma né organuli cellulari e, quindi, non sono capaci di replicarsi se non all'interno di una cellula ospite. Quando un virus infetta una cellula, aderisce alla sua superficie mediante le proteine del capsido e ciò spiega il motivo per il quale le infezioni virali non si trasferiscono facilmente da una specie all'altra: ogni virus, infatti, riesce ad infettare solo le cellule che possiedono sulla loro superficie recettori per le proteine del capsido.

All'interno della cellula il DNA o l'RNA virale si duplicano e dirigono la sintesi delle proteine virali. Quando le nuove particelle virali sono state assemblate vengono liberate attraverso la lisi cellulare, determinata da proteine codificate all'interno del cromosoma virale. Questo tipo di meccanismo riproduttivo del virus prende il nome di **ciclo litico**.

Tuttavia, esistono virus che si comportano in maniera diversa e dopo il loro ingresso nella cellula restano silenti per un lungo periodo di tempo. Questi, infatti, inseriscono il proprio cromosoma all'interno del cromosoma cellulare e si riproducono insieme alle cellule. Il ciclo litico viene innescato quando il cromosoma virale fuoriesce spontaneamente da quello della cellula ospite. In questo secondo caso si parla di **ciclo lisogeno**.

I virus capaci di inserire il proprio cromosoma all'interno del cromosoma eucariote sono denominati **provirus** e tra questi figurano *virus a DNA* e *retrovirus a RNA*, così chiamati perché prima di essere integrati nel cromosoma cellulare devono essere trasformati nel corrispondente filamento di DNA, operazione resa possibile dall'enzima **trascrittasi inversa**, presente nel capsido proteico. Si ricordi, altresì, che un esempio di retrovirus a RNA è il virus dell'HIV.

I virus possono costituire vettori di informazione genetica poiché nel ciclo litico, dopo la frammentazione del DNA della cellula ospite e l'assemblaggio delle nuove particelle virali, alcuni

frammenti del DNA cellulare possono essere casualmente inglobati nel capside proteico del virus e, quindi, trasferiti ad altre cellule. In tal caso i frammenti di DNA trasferiti sono del tutto casuali e si parla, quindi, di **trasduzione generale**. Anche nel ciclo lisogeno possono essere trasferiti frammenti di DNA cellulare, che in questo caso saranno quelli contigui al punto di inserzione del cromosoma virale, il quale, distaccandosi, può portare con sé frammenti del cromosoma cellulare. Per questo motivo si parla allora di **trasduzione specifica**. Mediante questi meccanismi i virus possono costituire veri e propri vettori attraverso i quali tratti di DNA vengono trasferiti da una cellula all'altra, provocando modificazioni permanenti nel DNA cellulare.

c) Trasposoni

Già negli anni Cinquanta del Novecento **Barbara McLintock**, studiando le variazioni presenti nei semi di granturco attraverso la mappatura genica, avanzò l'ipotesi che esistessero geni, da lei chiamati **trasposoni**, in grado di cambiare posizione all'interno del cromosoma o addirittura di saltare da un cromosoma all'altro. Solo trent'anni dopo, quando fenomeni analoghi furono riscontrati anche in organismi semplici come virus e batteri, il suo lavoro ebbe il giusto riconoscimento con il conferimento del premio Nobel.

Nei procarioti sono stati individuati due tipi di trasposoni: *semplici* e *complessi*. In entrambi i casi si tratta di frammenti di DNA capaci di spostarsi da una zona all'altra del cromosoma grazie alla presenza di un gene che codifica per un enzima, detto **trasposasi**.

In questo spostamento capita spesso che il trasposone non scompaia dal sito di provenienza, ma produca una nuova copia di sé che viene inserita nel nuovo sito. I *trasposoni semplici* sono piccoli e contengono soltanto i geni necessari al processo di trasposizione, mentre quelli *complessi* comprendono anche altri geni, il più delle volte situati tra due trasposoni semplici. Una volta inseriti nel DNA, i trasposoni possono inattivare geni codificanti, innestandosi all'interno della sequenza nucleotidica, oppure, se contengono sequenze di promotori, possono attivare la trascrizione di geni precedentemente inattivi. I trasposoni presenti negli eucarioti sono molto simili a quelli dei procarioti e, come questi, possono provocare mutazioni o alterazioni nel funzionamento dei geni.

1 La trasmissione dei caratteri mono e polifattoriali

Ogni *cellula* dell'organismo umano contiene nel suo nucleo **46 cromosomi** distribuiti in **23 coppie**; di questi:

- 44 sono **autonomi** e trasmettono i caratteri del *soma*;
- 2 sono **gonomi**, cioè *cromosomi del sesso*, e condizionano, quindi, il sesso dell'individuo.

All'interno dei *cromosomi* sono allineati longitudinalmente i **geni**, portatori dei caratteri ereditari; essi sono situati in ordine caratteristico, in modo che ciascuno occupi sempre lo stesso punto di un particolare cromosoma: questo punto prende il nome di «**locus**». Il *legame* tra i geni è più stretto in quelli più vicini fra loro ed è, pertanto, maggiore la probabilità che questi si trasmettano «in blocco», per diverse generazioni, senza che si verifichino fenomeni cosiddetti di «*crossing over*» o di «scambio».

Per comprendere il significato della *disposizione a coppia dei cromosomi* è necessario considerare quanto avviene per i **gameti**: durante la maturazione dello **spermatozoo** e della **cellula uovo**, una delle divisioni cellulari è costituita da un particolare processo, la **meiosi**, nel corso del quale ciascuna coppia di cromosomi si divide nei suoi due elementi senza che i cromosomi stessi si replichino, come avviene invece nelle normali divisioni cellulari o **mitosi**. Di conseguenza, le cellule originate dalla *meiosi* saranno **aploidi**, ossia possiederanno la *metà* (23 elementi) del patrimonio cromosomico originario, presente invece nelle cellule **diploidi** dalle quali derivano. Sarà la *fecondazione*, con l'incontro e l'unione dei cromosomi omologhi (l'uno di origine *paterna*, l'altro di origine *materna*), che darà luogo alla ricostruzione del nuovo *assetto diploide*. Le unità cellulari derivanti da quest'unione, dunque, conteranno un'altra volta 46 cromosomi.

Ogni carattere, allora, è condizionato da due geni, uno «paterno» e uno «materno», normalmente stabili, la cui «replicazione» è completa a ogni successiva *mitosi cellulare*. Per ciascun carattere esiste un gene o più geni **alleli**, affini tra loro ma non identici, che condizionano la variabilità del carattere stesso. Soltanto un gene di quella serie «*allelomorfica*» può occupare sul cromosoma il *locus* sede di quel particolare carattere.

Un soggetto, perciò, può possedere per un determinato carattere due geni identici, uno per ogni cromosoma della coppia, provenienti ciascuno da uno dei due genitori: in tal caso il soggetto sarà **omozigote** per quel determinato carattere. Se invece sui due cromosomi saranno presenti due differenti *alleli*, il soggetto sarà **eterozigote** per quel carattere. Se indichiamo con le lettere *H* e *K* due geni che possono occupare un dato *locus*, potremo avere tre diverse combinazioni e, per questo motivo, tre diversi individui:

- HH;
- HK;
- KK.

A volte non è facile riconoscere le tre classi di individui perché uno dei due geni (ad esempio l'*H*) domina l'altro (il *K*): ossia la presenza del primo, che viene chiamato **dominante**, impedisce al secondo, detto **recessivo**, di manifestarsi. In questo caso distingueremo soltanto la *classe H*, nella quale saranno compresi sia gli individui *HH* sia quelli *HK*, e indicheremo con l'espressione «**fenotipo**» tale classe. Per *fenotipo* s'intende, infatti, l'insieme dei «caratteri esteriori» di un

individuo: esso non tiene conto, cioè, dei «caratteri recessivi» e perciò può non corrispondere al «**genotipo**», il quale rispecchia, invece, la costituzione genetica derivante dall'*insieme dei geni dominanti e recessivi*, e mostra quindi l'intero patrimonio genetico dell'individuo.

Va precisato, inoltre, che la grande maggioranza dei **caratteri somatici** è condizionata da più geni, i quali, pur concorrendo insieme a determinare la morfologia di un carattere che appare singolo alla nostra osservazione, sono mal identificabili nella loro precisa funzione: tale condizione genetica viene indicata con il nome di «**polimeria**».

Un'altra *condizione genetica* che può rendere difficile l'interpretazione dell'*assetto genetico* di un soggetto è la **pleiotropia**, per la quale un solo gene, modificando particolari metabolismi intermedi, induce la comparsa di una serie di caratteri che solo apparentemente non hanno tra loro alcun rapporto (ad esempio, l'albinismo e le malformazioni che lo accompagnano).

2 Le malattie genetiche ereditarie

Esistono alcune malattie dell'uomo che sono dovute ad **alterazioni geniche o cromosomiche**. Una **mutazione**, ovvero una *variazione della sequenza di base del DNA di uno specifico gene*, porta alla sintesi di un *m-RNA anomalo* e, di conseguenza, di una *proteina anomala*. A seconda della localizzazione della mutazione nel gene, la malattia può presentarsi in forma più o meno grave.

Molte malattie ereditarie sono spesso legate ai *geni recessivi* e quindi non si manifestano finché incontrano un gene dominante. Se però un individuo riceve da entrambi i genitori i due *alleli recessivi*, il carattere si manifesta e compare la malattia: da due genitori sani (portatori del gene mutato) può nascere, pertanto, un figlio malato. Quando il gene responsabile della malattia è *dominante*, anche una sola copia del gene difettoso è sufficiente a provocarla.

Tra le **malattie genetiche ereditarie** si ricordano:

- la *talassemia*;
- l'*emofilia*;
- il *daltonismo*;
- la *fibrosi cistica*;
- l'*albinismo*;
- la *distrofia muscolare*;
- la *neurofibromatosi*.

a) Talassemia (o «anemia mediterranea»)

Con il termine di «*talassemia*», o «*anemia mediterranea*», s'intende un **gruppo di malattie ereditarie, caratterizzate dalla produzione di globuli rossi «anormali»**. Tale alterazione del sangue è più o meno grave, per cui porta a danni clinici molto diversi, arrivando, attraverso stadi intermedi, dallo stato di portatore sano, il quale ha una normale aspettativa di vita, all'***alfa talassemia maior***, che conduce inevitabilmente alla morte già nel corso della vita intrauterina.

I *globuli rossi* dei talassemici sono, appunto, «anormali», ovvero più piccoli (la malattia è detta perciò «**microcitemia**», dal greco *micros*, «piccolo»), incapaci di trasportare in maniera ottimale l'ossigeno e, per di più, sopravvivono meno del consueto. L'organismo cerca allora di «compensare» tale situazione producendone un numero più alto, espandendo il midollo osseo e aumentando l'assorbimento intestinale del *ferro*. Nel caso di individui che hanno ereditato la malattia da *uno solo* dei genitori, questo sistema funziona e allora si parla di ***beta talassemia minor***: il paziente è talvolta pallido e astenico (cioè un po' più debole del normale) ed è, soprattutto, un **portatore sano**.

Nei casi più gravi, **morbo di Cooley o *beta talassemia maior***, ossia in quei malati che hanno ricevuto i geni della patologia da *entrambi* i genitori, questo tentativo di «bilanciamento» non riesce a correggere l'*anemia*, anzi, a sua volta, provoca ulteriori problemi all'organismo, in quanto

l'espansione midollare, oltre certi limiti, deforma le ossa; inoltre, il *ferro* assorbito in eccesso si deposita in diversi punti del corpo umano, danneggiando il fegato, le ghiandole endocrine, il pancreas, nonché favorendo il diabete e le complicazioni cardiache, che rappresentano attualmente la causa più comune di morte.

Questi pazienti manifestano i segni della malattia già nei primi mesi di vita e se mal curati, o addirittura non curati, muoiono in tenerissima età, spesso non riuscendo a superare neppure la pubertà.

b) Emofilia

L'*emofilia* è un **carattere recessivo legato al sesso**. La malattia, infatti, compare nei maschi della famiglia e viene trasmessa dalle femmine. Se la donna possiede un gene difettoso in uno dei suoi due *cromosomi X*, allora ella sarà protetta dalle conseguenze di tale condizione dal gene «normale» presente nel suo secondo cromosoma X. Se l'uomo, dal canto suo, ha un *mutante X* e un *normale cromosoma Y*, egli sarà invece affetto dalla malattia.

Il *figlio* la cui madre ha due normali *alleli* non sarà colpito dall'*emofilia* anche se il padre ha la malattia e il gene difettoso; la *figlia* dei medesimi genitori sarà una *portatrice sana*. La madre portatrice e il padre normale trasmettono il gene dell'*emofilia* a circa la metà dei figli: metà delle figlie femmine saranno «portatrici sane» e metà dei figli maschi saranno «emofiliaci»; il resto dei figli saranno «normali». Le figlie, a meno che un genitore sia sano, possono essere solo portatrici. Se il figlio riceve il gene difettoso dalla madre, vuol dire che sarà emofiliaco.

Le due forme più frequenti — riguardanti la *deficienza del fattore VIII* e del *fattore IX*, e denominate, rispettivamente, «**emofilia di tipo A**» ed «**emofilia di tipo B**» — sono responsabili di più del 90% delle malattie emorragiche ereditarie. La forma più diffusa è quella di *tipo A* o **emofilia «classica»**. La possibilità che la patologia colpisca anche le donne è puramente teorica: sembra che, infatti, essa sia letale prima della nascita.

c) Daltonismo

Il *daltonismo* è un **difetto della vista per cui la persona non riesce a distinguere il «rosso» dal «verde»**.

La malattia del *daltonismo* si trova nel *cromosoma X*, unico negli uomini e doppio nelle donne: quindi, mentre gli *uomini* manifestano questo carattere, le *donne* ne sono solo «portatrici», perché l'altro cromosoma X, fortunatamente «sano», domina su quello «difettoso». Qualora la donna ereditasse entrambi i cromosomi X difettosi dai genitori, sarebbe anch'essa daltonica: ma questo avviene raramente.

L'incapacità di distinguere i colori può manifestarsi in modo diverso; si possono riconoscere *tre forme* di daltonismo, ossia:

- l'**acromatopsia totale**, un difetto estremamente raro che consiste nell'*assoluta incapacità di percepire i colori*, mentre l'acuità visiva è conservata. Tutti gli oggetti appaiono purtroppo, a questi pazienti, solo nelle sfumature del bianco e del nero;
- la **dischromatopsia**, nella quale *manca la percezione di uno dei tre colori fondamentali* per l'occhio: dunque o del rosso, o del verde, o del blu. Si distinguono così forme di cecità:
 - al rosso (**anerotropsia**),
 - al verde (**acloropsia**),
 - al blu (**acianoblepsia**).

Spesso la cecità al rosso e quella al verde sono associate: ed è proprio a questo disturbo che dovrebbe essere strettamente riservato, appunto, il termine «daltonismo»;

- la **tricomatopsia anomala**, contraddistinta da una *minore sensibilità a uno solo dei tre colori fondamentali*; tale disturbo altera comunque la normale percezione dei colori.

d) Fibrosi cistica

La *fibrosi cistica* è una **malattia molto grave, monogenetica, i cui primi sintomi si hanno di solito nell'infanzia**, anche se in circa il 3% dei pazienti la diagnosi viene fatta soltanto in età adulta.

La patologia è caratterizzata da:

- infezione cronica delle vie aeree;
- insufficienza pancreatica esocrina;
- anormale degenerazione delle ghiandole sudoripare;
- disfunzione urogenitale.

Si tratta di una malattia **autosomica recessiva**, dovuta a mutazioni in un gene localizzato sul *cromosoma 7*.

La prevalenza della *fibrosi cistica* varia a seconda dell'origine etnica di una popolazione; essa viene diagnosticata in circa:

- 1 su 2.500 nati vivi nella popolazione caucasica dell'America e dell'Europa settentrionali;
- 1 su 17.000 nati vivi negli Stati afro-americani;
- 1 su 90.000 nati vivi nella popolazione asiatica delle Hawaii.

e) Albinismo

L'*albinismo* è un'**anomalia ereditaria rara, caratterizzata dall'incapacità di produrre «melanina»**, vale a dire quel pigmento che dà il colore:

- ai peli;
- alla pelle;
- all'iride dell'occhio.

La patologia è dovuta a una combinazione di *alleli recessivi*, i quali determinano la scarsa produzione di un *enzima* necessario per la sintesi della *melanina*, la cui quantità, nell'organismo sano, varia da persona a persona. Senza questo pigmento, che normalmente «filtra» le radiazioni solari, la pelle è soggetta a un rischio più alto di cancro e a un rapido invecchiamento; gli stessi occhi poi, non essendo adeguatamente protetti, risultano estremamente esposti a pericolosi danni.

L'*albinismo* può presentarsi in forma *totale* o *parziale*:

- l'**albinismo totale** è piuttosto raro ed è contraddistinto da:

- pelle bianchissima,
- capelli quasi bianchi o giallo paglierino, di consistenza setosa,
- occhi grigio-bluastri o addirittura rosei.

Proprio a causa della mancanza di melanina, gli «albini»:

- a) manifestano disturbi più o meno gravi della vista,
- b) non riescono a sopportare la luce e la loro pelle non è in grado di restare esposta ai raggi del sole che per brevissimi periodi;

- l'**albinismo parziale**, al contrario, è relativamente frequente; in esso la mancanza di pigmentazione è di solito limitata:

- a un solo ciuffo di capelli,
- a uno degli occhi,
- a una ristretta zona cutanea.

f) Distrofia muscolare di Duchenne

La *distrofia muscolare di Duchenne* è un **disordine recessivo legato al sesso**, talora definito «**distrofia muscolare pseudoipertrofica**»; si manifesta con un'incidenza di circa *30 nati maschi su 100.000 nati vivi*. Il grave disturbo porta il nome del grande neurologo francese *Duchenne*, il primo a stabilirne i criteri diagnostici e a descriverne gli aspetti morfologici.

La condizione «distrofica» è presente sin alla nascita, ma la malattia si rende di solito evidente fra i 3 e i 5 anni di età.

I bambini che ne sono colpiti:

- cadono di frequente;
- hanno difficoltà a giocare con i coetanei;
- corrono e saltano (attività fondamentali per un fanciullo) con estrema fatica.

Attorno ai 5 anni i sintomi si fanno più chiari ed evidenti: nel rialzarsi dal pavimento, ad esempio, il paziente distrofico è costretto ad aiutarsi con le braccia.

Nei bambini più piccoli, i muscoli del polpaccio possono apparire:

- di consistenza aumentata;
- quasi gommosi;
- di dimensioni maggiori per fenomeni di «ipertrofia» muscolare effettiva.

In seguito, il muscolo viene rimpiazzato da tessuto adiposo e connettivo.

g) Neurofibromatosi

Con il termine di «*neurofibromatosi*» s'intendono almeno **sette diverse malattie genetiche accomunate dalla presenza di tumori «benigni» che si sviluppano a livello dei nervi**. Ciascuna delle patologie ha però distinte peculiarità diagnostiche, importanti ai fini di una specifica sorveglianza periodica e differenziata.

La forma più frequente è la **NF1**, che interessa oltre il 90% dei casi e colpisce *1 nato ogni 3.000*. Essa è facilmente riconoscibile sin dai primi anni di vita per la presenza sulla pelle di almeno sei *macchie color caffè-latte*. Con il passare del tempo compaiono i «**neurofibromi**», nella forma di noduli localizzati in varie parti del corpo.

A parte questi problemi prevalentemente estetici e a lenta evoluzione, la *NF1* ha degli sviluppi che possono essere per molti altri aspetti *imprevedibili*. Nel 20% dei casi si associano, infatti, alcune gravi manifestazioni come *tumori cerebrali ed extracerebrali*, mentre il 30-40% dei soggetti presenta dei *disturbi di apprendimento* — quali scarsa memoria, difficoltà di lettura o di concentrazione, di scrittura o di linguaggio — e necessita di un supporto adeguato sin dalla prima infanzia.

La **NF2** è molto meno diffusa ma, di regola, più grave, in particolare per la costante presenza di *tumori che colpiscono il nervo acustico*, di uno o di entrambi i lati del viso, e/o *un'altra parte del cervello o del midollo spinale*. A seconda della localizzazione, possono verificarsi sordità assieme a diversi e seri disturbi neurologici. Molti pazienti presentano, inoltre, una sorta di «cataratta», nonché un numero variabile di macchie caffè-latte e di neurofibromi.

3 DNA ricombinante e sue possibili applicazioni

a) Biotecnologia

La *biotecnologia* **comprende una serie di manipolazioni genetiche che servono per la produzione di beni e servizi**. Già nell'antichità venivano selezionati animali e piante per soddisfare meglio i bisogni dell'uomo: ad esempio, bovini da lavoro oppure adatti alla produzione di carne

e di latte; grano con la spiga non caduca etc. Gli agricoltori e gli allevatori facevano riprodurre gli animali con le caratteristiche più adatte a venire incontro alle loro necessità.

Dopo **Mendel**, con il progredire della *scienza genetica*, vennero utilizzate tecniche nuove: infatti, nel secondo dopoguerra fu scoperto il cosiddetto «*vigore degli ibridi*», che condusse la produzione non verso le «linee pure» ma nella direzione dei «prodotti d'incrocio».

Nel 1953, con la scoperta dell'**acido desossiribonucleico** (il famigerato **DNA**), si poté indagare a fondo sulle basi molecolari della vita. Nei primi anni Settanta furono elaborate delle «tecniche sugli esseri viventi», come quella del «taglia e cuci»: da un organismo vivente, cioè, si estrasse un pezzo di DNA e lo si impiantò in un altro organismo, producendo così una **combinazione genetica nuova**.

Nacquero in tal modo le **biotecnologie avanzate**, le quali consentirono di utilizzare gli organismi viventi in vari e differenti processi industriali. In particolare, si sviluppò la cosiddetta «**ingegneria genetica**», ossia quell'*insieme di operazioni che portano alla «costruzione» di DNA ricombinati* e, appunto, alla realizzazione di nuove combinazioni genetiche (uno dei settori maggiormente interessati da queste dinamiche è, ad esempio, quello farmaceutico, nel quale si assiste alla commercializzazione di prodotti e composti chimici innovativi).

Grazie alle moderne tecniche di manipolazione è oggi possibile inserire, modificandoli se necessario, geni provenienti da una determinata specie nell'informazione genetica di un'altra specie completamente diversa — geni animali in batteri o in piante, geni umani negli animali e così via — producendo piante e animali comunemente definiti «**transgenici**».

b) Ingegneria genetica

Considerata la più moderna tra le «biotecnologie», l'*ingegneria genetica* consiste nell'**impiego di tecniche capaci di modificare in modo predeterminato il patrimonio genetico di un organismo**: non si tratta, però, di metodi di pura «invenzione» umana, infatti i ricercatori e gli studiosi del settore non fanno altro che perfezionare tecniche biologiche già «naturalmente» messe a punto dalle cellule stesse nel corso dell'evoluzione. La modifica stabile di un **genoma cellulare** è un fenomeno presente in natura: se ne avvalgono i **virus**, che inseriscono il proprio DNA in quello della cellula batterica ospitante per «costringerla» a riprodurre nuove particelle virali.

Gli studi condotti sui meccanismi d'infezione virale hanno portato alla scoperta di molecole di estremo interesse genetico: i cosiddetti «**plasmidi**». Si tratta di *piccole molecole circolari di DNA*, presenti nei *batteri* esternamente al cromosoma e dotate di un meccanismo di riproduzione autonoma. In natura funzionano quindi da «vettori genetici» e allo stesso scopo sono utilizzati nell'ingegneria genetica per trasferire geni da un tipo di organismo a un altro.

Gli **enzimi di restrizione** sono in grado di «tagliare» la molecola di DNA in corrispondenza di determinate sequenze. Dopo il taglio da essi effettuato, l'opera di ricucitura della molecola di DNA viene eseguita da altri *enzimi*: le **DNA ligasi**.

La **trascrittasi inversa** è un enzima, presente nei virus a **RNA**, capace di catalizzare la «trascrizione» di una molecola di DNA a partire da una di RNA. I biologi molecolari adoperano questo tipo di enzima per sintetizzare il gene codificante per una certa proteina, a partire dalla corrispondente molecola di **mRNA** presente nel *citoplasma* cellulare.

Servendosi adeguatamente delle «competenze» biochimiche di queste molecole, è possibile manipolare l'informazione genetica di una cellula producendo **DNA ricombinante**, *ottenuto cioè dall'unione di materiale ereditario proveniente da individui geneticamente diversi*. È possibile quindi, ricombinando in maniera mirata il DNA di una cellula, farle produrre ad esempio *proteine*, per la sintesi delle quali essa non era naturalmente programmata. Dopo aver isolato il gene che interessa, da un organismo animale o vegetale, lo si fa «combinare» con un *plasmide*, ottenendo DNA ricombinato, il quale, inserito in una cellula batterica ospite, induce la sintesi della protei-

na voluta. La cellula ospite trasformata, duplicandosi, genera un clone di individui tutti uguali, contenenti il *plasmide modificato*. Operando in tal modo è possibile ottenere artificialmente notevoli quantità di proteine d'interesse terapeutico, quali l'**insulina**, l'**ormone della crescita** o il **fattore VIII**, indispensabile alla coagulazione del sangue e mancante nei soggetti emofiliaci.

c) Geni, piante e animali

La *modificazione genetica* nelle **piante**, attraverso geni prelevati da altri organismi, consiste:

- nell'introdurre nuove proprietà, al fine di renderle più resistenti ai parassiti e alla siccità;
- nel dotarle di un gusto più gradevole o di una maggiore concentrazione di sostanze nutritive.

Nella prima metà degli anni Ottanta del XX secolo, l'americano **Kary B. Mullis** ha scoperto un metodo che permette a frammenti di DNA di duplicarsi più velocemente. Gli scienziati:

- prelevano i geni responsabili dello sviluppo di proprietà e funzioni utili;
- ne producono molte copie;
- inseriscono queste ultime nella pianta che vogliono modificare.

L'**Unione Europea autorizza solo cinque vegetali geneticamente modificati** e la loro produzione è consentita dal 1993; essi sono:

- il *mais*;
- la *soia*;
- la *colza*;
- il *radicchio*;
- il *tabacco*.

Le *biotecnologie* che fanno uso di specie **animali** devono il proprio sviluppo al fatto che questi, sebbene siano più costosi di *piante* e *microrganismi*, producono le sostanze e i composti d'interesse in modo più simile a quanto avviene all'interno del corpo umano. Esattamente come i microrganismi e le specie vegetali, infatti, anche gli animali possono essere alterati geneticamente, introducendo nuovi geni a livello embrionale.

Per i *vertebrati*, e per i *mammiferi* in particolare, la **clonazione artificiale** richiede un procedimento — realizzato per la prima volta negli anni Sessanta del Novecento dal biologo inglese John B. Gurdon sulle *rane* — di trasferimento del nucleo di una *cellula somatica* (cioè «del corpo», non destinata quindi alla riproduzione) all'interno di una *cellula uovo*, svuotata del proprio nucleo. Nel 1997 è arrivato l'annuncio della clonazione della **pecora Dolly** (nata nel luglio 1996), cui segue quella di alcune varietà di topi e, nel 1998, la realizzazione di un *processo di clonazione umana*.

Usare *tecniche biotecnologiche* può produrre una serie di *vantaggi* e di *svantaggi*; ne elenchiamo brevemente i principali.

Vantaggi:

- aumento della produttività e, contemporaneamente, minor utilizzo di mezzi tecnici (fertilizzanti o antiparassitari), diminuendo così la percentuale di sostanze indesiderate nei cibi;
- abbassamento dei costi sia per gli agricoltori sia per i consumatori, a patto che i benefici economici non ricadano sui privati titolari di brevetti;
- riduzione dei trattamenti chimici in agricoltura e relativo calo dei costi di produzione e di manodopera;
- notevole diminuzione dei tempi d'irrigazione;
- cospicuo sfruttamento dei territori in ambienti ostili;
- piante più forti, più resistenti e più produttive.

Svantaggi:

- costi ancora troppo elevati, in quanto la ricerca risulta particolarmente costosa;
- accentuazione della dipendenza degli agricoltori da chi detiene il monopolio di conoscenze nel settore delle biotecnologie, nonché concentrazione nelle medesime mani delle tecnologie di ricerca e del sapere;
- possibilità che i prodotti «transgenici» arrechino danni alla salute dell'uomo, poiché le proteine codificate dai nuovi geni svolgono un'azione tossica e/o allergenica;
- sia nella fase di sperimentazione sia in quella di distribuzione non è mai stato eseguito un controllo accurato sulla commercializzazione di questi prodotti;
- prodotti importati senza l'obbligo di etichettatura, dunque impossibilità di sapere cosa mangiamo;
- diffusione dei «transgeni» attraverso il polline e, conseguentemente, presenza di fenomeni d'inquinamento genetico di altre piante della stessa specie, coltivate in aree adiacenti;
- squilibrio negli ecosistemi per la presenza di geni tossici nei confronti d'insetti «fitopatogeni», che colpiscono altre specie;
- resistenza agli antibiotici contenuti nelle piante, che potrebbero entrare nel nostro apparato digerente e causare disturbi non indifferenti;
- rapida riduzione della «biodiversità» delle piante d'interesse alimentare.

d) Applicazioni e conseguenze sull'ambiente

Per «degradare» le *sostanze inquinanti* presenti nell'*ambiente* e soprattutto nel *suolo* vengono introdotti particolari *microrganismi*, la cui crescita a volte può essere favorita facendo penetrare determinate sostanze nutritive nell'aria e nel suolo; a tale scopo vengono impiegati perlopiù microrganismi scelti appositamente per la loro capacità di «detossificazione». Talvolta si può addirittura rimuovere il suolo contaminato ed esporlo, in condizioni controllate, all'azione dei «microrganismi degradatori», per riposizionarlo poi nel luogo d'origine.

In molti Paesi i batteri sono adoperati per estrarre metalli come il *ferro*, lo *zinco* e l'*uranio* da giacimenti difficilmente accessibili o in cui il minerale d'interesse non è «puro». Le estrazioni minerarie che utilizzano queste tecnologie stanno diventando via via più importanti, a mano a mano che i depositi di facile accesso e ricchi di minerali cominciano a esaurirsi.

Il rapido sviluppo economico della nostra società ha causato notevoli cambiamenti nell'ambiente naturale, con gravi ripercussioni sulla qualità dell'atmosfera, del suolo e delle acque.

Gli **inquinanti** principali sono:

- fitofarmaci;
- metalli pesanti;
- idrocarburi clorurati e aromatici;
- nitrati.

Dal gennaio del 2000 ogni Nazione ha il diritto di respingere l'importazione di cibi geneticamente modificati; dal canto loro, i Paesi esportatori devono fornire informazioni dettagliate sui prodotti e indicare sull'etichetta la presenza di «**organismi geneticamente modificati**» (**OGM**). Dall'aprile dello stesso anno è obbligatorio segnalare, sempre sull'etichetta, la presenza di «composti transgenici».

1 Le teorie di Lamarck e di Darwin

Nel XIX secolo **Jean-Baptist de Lamarck** elaborò una *teoria dell'evoluzione* in base alla quale tutte le specie discendono da specie preesistenti, in seguito a un processo che si basa su due principi fondamentali: il primo sostiene che gli animali possono acquisire determinate caratteristiche nel corso della vita a causa di *stimoli ambientali* e che le caratteristiche acquisite sono ereditabili dalla progenie; il secondo invoca l'esistenza di un *impulso universale* che spinge ogni essere vivente ad evolversi tendendo a raggiungere l'apice della *scala naturae*, la scala gerarchica ipotizzata da Aristotele, che vede ai gradini inferiori gli esseri viventi più semplici e al vertice l'uomo.

Charles Darwin è considerato a ragione il padre della moderna **teoria dell'evoluzione**, in quanto, oltre a fornire un'interminabile serie di argomentazioni difficilmente confutabili, contrariamente ai suoi predecessori, fu il primo ad intuire il vero meccanismo alla base dei processi evolutivi.

La *teoria darwiniana* si basa su cinque premesse:

1. gli organismi si riproducono generando organismi simili;
2. il numero di individui di una specie che sopravvive è molto più piccolo del numero di individui nati;
3. organismi simili possiedono differenze innate ed ereditabili;
4. gli individui che riescono a sopravvivere sono quelli dotati delle caratteristiche più «favorevoli» rispetto all'ambiente in cui vivono. Tali caratteristiche tendono a diventare più frequenti da una generazione all'altra, in seguito a un processo di **selezione naturale**;
5. agendo per tempi molto lunghi, la selezione naturale può accumulare cambiamenti che portano alla distinzione di gruppi di organismi diversi a partire da un progenitore comune.

In effetti i cinque principi ipotizzati da Darwin sono tutti validi. Ciò che mancava alla sua teoria era la spiegazione del meccanismo dell'*ereditarietà dei caratteri* che **Gregor J. Mendel** stava studiando proprio in quegli anni, ma che avrebbe dovuto attendere gli inizi del Novecento per essere riconosciuta dalla comunità scientifica.

2 La teoria dell'evoluzione

L'incontro della teoria evoluzionistica di Darwin con la genetica mendeliana diede vita, agli inizi del Novecento, alla **genetica di popolazione**, una branca della biologia che focalizza l'attenzione non sul singolo organismo, bensì su una popolazione di organismi. Per «**popolazione**» si intende un gruppo di organismi della stessa specie, che vivono in un certo spazio e in un certo tempo e che si riproducono tra loro. Ogni popolazione è caratterizzata da un proprio **pool genico**, che rappresenta l'insieme di *tutti gli alleli di tutti i geni di tutti gli individui* che costituiscono la popolazione; il singolo individuo della popolazione è il depositario di una piccola parte del pool genico, che trasmetterà alla sua progenie.

La genetica di popolazione si occupa dei cambiamenti che si verificano nei pool genici delle popolazioni, dei fattori che li producono e di stabilire quanto sia ampia la **variabilità** all'interno del pool genico (ad esempio, da quanti alleli sia costituito). La frequenza di un allele nel pool genico è pari al numero di individui che possiedono l'allele rispetto al totale e può variare da una generazione all'altra, poiché individui con caratteristiche più favorevoli rispetto all'ambiente

hanno una maggiore probabilità di sopravvivere e riprodursi e quindi i loro alleli saranno presenti con una maggiore frequenza nella generazione successiva. La **fitness riproduttiva**, o **successo biologico**, di un individuo corrisponde proprio al numero relativo di discendenti che sopravvivono.

3 Variabilità e selezione naturale

Uno dei principi su cui si basa la teoria evolutiva di Darwin è che *gli organismi generano altri organismi simili*: ebbene, alla luce delle scoperte della genetica molecolare, oggi sappiamo che ciò si deve all'esattezza con cui una molecola di DNA produce una copia di sé stessa, cosicché il patrimonio genetico di un individuo è una copia esatta di quello che ha ricevuto dal padre e dalla madre. Tuttavia, affinché possa esserci evoluzione, gli individui devono mostrare una varietà che consenta loro di avere una *fitness* migliore o peggiore di altri e, quindi, di essere selezionati in senso positivo o negativo dalle pressioni esercitate dall'ambiente. In pratica la variabilità del pool genico è il materiale grezzo su cui agiscono le pressioni evolutive, grazie alla quale può verificarsi l'evoluzione.

La determinazione della variabilità è cosa abbastanza complessa e può essere realizzata in vari modi. Esempari, in tal senso, gli studi condotti su *Drosophila*: partendo da un ceppo di moscerini dotati di 36 setole sulla superficie ventrale, è stato prodotto un ceppo con un numero di setole minore, selezionando e incrociando tra loro gli individui che presentavano il minor numero di setole, e con la stessa procedura è stato prodotto un ceppo dotato di un numero di setole maggiore. Dalle 36 setole originariamente presenti si è giunti a due ceppi con un numero di setole rispettivamente pari a 30 e 56, il tutto senza introdurre alcun materiale genetico nuovo: ciò significa che nel pool genico era già latente il potenziale per una così grande variabilità.

Un altro metodo per determinare l'ampiezza della variabilità del pool genico consiste nell'analisi della varietà di proteine prodotte da una popolazione.

J.L. Hubby e **R.C. Lewontin**, esaminando le proteine prodotte da una popolazione di moscerini, isolarono diciotto enzimi diversi e attraverso l'analisi elettroforetica evidenziarono che almeno nove di questi erano costituiti in realtà da una miscela di proteine aventi strutture lievemente diverse, prodotte evidentemente da alleli diversi di uno stesso gene. Uno degli enzimi risultò addirittura una miscela di sei molecole differenti, indicando che nel pool genico della popolazione esistevano almeno sei alleli. Studi analoghi condotti sui tessuti umani (sangue, placenta etc.) hanno permesso di stabilire che non meno del 25% dei geni di una popolazione è rappresentato da almeno due o più alleli e che ogni individuo è eterozigote per almeno il 7% dei suoi geni.

4 L'equilibrio di Hardy-Weinberg

Una volta individuati i meccanismi attraverso i quali vengono trasmessi i caratteri ereditari, sorse il problema di spiegare il motivo per il quale nel pool genico di una popolazione continuassero ad essere mantenuti gli alleli recessivi, senza che venissero gradualmente eliminati dagli alleli dominanti.

Il matematico britannico **Godfrey H. Hardy** e il medico tedesco **Wilhelm Weinberg** fornirono una risposta a questo interrogativo studiando il *comportamento delle frequenze alleliche in una popolazione ideale*, nella quale:

- non si verificano mutazioni;
- non c'è ingresso o perdita di geni conseguente all'immigrazione o all'emigrazione;
- il numero di individui è sufficientemente ampio da poter applicare criteri statistici;
- l'accoppiamento è casuale;
- tutti gli alleli hanno il medesimo successo riproduttivo.

I due ricercatori dimostrarono che, in queste condizioni, la composizione del pool genico non cambia semplicemente per frutto del caso, ossia per effetto della normale ricombinazione genica, e la condizione di equilibrio per un gene che possieda due alleli differenti, A e a , rispettivamente dominante e recessivo, a cui sono associate rispettivamente le probabilità p e q , è espressa dall'equazione:

$$(p + q)^2 = p^2 + 2pq + q^2 = 1$$

dove:

$$(p + q) = 1$$

Il termine p^2 rappresenta la frequenza degli individui omozigoti dominanti, q^2 quella degli omozigoti recessivi e $2pq$ la frequenza degli eterozigoti. Questa equazione è particolarmente importante non tanto perché applicabile a un gran numero di casi reali, visto che le condizioni richieste affinché si abbia una situazione di equilibrio del pool genico sono piuttosto difficili da realizzare contemporaneamente, quanto perché consente di evidenziare le deviazioni dall'equilibrio e di ricercarne le cause.

Tra queste vi sono le **mutazioni**, cioè i cambiamenti ereditabili del genotipo di un individuo, consistenti in una delezione, trasposizione o duplicazione di un tratto di DNA o nella sostituzione, delezione o inserzione di un nucleotide. Le mutazioni sono casuali, possono riguardare sia geni strutturali sia geni di regolazione e vengono indotte da fattori esterni, quali raggi X o ultravioletti, composti radioattivi, nonché da alcune sostanze chimiche. In genere non hanno frequenze elevate (si consideri che ogni essere umano, possessore di circa 100.000 geni, è portatore di due nuove mutazioni), ma in ogni caso costituiscono il materiale grezzo su cui agiscono i cambiamenti evolutivi.

Anche il **flusso genico**, ossia la perdita o l'acquisizione di nuovo materiale genetico in seguito a fenomeni di emigrazione o immigrazione, può determinare la variazione delle frequenze alleliche o la modificazione della composizione del pool genico.

È ancora dibattuta, invece, l'effettiva rilevanza della **deriva genica**, ossia dei cambiamenti nella composizione del pool genico dovuti all'azione del caso. Questa può essere considerata trascurabile nell'ambito delle grandi popolazioni, che normalmente obbediscono all'equilibrio di Hardy-Weinberg, mentre assume maggiore importanza nelle piccole popolazioni. In particolare la deriva genica diventa significativa in occasione della drastica riduzione del numero di individui di una popolazione, come accade quando qualche evento esterno determina l'eliminazione di un gran numero di individui (fenomeno indicato come *collo di bottiglia*) oppure in seguito al distaccamento e all'isolamento di una piccola parte di una popolazione (il cosiddetto *effetto del fondatore*). In entrambi i casi possono verificarsi significative variazioni del pool genico: la porzione di popolazione considerata potrebbe essere geneticamente non rappresentativa della popolazione di provenienza, presentando, per puro effetto del caso, frequenze alleliche sensibilmente diverse rispetto alla popolazione di provenienza. Un ulteriore fattore in grado di alterare l'equilibrio di Hardy-Weinberg è infine l'**accoppiamento non casuale**, pratica comune fra gli allevatori che tentano di ottenere specie con caratteristiche prestabilite (*selezione artificiale*).

5 Il mantenimento della variabilità

Da quanto detto appare chiaro che l'ampiezza della variabilità del pool genico di una popolazione rappresenta una ricchezza per la popolazione stessa, poiché costituisce una sorgente potenziale di variazioni che consentono alla specie di rispondere alle modificazioni ambientali.

La variabilità negli eucarioti è mantenuta attraverso diversi tipi di strategie, la più importante delle quali è senza dubbio la **riproduzione sessuata**, che favorisce la variabilità nei seguenti modi:

- mediante l'assortimento indipendente durante la meiosi;
- mediante il *crossing-over* e la ricombinazione genica;
- mediante la combinazione di due genomi parentali al momento della fecondazione.

In molti organismi che si riproducono per via sessuata esistono meccanismi atti a favorire le nuove combinazioni: nelle piante, ad esempio, esiste un'ampia varietà di strategie volte ad impedire l'autoimpollinazione, così come negli animali determinate strategie comportamentali, fra cui l'allontanamento del giovane maschio dalla famiglia, tendono a favorire l'**esincrocio**.

La **diploidia** costituisce un altro fattore che mantiene la variabilità negli eucarioti. Infatti, un allele recessivo, che non si manifesta a livello fenotipico e quindi non è immediatamente soggetto a selezione, si conserva nel pool genico di una popolazione. Inoltre, minore è la sua frequenza v , minore è la frequenza della condizione di omozigosi v^2 e quindi la probabilità che l'allele sia soggetto a selezione naturale.

In alcuni casi, infine, un allele recessivo, anche se dannoso in condizione di omozigosi, può risultare favorito in condizione di eterozigosi, fenomeno indicato come **superiorità dell'eterozigote**. Ne costituisce un esempio l'allele responsabile dell'anemia falciforme, che nella condizione di omozigosi determina una malattia mortale, tanto che gli individui omozigoti non arrivano a riprodursi. Nonostante la perdita di alleli dovuta agli omozigoti, in alcune tribù africane quasi il 45% della popolazione è eterozigote per il carattere falciforme, in virtù del fatto che l'eterozigote ha il vantaggio selettivo, rispetto all'omozigote normale, di essere meno sensibile alla malaria, che in alcune regioni dell'Africa rappresenta la prima causa di morte.

6 La selezione naturale

La forza principale dell'evoluzione è la **selezione naturale**, che può essere considerata il risultato dell'interazione fra i singoli organismi e il loro ambiente, esteso sia alla componente *abiotica* sia a quella *biotica*. Tale forza può determinare variazioni nel pool genico di una popolazione, senza per questo eliminare la variabilità. Ciò si deve al fatto che le pressioni evolutive agiscono sull'intero fenotipo di un organismo e questo può essere realizzato attraverso molte combinazioni alleliche diverse. Inoltre, parlando di fenotipo in senso evolucionistico, ci si riferisce non solo alle caratteristiche esteriori, bensì a tutto l'insieme degli aspetti fisici, fisiologici e comportamentali che, nel loro complesso, determinano il successo dell'individuo. Quindi, un individuo svantaggiato per una caratteristica può essere favorito per altre e l'azione contemporanea di pressioni selettive diverse contribuisce a mantenere la variabilità.

Nell'ambito della selezione naturale si possono distinguere *cinque grandi categorie*:

1. la **selezione stabilizzante**, che tende ad eliminare gli individui con caratteristiche estreme;
2. la **selezione divergente**, la quale, invece, aumenta la frequenza delle caratteristiche estreme a sfavore di quelle intermedie;
3. la **selezione direzionale**, che agisce aumentando il numero di individui con una caratteristica specifica;
4. la **selezione frequenza-dipendente**, che tende a ridurre la frequenza dei fenotipi più comuni, aumentando quella dei fenotipi meno comuni;
5. la **selezione sessuale**, basata sulla competizione tra membri dello stesso sesso (generalmente quello maschile) per la conquista dell'altro. Si ritiene che il «*dimorfismo sessuale*», ossia la diversità esistente tra maschi e femmine di una stessa specie, che spesso rende molto più vistoso un sesso rispetto all'altro (si pensi, ad esempio, al pavone), dipenda proprio da questo tipo di selezione.

7 L'adattamento

L'evoluzione nel tempo produce *adattamento*, ovvero un graduale cambiamento che rende gli organismi più adatti al loro ambiente.

L'**adattamento all'ambiente fisico** può produrre i cosiddetti **clini**, cioè variazioni graduali del fenotipo in conseguenza di graduali variazioni delle condizioni ambientali: i passeri comuni, ad esempio, tendono ad assumere dimensioni minori nelle regioni calde, maggiori in quelle fredde, poiché aumentando le dimensioni il rapporto superficie/volume diminuisce e la capacità dell'animale di mantenere il calore corporeo aumenta. Quando una stessa specie occupa habitat molto diversi, arrivando ad assumere in ognuno fenotipi differenti, si parla di **ecotipi** diversi della stessa specie, che corrispondono a variazioni genetiche.

Anche l'**adattamento all'ambiente biologico** può produrre fenomeni interessanti, come quello della **coevoluzione**, che si verifica quando due o più specie interagiscono così fortemente tra loro da costituire un notevole fattore selettivo l'una per l'altra. Esempi tipici in questo senso sono forniti dai fiori e dai loro impollinatori.

Altre forme di adattamento all'ambiente biologico prevedono lo sviluppo di particolari **meccanismi di difesa**: la linfa amara di alcune piante, ad esempio, ha il compito di agire come deterrente nei confronti degli insetti. La farfalla monarca, al contrario, si ciba di questo tipo di piante e accumula la sostanza tossica in modo da essere sgradita ai predatori. Inoltre, per essere riconosciuta dal predatore, prima di essere aggredita adotta caratteristici colori di avvertimento.

Da questo tipo di adattamento discendono alcune forme di **mimetismo**, come quello messo in atto da specie non imparentate tra loro e dotate di meccanismi di difesa simili, che assumono lo stesso aspetto esteriore in modo da essere riconosciute come appartenenti alla stessa famiglia: ne sono esempio api, vespe e calabroni. Questo tipo di mimetismo, detto **mimetismo mulleriano**, risulta vantaggioso per tutte le specie coinvolte perché ognuna trae vantaggio dall'esperienza che il predatore ha già avuto con l'altra. Successivamente a questo tipo di strategia adattativa sono comparsi sulla scena evolutiva casi indicati col nome di **mimetismo batesiano**, in cui alcune specie assumono l'aspetto di altre, senza essere dotate delle stesse strategie difensive: molte farfalle, ad esempio, hanno assunto colorazioni simili alla farfalla monarca, senza avere lo stesso sapore sgradevole. Questa strategia è favorevole solo alle «imitazioni» e si rivela efficace fintantoché il loro numero si mantiene esiguo rispetto al modello.

8 La speciazione

L'evoluzione determinata dalla selezione naturale può essere ricondotta fondamentalmente a due modelli: quello **convergente**, in cui l'azione delle stesse pressioni evolutive su organismi diversi produce i medesimi tipi di adattamento, tanto da generare fenotipi assai simili in specie diverse, e quello **divergente**, in virtù del quale il lungo isolamento di una popolazione in condizioni ambientali particolari può originare ecotipi locali o addirittura nuove specie. Una **specie** costituisce un insieme di popolazioni naturali dotate di un pool genico comune, interfeconde tra loro e non con altre.

Il meccanismo di **speciazione**, che presiede alla produzione di nuove specie, è conseguenza di un processo evolutivo di tipo *divergente*: quando una popolazione si trova isolata per lungo tempo può divergere geneticamente da quella di provenienza, in parte per l'effetto del fondatore, in parte per l'azione esercitata da pressioni selettive particolari. Se l'isolamento è prolungato nel tempo e le pressioni sono forti, il risultato è la produzione di una nuova specie, che non è più interfeconda con quella di provenienza. Tale fenomeno viene indicato come **speciazione allopatrica**, così chiamata perché deriva generalmente da un isolamento dovuto all'insorgere di barriere geografiche che isolano una porzione di popolazione. Un esempio famoso di questo

tipo di speciazione è rappresentato dai cosiddetti «fringuelli di Darwin»: nel suo viaggio alle isole Galàpagos, Darwin osservò tredici specie di fringuelli, distinte tra loro per le dimensioni e per la forma del becco, funzionale al tipo di alimentazione di ognuna, verosimilmente evolute da un antenato comune in seguito all'isolamento di piccoli gruppi in condizioni ambientali diverse.

Altri meccanismi di speciazione non sono riconducibili all'isolamento genetico: la **speciazione simpatica**, ad esempio, è un fenomeno dovuto al verificarsi di eventi di altro tipo, come la **poliploidia**, consistente in un repentino incremento del numero di cromosomi di una specie rispetto al normale assetto diploide. Di solito questo fenomeno è legato alla mancata disgiunzione dei cromosomi durante la *mitosi* o la *meiosi*, ma può verificarsi anche quando la duplicazione dei cromosomi non è seguita da *citodieresi*. In tal caso, il numero di cromosomi dell'individuo generato è doppio rispetto a quello dell'individuo di origine, cosicché si produce una nuova specie non più interfecunda con quella originaria.

A seguito dell'accoppiamento tra individui di specie diverse vengono prodotti **ibridi** che generalmente non sono fecondi, poiché i cromosomi mancano degli omologhi e non possono dare luogo alla meiosi. Tuttavia, se nell'ibrido si verifica il fenomeno della poliploidia, in modo da raddoppiare il numero dei cromosomi, l'ibrido diventa fecondo e si ha la creazione di una nuova specie.

9 L'isolamento genetico delle specie

Una volta avvenuto il processo di speciazione, diversi tipi di meccanismi tendono a mantenere l'*isolamento genetico tra specie diverse*, distinguibili in meccanismi di tipo *prezigotico* o *postzigotico*.

I **meccanismi prezigotici** intervengono prima dell'accoppiamento e sono riconducibili fondamentalmente a rituali comportamentali o a segnali convenzionali che permettono il riconoscimento tra individui della stessa specie: le lucciole, ad esempio, hanno tipi di lampeggiamento specifici per ogni specie; il canto degli uccelli e il gracidiare delle rane costituiscono vere e proprie forme di richiamo sessuale. Alcuni animali utilizzano come richiamo sostanze chimiche altamente specifiche per il tipo di specie, che a volte svolgono anche la funzione di indurre la produzione dei gameti. Infine, molte specie di piante e di animali hanno periodi di fertilità ben definiti, tali da determinarne l'isolamento riproduttivo.

I **meccanismi postzigotici**, invece, entrano in gioco durante o dopo l'accoppiamento, oppure dopo la fecondazione. Tali sono, ad esempio, le eventuali incompatibilità anatomiche che impediscono l'inseminazione, l'impossibilità degli spermatozoi di sopravvivere nel canale riproduttivo femminile o di fecondare l'uovo, il mancato sviluppo dell'embrione o la nascita di prole infeconda.

10 Le prove a sostegno dell'evoluzione

Le prove più consistenti a sostegno della teoria evoluzionistica non derivano tanto dall'osservazione degli esseri viventi che attualmente popolano la Terra, quanto dallo *studio dei reperti fossili*, che ha rivelato una successione nel tempo delle forme viventi dalle più semplici alle più complesse, evolute attraverso *processi di speciazione riconducibili a tre modelli*:

1. il **cambiamento filetico**, risultato di un processo di selezione divergente, il quale determina l'accumulo in una specie di un così gran numero di cambiamenti che la specie di arrivo è totalmente diversa da quella di partenza;
2. la **cladogenesi**, ossia la divisione di una linea evolutiva di organismi in due o più specie, discendenti contemporanee di un antenato comune. Si osservi che il cambiamento filetico e

la cladogenesi non sono meccanismi in competizione tra loro, bensì processi complementari che agiscono insieme, determinando l'evoluzione;

3. la **radiazione adattativa**, consistente nella produzione contemporanea di un gran numero di nuove specie a partire da un antenato comune. Questo tipo di «esplosione evolutiva» si verifica in conseguenza della disponibilità di molte nuove opportunità per gli esseri viventi, come accadde per i rettili quando l'evoluzione dell'uovo dotato di guscio fornì loro la possibilità di colonizzare le terre emerse. Tale fenomeno, più volte testimoniato dai reperti fossili, suggerisce che *l'evoluzione non sia stata sempre il frutto di cambiamenti gradualisti, ma sia spesso proceduta «per salti»*.

Eventi ampiamente documentati nei reperti fossili, che insieme ai processi di speciazione hanno determinato il percorso dell'evoluzione, sono i fenomeni di **estinzione**, termine indicante la repentina e completa scomparsa di una o più specie, o addirittura di interi gruppi di specie.

I fossili evidenziano tassi di estinzione più o meno costanti nel tempo, interrotti da *estinzioni di massa* come quella avvenuta circa 65 milioni di anni fa, alla fine del Cretaceo, che causò la scomparsa dei dinosauri e di numerose altre specie animali, lasciando indenni solo piante tropicali e piccoli organismi acquatici e terrestri. Si ipotizza che quell'evento sia stato determinato dalla caduta sulla Terra di un enorme meteorite, che, provocando il sollevamento di una gigantesca nube di detriti, tale da mantenere il pianeta in uno stato di oscurità per mesi, avrebbe bloccato i processi di fotosintesi e causato cambiamenti climatici di portata tale da innescare l'estinzione di quasi tutte le forme viventi. Tale ipotesi, peraltro ancora dibattuta, sembra trovare conferma nel ritrovamento di strati di argilla, risalenti al passaggio dal Cretaceo al Terziario, contenenti un'elevata quantità di iridio, metallo raro nella crosta terrestre, ma abbondante nei meteoriti. Inoltre, nel 1990 è stato scoperto in Messico un cratere da impatto del diametro di 300 km, compatibile con l'ipotesi dell'evento in questione.

In ogni caso, qualunque sia il motivo della ricorrenza delle estinzioni di massa, è certo che esse producono svolte decisive nel percorso evolutivo, eliminando completamente alcune specie ed offrendo nuove opportunità agli organismi sopravvissuti.

C'è da aggiungere, infine, che la storia della vita sulla Terra così come viene ricostruita tramite i fossili sembra avere qualche carenza, come se i fossili risalenti ad alcuni periodi fossero totalmente assenti. Pensando al processo evolutivo come ad un cambiamento lento e continuo, senza dubbio questa sembrava essere l'ipotesi più plausibile. Tuttavia, i fossili mostrano rare testimonianze di cambiamenti filitici, mentre i processi evolutivi sembrano generalmente procedere «per salti», con l'improvvisa comparsa negli strati fossili di alcune specie che persistono per milioni di anni e poi scompaiono altrettanto repentinamente. Pertanto si è fatta largo l'idea che l'evoluzione proceda per **equilibri intermittenti**, ossia attraverso lunghe fasi di stabilità seguite da periodi di cambiamenti repentini, e che la selezione naturale abbia sulle specie gli stessi effetti che produce sui singoli individui.

1 La sistematica scientifica

La nascita della **sistematica scientifica**, scienza che si occupa della classificazione dei viventi, viene fatta coincidere con l'opera di **Linneo**, nome latinizzato del botanico svedese **Carl von de Linnè** (1707-1778). Nel *Systema Naturae* egli effettuò la classificazione di oltre 4.000 animali basandosi su caratteristiche evidenti, come la disposizione dei denti nei mammiferi, il tipo di becco e zampe negli uccelli, la posizione delle pinne nei pesci, raggruppando insieme organismi aventi almeno un carattere comune facilmente riscontrabile, pur differendo tra loro per altri particolari. A Linneo si deve inoltre la **nomenclatura binomiale**, ancora in uso, in base alla quale ogni organismo è designato attraverso una coppia di nomi, di cui il primo, scritto con la lettera maiuscola, fa riferimento al genere, mentre il secondo alla specie.

L'approccio di Linneo nella *classificazione delle specie* ha dominato in biologia per quasi due secoli. Oggi, però, prevalgono altri criteri di classificazione, i quali partono dall'idea che questa debba riflettere la *filogenesi*, ossia la storia evolutiva delle diverse unità tassonomiche, in modo tale che ognuna di esse includa una singola sezione dell'albero filogenetico.

Nella **tassonomia evolutiva classica**, che prende in considerazione soprattutto somiglianze e differenze morfologiche, è fondamentale stabilire se alla base delle somiglianze riscontrabili in organismi diversi vi siano processi di adattamento ad ambienti simili da parte di organismi che non hanno antenati comuni o se, al contrario, le differenze osservabili derivino da adattamenti ad ambienti diversi di organismi strettamente imparentati tra loro. In quest'ambito sono quindi nati i concetti di **omologia** e **analogia**: si dicono *omologhe* le strutture appartenenti ad organismi diversi aventi diversa funzione, ma dotate della stessa struttura di base, mentre sono dette *analoghe* le strutture che assolvono alla stessa funzione, ma hanno caratteristiche strutturali completamente diverse (Fig. 1).

Non c'è dubbio che la presenza di strutture omologhe in organismi diversi denunci un collegamento evolutivo tra questi, essendo improbabile l'ipotesi che strutture complesse comuni a diverse specie, caratterizzate da numerose parti distinte, si siano evolute indipendentemente in ognuna di esse. Un tipico esempio di omologia è fornito dalle ossa degli arti anteriori dei vertebrati; viceversa, le ali degli uccelli e degli insetti costituiscono esempi di strutture analoghe.

A questo tipo di considerazioni se ne sono aggiunte altre di tipo embriologico, orientate a riconoscere similitudini nelle prime fasi dello sviluppo embrionale, oppure considerazioni di tipo biochimico, in cui si stabiliscono relazioni fra gli organismi sulla base di somiglianze o differenze riscontrabili nel tipo di proteine prodotte.

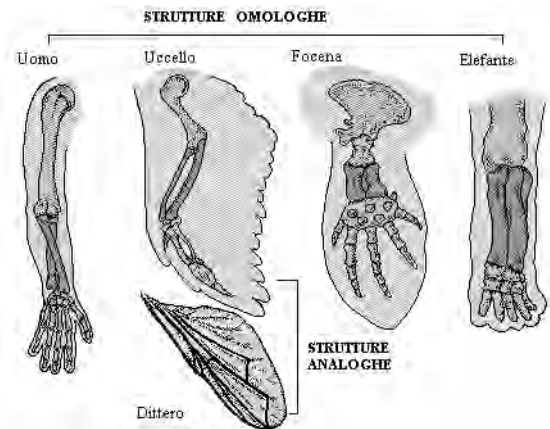


Fig. 1 - Strutture omologhe e analoghe

2 La classificazione gerarchica

In base a quanto detto finora, la classificazione dovrebbe, quindi, rispecchiare le relazioni filogenetiche. La **tassonomia** (dal greco *taxis*, «ordine», e *nomos*, «regola») è la scienza che si occupa di raggruppare i viventi in categorie tassonomiche. Quelle attualmente utilizzate, disposte secondo un ordine gerarchico di ampiezza crescente, sono:

- specie;
- genere;
- famiglia;
- ordine;
- classe;
- *phylum* o divisione;
- regno;
- dominio.

Dunque, la specie è l'unità tassonomica più circoscritta. Va sottolineato che non esiste un unico tipo di classificazione perché la sistematica si evolve nel tempo, parallelamente allo sviluppo di nuove tecnologie e all'acquisizione di ulteriori informazioni.

Un tipo di classificazione generalmente accettata è quella del biologo statunitense **Robert H. Whittaker** (1969), che prevede la suddivisione dei viventi in *due domini*, procarioti ed eucarioti, e *cinque regni*. La suddivisione degli organismi e le loro caratteristiche principali sono brevemente riassunte nella tabella che segue.

Regno	Tipo di cellula	Organizzazione cellulare	Modalità di nutrizione	Modalità di riproduzione	Modalità di movimento
Monera	procariota	unicellulare o coloniale	autotrofa (fotosintetica o chemiosintetica), eterotrofa (per assorbimento)	asessuata (per scissione)	mobili (flagelli)
Protista	eucariota	unicellulare o coloniale	autotrofa (fotosintetica), eterotrofa (per assorbimento o ingestione)	asessuata e sessuata	mobili (ciglia e flagelli), immobili
Plantae	eucariota con parete (cellulosa)	pluricellulare con tessuti	autotrofa (fotosintetica)	asessuata e sessuata	immobili
Fungi	eucariota con pareti (chitina)	sinciziale (senza confini definiti tra le cellule)	eterotrofa (per assorbimento)	asessuata e sessuata	immobili
Animalia	eucariota senza pareti	pluricellulare con tessuti	eterotrofa (per ingestione)	generalmente sessuata	mobili (fibre contrattili)

Tab. 1 - Classificazione di Whittaker

Successivamente sono state proposte altre classificazioni, come quella di **Carl Richard Woese** (1990), che prevede la suddivisione dei viventi in *tre domini* e *sei regni*.

Dominio	Regno
Bacteria	Eubacteria
Archea	Archeabacteria
Eukarya	Protista
	Fungi
	Plantae
	Animalia

Tab. 2 - Classificazione di Woese

Rispetto alla precedente divisione in cinque regni, è stato dunque introdotto il regno degli *archebatteri*, che potrebbero costituire le prime forme di vita comparse sulla Terra. Si tratta di batteri che vivono in condizioni del tutto particolari: possono sopportare la disidratazione (tanto da vivere in acque sature di sale), nonché resistere a temperature e a pressioni elevatissime (infatti sono stati rinvenuti sui fondali oceanici, accanto a bocche eruttive). Contrariamente agli *eubatteri*, hanno pareti cellulari prive di peptidoglicani e membrane cellulari costituite da catene di isoprene e non da fosfolipidi. Inoltre, sono dotati di un pigmento sensibile alla luce rossa, la *alorodopsina*, molto simile alla *rodopsina* presente nella retina dei vertebrati, che utilizzano per ricavare energia.

1 Il regno delle monere

Il *regno delle monere* comprende tutti i **procarioti**, i primi organismi comparsi sulla Terra, i più semplici e, attualmente, anche i più abbondanti (un grammo di suolo fertile può contenere fino a due miliardi e mezzo di batteri).

Il loro successo biologico è dovuto da un lato alla velocità di riproduzione, dall'altro alla versatilità metabolica, che ha consentito la loro sopravvivenza anche in condizioni ambientali avverse: alcuni di essi, infatti, sono capaci di produrre spore in grado di rimanere quiescenti fino allo stabilirsi di condizioni di vita più favorevoli.

La classificazione dei procarioti è tradizionalmente basata sulle differenze riguardanti le dimensioni, la forma, l'aspetto delle colonie e le modalità di nutrizione. In epoca più recente, a seguito di studi di biochimica cellulare, è stata individuata una ulteriore classe di batteri, dotati di sequenze nucleotidiche molto diverse dagli altri e di sequenze metaboliche del tutto particolari, che li rendono idonei a sopravvivere in ambienti estremi (acque sature di sale, pressioni e temperature elevate), molto simili alle condizioni della Terra primordiale.

Tali batteri sono stati riuniti nella famiglia degli **archebatteri**, che nelle classificazioni più recenti è considerata un regno a sé stante. Tra questi si distinguono:

- **metanogeni**, capaci di sintetizzare metano a partire da anidride carbonica e idrogeno;
- **aloofili**, in grado di vivere in ambienti ricchi di sali;
- **termoacidofili**, che invece vivono in ambienti fortemente acidi e ad alte temperature.

Tutti gli altri batteri vengono classificati come **eubatteri** e comprendono specie *autotrofe* ed *eterotrofe*.

I **batteri eterotrofi** possono essere distinti in:

- **patogeni**, che si nutrono ricavando energia dai tessuti e dai liquidi corporei dell'organismo ospite, a cui possono arrecare danni, fino a provocargli malattie;
- **non patogeni**, i quali, pur nutrendosi dei tessuti dell'ospite, non gli arrecano danni, ma possono, in alcuni casi, essergli di utilità. I ruminanti, ad esempio, sono in grado di utilizzare la cellulosa come fonte di energia grazie alla presenza, nel loro apparato digerente, di microrganismi capaci di idrolizzarla;
- **saprobi**, i più numerosi, che normalmente vivono sul materiale organico morto, operando la demolizione e quindi il riciclaggio delle sostanze organiche presenti nel suolo, rendendole nuovamente disponibili per piante e animali. Ogni gruppo di questi batteri è specializzato nella demolizione di un particolare tipo di composti (cellulosa, amidi, proteine etc.) e proprio questa specificità consente la convivenza, senza competizione e con un rapporto di mutua assistenza, di tante forme diverse nella stessa area.

I **batteri autotrofi**, invece, si distinguono in:

- **chemiosintetici**, ossia organismi che ricavano energia dall'ossidazione di sostanze inorganiche come lo zolfo, il ferro, l'azoto. I batteri **nitrificanti**, in particolare, svolgono un ruolo fondamentale nel ciclo dell'azoto;

- **fotosintetici**, fra i quali vi sono i **cianobatteri**, che conducono la fotosintesi in modo assai simile alle piante, ossidando l'acqua e producendo ossigeno. Tali batteri vivono nelle acque dolci, isolati o in gruppi costituiti da filamenti simili ad alghe, e possiedono la clorofilla, come le piante, che tuttavia non si trova in strutture analoghe ai cloroplasti, bensì in semplici vescicole denominate *cromatofori*. Altri batteri fotosintetici sono rappresentati dai **batteri verdi** e dai **batteri purpurei**, che possiedono pigmenti diversi e nella reazione di fotosintesi ossidano l'acido solfidrico o altri composti, senza produrre ossigeno.

2 Il regno dei protisti

Il *regno dei protisti* comprende tutti gli **organismi eucarioti unicellulari** e alcuni **semplici organismi pluricellulari**. Si pensa che i primi organismi eucarioti abbiano avuto origine in seguito all'ingresso di organismi procarioti in cellule più grandi, con le quali hanno stabilito un rapporto di simbiosi: in particolare, i procarioti fotosintetici avrebbero dato origine agli attuali *cloroplasti*, mentre organismi procarioti eterotrofi avrebbero originato i *mitocondri*. Questa ipotesi, nota come **teoria dell'endosimbionte**, sembra avvalorata dalla somiglianza riscontrata tra le membrane di alcuni batteri e quelle mitocondriali, dalla somiglianza di alcune sequenze di r-RNA e, soprattutto, dal fatto che sia i mitocondri sia i cloroplasti possiedono un cromosoma proprio, molto simile al cromosoma batterico in cui sono codificate alcune delle proteine mitocondriali. Mitocondri e cloroplasti, inoltre, possono riprodursi solo a partire da altri mitocondri e cloroplasti preesistenti.

I protisti possono essere *autotrofi*, *eterotrofi* o *versatili* (cioè organismi autotrofi ed eterotrofi contemporaneamente), unicellulari o pluricellulari, e sono capaci di raggiungere notevoli gradi di complessità, visto che tutte le funzioni vitali sono svolte da un'unica cellula.

Vengono tradizionalmente suddivisi in:

- **protozoi**, eterotrofi unicellulari in grado di condurre vita libera o parassita e di riprodursi generalmente per via asessuata, anche se in alcuni casi possono avere cicli sessuati. In base al tipo di mobilità, si distinguono in:
 - **mastigofori**: dotati di uno o più flagelli a seconda delle dimensioni, vivono per la maggior parte dentro o sopra altri organismi (*Trichonympha*, ad esempio, è un protozoo che vive nel tubo digerente delle termiti, rendendole capaci di digerire la cellulosa),
 - **sarcodini**: sono dotati di pseudopodi, estensioni citoplasmatiche che consentono gli spostamenti. Comprendono amebe e organismi ameboidi; alcuni sono forniti di conchiglia esterna, come i *foraminiferi*, i cui gusci, accumulati nel tempo, danno origine alle scogliere calcaree,
 - **ciliofori**: dotati di ciglia, sono i protozoi più complessi. Hanno due tipi di nuclei, i macro-nuclei e i micronuclei, trasferibili da una cellula all'altra,
 - **opalinidi**: dotati di ciglia come i ciliofori, sono tutti parassiti e vivono nel tubo digerente di vari animali (rane, pesci, rettili),
 - **sporozi**: sono parassiti privi di ciglia o flagelli, caratterizzati da cicli vitali complessi, come il *Plasmodium vivax*, protozoo responsabile della malaria;
- **funghi mucilluginosi**, rappresentati da organismi pluricellulari o plurinucleati, cioè dotati di più nuclei. Si riproducono mediante *spore*, cellule riproduttive aploidi capaci di generare nuovi organismi senza fondersi con altre cellule, prodotte per meiosi dagli sporangi;
- **muffe d'acqua**, costituite da cellule plurinucleate, come i funghi mucilluginosi, ma con la peculiarità di possedere cellule riproduttive flagellate. Sono saprobie o parassite, come la *Phytophthora infestans*, che causa la peronospora nella patata, responsabile di gravi carestie;

- **alghe**, distinte in sei divisioni in base al tipo di pigmenti fotosintetici accessori, al tipo di riserva alimentare accumulata e alla presenza di parete cellulare. Tutte possiedono come principale pigmento fotosintetico la clorofilla *a*, mentre gli altri pigmenti possono essere sfruttati per aumentare la porzione di luce solare utilizzabile.

In particolare:

- **Euglenophyta**, **Crisophyta** e **Dinoflagellata** sono alghe prevalentemente unicellulari, che vivono in prossimità della superficie del mare o delle acque dolci, sfruttando la luce solare, l'anidride carbonica e i minerali disciolti nell'acqua. Costituiscono il *fitoplancton*, la parte fotosintetica del plancton, che compie gran parte dell'attività fotosintetica avvenute luogo negli oceani, grazie alla quale l'atmosfera viene rifornita di grandi quantità di ossigeno;
- **Chlorophyta**, **Phaeophyta** e **Rhodophyta** sono, invece, alghe pluricellulari adattatesi a vivere in acque poco profonde o lungo le coste, dove il martellamento continuo dei frangenti ha determinato, in alcune alghe, lo sviluppo dell'aptere, una struttura che consente l'ancoraggio dell'alga alla roccia.

L'acqua fa da filtro alla luce solare e assorbe le lunghezze d'onda maggiori (luce rosso-arancio). Per questo, alcune *alghe* hanno sviluppato pigmenti accessori che consentono di assorbire e utilizzare una porzione più ampia dello spettro solare, assorbendo anche lunghezze d'onda minori (luce verde-azzurra). Le *alghe rosse* o *rodofite* hanno come pigmenti la clorofilla *a*, i carotenoidi, la ficoeritrina ed assorbono la luce blu; le *alghe brune* o *feofite* hanno come pigmenti la clorofilla *a* e *c*, e i carotenoidi; le *alghe verdi* o *clorofite* contengono la clorofilla *a* e *b*, ed assorbono principalmente la luce rossa.

Le alghe verdi sono interessanti dal punto di vista evolutivo perché esistono in forma sia unicellulare sia pluricellulare, nonché in un certo numero di forme intermedie, che devono aver segnato il passaggio dall'una all'altra. Tra queste vi sono le **colonie**, in cui gruppi di cellule associate e collegate da filamenti citoplasmatici vivono insieme, mantenendo ognuna la propria autonomia. Esistono forme **plurinucleate**, un tipo di forma intermedia tra l'unicellularità e la pluricellularità, in cui devono essersi verificate varie divisioni nucleari in successione, alle quali, però, non è seguita una divisione citoplasmatica. Infine, ci sono vere e proprie forme **pluricellulari**, in cui alla divisione nucleare seguono normalmente la divisione citoplasmatica e la formazione della parete, senza che le cellule si separino l'una dall'altra. Nelle alghe verdi si osserva, altresì, una certa varietà di cicli vitali, da quello tipico dei protisti e dei funghi, in cui l'organismo è aploide per quasi tutto il ciclo e la meiosi avviene subito dopo la fecondazione, a quello caratteristico delle piante, in cui si osserva la già citata «alternanza di generazioni». La comunanza dei cicli vitali, il possesso degli stessi pigmenti fotosintetici (clorofilla *a* e *b*), la similitudine delle pareti cellulari e del meccanismo di citodieresi, nonché l'utilizzo dello stesso polimero per l'accumulo delle sostanze, l'amido, lasciano pensare che le piante si siano evolute a partire dalle alghe verdi.

3 Il regno dei funghi

Il *regno dei funghi* si compone di **organismi unicellulari o pluricellulari** le cui caratteristiche sono tanto diverse da quelle di piante ed animali da essere considerati un regno a sé stante. I funghi sono costituiti da filamenti, detti *ife*, le cui pareti contengono *chitina*, un polisaccaride assente nelle piante, ma presente nell'esoscheletro di insetti e artropodi. Le ife sono a loro volta riunite a formare il *micelio*. Sono organismi *eterotrofi*, *saprobi* o *parassiti*, che riversano enzimi digestivi sulle sostanze organiche fino a decomporle in molecole più piccole, destinate ad essere assorbite. Insieme ai batteri rappresentano i principali organismi decompositori della materia

organica e per questo crescono in qualsiasi ambiente in cui questa sia disponibile (acqua, suolo etc.). Si tratta di organismi immobili, il cui unico mezzo di «locomozione» è l'accrescimento.

I **lieviti** sono *funghi unicellulari* ampiamente utilizzati nella preparazione di *alimenti*, quali pane, vino, birra, formaggi, e *antibiotici*, come nel caso del fungo *Penicillium*, da cui si ottiene la penicillina. Alcuni possono provocare malattie in piante, uomini e animali.

I *funghi pluricellulari*, invece, vengono raggruppati, a seconda delle caratteristiche riproduttive, in quattro divisioni:

- **ascomiceti, basidiomiceti, zigomiceti**, capaci di riprodursi per via sessuata o asessuata;
- **deuteromiceti**, in grado di riprodursi solo per via asessuata.

La *riproduzione asessuata* avviene in alcuni casi per frammentazione delle ife (per cui ogni frammento produce un nuovo individuo), oppure, in alcuni funghi, attraverso la creazione di spore asessuate, gli *sporangî*, prodotte in apposite ife da strutture denominate *sporangiofori*.

La *riproduzione sessuata*, invece, ha luogo mediante il contatto fra ife di due diversi ceppi di accoppiamento, che prima o dopo la fusione sviluppano i *gametangi*, strutture specializzate in cui più coppie di nuclei aploidi si fondono per formare zigoti diploidi. Ne deriva una cellula plurinucleata attorno alla quale si costruisce una parete rigida, lo *zigosporangio quiescente*, dopodiché, al termine del periodo di quiescenza, si ha la germinazione e i nuclei diploidi effettuano la meiosi, producendo spore aploidi.

I funghi possono dare origine anche ad **associazioni simbiotiche**, cioè associazioni a lungo termine tra due organismi, fra i quali può sussistere una *relazione parassitaria*, se uno dei due trae vantaggio dall'associazione a scapito dell'altro, o di reciproco vantaggio, caso in cui si generano individui con caratteristiche diverse da entrambi gli organismi d'origine. I **licheni**, ad esempio, sono associazioni simbiotiche tra un fungo e un'alga verde o un cianobatterio. Estremamente diffusi in natura, costituiscono i primi colonizzatori delle pareti rocciose nude, in quanto non necessitano di fonti di nutrienti organici, grazie alla presenza dell'organismo autotrofo, né di acqua, giacché riescono ad assorbire quella piovana. Si riproducono per via asessuata, mediante la separazione di ife fungine contenenti cellule fotosintetiche, oppure in seguito alla cattura di un'alga o di un cianobatterio da parte della componente fungina staccatasi dal lichene.

Le **micorrize**, invece, sono associazioni tra un fungo e la radice di piante vascolari, importantissime per la colonizzazione dei suoli da parte delle piante, poiché facilitano l'assorbimento dei nutrienti presenti nel suolo. Si distinguono in *endomicorrize*, in cui le ife fungine penetrano nella radice, formando spirali o ramificazioni ed estendendosi, contemporaneamente, nel suolo circostante, ed *ectomicorrize*, in cui le ife fungine costruiscono una sorta di guaina attorno alla radice. In entrambi i casi, le micorrize corrispondono a una sorta di estensione dell'apparato radicale della pianta, alla quale vengono messe a disposizione sostanze organiche in decomposizione presenti nel suolo, mentre il fungo, a sua volta, può assorbire sostanze dalla pianta.

4 Il regno delle piante

Le *piante sono organismi fotosintetici pluricellulari* che si sono adattati a vivere sulle terre emerse, originandosi, probabilmente, dalle alghe verdi, con le quali condividono molte caratteristiche. Rispetto a quello acquatico, l'ambiente terrestre offre il vantaggio di una maggiore disponibilità di luce, che non viene assorbita dall'acqua, e di elevate quantità di anidride carbonica nell'atmosfera.

Lo spostamento sulla terraferma ha comportato una serie di adattamenti comuni a tutte le piante terrestri, fra i quali:

- la formazione della **cuticola**, un rivestimento ceroso che serve a preservare le parti aeree della pianta dalla perdita d'acqua;

- la formazione degli **stomi**, aperture presenti sulle foglie, attraverso le quali avvengono gli scambi gassosi;
- lo sviluppo di **organi riproduttivi**, come gli sporangi e i gametangi.

Dopo lo spostamento sulla terraferma, le piante si differenziarono in due linee evolutive distinte: le *brيوفite*, generalmente di piccole dimensioni, prive di apparati vascolari e radicali, e le *piante vascolari*, che, disponendo di apparati per la distribuzione di acqua, minerali e nutrienti, hanno potuto assumere dimensioni maggiori.

a) Briofite

Le *briofite* sono piante che restano ancorate al terreno mediante i *rizoidi*, costituiti da cellule allungate o filamenti di cellule. Hanno un ciclo vitale con alternanza di generazioni, che ha inizio con la germinazione della spora aploide, la quale produce una serie di filamenti da cui si originano i *gametofiti*. All'interno del gametofito, il gametangio femminile, l'*archegonio*, crea la cellula uovo, mentre quello maschile, l'*anteridio*, produce le cellule spermatiche, che, in condizioni di umidità sufficiente, vengono liberate e possono raggiungere l'archegonio, dove ha luogo la fecondazione. Lo zigote si sviluppa all'interno dell'archegonio formando lo *sporofito*, un individuo diploide che, maturando, produce per meiosi spore aploidi, da cui poi il ciclo ricomincia.

Una particolarità delle briofite è quella di avere un gametofito aploide più grande dello sporofito diploide, indipendente dal punto di vista nutritivo. Lo sporofito, invece, non essendo fotosintetico, si forma e rimane all'interno del gametofito, da cui dipende sotto il profilo nutritivo. Si ricordi, infine, che le briofite possono riprodursi anche per via asessuata, tramite frammentazione.

b) Piante vascolari

La prima *pianta vascolare*, la *Rhynia maior*, era molto simile esteticamente ad una briofita, ma possedeva, all'interno del fusto, un cilindro di tessuto vascolare specializzato nel trasporto di sostanze.

Attualmente, tutte le piante terrestri superiori sono piante vascolari o **tracheofite**, complessivamente divise in *nove raggruppamenti*.

Una prima distinzione riguarda le piante con seme e quelle che invece ne sono prive. Nell'ambito delle piante con seme, si distinguono ulteriormente **gimnosperme** (piante con il seme nudo) e **angiosperme** (piante con il seme protetto). Rispetto all'antenata *Rhynia maior*, queste piante hanno sviluppato varie strutture comuni:

- la **radice**, la prima innovazione comparsa nelle piante vascolari, indispensabile per l'ancoraggio al terreno e per l'assorbimento dell'acqua;
- la **foglia**, struttura specializzata nella fotosintesi;
- il **sistema vascolare**, distinto in **xilema**, deputato al trasporto di acqua e ioni (linfa grezza) dalle radici alle foglie, e **floema**, il cui compito consiste nel distribuire i prodotti della fotosintesi dalle foglie alle parti non fotosintetiche della pianta;
- il **gametofito ridotto**, che nelle piante vascolari, al contrario di quanto accade nelle briofite, è più piccolo dello sporofito. Nelle piante prive di semi, come le felci, il gametofito continua ad essere indipendente dallo sporofito sotto il profilo nutritivo; nelle piante con semi, invece, è presente sempre nelle due forme maschile e femminile, entrambe di dimensioni ridottissime e dipendenti dallo sporofito su cui si sono originate;
- il **seme**, la più recente innovazione delle piante vascolari, in cui viene custodito e protetto l'embrione, ossia il giovane sporofito. Nel tegumento del seme (guscio) l'embrione può rimanere quiescente anche per anni, fino al sopraggiungere di condizioni favorevoli alla sua germinazione.

La comparsa delle prime piante con semi risale al *periodo Carbonifero* (360-286 milioni di anni fa), al quale seguirono glaciazioni e periodi di siccità che sottoposero piante e animali a forti pressioni selettive. In quell'occasione le piante con semi, capaci di sopravvivere in condizioni avverse, acquistarono un forte vantaggio sulle altre.

Durante il *Permiano* (286-245 milioni di anni fa) avvenne poi la diversificazione delle gimnosperme, attualmente divise in *tre gruppi*, il più abbondante dei quali è quello delle **conifere** (pini, abeti, cipressi, ginepri, sequoie giganti), così chiamate perché dotate di organi riproduttori a forma di cono: i coni maschili portano sporangi che producono granuli pollinici, da cui si originano gametofiti maschili, mentre i coni femminili, più grandi e legnosi, racchiudono ovuli atti a produrre gametofiti femminili.

Nelle *gimnosperme* l'incontro tra la cellula spermatica di una pianta e la cellula uovo di un'altra è affidato all'azione del vento: per questo motivo l'ovulo spesso trasuda un liquido viscoso, deputato a catturare i granuli pollinici per facilitare la fecondazione.

I granuli pollinici e il fluido viscoso dal sapore zuccherino costituirono certamente una fonte di cibo per gli insetti, che inavvertitamente trasportavano polline da una pianta all'altra. Questo tipo di impollinazione, ben più efficace di quella affidata al vento, sicuramente costituì una forte spinta evolutiva per le piante impollinate da insetti, cosicché, durante il *Cretaceo* (144-66 milioni di anni fa), dalle gimnosperme ebbero modo di evolversi le *angiosperme*. In queste piante comparvero due nuove strutture, il *fiore* e il *frutto*, utili per indurre gli animali a realizzare le strategie riproduttive della pianta. Si sviluppò quindi il *nettario*, una struttura secernente nettare, racchiusa in un fiore che richiamava gli insetti impollinatori e, con le sue forme particolari, limitava le visite solo ad alcuni di essi.

La coevoluzione di piante ed insetti impollinatori fece in modo di limitare la dispersione del nettare su piante di altre specie, mentre gli insetti avevano il vantaggio di disporre di una fonte di cibo praticamente esclusiva.

I *frutti*, invece, sono le strutture attraverso le quali la pianta ha risolto il problema della dispersione dei semi: il nuovo embrione, contenuto e protetto nel seme, per potersi sviluppare ha bisogno di essere trasportato lontano dalla pianta d'origine, là dove siano disponibili spazio e luce solare. I semi vengono ingeriti insieme al frutto e trasportati dall'animale; in alcuni casi il transito nel canale digerente e il contatto del tegumento del seme con i succhi digestivi costituiscono condizioni necessarie alla germinazione.

5 Il regno animale

Gli *animali* sono *organismi pluricellulari eterotrofi*, accomunati dalla caratteristica di doversi procacciare il cibo. La modalità di nutrimento è principalmente l'ingestione, attraverso la quale il cibo è introdotto in cavità interne, dove viene digerito. La necessità di procurarsi il cibo ha determinato negli animali lo sviluppo di organi di locomozione di tipo diverso a seconda della specie, che in ogni caso richiedono la presenza di cellule contrattili (cellule muscolari) e di sistemi nervosi per la trasmissione degli impulsi e la coordinazione dei movimenti.

La necessità di movimento pone anche un limite alla massa raggiungibile dall'animale, cosicché contrariamente alle piante, che continuano ad accrescersi per tutta la vita, la maggior parte degli animali, raggiunta l'età adulta, arresta la propria crescita. Inoltre, al fine di limitare il peso corporeo, le sostanze nutritive in eccesso vengono accumulate sotto forma di glicogeno e di grassi che, a parità di peso, consentono di immagazzinare una quantità di energia maggiore rispetto all'amido.

I fossili dei primi animali, risalenti a 670 milioni di anni fa, evidenziano come questi avessero strutture piatte e fogliiformi, al fine di conseguire un sufficiente rapporto superficie/volume. Gli animali moderni, come vedremo, sono invece dotati di strutture tubolari, all'interno delle quali sono presenti cavità.

a) Invertebrati

Gli *invertebrati* rappresentano il 95% di tutti gli animali esistenti e mostrano un'enorme varietà, passando da forme microscopiche ad altre molto grandi. Si dividono in: **poriferi o spugne, cnidari, platelminti, nemertini, nematodi, molluschi, anellidi, artropodi, echinodermi.**

• Poriferi o spugne

Sono organismi sessili in età adulta, che vivono in acque poco profonde e si nutrono filtrando l'acqua. Hanno un livello di organizzazione interna che si può considerare intermedio tra una colonia e un organismo pluricellulare, poiché le cellule non sono organizzate in tessuti o organi, ma si riconoscono reciprocamente mantenendo un certo assetto e, contrariamente a quanto accade nelle colonie, sono differenziate tra loro.

La *struttura* delle spugne è piuttosto semplice, costituita da uno strato esterno di **cellule epiteliali** contenenti fibre contrattili, uno più interno di **amebociti**, dotati di spicole inorganiche di silice o di carbonato di calcio, che costituiscono una struttura di sostegno, e infine da uno strato ancora più interno di **coanociti**, cellule finalizzate al nutrimento e dotate di ciglia e flagello. L'acqua fa ingresso nella spugna dai pori della parete e si sposta verso l'*osculo*, il foro superiore da cui poi fuoriesce, spinta dai flagelli dei coanociti che ne dirigono il flusso, mentre le ciglia trattengono le sostanze nutritive. La *digestione* ha luogo nella singola cellula, per cui anche le spugne di grandi dimensioni possono nutrirsi unicamente di particelle microscopiche.

La *riproduzione* può essere *asessuata*, cioè per distacco di frammenti, oppure sessuata. Si tenga presente che molte spugne, essendo organismi sessili, sono *ermafrodite*, cioè possiedono strutture riproduttive sia maschili sia femminili.

• Cnidari

Sono *animali acquatici a simmetria radiale*, cioè simmetrici rispetto ad un asse centrale. Hanno cellule organizzate in tessuti e strutture di tipo polipoide (a forma di vaso) o medusoide (a forma di campana). In entrambe le forme si riconosce uno strato esterno, l'**epidermide**, e uno interno, il **gastroderma**, che riveste la cavità gastrovascolare a fondo cieco, mentre fra i due strati si colloca una sostanza gelatinosa, detta **mesoglea**. Meduse e polipi possiedono tentacoli con i quali catturano il cibo e lo introducono nella cavità gastrovascolare, dove appositi enzimi iniziano la digestione, che si completa all'interno delle cellule. La parete esterna è rivestita da **cnidociti**, cellule dotate di una cavità interna in cui è situata la *nematocisti*, un tubo avvolto a spirale che può fuoriuscire a seguito di stimoli meccanici, liberando una sostanza urticante.

Gli cnidari sono dotati di *sistemi nervosi* primitivi, costituiti da una rete di cellule nervose poste al di sotto della superficie esterna, e di rudimentali organi di senso, come le **macchie oculari**, organi fotorecettori. Possiedono inoltre un organo sensibile alla gravità, la **statocisti**, formata da una cavità rivestita di cellule ciliate, in cui è contenuto un granello di sale di calcio, detto *statolite*: questo, toccando le cellule della parete, stimola le cellule nervose, fornendo il senso della posizione.

Il ciclo vitale è caratterizzato da una rapida *riproduzione* asessuata durante la forma polipoide e dalla ricombinazione genica nella forma medusoide. Dall'unione dei gameti prodotti dalle forme medusoidi si origina la *planula*, una forma larvale liberamente natante, la quale, una volta ancoratasi al fondo, si sviluppa in forma polipoide. Il polipo, a sua volta, genera per via asessuata grosse colonie, per poi evolvere nuovamente in forma medusoide.

• Platelminti

Denominati anche **vermi piatti**, sono animali a *simmetria bilaterale*, ossia simmetrici rispetto ad un asse longitudinale, struttura che consente una migliore capacità di movimento rispetto

al gruppo degli cnidari. In essi sono, quindi, distinguibili due superfici (dorsale e ventrale) e due estremità (cefalica e caudale). Tipico degli animali molto attivi è il possesso di un'estremità anteriore, che nei plattelminti ospita sia rudimentali organi di senso (macchie sensibili alla luce e sensori di sostanze chimiche), sia cellule nervose raggruppate in gangli (precursori del cervello).

In questi organismi sono presenti tre strati di tessuto embrionale: l'**ectoderma**, da cui si originano tessuti di rivestimento interni ed esterni, e i tessuti nervosi; il **mesoderma**, da cui derivano i muscoli; l'**endoderma**, da cui si forma la cavità digerente. Inoltre, inizia a comparire l'associazione di tessuti diversi per la formazione di rudimentali organi. La cavità gastrovascolare è a fondo cieco, cioè possiede un'unica apertura, la bocca, per mezzo della quale vengono aspirati frammenti di animali morti ed eliminati i residui non utilizzati. Il movimento è assicurato da una complessa muscolatura. I plattelminti conducono generalmente vita libera, anche se esistono forme parassite, come la *tenia*, che, assumendo sostanze nutritive dall'organismo ospite, hanno perso il tubo digerente.

- Nemertini

Detti anche **vermi filiformi**, sono organismi interessanti dal punto di vista evolutivo, perché rispetto ai plattelminti hanno un *apparato digerente* a senso unico, dotato di bocca e ano. Ciò consente una maggiore efficienza nell'alimentazione, che, oltre a poter essere continua, in quanto il cibo si sposta sempre nella stessa direzione, permette la «specializzazione» dei diversi tratti del tubo digerente. Inoltre sono dotati di una *proboscide retrattile*, con cui afferrano e ingoiano il cibo, e di un rudimentale *apparato circolatorio*, con un vaso dorsale e due laterali in cui circola sangue incolore.

- Nematodi

Sono i cosiddetti **vermi cilindrici**, dotati di una struttura simile ai nemertini, con un'organizzazione corporea in tre strati di tessuto e un tubo digerente a senso unico. Sono significativi dal punto di vista evolucionistico per la comparsa di uno **pseudoceloma**, una cavità colma di liquido presente tra endoderma e mesoderma, che funziona da scheletro idrostatico, aiutando l'animale a riacquistare la propria forma originaria dopo la contrazione muscolare. Presentano, inoltre, la particolarità di essere rivestiti da una cuticola esterna, che viene periodicamente cambiata in seguito all'accrescimento dell'animale.

- Molluschi

Sono contraddistinti da un *corpo molle* e da una *conchiglia di protezione*. Si distinguono in: **bivalvi**, dotati di due conchiglie unite mediante un perno, come ostriche, vongole e mitili; **gasteropodi**, dotati di una conchiglia unica, come le lumache; **cefalopodi**, come polpi, seppie e calamari, nei quali la conchiglia è del tutto scomparsa o rientrata all'interno del corpo.

Muniti di una complessa *organizzazione corporea*, i molluschi sono organismi in cui compaiono quasi tutti gli apparati che caratterizzano gli animali superiori. Il corpo può essere distinto in tre parti: un **capo-piede**, contenente gli organi sensoriali e motori; una **massa viscerale**, in cui sono ospitati gli organi per la digestione, l'escrezione e la riproduzione; un **mantello**, che riveste la massa viscerale, secerne la conchiglia e contiene la cavità viscerale ospitante le branchie, nelle quali scaricano gli apparati digerente, escretore e riproduttore.

Ad eccezione dei bivalvi, che sono organismi filtratori, tutti i molluschi possiedono la *radula*, una sorta di spazzola dentellata posta nel primo tratto dell'*apparato digerente*, a sua volta diviso in due parti: la prima preposta alla digestione e all'assorbimento del cibo; la seconda alla compressione dei rifiuti e al loro «impacchettamento» in grumi fecali, espulsi nella cavità del mantello, da cui poi vengono asportati dalle correnti d'acqua.

Questi animali possiedono anche un efficiente *apparato circolatorio*, dotato di cuori che pompano il sangue e di vasi che lo trasportano ai tessuti. In alcuni la circolazione è di tipo aperto: il sangue, cioè, non circola unicamente all'interno dei vasi, ma viene liberato negli spazi fra i tessuti, per poi essere raccolto nelle branchie, ossigenato e portato nuovamente al cuore, che lo distribuisce al resto del corpo attraverso i vasi. Il cuore è generalmente costituito da tre comparti: due atri e un ventricolo. Le *branchie* corrispondono a polmoni primitivi, in cui la diffusione di ossigeno e anidride carbonica è resa più efficiente mediante l'aumento della superficie di scambio tra acqua e sangue. I molluschi sono inoltre dotati di un *apparato escretore*, formato da uno o due *nefridi*, tubuli dotati di un'apertura in prossimità del cuore, da cui entra il liquido corporeo, e da una cavità del mantello, dalla quale vengono eliminate le sostanze di rifiuto. Infine, in essi è presente anche un *celoma*, seppure di dimensioni assai ridotte.

• Anellidi

Appartengono a questo *phylum* molte specie di *vermi marini, terrestri e di acqua dolce*, tutti caratterizzati dalla segmentazione del corpo in anelli, visibili anche dall'esterno e divisi internamente da pareti. In questi animali è presente un *celoma*, una cavità interna al corpo che ne occupa una parte significativa, in cui si trovano organi ed apparati sospesi da doppi strati di mesoderma, i *mesenteri*. Lo sviluppo del celoma è stato un evento importante dal punto di vista evolutivo, perché ha permesso un migliore «impacchettamento» degli apparati, che, incurvandosi e sovrapponendosi tra loro, hanno potuto aumentare l'efficienza degli scambi. Negli anellidi il celoma svolge, altresì, una funzione fondamentale nel movimento, assicurato da muscoli distinti in due serie, una longitudinale e l'altra circolare, e divisi per ogni segmento in modo tale che ognuno possa contrarsi indipendentemente dall'altro. A sua volta, ciascun segmento è munito di setole per l'ancoraggio al suolo o di *parapodi* (per le forme marine).

L'*apparato digerente*, assai evoluto, è differenziato in *bocca, faringe muscolare* (che aspira le particelle di cibo), *ingluvie* (che immagazzina il cibo stesso), *ventriglio* (che sminuzza il cibo grazie all'azione meccanica del suolo) e *intestino* (sede delle attività di digestione e assorbimento). Il *sistema circolatorio* è formato da un lungo vaso muscolare dorsale e da lunghi vasi longitudinali ventrali, che percorrono il corpo da un'estremità all'altra, in cui confluiscono i vasi parietali, deputati a raccogliere il sangue dai tessuti per riportarlo al vaso dorsale. Nei primi cinque segmenti sono presenti altrettanti cuori, costituiti semplicemente da zone muscolari pulsanti dei vasi. Il *sistema nervoso*, strutturato a forma di catena, in modo da permettere il controllo di ogni segmento corporeo, possiede cellule sensoriali, per il tatto, il gusto, il grado di umidità, e cellule fotosensibili. Il *sistema escretore* è costituito da due coppie di nefridi per ogni segmento.

• Artropodi

Così denominati perché muniti di zampe articolate, sono *animali segmentati* (come gli anellidi), nei quali il numero dei segmenti si è notevolmente ridotto nel corso dell'evoluzione e vari segmenti si sono fusi fino a formare, generalmente, tre regioni distinte: capo, torace, addome. Il grado di fusione dei segmenti, insieme alle caratteristiche delle appendici articolate, specializzate in diverse funzioni, permette di distinguere gli artropodi in **aracnidi** (acari, ragni, scorpioni), **crostacei** (granchi, aragoste) e **insetti**.

Una caratteristica comune a tutti è la presenza dell'**esoscheletro**, un rivestimento esterno rigido costituito da chitina, che ha funzioni di protezione, limita la perdita d'acqua per le forme terrestri ed è articolato in modo da consentire il movimento. L'**esoscheletro** non si ingrandisce insieme all'animale, che quindi, nel corso dell'accrescimento, subisce diverse mute, durante le quali vengono secreti enzimi che ne provocano la rottura. Appena abbandonato l'**esoscheletro**, l'animale si accresce rapidamente, assorbendo acqua ed aria prima che il nuovo, già in forma-

zione, si indurisce. Negli insetti l'esoscheletro riveste anche le trachee, dotti che portano l'aria alle diverse regioni del corpo. Gli artropodi terrestri possiedono, invece, polmoni a libretto.

Un'altra struttura distintiva di questo *phylum* è costituita dai *tubuli malpighiani*, preposti all'*escrezione* e collegati all'intestino posteriore. Il *sistema circolatorio* è aperto e dotato di un cuore tubolare; gli *organi sensoriali* sono estremamente sensibili; il *sistema nervoso* è complesso e segmentato, in modo che alcune attività siano controllate a livello locale, altre dal cervello, a sua volta costituito dalla fusione, nella parte anteriore del corpo, di coppie di gangli dorsali.

Il successo biologico degli insetti, testimoniato dall'abbondanza e dalla varietà degli individui esistenti, è dovuto sia all'elevata varietà di microambienti forniti dalle piante vascolari, sia alle caratteristiche del loro ciclo vitale, che prevede, per la maggior parte di essi, quattro forme: **uovo, larva, pupa, adulto**. Allo stadio larvale l'insetto, dotato di scarsissima mobilità, si accresce in prossimità del luogo di deposizione delle uova. Nella fase di sviluppo, quasi tutti gli insetti subiscono una metamorfosi che determina un profondo rimodellamento dell'organismo, a seguito del quale l'insetto assume un aspetto completamente diverso. Il momento di passaggio dalla *forma larvale* a quella *adulta* corrisponde allo stadio di *pupa*, che, insieme all'uovo, costituisce una forma resistente a condizioni ambientali avverse, come caldo e freddo intensi.

- Echinodermi

Sono animali a *simmetria radiale*, con il corpo suddiviso in un numero di sezioni pari a cinque o multiplo di cinque, detta per questo *simmetria radiale pentamera*, acquisita in uno sviluppo secondario, poiché la larva possiede simmetria bilaterale. Appartengono a questo *phylum* gigli di mare, stelle marine, cetrioli di mare, tutti accomunati dalla presenza di un **endoscheletro**, cioè uno scheletro interno munito di spine protettive e formato da piastre separate, tenute insieme da tessuto muscolare o epidermico. Un'altra caratteristica del *phylum* è la presenza di un sistema acquifero, cavità celomatica modificata che fa da sostegno ai pedicelli ambulacrali, strutture con cui gli echinodermi aderiscono al substrato, si muovono e catturano le prede.

b) Vertebrati

I *vertebrati* appartengono al *phylum* dei **cordati** e sono costituiti da quattro componenti fondamentali: la **notocorda**, un cordone cartilagineo che si estende per tutta la lunghezza del corpo; il **cordone nervoso dorsale cavo**, un tubo pieno di liquido che corre sotto la superficie dorsale dell'animale; la **faringe**, dotata di fessure branchiali; la **coda**, posteriore all'ano.

Questi animali, la cui evoluzione dai cordati è testimoniata da numerosi reperti fossili, sono caratterizzati sia da un *endoscheletro osseo*, che cresce insieme all'organismo, sia dal possesso di una *colonna vertebrale flessibile*, che in genere costituisce l'asse del corpo. La colonna, formata da vertebre intercalate con dischi cartilaginei, che le conferiscono flessibilità, circonda e protegge il cordone nervoso, mentre il cervello di solito è racchiuso tra piastre ossee. Fra i vertebrati si distinguono le classi di: **pesci, anfibi, rettili, uccelli e mammiferi**.

- Pesci

I primi esemplari di questa classe erano privi di mascelle, caratteristica mantenuta attualmente solo da lamprede e missinoidi. La trasformazione degli archi branchiali anteriori in mascelle ha costituito un fattore evolutivo di grande importanza, poiché ha permesso un notevole aumento delle dimensioni dei *pesci*, mettendo a loro disposizione una maggiore quantità di risorse alimentari. Probabilmente questi animali hanno trascorso gran parte della loro storia evolutiva nelle acque dolci e si sono adattati a vivere in mare solo in un periodo successivo. In alcuni di essi, le acque dolci poco profonde, spesso carenti di ossigeno, determinarono lo sviluppo dei *polmoni* o di strutture simili, che agissero di supporto alle *branchie* in condizioni di necessità.

Dagli antichi pesci polmonati si sono poi sviluppati i pesci polmonati attuali, mentre in altri il polmone si è evoluto in strutture diverse: in alcuni, ad esempio, è diventato una *vescica natatoria* che funziona da camera di galleggiamento, consentendo al pesce di mantenersi a una certa profondità tramite l'aggiunta o l'eliminazione di aria. In alcuni pesci polmonati si svilupparono anche *sostegni scheletrici* per il torace, che permettevano di spostarsi sui fondali fangosi dei torrenti o addirittura sul terreno, per andare in cerca di acque più profonde: il passaggio sulla terra emersa potrebbe essersi verificato proprio in questo modo.

• Anfibi

Si sono evoluti dai pesci polmonati che respiravano aria e comprendono *rane*, *rospi*, *salamandre*. Queste ultime sono facilmente distinguibili dai rettili perché prive di squame o scaglie e dotate di una pelle molto sottile, che funge da organo respiratorio: alcune respirano unicamente attraverso la pelle e le mucose della bocca. Le rane, invece, sono munite anche di polmoni contenenti aria e hanno due stadi di vita, da cui il nome di *anfibi*: le uova vengono deposte nell'acqua e fecondate esternamente, dopodiché dall'uovo fecondato nasce una larva munita di branchie, il **girino**, che a sua volta, trasformandosi in adulto, perde le branchie e sviluppa i polmoni.

• Rettili

Si sono evoluti dagli anfibi alla fine del Carbonifero e il numero di specie aumentò notevolmente durante il Permiano, periodo in cui si verificò un forte sviluppo anche delle conifere, segno che il clima asciutto stava agendo come forza selettiva.

La conquista delle terre emerse da parte dei rettili è da attribuire all'evoluzione dell'**uovo amniotico**, contenente una riserva d'acqua che gli permette di sopravvivere al di fuori di questa. L'uovo dei rettili è costituito da una grande quantità di *tuorlo*, che serve da nutrimento per l'embrione in via di sviluppo, e da abbondante *albume*, in cui si trovano acqua e nutrienti. L'embrione è circondato dall'*amnios*, una membrana colma di liquido che riproduce l'ambiente acquatico, e passa attraverso uno stadio branchiale, da trascorrere nell'uovo o nell'apparato riproduttivo materno.

Per adattarsi a vivere in ambienti molto asciutti i rettili hanno sviluppato una pelle secca ricoperta di *squame*, che limita la traspirazione dell'acqua. Questi animali sono tutti *ectotermi*, cioè mantengono la propria temperatura corporea scambiando calore con l'ambiente, ad eccezione di una specie di tartarughe moderne, che mostra segni di endotermia. Sebbene i *serpenti* e alcune *lucertole* abbiano perso le zampe, i rettili sono tipicamente tetrapodi. Il perfezionamento dei mezzi di locomozione e una forte spinta al bipedismo determinarono l'evoluzione degli **archeosauri**, distinti in tre gruppi principali: **pterosauri** o rettili volanti; **coccodrilli**, in grado di mantenere la condizione di quadrupedi; **dinosauri**, che dominarono la vita sulla Terra per 150 milioni di anni, fino alla misteriosa estinzione di 65 milioni di anni fa.

• Uccelli

Rappresentano l'unica linea di discendenza lasciata dai dinosauri e sono, quindi, *rettili specializzati* nel volo: le zampe ricoperte da squame sono un segno della loro origine. La struttura corporea è completamente adattata a questa funzione: i loro corpi contengono sacchi aerei, le ossa sono cave, perfino l'apparato riproduttore è ridotto ad un'unica ovaia, che si ingrandisce solo nella stagione degli accoppiamenti. L'osso più grande dello scheletro è lo sterno, su cui si inseriscono i grossi muscoli che azionano le ali.

Sono animali *endotermi*, in grado di produrre calore e l'energia necessaria al volo attraverso il metabolismo interno, mentre il *piumaggio*, oltre ad essere indispensabile al volo stesso, è un isolante termico che diminuisce la dispersione di calore. Molti uccelli nascono in uno stadio di sviluppo prematuro e, dunque, hanno bisogno di prolungate cure parentali.

- **Mammiferi**

Anch'essi, come gli uccelli, *discendono dai rettili*. Sono animali *endotermi* ricoperti di pelo, che funziona da isolante termico come il piumaggio dei volatili. Quasi tutti i mammiferi generano figli vivi e sono dotati di ghiandole specializzate nella produzione del latte, con cui alimentano i piccoli.

Quelli più comuni sono i **placentati**, che custodiscono i figli nel corpo materno fino ad uno stadio di sviluppo molto avanzato, nutrendoli attraverso la placenta. I **marsupiali**, invece, partoriscono figli immaturi, custodendoli in una tasca protettiva e allattandoli fino al completamento dello sviluppo. I **monotremi**, infine, depongono uova con guscio e allattano i piccoli dopo la schiusa.

1 Le funzioni degli organismi viventi

Tutti i viventi, dai semplici organismi unicellulari fino agli animali complessi come l'uomo, hanno dovuto risolvere una serie di problematiche legate alle relazioni dell'organismo con l'ambiente. Ogni essere vivente è caratterizzato da un *alto grado di organizzazione interna*, evidenziabile già a livello della singola cellula, che negli animali complessi si riproduce a livelli di organizzazione più alti: infatti, insiemi di cellule che assolvono alla stessa funzione sono organizzati in **tessuti**; tessuti diversi, uniti e integrati tra loro, formano gli **organi**; gruppi di organi con funzioni correlate e integrate generano gli **apparati**. Tutti gli apparati, lavorando insieme, consentono all'organismo di realizzare alcune condizioni indispensabili per la vita di ciascun organismo vivente, a prescindere dal suo grado di complessità.

Ogni vivente ha elaborato un proprio modo di rispondere a queste esigenze che porta codificato nel suo *patrimonio genetico*, in modo da poterlo trasmettere alla discendenza. La *selezione naturale* ha poi provveduto a individuare gli organismi capaci di trovare la risposta più efficiente a questo tipo di problemi.

a) Omeostasi

Ogni organismo vivente ha la necessità di conservare un ambiente interno controllato e relativamente costante, nel quale possano svolgersi correttamente i processi fisiologici. Ciò richiede un accurato controllo della temperatura, dell'ambiente chimico, nonché la protezione da agenti esterni potenzialmente nocivi. Probabilmente il mantenimento dell'*omeostasi* costituisce un problema più per i piccoli organismi unicellulari o pluricellulari che per gli organismi complessi, dotati di un minore rapporto superficie/volume e spesso di rivestimenti di protezione dall'ambiente esterno. In organismi complessi come quello umano, tutti gli apparati insieme concorrono al mantenimento dell'omeostasi.

b) Metabolismo

L'alto grado di ordine e di organizzazione interna che contraddistingue gli esseri viventi contrasta con la naturale tendenza dell'Universo all'aumento dell'entropia («disordine») e può essere conservato solo attraverso un continuo dispendio di energia: in tal senso gli esseri viventi possono essere considerati **sistemi dissipativi**.

Gli animali eterotrofi ricavano energia dal cibo, che, trasformato in sostanze utilizzabili dalle cellule tramite i processi digestivi, viene demolito in presenza di ossigeno nel processo di *respirazione cellulare*. Negli animali di grandi dimensioni sorge la necessità di distribuire le sostanze nutritive a tutte le cellule del corpo, assicurando loro un sufficiente rifornimento di ossigeno, e di eliminare le sostanze di rifiuto. Queste funzioni vengono svolte attraverso la «*cooperazione*» di quattro apparati: *digerente, respiratorio, circolatorio, escretore*.

c) Difesa

Il mantenimento dell'equilibrio interno e dell'integrità dell'organismo è spesso legato anche alla capacità di difendersi da agenti potenzialmente lesivi, sia esterni, come i microrganismi

patogeni, sia interni, come le cellule malate dell'organismo stesso. Tale funzione è assolta dal *sistema immunitario*, le cui cellule, prodotte dal midollo osseo (*sistema scheletrico*), vengono trasportate al resto dell'organismo mediante il torrente ematico (*sistema circolatorio*).

d) Regolazione e controllo

Si è visto come la soluzione ad ogni esigenza dell'organismo sia il risultato dell'azione congiunta di più apparati, le cui azioni devono essere coordinate e modificate a seconda delle momentanee e specifiche necessità. Appare dunque evidente il bisogno di un *sistema di integrazione e di controllo* di tutte queste operazioni, normalmente affidato al **sistema endocrino** e al **sistema nervoso**, i quali, difficilmente separabili dal punto di vista sia funzionale sia anatomico, cooperano alla regolazione degli altri apparati.

Il sistema endocrino agisce principalmente sulle regolazioni a lungo termine (da minuti a mesi), mentre quello nervoso svolge un'azione molto più rapida, producendo risposte nell'arco di millisecondi.

Il *sistema nervoso* ha una complessa organizzazione e può essere distinto in **sistema nervoso somatico o volontario**, che innerva la muscolatura scheletrica ed è responsabile del movimento, e **sistema nervoso autonomo o involontario**, che innerva la muscolatura liscia, il cuore, le ghiandole. Nell'ambito del sistema nervoso autonomo, si distinguono ulteriormente il **sistema simpatico**, deputato alle risposte rapide in momenti di stress o di pericolo, e il **sistema parasimpatico**, che invece regola le funzioni fisiologiche ordinarie.

I *sistemi di regolazione* presenti nell'organismo umano sono abbastanza precisi e funzionano attraverso meccanismi di *feedback* (retroazione) positivi o negativi. In pratica, questi sistemi producono una risposta contraria alla deviazione che si sta verificando nell'organismo rispetto a una condizione prefissata. Un esempio è rappresentato dalla regolazione del livello di glucosio nel sangue, realizzata da due ormoni prodotti dal *pancreas*, l'*insulina* e il *glucagone*. Quando il livello di glucosio nel sangue aumenta, come accade dopo un pasto, il pancreas secerne insulina, che stimola le cellule ad assorbire e utilizzare il glucosio; viceversa, quando il livello di glucosio si abbassa, il pancreas produce glucagone, che stimola il rilascio di glucosio da parte del fegato.

e) Riproduzione

Un'altra esigenza irrinunciabile per tutti gli organismi viventi è la *riproduzione*, a cui è legata la sopravvivenza della specie. A tal fine ognuna ha messo a punto gli organi, gli apparati e le strategie riproduttive più adatte al proprio tipo di ambiente.

2 Dalle cellule ai tessuti

Si è già detto che gli organismi viventi sono caratterizzati da un'organizzazione gerarchica interna che vede l'unione di *cellule in tessuti*, di *tessuti in organi*, di *organi in apparati*.

Nel corpo umano sono presenti oltre duecento tipi di cellule diverse, che tuttavia possono essere organizzate in tessuti riconducibili a quattro tipologie fondamentali:

- tessuto epiteliale;
- tessuto connettivo;
- tessuto muscolare;
- tessuto nervoso.

a) Tessuto epiteliale

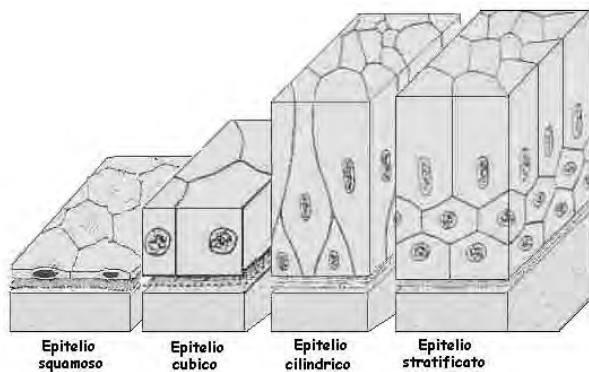


Fig. 1 - Diversi tipi di tessuto epiteliale

Il **tessuto epiteliale** è quello che *riveste e protegge il corpo* sia esternamente (pelle) sia internamente, ricoprendo organi, canali e cavità. A seconda della forma delle cellule che lo compongono, il tessuto epiteliale viene distinto in **squamoso** (costituito da cellule piatte), **cubico** o **cilindrico** (Fig. 1). Inoltre, in base al numero di strati delle cellule, si distinguono tessuti epiteliali **semplici** (*monostratificati*, come quelli dei capillari e degli alveoli polmonari) e tessuti epiteliali **composti** (*pluristratificati*, come la pelle o il rivestimento interno della bocca).

Lo strato di tessuto epiteliale è generalmente attaccato a uno strato di tessuto connettivo sottostante, chiamato **membrana basale**. Dovendo impedire l'infiltrazione di sostanze tra una cellula e l'altra, le cellule che formano i tessuti epiteliali sono strettamente unite tra loro mediante le cosiddette **giunzioni occludenti**, saldature continue tra le membrane, che appaiono quasi fuse tra loro. Nei tessuti sottoposti a sollecitazioni meccaniche, come la pelle, fra cellule adiacenti si trovano veri e propri punti di saldatura, detti **desmosomi**, costituiti da placche di materiale fibroso. Infine, le cellule degli epiteli delle cavità interne risultano modificate per secernere muco, necessario alla lubrificazione delle superfici, mentre altre cellule epiteliali, specializzate nella secrezione di sostanze specifiche (sudore, saliva, ormoni), sono raggruppate a formare *ghiandole*.

b) Tessuto connettivo

Contrariamente ai tessuti epiteliali, in cui le cellule sono strettamente addossate le une alle altre, quelli *connettivi* sono costituiti da cellule immerse in un'abbondante matrice extracellulare. Essi hanno la *funzione di connettere*, appunto, e *sostenere gli altri tessuti*, e sono dotati di un diverso grado di rigidità in base alla natura della matrice extracellulare, composta da una sostanza fluida e da fibre di diverso tipo, come fibre collagene, fibre elastiche e fibre reticolari.

Si distinguono due tipi di tessuto connettivo, differenti per la quantità e la disposizione delle fibre presenti nella matrice:

- il **tessuto connettivo lasso**, quello più diffuso, caratterizzato da un'abbondante matrice extracellulare, nonché da fibre disposte in modo irregolare e intrecciate fra loro;
- il **tessuto connettivo denso o compatto**, contraddistinto da una netta preponderanza di fibre collagene rispetto alla matrice extracellulare e alle cellule, disposte in modo parallelo a formare una struttura regolare e compatta, oltre che notevolmente resistente.

I principali tessuti connettivi presenti nell'organismo umano sono il *tessuto osseo*, il *sangue* e la *linfa*.

Il **tessuto osseo**, in particolare, è formato da *osteociti*, *osteoblasti* (che rappresentano una fase immatura degli osteociti) e *osteoclasti* (deputati al riassorbimento del tessuto invecchiato), dispersi in una matrice contenente una notevole quantità di fibre collagene e un'abbondante componente inorganica, soprattutto fosfato e carbonato di calcio, che nell'adulto arrivano al 65% del peso secco.

È organizzato in *lamelle*, piccole strutture appiattite in cui si dispongono le cellule, le fibre collagene e la matrice intercellulare, e può essere distinto in:

- **tessuto osseo spugnoso o trabecolare**, in cui le lamelle sono disposte in modo da creare un reticolo ricco di piccole cavità, che gli conferiscono l'aspetto di una spugna;
- **tessuto osseo compatto**, nel quale, invece, si distinguono strutture, dette *osteoni*, in cui le lamelle sono poste concentricamente attorno ai cosiddetti **canali di Havers**, generando una struttura più compatta (Fig. 2).

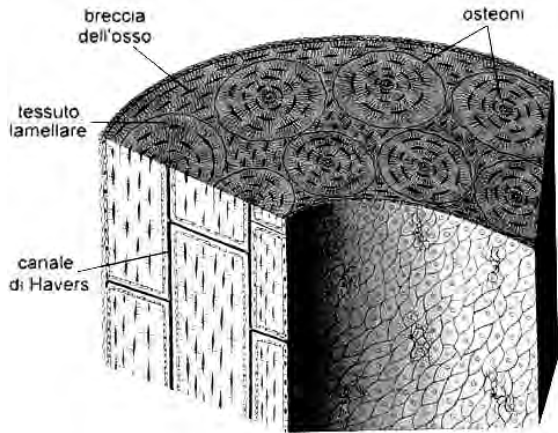


Fig. 2 - Conformazione del tessuto osseo

c) Tessuto muscolare

Quello *muscolare* è il *tessuto responsabile del movimento*, costituito da cellule in grado di contrarsi grazie alla presenza delle proteine contrattili actina e miosina. Si distingue in:

- **tessuto muscolare liscio**, tipico della muscolatura involontaria che circonda le pareti degli organi interni, dei vasi sanguigni etc.;
- **tessuto muscolare scheletrico**, tipico della muscolatura volontaria;
- **tessuto muscolare cardiaco**.

Sia nel tessuto scheletrico che in quello cardiaco le fibre di actina e miosina sono raggruppate in fasci che determinano caratteristiche striature, da cui il nome di **tessuto muscolare striato**. I muscoli scheletrici, connessi alle ossa mediante robuste strisce di tessuto connettivo, i **tendini**, sono formati da fasci di fibre muscolari tenute insieme da tessuto connettivo, ciascuna delle quali è costituita da un'unica cellula polinucleata, che può avere un diametro di 50-100 micron e una lunghezza di qualche centimetro.

d) Tessuto nervoso

Il *tessuto nervoso* è formato da due tipi di cellule: i **neuroni**, in grado di trasmettere l'impulso nervoso, e le **cellule gliali**, le quali isolano, sostengono e nutrono i neuroni.

Pur esistendo in molte varianti diverse, i neuroni hanno una struttura di base costituita da un **corpo**, contenente il nucleo e gli organuli cellulari; dai **dendriti**, corte estensioni citoplasmatiche, presenti in gran numero, deputate a raccogliere gli stimoli dalle altre cellule; e da un **assone** o **fibra nervosa**, un prolungamento che permette la trasmissione a distanza dell'impulso nervoso, che può raggiungere lunghezze notevoli: l'assone di un neurone motorio situato nel midollo spinale può arrivare fino alla punta del piede (Fig. 3).

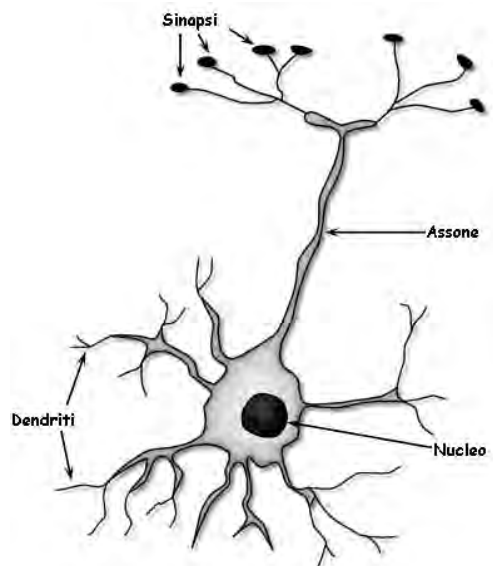


Fig. 3 - Struttura di base di un neurone

A seconda della collocazione e della funzione svolta nel sistema nervoso, si distinguono:

- **neuroni sensoriali**, che ricevono le informazioni dall'esterno e le trasmettono al sistema nervoso centrale;
- **interneuroni**, i quali trasmettono i segnali da un neurone all'altro all'interno del sistema nervoso centrale;
- **neuroni di associazione**, che trasmettono i segnali da una regione all'altra del sistema nervoso centrale;
- **neuroni motori**, incaricati di trasmettere i segnali dal sistema nervoso centrale agli organi effettori.

I neuroni possono essere collegati tra loro in vario modo, dai *semplici archi riflessi* alle *complesse connessioni del cervello*. Gli **archi riflessi** semplici consentono di rispondere in modo praticamente istantaneo a stimoli provenienti dall'ambiente; sono costituiti da un *neurone sensoriale* che trasmette il segnale a un *interneurone* situato nel sistema nervoso centrale, ad esempio nel midollo spinale, a sua volta collegato a un *neurone motorio* che stimola l'effettore.

1 Lo scheletro umano

Nell'embrione lo *scheletro* è costituito da circa 350 ossa, per la maggior parte di natura cartilaginea, alcune delle quali si fondono con la crescita e si trasformano in tessuto osseo, più ricco di minerali e quindi più rigido. Lo scheletro raggiunge la maturità intorno ai 25 anni, momento in cui la *cartilagine* è quasi scomparsa, permanendo solo nel naso, nell'orecchio e sulle superfici articolari.

In generale lo scheletro umano (Fig. 1) può essere diviso in:

- **scheletro assiale**, disposto lungo l'asse longitudinale del corpo, che comprende *cranio*, *colonna vertebrale* e *gabbia toracica*, e svolge prevalentemente una funzione di sostegno e protezione degli organi interni;
- **scheletro appendicolare**, comprendente gli *arti superiori* e *inferiori*, nonché i *cinti scapolare* e *pelvico*, le cui funzioni prevalenti riguardano il sostegno e il movimento.

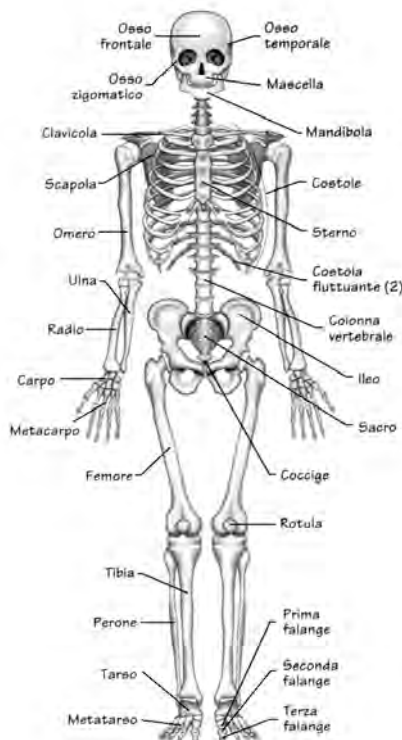


Fig. 1 - Scheletro umano

a) Ossa

A seconda della loro forma, le ossa sono classificate in:

- **ossa lunghe**, che possiedono una dimensione prevalente sulle altre, come il *femore* (Fig. 2). Sono composte da una parte cilindrica centrale, detta *diafisi*, costituita da tessuto osseo compatto, che delimita la cavità midollare, e da due parti terminali rigonfie, le *epifisi*, formate da tessuto osseo spugnoso;
- **ossa corte**, che invece possiedono tre dimensioni più o meno equivalenti, come accade per le *vertebre* (Fig. 3);
- **ossa piatte**, dotate di uno spessore minimo, come la *scapola* (Fig. 4) o le ossa del cranio.



Fig. 2 - Femore

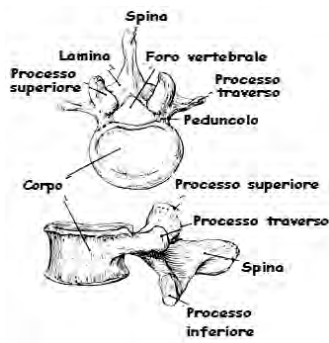


Fig. 3 - Vertebra

Sia le ossa corte sia quelle piatte sono costituite principalmente da *tessuto osseo spugnoso*. La superficie delle ossa non è perfettamente liscia, ma presenta sporgenze (*apofisi*), rilievi (*tubercoli*), spine e depressioni (solchi e fosse arrotondate, alcune delle quali modellate dal movimento dei muscoli), che permettono l'inserzione dei muscoli e dei tendini.

Strutturalmente l'osso si compone di più strati. Il **periostio** è la membrana fibrosa esterna, vascularizzata e innervata, che avvolge l'osso e termina al limite delle zone articolari. L'**endostio** è una membrana microscopica, situata all'interno del periostio, che avvolge il tessuto osseo, contraddistinto da uno strato esterno molto compatto, all'interno del quale si trova il tessuto spugnoso pervaso dal **midollo osseo**.

Quest'ultimo è un tessuto in continua riproduzione, che può essere distinto in **midollo giallo**, ricco di tessuto adiposo, e **midollo rosso**, costituito da *osteoblasti*, *osteociti* ed *osteoclasti*, le cellule che «costruiscono» le ossa e producono la maggior parte degli elementi figurati del sangue.

All'interno della struttura ossea compare, infine, una fitta rete di **canali di Havers**, che provvedono alla distribuzione dei nutrienti e attorno ai quali le cellule sono disposte concentricamente.



Fig. 4 - Scapola

b) Articolazioni

Le ossa sono connesse tra loro per mezzo delle *articolazioni* che, a seconda del grado di mobilità, si dividono in:

- articolazioni *mobili* o **diartrosi**, di cui sono esempi la *spalla* e il *ginocchio*;
- articolazioni *semimobili* o **anfiartrosi**, come la *sinfisi pubica* e le *articolazioni della spina dorsale*;
- articolazioni *fisse* o **sinartrosi**, come *quelle della scatola cranica*.

In riferimento al tipo di tessuto che collega le ossa, invece, le articolazioni si distinguono in:

- **suture**, in cui le ossa sono collegate da tessuto connettivo denso;
- **sinfisi**, nelle quali è presente un tessuto cartilagineo più o meno fibroso.

Le parti dell'osso che partecipano all'articolazione sono rivestite di *cartilagine* e nelle articolazioni mobili, al fine di ridurre l'attrito, sono ricoperte da una *membrana sinoviale*, all'interno della quale è presente il liquido sinoviale o **sinovia**. I capi articolari sono rivestiti da un mani-

cotto fibroso, la **capsula articolare**, composta da tessuto connettivale denso, il quale, inserito ai margini delle cartilagini, prosegue nei due periosti. Al collegamento fra le ossa provvedono la capsula articolare e i **legamenti**, i quali sono costituiti da fasce di tessuto connettivo fibroso e possono essere *interni* all'articolazione, anche se in realtà isolati dalla membrana sinoviale, *periferici*, se inseriti in sede articolare o periarticolare, o *a distanza*, quando risultano inseriti sulle ossa, anche a notevole distanza dall'articolazione.

c) Scheletro assiale

Lo *scheletro assiale* comprende il *cranio*, la *colonna vertebrale* e la *gabbia toracica*.

Il **cranio** è formato da ossa piatte, connesse tra loro da *suture*, cioè articolazioni immobili, fatta eccezione per le articolazioni mandibolare e cervicale. Dovendo proteggere il cervello, le ossa craniche hanno una particolare struttura a strati, che ne determina la notevole resistenza agli urti. Sono costituite da uno strato interno di tessuto osseo compatto, uno intermedio di tessuto osseo spugnoso, ricco di cavità e attraversato da vasi sanguigni, e uno esterno di tessuto osseo compatto, più spesso di quello interno. Anatomicamente, il cranio si compone di una *scatola cranica*, formata da 8 ossa piatte, che ospita e protegge il cervello, e di un *cranio viscerale*, formato da 14 ossa che costituiscono la faccia e ospitano gli organi di senso (occhi, orecchie, naso, bocca).

La **colonna vertebrale** è composta da 33 vertebre, articolate una sull'altra e suddivise in cinque regioni:

1. **zona cervicale**, costituita da 7 vertebre che consentono la rotazione della testa;
2. **zona toracica**, formata da 12 vertebre, su cui si articolano le costole;
3. **zona lombare**, composta da 5 vertebre di dimensioni maggiori rispetto alle altre, perché sostengono la maggior parte del peso corporeo;
4. **zona sacrale**, costituita da 5 vertebre saldate tra loro a formare l'osso sacro, su cui si articolano le ossa del bacino e il coccige;
5. **coccige**, formato da 4 vertebre di dimensioni ridottissime, che nei vertebrati sostengono la coda.

Le **vertebre** sono ossa corte, composte da un corpo cilindrico, disposto anteriormente, e da un arco, situato posteriormente, da cui si dipartono due sporgenze, dette **processi trasversi**. Su questi si inseriscono i muscoli e i legamenti del tronco, oppure, nella zona toracica, si articolano le costole. Le *vertebre di tessuto osseo* conferiscono potenza alla colonna vertebrale che deve sostenere il peso del corpo, mentre i *dischi cartilaginei intervertebrali* le garantiscono la necessaria flessibilità.

La colonna vertebrale, il cui canale interno contiene e protegge il midollo spinale, presenta ampie curvature (lordosi cervicale, cifosi dorsale, lordosi lombare), finalizzate ad ottimizzare la distribuzione del peso corporeo.

La **gabbia toracica**, infine, è costituita da 12 costole che si articolano posteriormente sulla colonna vertebrale. Le prime sette paia sono articolate anteriormente sullo sterno, mentre le coppie dall'ottava in poi sono legate alla cartilagine della coppia precedente. L'undicesima e la dodicesima costola sono dette *fluttuanti*, perché sono articolate solo posteriormente alla colonna vertebrale e non si chiudono anteriormente. La gabbia toracica ospita e protegge cuore e polmoni, di cui determina la dilatazione e la contrazione allargandosi e restringendosi sotto l'azione dei muscoli costali, e costituisce il supporto dei muscoli che sostengono gli organi addominali.

d) Scheletro appendicolare

Lo *scheletro appendicolare* comprende gli *arti* e i *cinti*, cioè le strutture per mezzo delle quali gli arti sono collegati allo scheletro assiale. Il **cinto scapolare** è formato dalla *scapola*, un osso piatto situato posteriormente alla gabbia toracica, e dalla *clavicola*, posta anteriormente a que-

sta. Le ossa di questo cinto non sono articolate allo scheletro assiale, ma strettamente legate ad esso tramite muscoli e legamenti. Inoltre, attraverso l'articolazione della spalla, la clavicola e la scapola sono collegate all'osso lungo dell'arto superiore: l'*omero*.

L'**arto superiore** si distingue in *braccio* (formato dall'*omero*), *avambraccio* (composto da *radio* e *ulna*) e *mano* (suddivisa in *carpo*, *metacarpo* e *falangi*). L'omero si articola con radio e ulna attraverso il *gomito*, a sua volta collegato alla mano tramite il *polso*. Spalla, gomito e polso sono articolazioni dotate della massima mobilità, meno forti rispetto alle corrispondenti articolazioni dell'arto inferiore, in quanto devono garantire la massima capacità di movimento, più che la resistenza agli sforzi.

Il **cinto pelvico o bacino** è composto dalle tre ossa dell'*anca* (*ileo*, *ischio* e *pube*), articolate posteriormente sul coccige e anteriormente tra loro. Si tratta di un complesso osseo piuttosto rigido e massiccio, capace di sostenere gli organi addominali e di garantire una salda articolazione agli arti inferiori. Il cinto pelvico presenta forti differenze tra uomini e donne, legate soprattutto alla riproduzione: il bacino femminile, infatti, è maggiormente sviluppato in larghezza e ha lo stretto superiore più ampio e di forma ovale, per riuscire a sopportare il peso del feto e consentirne l'uscita durante il parto.

Ogni osso dell'anca è articolato con il *femore*, l'osso lungo dell'**arto inferiore**, a sua volta distinto in *coscia* (formata dal femore), *gamba* (costituita da *tibia* e *perone*) e *piede* (diviso in *tarso*, *metatarso* e *falangi*). Il femore si articola con tibia e perone tramite il ginocchio, l'articolazione più complessa del nostro corpo. La *caviglia*, invece, è l'articolazione che collega la gamba col piede, caratterizzata da mobilità e resistenza elevata, necessaria per sostenere il carico corporeo.

2 Il sistema muscolare

a) Classificazione dei muscoli

Il *sistema muscolare* rappresenta la parte attiva dell'apparato locomotore. I **muscoli**, formati da tessuti contrattili, sono in grado di determinare il movimento delle ossa, unite da articolazioni mobili e semimobili, attraverso movimenti di contrazione e di estensione.

I **muscoli scheletrici**, costituiti da tessuto muscolare striato, sono di tipo **volontario**, se controllati dal sistema nervoso centrale, o **involontario**, se controllati dal sistema nervoso periferico (come i muscoli che regolano la postura): i primi sviluppano un'elevata potenza in tempi brevi, i secondi una potenza minore per tempi prolungati.

In base alla *forma*, i muscoli scheletrici si distinguono in:

- muscoli **fusiformi**, le cui estremità sono collegate alle ossa. Quando un muscolo collega due o più ossa si dice che «origina» in una e «si inserisce» nell'altra. L'origine è riconoscibile dal fatto che le fibre muscolari partono dal periostio, mentre l'inserzione è caratterizzata dall'assottigliamento del muscolo che termina con un tendine, costituito da tessuto connettivo duro e non elastico, il quale può avere forma allungata e collegarsi ad un punto preciso dell'osso, o forma appiattita. Alcuni muscoli, come il *bicipite* e il *quadricipite*, si originano in più ossa;
- muscoli **larghi e piatti**, di cui sono esempi il *deltoide* della spalla e il *grande pettorale*;
- muscoli **nastriiformi**;
- muscoli **cutanei o pellicciai**, inseriti sotto la cute e dotati di forme varie, responsabili della mimica facciale.

Oltre a congiungere due ossa, i muscoli scheletrici possono inserirsi uno nella guaina fibrosa dell'altro, stratificandosi in più livelli, così da potenziare una l'azione dell'altro, oppure possono collegarsi direttamente al tessuto sottocutaneo, come i muscoli del viso.

A seconda della *disposizione delle fibre muscolari rispetto alla direzione di contrazione*, i muscoli si differenziano in:

- muscoli **lunghi**, dotati di fibre disposte parallelamente tra loro e alla linea di trazione, in modo da contrarsi tutte nella stessa direzione. Questo tipo di disposizione realizza il massimo accorciamento possibile;
- muscoli **pennati**, muniti di fibre brevi e raccolte lungo fasci inclinati rispetto alla linea di trazione, cosicché la contrazione risulta molto potente, ma con un minore accorciamento;
- muscoli **di torsione**, con fibre disposte a spirale;
- muscoli **a ventaglio**, composti da fibre con una disposizione triangolare, poste obliquamente rispetto alla linea di trazione, così da concentrare tutta la potenza della contrazione in un unico punto.

Infine, considerando il *tipo di movimento che determinano*, i muscoli scheletrici si distinguono in:

- muscoli **adduttori**, che permettono l'avvicinamento di due parti del corpo;
- muscoli **abductori**, che, al contrario, consentono l'allontanamento di due parti corporee;
- muscoli **estensori**, che rendono possibile la distensione di un arto piegato;
- muscoli **flessori**, che permettono il piegamento di un arto disteso;
- muscoli **rotatori**, che consentono la rotazione di una parte del corpo.

Una caratteristica dei muscoli scheletrici è quella di lavorare sempre in **coppie di antagonisti**: la contrazione di uno dei muscoli della coppia, cioè, è accompagnata dalla distensione dell'altro. Negli arti superiori, ad esempio, la contrazione del bicipite e la contemporanea distensione del tricipite provocano la flessione dell'avambraccio sul braccio. Viceversa, la contrazione del tricipite e la distensione del bicipite permettono la distensione dell'arto. Ciò si deve al fatto che il muscolo, attraverso la contrazione, può esercitare solo un'azione di trazione sulle ossa, cioè può tirare l'osso ma non spingerlo.

b) Fisiologia della contrazione muscolare

I muscoli scheletrici sono formati da fasci di **fibre muscolari** tenute insieme da tessuto connettivo. Ciascuna fibra è costituita da un'unica cellula, che può arrivare a diversi centimetri di lunghezza e a 100μ di diametro, avvolta da una membrana esterna denominata **sarcolemma**. Nel citoplasma delle fibre muscolari sono contenute da 1.000 a 2.000 **miofibrille**, disposte parallelamente rispetto alla lunghezza della cellula, i cui nuclei sono confinati nella parte periferica, accanto al sarcolemma. Ogni miofibrilla è circondata da una struttura membranosa, il **reticolo sarcoplasmatico**, i cui vacuoli, contenenti ioni Ca^{2+} , svolgono un ruolo fondamentale nella contrazione muscolare.

Le miofibrille derivano dalla ripetizione di un'unità funzionale, il **sarcomero**, lungo $2-3\mu$, formato da un insieme ordinato di filamenti proteici, alcuni più spessi, costituiti da molecole di **miosina**, altri più sottili, costituiti da molecole di **actina** (Fig. 5). Ciascun filamento di miosina è circondato da circa sei filamenti di actina, le cui estremità sono collegate a uno strato proteico che prende il nome di **linea zeta**. La struttura delle molecole di actina e miosina è tale che tra esse possono formarsi legami a ponte temporanei, i quali determinano lo scivolamento dei filamenti di actina su quelli di miosina e la conseguente «chiusura del sarcomero», che risulta accorciato.

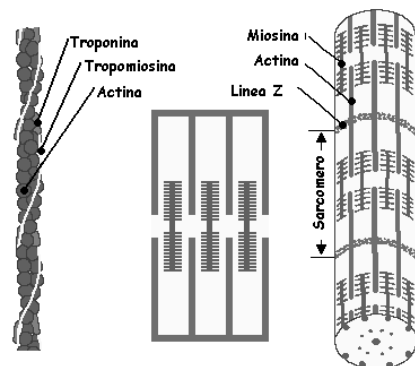


Fig. 5 - Struttura della miofibrilla

In condizioni di riposo, sulle molecole di actina si trovano avvolte molecole di *tropomiosina*, che bloccano i siti di legame con la miosina, impedendo la formazione dei ponti trasversali. La *tropomiosina* porta legate, a intervalli regolari, molecole di *troponina*. Quando la cellula viene stimolata da un impulso nervoso, gli ioni calcio penetrano al suo interno e, legandosi alla *troponina*, determinano lo spostamento delle catene di *tropomiosina*, che lasciano scoperti i siti di legame dell'actina con la miosina. In tal modo si formano i legami a ponte trasversali, che causano l'accorciamento del sarcomero e, in ultima analisi, del muscolo. Al termine dell'impulso, gli ioni calcio vengono pompato fuori dalla cellula attraverso un meccanismo di *trasporto attivo*, le molecole riacquistano le conformazioni iniziali e il muscolo si distende.

1 Alimentazione e digestione

L'uomo, come tutti gli organismi eterotrofi, acquista energia e materiali di base per la sintesi delle proprie biomolecole attraverso il cibo. L'*alimentazione* umana è costituita per la maggior parte da tessuti animali e vegetali che non possono essere utilizzati come tali, ma devono subire un processo di *digestione*, cioè di scomposizione in sostanze semplici, utilizzabili dalle cellule.

Come si vedrà studiando l'*apparato digerente*, le esigenze energetiche dell'organismo possono essere soddisfatte mediante l'apporto di *carboidrati*, *proteine* o *grassi*: il fegato, infatti, converte le proteine eccedenti in glucosio, per cui, all'occorrenza, anch'esse possono essere sfruttate come fonte di energia.

Tuttavia, è indispensabile che nella dieta siano presenti, in modo equilibrato, tutti i principi nutritivi, rappresentati da:

- **carboidrati**, i quali costituiscono la fonte energetica più prontamente utilizzabile dall'organismo e devono essere introdotti in quantità tali da coprire circa il 55-60% del fabbisogno energetico giornaliero;
- **proteine**, fonti indispensabili di amminoacidi. In particolare, la dieta dovrebbe garantire un apporto minimo degli otto amminoacidi essenziali, che l'organismo non è in grado di sintetizzare a partire da altre molecole organiche. Nella dieta, la componente proteica dovrebbe costituire circa il 10-15% del fabbisogno calorico giornaliero;
- **grassi**, necessari per introdurre gli acidi grassi essenziali, precursori di molti ormoni (come le prostaglandine) e dei fosfolipidi di membrana, che non dovrebbero superare il 30% del fabbisogno calorico totale;
- **vitamine**, un altro importante gruppo di nutrienti essenziali, indispensabili all'organismo, seppure in quantità nettamente inferiori rispetto agli altri. Le vitamine, infatti, svolgono sostanzialmente funzioni di regolazione (molte di esse sono *coenzimi*) e sono sufficienti anche in piccole quantità. Al contrario, quantità eccessive di vitamine, in particolare quelle liposolubili, che non possono essere eliminate con le urine e tendono ad accumularsi nell'organismo, possono addirittura avere effetti tossici;
- **sali minerali**, anch'essi nutrienti essenziali, tra cui figurano il *calcio* e il *fosforo*, che partecipano alla composizione delle ossa, il *ferro*, costituente dell'emoglobina e dei citocromi, il *sodio* e il *potassio*, essenziali per il mantenimento dell'equilibrio osmotico delle cellule.

2 L'apparato digerente

L'uomo e i vertebrati possiedono un **tubo digerente** che si estende dalla bocca all'ano e può essere considerato una cavità esterna al corpo. Infatti, l'ingresso effettivo del cibo nell'organismo ha luogo quando le sostanze nutritive, digerite e scomposte in molecole semplici, vengono assorbite nel lume intestinale.

Per tutta la sua lunghezza, il canale digerente è costituito da quattro strati:

1. la **mucosa**, corrispondente allo strato più interno, composta da tessuto epiteliale e dalla sottostante membrana basale, e caratterizzata dalla presenza di numerose cellule secernenti muco e ghiandole che producono enzimi;

2. la **sottomucosa**, costituita da tessuto connettivo, fibre nervose, vasi sanguigni e linfatici;
3. la **tonaca muscolare**, formata da due strati di tessuto muscolare liscio: quello *interno*, composto da fibre disposte circolarmente, e quello *esterno*, costituito da fibre disposte longitudinalmente. La contrazione di queste fibre genera un movimento ondulatorio, detto **peristalsi**, che consente la progressione del cibo lungo tutto il canale. In alcuni punti questa muscolatura risulta ispessita in modo da formare gli **sfinteri**, i quali costituiscono una sorta di valvole che separano i diversi comparti del tubo digerente;
4. la **sierosa**, il quarto e ultimo strato, che costituisce il rivestimento esterno di tessuto connettivo.

a) Bocca

L'apparato digerente (Fig. 1) comincia con la **bocca**, all'interno della quale si trovano, nella maggior parte dei mammiferi, i **denti**, necessari a frantumare il cibo, e la **lingua**, la cui funzione è quella di rimescolare il cibo, amalgamandolo alla **saliva**. Quest'ultima, prodotta da tre ghiandole «salivari» (parotide, sottomandibolare e sottomandibolare), ha il compito di ammorbidire il cibo per facilitarne la deglutizione e contiene la **ptialina**, un enzima che già nella bocca dà inizio alla digestione degli amidi. La salivazione è sotto il controllo del sistema nervoso autonomo e viene indotta dalla stimolazione dei bottoni gustativi o, talvolta, semplicemente dall'odore o dalla vista del cibo.

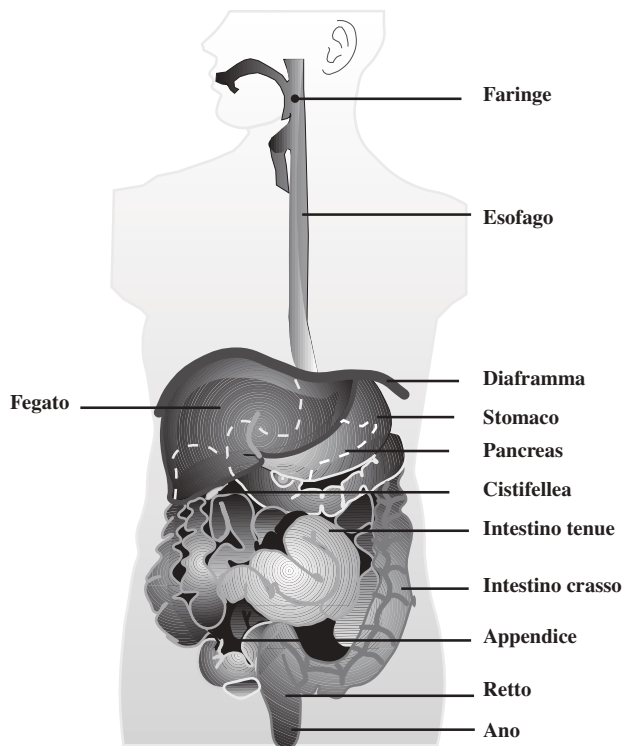


Fig. 1 - Apparato digerente umano

b) Esofago

In seguito alla deglutizione, il cibo passa dalla bocca nell'**esofago**, un canale lungo circa 25 cm, attraverso il quale il *bolo alimentare* transita grazie ai movimenti peristaltici: il tratto che precede il bolo si contrae, mentre quello successivo si rilassa, in modo da determinarne il rapido spostamento.

Dall'esofago il bolo passa nello stomaco tramite il **cardias**, una struttura valvolare che ha la funzione di impedire il reflusso del contenuto gastrico nell'esofago, la cui mucosa non è in grado di tollerarne l'elevata acidità.

c) Stomaco

Lo *stomaco* è un organo a forma di sacco, normalmente ripiegato su sé stesso, con una capacità variabile dai 2 ai 4 litri. È dotato di una tunica muscolare a tre strati, con fibre disposte longitudinalmente nel primo, concentricamente all'asse nel secondo, obliquamente nel terzo, in modo da consentire movimenti complessi.

Lo strato mucoso interno è molto spesso e contiene numerose fossette che secernono i **succhi gastrici**, costituiti principalmente da *muco*, *acido cloridrico* e *pepsinogeno*.

Il pH nello stomaco è molto basso (compreso tra 1,5 e 2,5), tanto da eliminare la maggior parte dei batteri introdotti con il cibo e garantire un ambiente in cui possano sciogliersi le componenti fibrose degli alimenti. Inoltre, in ambiente fortemente acido, il pepsinogeno viene trasformato in **pepsina**, l'enzima preposto all'idrolisi delle proteine. La regolazione della secrezione gastrica dipende dal sistema nervoso e da quello endocrino: in particolare, la vista del cibo e la sua presenza nello stomaco determinano la liberazione di **gastrina**, un ormone che, riversato nel torrente ematico, aumenta la secrezione di succo gastrico e stimola i movimenti peristaltici. Dallo stomaco, il cibo, ormai trasformato in una massa semiliquida denominata **chimo**, passa nell'intestino tenue attraverso lo **sfintere pilorico**.

d) Intestino

Si è soliti suddividere l'*intestino* in diversi tratti. Il primo, lungo circa 7 metri, è detto **intestino tenue**. Qui si completa la digestione e vengono assorbite le sostanze nutritive. Per aumentare l'efficienza del processo di assorbimento dei nutrienti, questa porzione di intestino è caratterizzata da un elevato sviluppo superficiale, dovuto alla presenza di pieghe nella sottomucosa e di estroflessioni della mucosa, dette **villi**. Inoltre, ogni cellula dell'epitelio possiede microscopiche proiezioni citoplasmatiche, i **microvilli**, che aumentano enormemente lo sviluppo superficiale.

I primi 25 centimetri dell'intestino tenue costituiscono il **duodeno**, la parte più attiva nei processi digestivi, in cui si riversano i prodotti del pancreas, costituiti da enzimi pancreatici e da un liquido alcalino che neutralizza l'acidità del chimo, e la *bile*, prodotta dal fegato e accumulata nella cistifellea.

L'attività dell'intestino tenue è sotto il controllo del sistema endocrino: la presenza del chimo nel lume intestinale provoca la secrezione da parte della mucosa duodenale di **secretina**, un ormone che stimola il fegato e il pancreas a secernere liquidi alcalini. La presenza di grassi e amminoacidi stimola, invece, la produzione di **colecistochinina**, che induce, a sua volta, la secrezione di enzimi pancreatici e lo svuotamento della cistifellea. Le contrazioni intestinali sono influenzate anche dal sistema nervoso autonomo, inibite dal sistema simpatico e aumentate dal parasimpatico.

Nel duodeno si completa la digestione degli zuccheri ad opera dell'amilasi pancreatica; i grassi, ridotti in minuscole gocce grazie all'azione dei sali biliari, vengono idrolizzati a glicerolo e ad acidi grassi dalle lipasi; le proteine, infine, vengono attaccate da una combinazione di enzimi, alcuni attivi solo su legami peptidici tra specifici amminoacidi, altri solo sulle estremità delle catene da cui staccano dipeptidi, altri ancora su specifici dipeptidi, che le riducono ad amminoacidi semplici.

Nei tratti successivi, il **digiuno** e l'**ileo**, avviene l'assorbimento dei nutrienti. Monosaccaridi e amminoacidi semplici vengono assorbiti per trasporto attivo nei capillari dei villi ed entrano nella circolazione sanguigna. Gli acidi grassi di piccole dimensioni entrano direttamente nei vasi sanguigni, mentre quelli di dimensioni maggiori sono assorbiti dalle cellule epiteliali, ricomposti in grassi, avvolti con un rivestimento proteico e secreti nei vasi linfatici, attraverso i quali raggiungono il sistema circolatorio.

Il riassorbimento dell'acqua e dei minerali ha luogo nell'**intestino crasso**, suddiviso in *intestino cieco*, immediatamente successivo alla valvola ileo-cecale, *colon ascendente*, *traverso* e *discendente*, *colon sigmoideo* e *retto*. L'intestino crasso ospita batteri simbiotici che si nutrono delle sostanze non digerite e producono amminoacidi e vitamine (ad esempio, la vitamina K) destinate ad essere assorbite dal sangue. La massa fecale, costituita dall'insieme delle sostanze non digerite (principalmente fibre di cellulosa), viene temporaneamente accumulata nel retto, per poi essere espulsa dall'ano.

e) Fegato e pancreas

Il **fegato** è la ghiandola più grande del nostro organismo. Oltre a produrre la *bile* (la quale emulsionando i grassi ne permette l'assorbimento da parte dell'intestino), che viene accumulata nella **colecisti** o **cistifellea**, assolve a numerose funzioni metaboliche di estrema importanza.

Al fegato giungono il glucosio e gli altri monosaccaridi assorbiti dal sangue attraverso la vena porta epatica, che vengono trasformati in glicogeno e grassi, poi accumulati nelle cellule adipose. Ad esso spetta anche il compito di scindere e trasformare in glucosio gli amminoacidi in eccesso, producendo come sottoprodotto azotato l'*urea*.

Il **pancreas**, oltre a produrre gli enzimi necessari alla digestione duodenale di proteine, grassi e zuccheri, è la ghiandola che provvede, insieme al fegato, alla regolazione degli zuccheri nel sangue: le cellule delle **isole di Langerhans**, infatti, secernono *insulina*, *glucagone* e *somatostatina*, gli ormoni che regolano la concentrazione ematica del glucosio.

3 Respirazione e scambi gassosi

Gli organismi che vivono in ambiente terrestre ricavano l'ossigeno necessario alla *respirazione cellulare* direttamente dall'aria. Gli scambi gassosi tra organismo e ambiente avvengono tutti attraverso processi di diffusione, ovvero in seguito allo spostamento spontaneo dei gas da una zona a concentrazione maggiore ad una a concentrazione minore: l'ossigeno entra nelle cellule per diffusione e analogamente l'anidride carbonica, prodotta come rifiuto della respirazione cellulare, esce dalle cellule attraverso lo stesso tipo di meccanismo.

Tuttavia, la diffusione è un processo lento, che avviene con una certa velocità solo su brevi distanze e risulta sufficiente al rifornimento di ossigeno solo per organismi di piccole dimensioni. Gli organismi complessi hanno dovuto sviluppare apparati respiratori e circolatori che garantissero un rapido rifornimento di ossigeno a tutte le cellule, ricorrendo a soluzioni diverse a seconda delle loro dimensioni e del loro habitat.

I **vermi**, ad esempio, hanno sviluppato una *rete di capillari* sotto l'addome affinché il sangue, che trasporta ossigeno e lo distribuisce a tutte le cellule dell'organismo, potesse scorrere vicino alla superficie esterna, scambiando ossigeno e anidride carbonica attraverso la superficie umida del corpo.

Gli **insetti**, invece, respirano tramite un *sistema di tubuli rivestiti di chitina*, assimilabili a una rete di trachee, nei quali l'aria transita sotto la spinta dei movimenti corporei. Questo tipo di soluzione ha un'efficienza piuttosto scarsa e ha imposto severe limitazioni alle dimensioni raggiungibili da questi animali.

Negli **organismi acquatici** si sono sviluppate le *branchie*, che, analogamente ai polmoni, sono strutture finalizzate ad aumentare la superficie degli scambi gassosi. Tali organismi sono svantaggiati rispetto a quelli terrestri, sia perché l'acqua contiene una piccola quantità di ossigeno disciolto (pari a circa lo 0,5%), sia perché la maggiore densità dell'acqua rispetto all'aria aumenta notevolmente il lavoro associato alla respirazione e rallenta i processi di diffusione.

Per ottimizzare l'efficienza degli scambi, l'acqua transita nelle branchie spostandosi in senso opposto rispetto alla direzione del flusso sanguigno, realizzando una vera e propria «estrazione

in controcorrente» dell'ossigeno dall'acqua, così chiamata perché i due fluidi scorrono in senso inverso; ciò consente di mantenere la differenza di concentrazione al massimo valore possibile, aumentando l'efficienza della diffusione.

Gli **organismi terrestri**, infine, hanno sviluppato i *polmoni*, cavità interne al corpo in cui entra l'aria, che offrono il vantaggio di aumentare le superfici di scambio e di minimizzare la perdita di acqua dovuta all'evaporazione.

In tutti i vertebrati, a prescindere dalle diversità anatomiche, l'ingresso dell'aria nei polmoni avviene in seguito alla variazione di volume della cavità polmonare, determinata dalla contrazione e dal rilassamento di specifici muscoli. Di qui l'ossigeno, si diffonde nei capillari sanguigni attraverso l'epitelio umido e viene poi trasportato dal sangue per flusso di massa, finché dal sangue passa per diffusione nei liquidi interstiziali e di qui nelle cellule.

4 L'apparato respiratorio

a) Struttura e organi costitutivi

Il *sistema respiratorio dell'uomo* (Fig. 2) inizia con le **cavità nasali**, da cui entra l'aria, che, scaldata e purificata dalle cellule ciliate, passa poi nella **faringe** e quindi nella **trachea**, il canale anteriore all'esofago, dotato di anelli cartilaginei che lo mantengono costantemente aperto. La trachea si divide poi nei due **bronchi**, *sinistro* e *destro*, da cui si dipartono i due alberi bronchiali, generati dalla divisione dei bronchi in ramificazioni sempre più piccole, dette **bronchioli**.

Intorno alle estremità dei bronchioli più piccoli si sviluppano gli acini polmonari (Fig. 3), costituiti da **alveoli** raggruppati come acini d'uva. Ogni alveolo misura 0,1-0,2 mm ed è circondato da capillari, dai quali è separato da un unico strato di cellule epiteliali appiattite, dello spessore di $0,5\mu$, che costituisce la sottilissima parete attraverso la quale avvengono gli scambi gassosi. Grazie a questa struttura così complessa, la superficie di scambio polmonare arriva a $70-80\text{ m}^2$, pari a circa 40 volte la superficie corporea. Alcune delle cellule epiteliali che rivestono le vie respiratorie secernono muco, il quale trattiene le particelle estranee penetrate con l'aria, mentre altre cellule dotate di ciglia ripuliscono la superficie spingendo il muco verso la faringe, dove viene deglutito.

L'ingresso e l'uscita dell'aria dai polmoni (*ventilazione*) si verificano in seguito alla differenza di pressione tra l'aria contenuta negli alveoli e la pressione esterna, determinata dalla variazione di volume delle cavità polmonari, che aumenta e diminuisce contestualmente alla dilatazione e

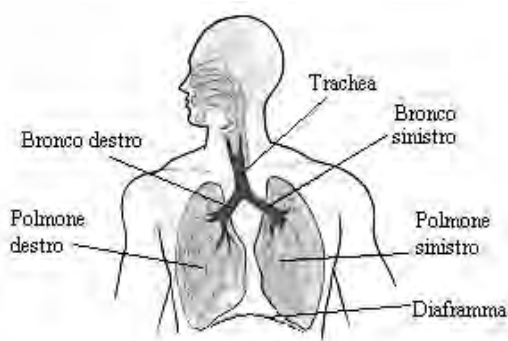


Fig. 2 - Apparato respiratorio umano

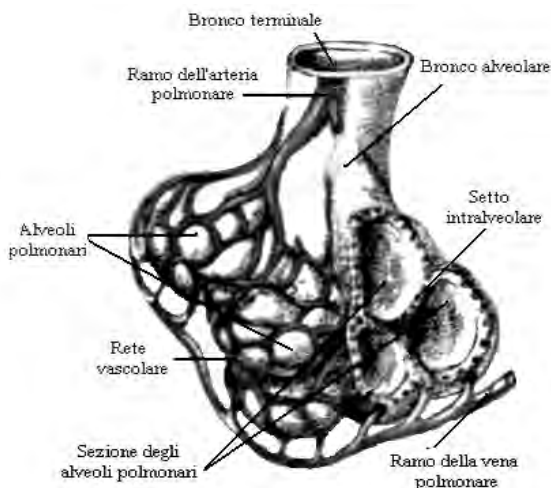


Fig. 3 - Particolare dell'apparato respiratorio

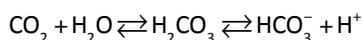
alla contrazione della gabbia toracica. Durante l'**inspirazione** il diaframma si contrae e si abbassa, ampliando la cavità toracica, e contemporaneamente si contraggono i muscoli intercostali, che alzano e allargano la gabbia toracica; in tal modo il volume polmonare aumenta, la pressione alveolare diminuisce e l'aria entra nei polmoni. Viceversa, durante l'**espirazione** il diaframma e i muscoli intercostali si rilassano, la cavità toracica diminuisce di volume e l'aria fuoriesce dai polmoni in seguito all'aumento della pressione alveolare. Attraverso questo meccanismo, ad ogni respiro viene ricambiato solo il 10% dell'aria contenuta nei polmoni, percentuale che può tuttavia aumentare fino all'80% con respiri lunghi e profondi.

b) Trasporto dei gas

L'**ossigeno** è scarsamente solubile nel plasma sanguigno; per aumentarne la quantità scambiata nella respirazione e trasportata ai tessuti, pertanto, nel sangue sono presenti specifiche proteine trasportatrici. Nell'uomo e nei vertebrati la proteina trasportatrice di ossigeno è l'**emoglobina**, una molecola costituita da quattro subunità proteiche, due di tipo α e due di tipo β , ciascuna delle quali è legata a un gruppo non proteico, denominato «eme», in cui un atomo di ferro è in grado di legare una molecola di ossigeno. Complessivamente, quindi, ogni molecola di emoglobina può trasportare quattro molecole di O_2 . Nel sangue umano l'emoglobina si trova all'interno di cellule specializzate, i **globuli rossi**, ciascuno dei quali contiene circa 265 milioni di molecole di emoglobina. In tal modo il sangue riesce a trasportare una quantità di ossigeno notevolmente maggiore di quanto consentirebbe la sua solubilità.

Il legame dell'emoglobina con l'ossigeno negli alveoli polmonari e il successivo rilascio nei tessuti dipendono dalla concentrazione di ossigeno nel plasma: in presenza di alte concentrazioni di O_2 , come negli alveoli, la maggior parte dell'emoglobina si combina con l'ossigeno (acquistando una colorazione rosso brillante), mentre nei tessuti, dove la concentrazione di O_2 è bassa, l'emoglobina rilascia l'ossigeno, che si diffonderà nel plasma e da questo nei tessuti.

L'**anidride carbonica** prodotta dalla respirazione cellulare viene trasportata dal plasma sanguigno sia come tale, sia in forma di ione bicarbonato. Questa, infatti, essendo un acido debole, può dare luogo al seguente equilibrio, catalizzato dall'enzima anidrasi carbonica contenuto nei globuli rossi:



La mobilità dell'equilibrio fa sì che nei tessuti, dove la concentrazione della CO_2 è elevata, l'equilibrio si sposti verso destra, formando ioni bicarbonato e ioni idrogeno. Viceversa, negli alveoli polmonari, dove la concentrazione di anidride carbonica è bassa, gli ioni bicarbonato e gli ioni idrogeno si ricombinano liberando CO_2 , che dal plasma si diffonde negli alveoli.

c) Regolazione della respirazione

La *respirazione* è un *atto involontario* che avviene sotto il controllo dei neuroni respiratori del tronco cerebrale. La regolazione della frequenza e della profondità del respiro si avvale di diversi sistemi di controllo, integrati tra loro in modo da garantire un costante rifornimento di ossigeno a tutte le cellule del corpo, in particolare a quelle del cervello. I neuroni respiratori ricevono segnali da recettori sensibili all'anidride carbonica, all'ossigeno e alla concentrazione degli ioni idrogeno: quando la concentrazione dell'anidride carbonica e degli ioni idrogeno aumenta, la respirazione diventa più veloce e profonda, così da permettere la diffusione di una maggiore quantità di anidride carbonica dai capillari negli alveoli. I recettori sensibili all'ossigeno situati nelle arterie carotidi, che riforniscono il cervello di sangue ossigenato, esercitano una funzione di controllo su questo meccanismo di regolazione, per cui, in caso di diminuzione della concentrazione di ossigeno nel sangue, stimolano i neuroni respiratori ad incrementare la frequenza e la profondità del respiro.

1 Il sistema circolatorio negli animali

Gli *animali*, fatta eccezione per quelli di dimensioni molto piccole, hanno dovuto sviluppare sistemi circolatori attraverso i quali distribuire a tutte le cellule dell'organismo le sostanze nutritive prodotte dalla digestione e l'ossigeno introdotto con la respirazione. Inoltre, il sangue raccoglie dai tessuti le sostanze di rifiuto, per poi trasportarle agli organi deputati alla loro eliminazione. Le soluzioni adottate per far fronte a questo problema variano a seconda del tipo di animale e delle sue dimensioni. Ad esempio, il **lombrico** ha un sistema circolatorio chiuso, in cui il sangue scorre sotto la spinta di cuori costituiti da parti muscolari di alcuni vasi. I **molluschi** hanno un cuore diviso in camere come i vertebrati, ma il loro sistema circolatorio è aperto e il sangue si diffonde all'interno dei tessuti, da cui poi torna alle branchie.

I *vertebrati*, invece, hanno sistemi circolatori chiusi e cuori divisi in comparti. I **pesci** hanno un cuore diviso in un atrio e in un ventricolo che spinge il sangue nei capillari branchiali, da dove, una volta ossigenato, passa ai tessuti, per poi fare ritorno al cuore. Gli **anfibi** e i **rettili** hanno cuori con due atri e un solo ventricolo: l'atrio destro riceve il sangue deossigenato proveniente dai tessuti e l'atrio sinistro quello ossigenato proveniente dai polmoni. Entrambi gli atri sboccano nello stesso ventricolo, dotato di un setto interno che riesce a mantenere separati i due tipi di sangue, indirizzando il primo verso la *circolazione polmonare*, il secondo verso la *circolazione sistemica*.

Uccelli e mammiferi, infine, hanno il cuore diviso in due parti completamente separate: quella destra, che riceve il sangue dai tessuti e lo pompa verso i polmoni (*circolazione polmonare*), e quella sinistra, che riceve il sangue ossigenato dai polmoni e lo pompa verso i tessuti (*circolazione sistemica*). In figura 1 vengono messi a confronto i sistemi circolatori di pesci, anfibi e mammiferi.



Fig. 1

2 L'apparato circolatorio umano

a) Sangue

Il *sangue* è considerato un tessuto connettivo perché composto da cellule immerse in una matrice extracellulare liquida, il **plasma**. Quest'ultimo, costituito da *acqua* e *soluti disciolti* (sali, molecole proteiche, zuccheri, grassi, sostanze di rifiuto del metabolismo cellulare), rappresenta il 55% del sangue, mentre la **parte corpuscolare**, formata da *piastrine*, *globuli rossi* e *bianchi*, ne costituisce il restante 45%.

I **globuli rossi** o **eritrociti** sono presenti in quantità pari a 4-5 milioni per mm^3 . Durante la maturazione queste cellule perdono il nucleo e gli organuli, perché sono specializzate nel trasporto di ossigeno e contengono elevatissime quantità di emoglobina. Hanno dunque un tempo di vita limitato (120-130 giorni) e vengono continuamente rigenerati dal midollo osseo.

I **globuli bianchi** o **leucociti**, meno numerosi dei rossi (6-9 mila per mm^3), provvisti di nucleo e privi di emoglobina, vengono prodotti da midollo osseo e milza. La loro funzione consiste essenzialmente nel difendere l'organismo dall'attacco di virus, batteri o corpi estranei. A tal fine, si possono diffondere dai vasi sanguigni nei liquidi interstiziali, dove, spostandosi attraverso pseudopodi, fagocitano gli organismi invasori.

Le **piastrine**, infine, non hanno natura cellulare, ma sono semplicemente frammenti del citoplasma di cellule del midollo osseo, ricoperte da una membrana. Hanno un ruolo essenziale sia nella riparazione dei vasi sanguigni, sia nel processo di coagulazione del sangue. Vengono attivate nel momento in cui il sangue, fuoriuscendo dai vasi, viene a contatto con una proteina presente sulle cellule dei tessuti, il *fattore tissutale*, assente nelle cellule che ricoprono il lume interno dei vasi. In seguito a queste reazioni, una proteina plasmatica solubile, il *fibrinogeno*, viene trasformata in una proteina insolubile, la **fibrina**, responsabile della formazione del coagulo di sangue.

b) Vasi sanguigni

Il sangue pompato dal cuore circola costantemente all'interno di vasi che costituiscono un circuito chiuso. I vasi che trasportano il sangue dal cuore verso la periferia sono denominati **arterie** e si suddividono in vasi sempre più piccoli, le *arteriole* e i *capillari*. A livello dei **capillari**

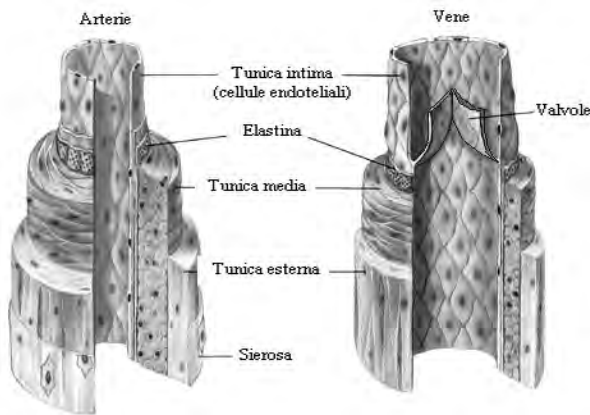


Fig. 2 - Conformazione di arterie e vene

avvengono gli scambi con i tessuti, dai quali il sangue passa nelle venule e poi nelle **vene**, che lo trasportano verso il cuore. Arterie, vene e capillari, oltre a differire per le dimensioni, si distinguono anche per la struttura. I capillari hanno pareti costituite da un unico strato di endotelio, il tessuto epiteliale di cui sono ricoperti i vasi sanguigni, che facilita la diffusione delle sostanze. Le arterie e le vene, invece, possiedono anche tessuti muscolari e di sostegno (Fig. 2), necessari a regolare la resistenza opposta al passaggio del sangue. Le vene, inoltre, sono compartimentate

internamente da valvole a mezzaluna, che si aprono per consentire il flusso del sangue unicamente in direzione del cuore e poi si richiudono per impedirne il riflusso in direzione opposta.

c) Cuore

Nell'uomo, come nei mammiferi, il **cuore** (Fig. 3) è un organo muscolare cavo, formato da un particolare tessuto, il **miocardio**, e rivestito da una membrana, detta **pericardio**.

Anatomicamente e funzionalmente è diviso in due parti distinte, il *cuore destro* e il *cuore sinistro*, ciascuna delle quali comprende una parte superiore, detta **atrio**, collegata alla parte inferiore, chiamata **ventricolo**, attraverso un orifizio atrio-ventricolare. Nell'atrio destro entra il sangue proveniente dai tessuti attraverso la vena cava superiore e inferiore, mentre l'atrio

sinistro si riempie del sangue proveniente dai polmoni attraverso le vene polmonari. Gli atri si contraggono insieme e il sangue passa nei ventricoli tramite le valvole atrio-ventricolari (*bicuspidale* e *tricuspidale*), che, quando i ventricoli si contraggono, si chiudono per effetto della pressione sanguigna, in modo da evitare il reflusso del sangue negli atri. Il ventricolo destro invia il sangue deossigenato verso i polmoni attraverso le arterie polmonari, mentre il ventricolo sinistro spinge il sangue ossigenato nell'*aorta*. Le arterie e i ventricoli sono separati da valvole a semiluna, che si chiudono dopo la contrazione ventricolare per evitare il reflusso del sangue nei ventricoli.

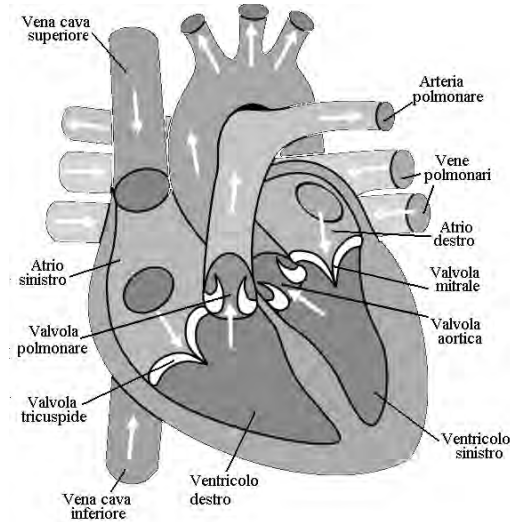


Fig. 3 - Cuore umano

d) Regolazione del battito cardiaco

In un uomo adulto, il *battito cardiaco* ha la frequenza media di 70 battiti al minuto. Contrariamente agli altri muscoli del corpo, che si contraggono in seguito alla stimolazione del sistema nervoso, lo stimolo alla contrazione cardiaca si origina direttamente nel cuore, in una regione posta nell'atrio destro, denominata **nodo senoatriale**. Qui nasce l'impulso elettrico che si propaga rapidamente agli atri, determinandone la contrazione simultanea; giunto al **nodo atrio-ventricolare**, l'impulso passa ai ventricoli attraverso il **fascio di His**. La conduzione dell'impulso a livello del nodo atrio-ventricolare è lenta, in modo da creare un ritardo tra la contrazione atriale e quella ventricolare. Anche se l'impulso elettrico che determina la contrazione cardiaca ha origine all'interno del cuore, la frequenza del battito cardiaco può essere influenzata da vari fattori esterni, al fine di adeguarla alle esigenze contingenti dell'organismo. Il sistema nervoso autonomo può contribuire a variarla: in particolare, il *sistema parasimpatico* agisce diminuendola, mentre la *stimolazione del simpatico* aumenta sia la frequenza sia la potenza delle contrazioni cardiache. Un'influenza analoga è esercitata dall'*adrenalina*, ormone prodotto dalla ghiandola surrenale.

e) Pressione sanguigna e centro di controllo cardiovascolare

Sotto la spinta del cuore il sangue entra nella circolazione sistemica, incontrando la resistenza dei vasi: ciò determina l'esistenza di una *pressione* che il sangue esercita sulle pareti dei vasi stessi e che varia da un valore massimo, durante la *sistole ventricolare* (contrazione), a un valore minimo, durante la *diastole* (rilassamento). In un individuo normoteso, questi valori corrispondono rispettivamente a 120 e 80mmHg all'incirca.

La velocità del sangue nei vasi è regolata dalla loro contrazione o dilatazione: in seguito alla *vasodilatazione* il flusso del sangue aumenta, come accade ai muscoli scheletrici nel corso dell'attività fisica, oppure nello stomaco durante la digestione, mentre la *vasocostrizione* fa diminuire il flusso, come si verifica a livello della pelle in conseguenza di una diminuzione della temperatura esterna.

L'azione dei muscoli lisci, che provocano la variazione del lume delle arteriole, è sotto il controllo del sistema nervoso autonomo, degli ormoni adrenalina e noradrenalina, e del *peptide* cardiaco. Le cellule del cuore, infatti, producono e rilasciano un peptide che, dotato di recettori nei reni, nei vasi sanguigni, nelle ghiandole surrenali e nel cervello, è chiaramente coinvolto nella regolazione della pressione del sangue.

La pressione arteriosa e l'attività cardiaca sono regolate mediante il **centro di regolazione cardiovascolare** situato nel midollo allungato, che controlla e coordina l'azione dei nervi del sistema nervoso simpatico e parasimpatico, del cuore e dei vasi sanguigni, in seguito alla ricezione continua di informazioni sulla pressione provenienti dai *barocettori* collocati nelle carotidi, nelle vene cave, nell'aorta e nel cuore. A una diminuzione di pressione corrisponde un aumento dell'attività cardiaca e un restringimento dei vasi, mentre a un incremento della pressione corrisponde una diminuzione dell'attività cardiaca e una dilatazione dei vasi (*feedback negativo*).

3 Regolazione dell'equilibrio chimico: l'apparato escretore

Abbiamo visto come il sangue garantisca alle cellule dell'organismo il necessario rifornimento di nutrienti ed ossigeno, raccogliendo, nello stesso tempo, le sostanze di rifiuto, l'escrezione delle quali è di fondamentale importanza per il benessere degli organismi. Gli organi preposti a tale funzione sono principalmente i *reni*, sebbene vi contribuiscano anche altri organi, come il *fegato*.

a) Funzionalità dei reni

I **reni** sono due piccoli organi a forma di fagiolo, situati posteriormente allo stomaco e al fegato. Le loro funzioni più rilevanti consistono nell'*eliminare le sostanze di rifiuto potenzialmente tossiche*, nel *regolare la concentrazione di elettroliti e sostanze nutritive nel sangue*, nel *mantenere l'equilibrio idrico*.

Le principali sostanze di rifiuto espulse dal metabolismo cellulare sono l'*anidride carbonica* (proveniente dall'ossidazione delle sostanze organiche ed eliminata attraverso l'apparato respiratorio) e l'*ammoniaca* (prodotta dal metabolismo delle sostanze azotate, dotata di un'elevata tossicità anche a basse concentrazioni). Negli organismi complessi, in cui l'eliminazione mediante processi di diffusione risulta insufficiente e troppo lenta, l'ammoniaca viene convertita in composti di tossicità ridotta: acido urico in uccelli, rettili e insetti; urea negli anfibi e nei mammiferi. Contrariamente all'acido urico, che ha scarsa solubilità, l'urea è solubile in acqua e viene trasportata dal sangue ai reni, dove poi viene eliminata con l'urina.

Attraverso il processo di filtrazione del sangue, il rene regola anche la concentrazione dei soluti disciolti (come sostanze nutritive costituite da glucosio, aminoacidi etc.) e degli ioni (Na^+ , K^+ , H^+ , HCO_3^-), che svolgono un ruolo fondamentale nei processi vitali, in quanto conservano l'equilibrio osmotico delle cellule, le condizioni di pH ottimali per il funzionamento degli enzimi etc.

Infine, il rene è l'organo deputato a regolare la quantità d'acqua presente nel corpo, che nell'uomo ha un bassissimo margine di oscillazione: si pensi, ad esempio, che la perdita di appena il 10% dell'acqua corporea provoca sordità, insensibilità al dolore, delirio. È dunque essenziale che possa realizzarsi un equilibrio tra la quantità d'acqua assunta attraverso cibi e bevande, nonché prodotta con i processi metabolici, e la quantità persa a seguito della respirazione (umidità del respiro), della sudorazione, dell'escrezione (feci e urina).

b) Anatomia e fisiologia del rene

L'unità funzionale del rene è il **nefrone**, composto da un ammasso di capillari, detto **glomerulo**, e da un **tubulo renale** molto lungo, formalmente distinto in tre parti: *tubulo prossimale*, *ansa di Henle* e *tubulo distale*, le cui pareti sono formate da un unico strato di epitelio specializzato nel trasporto attivo di sostanze. Il tubulo renale parte da una sorta di coppa che circonda il glomerulo, denominata **capsula di Bowman**, e termina in un dotto collettore che convoglia l'urina nella **pelvi renale**, una struttura a forma di imbuto da cui, per peristalsi, l'urina procede verso l'uretere, per poi raccogliersi nella vescica (per una schematizzazione dell'apparato escretore e della struttura del nefrone si osservino le figure 4 e 5).

Il sangue entra nel rene attraverso l'*arteria renale*, che si divide in arteriole progressivamente più piccole, ciascuna delle quali termina in un *nefrone*. I processi che avvengono all'interno del nefrone possono essere suddivisi in quattro fasi:

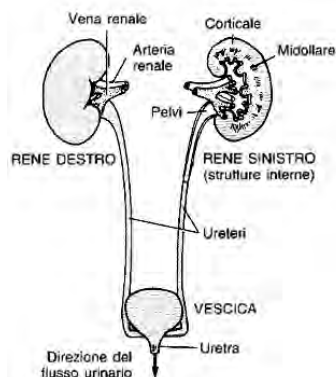


Fig. 4 - Apparato escretore

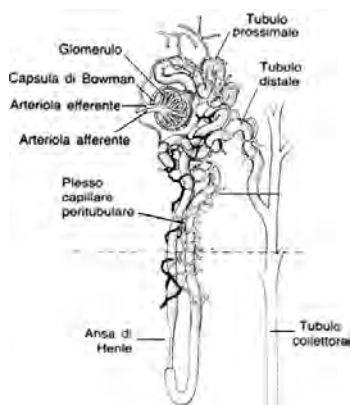


Fig. 5 - Struttura del nefrone

1. **filtrazione:** il glomerulo si colloca tra un'arteriola afferente e una efferente che si divide in una rete di capillari peritubulari, i quali si ricongiungono in una venula che porta il sangue alla vena renale. A causa della vasocostrizione delle arteriole afferente ed efferente, la pressione nel glomerulo risulta circa doppia rispetto a quella del sangue e ciò determina il passaggio di buona parte del plasma sanguigno nella capsula di Bowman. Questo filtrato, in cui sono assenti le proteine che non sono in grado di diffondersi attraverso le pareti dei capillari, ha una composizione molto simile a quella del plasma;
2. **secrezione:** durante il passaggio del filtrato attraverso il tubulo renale, alcune sostanze rimaste nel plasma arrivano al filtrato mediante trasporto attivo (vengono escreti in questo modo molti farmaci, tra cui la penicillina);
3. **riassorbimento:** contemporaneamente all'escrezione di specifiche sostanze, ha luogo il riassorbimento di gran parte dell'acqua e dei soluti (glucosio, aminoacidi, vitamine, sali etc.) passati nel filtrato primario;
4. **escrezione:** consiste nel passaggio del liquido rimanente, l'urina, nel dotto collettore e di qui nella pelvi renale.

La quantità d'acqua eliminata con l'urina è regolata da un sistema complesso e varia in risposta alle condizioni momentanee, tra cui la pressione del sangue e la concentrazione dei soluti nel plasma sanguigno. Quando l'organismo ha necessità di conservare l'acqua, il nefrone può eliminare un'urina ipertonica rispetto al plasma. La quantità d'acqua riassorbita è regolata dall'**ormone antidiuretico ADH**: quando questo è assente, le pareti del dotto collettore sono impermeabili all'acqua, che quindi non può essere riassorbita; l'urina eliminata, in questo caso, è *ipotonica* rispetto al plasma sanguigno. Al contrario, quando è presente l'ADH, le pareti del dotto collettore diventano permeabili all'acqua, che può essere riassorbita, producendo un'urina *ipertonica* rispetto al plasma sanguigno. L'ADH è un ormone ipotalamico accumulato e liberato dall'*ipofisi* in base ai valori della pressione sanguigna e della concentrazione dei soluti nel plasma.

L'*ipotalamo* riceve i segnali sia dei recettori osmotici sensibili alla concentrazione sia dei barocettori situati nel cuore, nelle carotidi e nell'aorta, ordinando il rilascio dell'ADH se la pressione sanguigna è bassa o la concentrazione dei soluti è elevata, come nel caso di emorragie o disidratazione (in tal modo, l'ADH induce il riassorbimento di acqua, che contrasta le variazioni registrate); viceversa, se la pressione del sangue è elevata o la concentrazione dei soluti è bassa, l'ipotalamo non ordina il rilascio di ADH dall'ipofisi, in modo da aumentare il flusso urinario.

Altri importanti ormoni coinvolti in questo equilibrio sono l'**aldosterone**, prodotto dalla corteccia surrenale, che aumenta le quantità di acqua e di sodio riassorbite a livello del tubulo distale, e il **peptide cardiaco**, rilasciato dagli atri del cuore, che invece inibisce il riassorbimento del sodio e dell'acqua a livello del tubulo distale, agendo sul nefrone sia in modo diretto sia in modo indiretto, inibendo la produzione di aldosterone.

1 I messaggeri chimici dell'organismo

Il mantenimento dell'omeostasi in un organismo animale coinvolge un gran numero di attività fisiologiche e richiede una continua comunicazione e integrazione di informazioni. Questa si realizza mediante segnali di tipo chimico, cioè attraverso sostanze dette **ormoni**, che, prodotte in seguito a uno stimolo, vengono trasferite alle cellule o agli organi bersaglio, al fine di ottenere una risposta adeguata. Moltissime cellule producono ormoni, ma, per quelle epiteliali delle ghiandole e per quelle neurosecretrici, la secrezione ormonale rappresenta l'attività principale.

Si dicono **endocrine** le ghiandole prive di dotti in cui raccogliere i prodotti della secrezione, le quali riversano gli ormoni direttamente nei liquidi extracellulari, da dove poi si diffondono nella corrente sanguigna. Sono dette **esocrine**, invece, le ghiandole che secernono i propri ormoni all'interno di dotti che sboccano su superfici come la pelle, le pareti dello stomaco etc.

Gli ormoni, che dal punto di vista chimico si distinguono in *steroidi*, *proteici* o *derivati dagli amminoacidi*, sono sostanze molto potenti, attive a bassissime concentrazioni, per cui, una volta esplicata la loro azione, vengono rapidamente demolite. Ciascun ormone è specifico per il tessuto o l'organo bersaglio, dotato dei corrispondenti recettori che possono essere *intracellulari* (se collocati nel nucleo o nel citoplasma, come accade per gli ormoni steroidi) oppure *inseriti nella membrana cellulare* (come, invece, si verifica per gli ormoni proteici e per quelli derivati dagli amminoacidi).

Gli *ormoni steroidi*, essendo di piccole dimensioni e di natura lipofila, possono diffondersi facilmente attraverso le membrane cellulari; all'interno della cellula si forma il complesso recettore-ormone che avvia la trascrizione nel nucleo di uno specifico tratto di DNA, producendo così la risposta desiderata. Gli *ormoni peptidici* e *proteici*, invece, non sono in grado di attraversare le membrane cellulari e hanno i loro recettori situati sulle membrane, esternamente alla cellula. Una volta formato, il complesso ormone-recettore può essere trasportato all'interno attraverso endocitosi mediata, oppure può mettere in moto un secondo messaggero interno alla cellula che attiva la serie di reazioni desiderate. Per numerosi ormoni questo secondo messaggero è l'AMP ciclico (adenosinmonofosfato).

2 Sistema endocrino e regolazione ormonale

Il *sistema endocrino* è quello anatomicamente più eterogeneo dell'organismo umano, essendo costituito da organi assai diversi sia dal punto di vista morfologico e funzionale, sia sotto il profilo della derivazione embrionale. Nonostante ciò, le ghiandole endocrine sono considerate, a ragione, un sistema perfettamente integrato, non solo perché hanno in comune la funzione di produrre ormoni, ma anche perché il loro funzionamento è strettamente interconnesso, in modo da generare una vera e propria «rete chimica» di comunicazione che controlla tutte le funzioni dell'organismo.

a) Regolazione ormonale

La *regolazione ormonale* è legata alle funzioni svolte dall'*ipofisi*, dalla *ghiandola pineale* e dall'**ipotalamo**. Quest'ultimo è il principale centro di integrazione fra il sistema nervoso centrale e quello endocrino, in quanto secerne ormoni indicati come **RF** (fattori di rilascio), che stimolano o

inibiscono la secrezione ormonale dell'adenoipofisi. Questi ormoni non entrano nella circolazione sistemica, poiché vengono trasportati all'ipofisi diffondendosi attraverso un sistema di capillari. L'ipotalamo produce inoltre l'*ossitocina*, responsabile delle contrazioni uterine durante il parto e della secrezione del latte dalle ghiandole mammarie, e l'ormone antidiuretico ADH, detto anche *vasopressina*, entrambi accumulati nella neuroipofosi, da cui successivamente vengono rilasciati.

L'**ipofisi** è divisa in due lobi strutturalmente e funzionalmente distinti: la **neuroipofosi**, la cui funzione è essenzialmente quella di accumulare gli ormoni prodotti dall'ipotalamo, e l'**adenipofisi**, che invece costituisce il bersaglio dei fattori di rilascio ipotalamici.

Gli ormoni prodotti dall'ipofisi sono:

- la **somatotropina** o **ormone della crescita**, che influenza sia lo sviluppo del tessuto osseo e muscolare sia il metabolismo del glucosio, inibendone l'assorbimento da parte delle cellule;
- la **prolattina**, responsabile della secrezione del latte dopo il parto, la cui produzione è normalmente controllata da un ormone inibitorio ipotalamico, a sua volta inibito dallo stimolo della suzione;
- l'**ormone tireotropo** (TSH), che stimola la tiroide a secernere tiroxina;
- l'**ormone adrenocorticotropo** (ACTH), che regola la produzione del cortisolo;
- le **gonadotropine** (LH e FSH), che agiscono sulle gonadi (testicoli e ovaie).

La **ghiandola pineale**, detta anche **epifisi**, è una piccolissima ghiandola, sito di produzione della **melatonina**, che sembra essere inibita dalla luce e stimolata dalla noradrenalina. Quest'ormone contribuisce a regolare il ciclo del sonno e della veglia, del riposo e dell'attività, e influisce sul ciclo ovarico.

b) Tiroide e paratiroidi

La **tiroide**, ubicata alla base del collo, è una delle ghiandole endocrine più importanti, al punto che il suo mancato o inadeguato funzionamento nell'età dello sviluppo può influenzare negativamente la crescita delle cellule cerebrali, fino a determinare ritardo mentale. In effetti la tiroide produce ormoni che regolano tutti i processi metabolici dell'organismo: la **tiroxina** agisce sul metabolismo energetico (aumentando la velocità della respirazione cellulare), sull'accrescimento, sulla termoregolazione, e la **calcitonina** diminuisce la concentrazione di calcio nel sangue, aumentandone l'escrezione renale, limitandone l'assorbimento intestinale e il rilascio dal tessuto osseo.

Le **paratiroidi**, invece, sono quattro piccole ghiandole grandi come un pisello, situate posteriormente alla tiroide (due in alto, due in basso) e più o meno integrate nel suo tessuto. La loro attività principale consiste nel produrre il **paratormone** (PTH), che aumenta la concentrazione di calcio nel sangue, svolgendo una funzione opposta a quella della calcitonina. In tal modo è possibile realizzare una fine regolazione della concentrazione di questo ione, che ha un ruolo di primo piano nelle contrazioni muscolari e nella trasmissione degli impulsi nervosi.

c) Pancreas

Il *pancreas* è una ghiandola sia esocrina sia endocrina, la cui attività esocrina è legata, in particolare, alla produzione di enzimi digestivi. Dispersi all'interno del tessuto esocrino si trovano raggruppamenti cellulari, denominati *isole di Langerhans*, che svolgono attività endocrina secernendo insulina, glucagone e somatostatina. L'**insulina** è l'ormone che determina l'abbassamento della concentrazione di glucosio nel sangue, stimolandone l'utilizzo da parte delle cellule e la sua trasformazione in glicogeno. Il **glucagone**, invece, svolge un'azione opposta all'insulina, poiché spinge il fegato a mobilitare il glucosio dalle riserve di glicogeno e ne riduce il consumo da parte delle cellule, stimolandole ad utilizzare grassi e proteine. La **somatostatina**, infine, è l'ormone che inibisce la produzione sia dell'insulina sia del glucagone.

L'aumento eccessivo del livello di glucosio nel sangue, patologia nota come **diabete**, è dovuto, in alcuni casi, alla scarsa produzione di insulina, come accade per il diabete giovanile, mentre in altri casi può essere provocato dalla scarsa sensibilità delle cellule all'insulina stessa, a causa della ridotta presenza di recettori per l'insulina sulle membrane cellulari, come spesso si verifica negli adulti.

d) Ghiandole surrenali

Sono situate al di sopra dei reni e divise in una parte *corticale*, più esterna, e una *midollare*, più interna.

La **parte corticale** produce ormoni come l'**aldosterone** e il **cortisolo**: il primo, come già visto, è un ormone mineralcorticoide coinvolto nella regolazione degli ioni sodio e potassio e nell'eliminazione di acqua a livello del tubulo distale, mentre il secondo è un ormone di tipo steroideo, che favorisce la trasformazione di grassi e proteine in glucosio e ne riduce l'utilizzo da parte delle cellule, ad eccezione di quelle cardiache e cerebrali.

Gli ormoni della famiglia del cortisolo sono soppressori delle risposte immunitarie ed infiammatorie e, probabilmente, costituiscono gli ordinari fattori di spegnimento della risposta immunitaria. Questa proprietà ne motiva l'utilizzo nella terapia di malattie autoimmuni o nel trattamento delle allergie.

La **parte midollare** del surrene, invece, produce **adrenalina** e **noradrenalina**, ormoni che agiscono aumentando sia la frequenza e la potenza del battito cardiaco sia la pressione del sangue, dilatando le vie respiratorie e accrescendo la concentrazione di glucosio nel sangue. Questa parte del surrene è stimolata dal sistema nervoso simpatico e svolge un'azione di rinforzo dell'attività simpatica.

3 Il sistema nervoso

Le cellule del sistema nervoso, i **neuroni**, hanno la capacità di trasferire, nei diversi distretti di un organismo, impulsi elettrici a gran velocità, in corrispondenza di particolari giunzioni cellulari, dette *sinapsi*, tramite segnali chimici. Il neurone, infatti, è una cellula dalla forma molto particolare, costituita da un **corpo cellulare**, da cui partono numerosi *dendriti*, e da un **assone**, che può raggiungere lunghezze notevoli.

Nel sistema nervoso degli invertebrati e dei vertebrati, i corpi cellulari dei neuroni sono spesso raggruppati in ammassi, denominati **gangli o nuclei**, a seconda che si trovino, rispettivamente, all'esterno o all'interno del sistema nervoso centrale (SNC). Gli assoni, invece, sono riuniti a formare dei fasci che prendono il nome di **nervi o tratti**, a seconda che si trovino, rispettivamente, all'esterno o all'interno del SNC. Alcuni di essi possiedono una *guaina di mielina*, sostanza isolante a cui si deve il caratteristico colore bianco.

Negli invertebrati il sistema nervoso è generalmente in posizione ventrale, costituito da una rete più o meno complessa, e in alcuni organismi i gangli possono raggiungere dimensioni elevate, tanto da poter essere considerati un encefalo primitivo.

Nei vertebrati, invece, il sistema nervoso è spostato in posizione dorsale ed è assai più complesso, caratterizzato dalla presenza di un encefalo e di un midollo, racchiusi e protetti da strutture ossee (cranio e spina dorsale). Può essere distinto in varie parti, sostanzialmente raggruppabili in un **sistema nervoso centrale (SNC)**, comprendente encefalo e midollo spinale, e un **sistema nervoso periferico (SNP)**, formato dalle vie sensoriali e motorie, che trasmettono e ricevono impulsi dal SNC.

a) Sistema nervoso centrale

Il *sistema nervoso centrale* è composto dall'**encefalo** e dal **midollo spinale**, che costituisce il collegamento tra l'encefalo e il resto del corpo. Visto in sezione, il midollo spinale è formato da una parte interna, la **sostanza grigia**, costituita principalmente da interneuroni, corpi cellulari dei neuroni motori e cellule gliali, e da una parte esterna, la **sostanza bianca**, composta dai tratti formati dagli assoni dei neuroni di associazione (cioè neuroni che collegano zone diverse del SNC).

b) Sistema nervoso periferico

Il *sistema nervoso periferico* comprende i neuroni i cui assoni si trovano al di fuori del SNC, ossia neuroni che ricevono o inviano segnali verso la periferia del corpo, definiti, rispettivamente, *sensoriali* e *motori*. I nervi di questi neuroni si dicono **cranici**, se sono in diretta comunicazione con l'encefalo, o **spinali**, se invece sono connessi al midollo spinale.

Le fibre motorie e sensoriali che innervano una certa regione del corpo viaggiano insieme fino al midollo spinale, dove poi si suddividono in una radice motoria e una sensoriale, che entrano entrambe nel midollo attraverso gli spazi intervertebrali: la prima sul

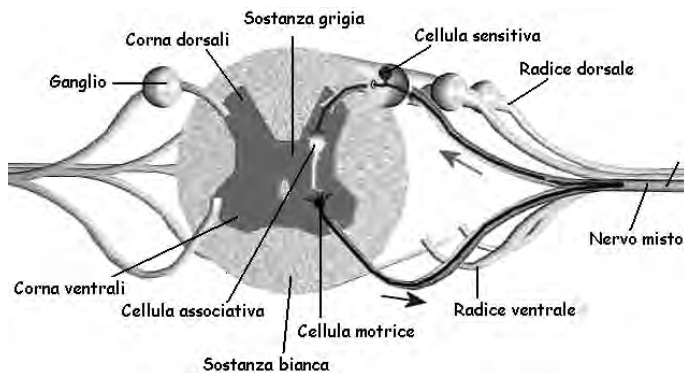


Fig. 1 - Particolare del sistema nervoso

lato ventrale, la seconda su quello dorsale. I corpi cellulari dei neuroni sensoriali si trovano fuori del midollo spinale, nei gangli della radice dorsale, mentre i neuroni motori hanno i corpi cellulari all'interno del midollo. Nel midollo spinale i neuroni sensoriali o motori possono essere in diretta comunicazione tra loro, oppure collegati con interneuroni o con neuroni di associazione, come mostrato in figura 1.

Il sistema nervoso periferico può essere ulteriormente distinto in due parti, differenti sia per struttura sia per funzioni: **sistema nervoso somatico** o «**volontario**» (*SNS*), che controlla i muscoli scheletrici, e **sistema nervoso autonomo** o «**involontario**» (*SNA*), che innerva la muscolatura liscia, il cuore, le ghiandole.

I neuroni del sistema somatico hanno corpi cellulari situati nel SNC e lunghi assoni che si estendono senza interruzione fino al muscolo innervato, detto *effettore*; al contrario, i neuroni del sistema autonomo hanno i corpi cellulari all'interno del SNC, ma i loro assoni non giungono direttamente all'organo o al tessuto bersaglio, bensì formano **sinapsi** (giunzioni tra due neuroni) fuori del SNC, situate nei gangli. Pertanto, si può distinguere un *neurone pregangliare*, con il corpo cellulare nel SNC, e un *neurone postgangliare*, che termina negli effettori. Questo sistema a due neuroni è tipico del SNA.

Il sistema nervoso autonomo può essere a sua volta suddiviso in **sistema nervoso simpatico** e **sistema nervoso parasimpatico**, distinti anatomicamente e funzionalmente (Fig. 2). Le fibre del simpatico si originano nella regione toracica e lombare del midollo spinale e possiedono gangli molto vicini al SNC, mentre le fibre del parasimpatico si originano nella regione sacrale e cranica e possiedono sinapsi vicine agli organi bersaglio. Sotto il profilo funzionale, i due sistemi svolgono azioni contrapposte: il *simpatico* ha generalmente una *funzione eccitatoria* (aumento

della frequenza cardiaca e della pressione sanguigna, aumento del rilascio di glucosio), poiché prepara il corpo all'azione e si attiva in situazioni di paura o di stress; il *parasimpatico* ha spesso una *funzione inibitoria* ed è coinvolto soprattutto nella regolazione di attività fisiologiche ordinarie, come la digestione o l'escrezione.

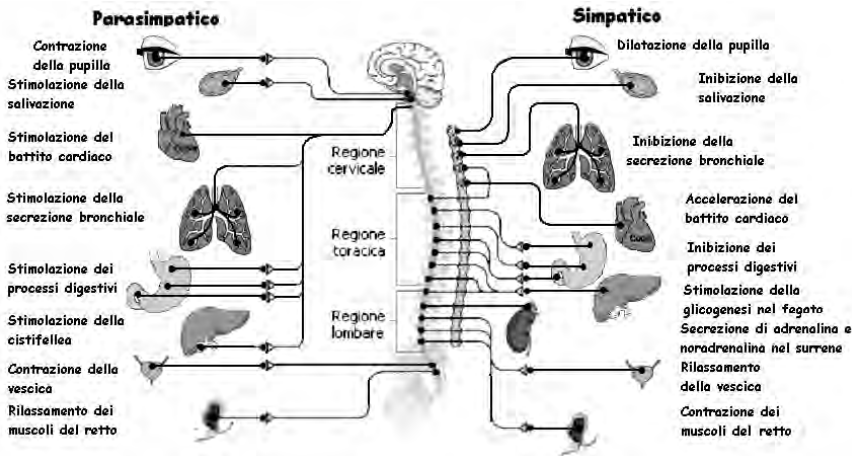


Fig. 2 - Sistema nervoso simpatico e parasimpatico

c) Trasmissione dell'impulso nervoso

• Potenziale d'azione

La capacità dei neuroni di trasferire impulsi elettrici a gran velocità dipende dall'esistenza di una differenza di potenziale tra il lato esterno e quello interno della membrana del neurone, generata dalla differenza delle concentrazioni degli ioni Na^+ e K^+ sui due lati della membrana. In condizioni di riposo, la concentrazione interna dello ione K^+ è trenta volte maggiore di quella esterna, mentre la concentrazione degli ioni Na^+ è dieci volte maggiore all'esterno che all'interno. Questa differenza di concentrazione è mantenuta dalla permeabilità selettiva della membrana, dotata di canali ionici per il passaggio degli ioni Na^+ e K^+ e impermeabile agli ioni negativi, e dall'azione di una proteina di membrana, la **pompa sodio/potassio**, che spinge gli ioni sodio all'esterno della cellula, scambiandoli con ioni potassio. In tal modo la membrana è **polarizzata** e la lieve prevalenza di cariche negative all'interno e positive all'esterno è tale da generare una differenza di potenziale indicata come **potenziale di riposo**, pari a 70mV.

In condizioni di riposo, i canali ionici del sodio sono chiusi, impedendo la diffusione secondo gradiente verso l'interno della cellula. Tuttavia, in seguito a uno stimolo di sufficiente intensità, i canali si aprono temporaneamente determinando l'ingresso del sodio. L'aumento di cariche positive all'interno

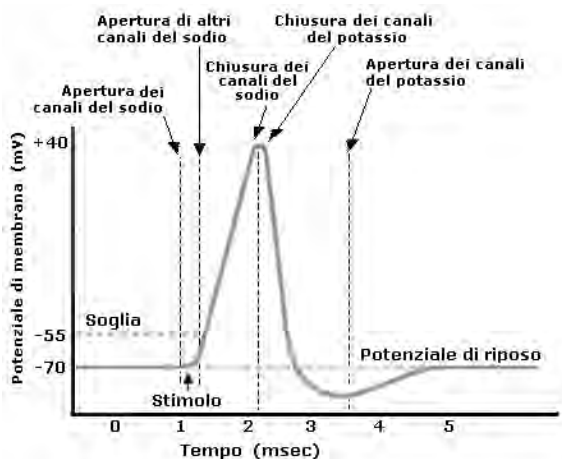


Fig. 3 - Schematizzazione dell'impulso nervoso

della cellula inverte momentaneamente la polarità della membrana (fase di *depolarizzazione*), producendo il cosiddetto **potenziale di azione**. Dopo circa mezzo millisecondo, i canali del sodio si chiudono e si aprono quelli del potassio che, spostandosi secondo gradiente, riporta il potenziale al valore del potenziale di riposo (fase di *ripolarizzazione*). Questa depolarizzazione, che procede rapidamente e con intensità costante per tutta la lunghezza dell'assone, costituisce l'**impulso nervoso** (Fig. 3).

Un'importante caratteristica dell'impulso nervoso è quella di procedere verso un'unica direzione, in quanto, dopo l'espressione del potenziale di azione, ciascun segmento di assone passa attraverso un periodo refrattario, in cui i canali del sodio non si aprono: in tal modo l'impulso non può tornare indietro.

La velocità di propagazione dell'impulso è aumentata dalle guaine mieliniche, dotate di interruzioni che prendono il nome di **nodi** (*nodi di Ranvier*). La *mielina* è una sostanza isolante e, quindi, gli ioni Na^+ e K^+ possono passare soltanto attraverso i nodi, per cui nelle fibre mieliniche l'impulso «salta» da un nodo all'altro arrivando alla velocità di 200 m/s, contro i pochi mm/s che raggiunge nelle fibre non mielinizzate.

Va sottolineato, infine, che i neuroni esprimono un potenziale d'azione sempre della stessa intensità per tutti gli stimoli superiori a un valore di soglia, al di sotto del quale non si ha alcuna risposta. All'aumentare dell'intensità dello stimolo, aumenta, invece, la frequenza con la quale vengono prodotti gli impulsi.

- **Sinapsi**

Con il termine *sinapsi* si indicano le strutture di connessione attraverso le quali due neuroni comunicano tra loro, senza venire a contatto diretto, ma rimanendo divisi da uno **spazio intersinaptico** (Fig. 4). La propagazione dell'impulso nervoso dal neurone presinaptico a quello postsinaptico avviene mediante il rilascio nello spazio intersinaptico di appositi **neurotrasmettitori**,

molecole che, oltre a poter fornire un segnale di intensità proporzionata allo stimolo, sono in grado di produrre effetti eccitatori o inibitori sulla cellula postsinaptica.

I neurotrasmettitori sono sintetizzati nei neuroni e accumulati in vescicole nelle terminazioni assoniche. Quando l'impulso nervoso giunge alla terminazione assonica, le vescicole si fondono con la membrana e rilasciano il neurotrasmettitore nello spazio intersinaptico, dove si combina con il recettore presente sulla membrana del neurone postsinaptico.

La formazione del complesso neurotrasmettitore-recettore determina, attraverso meccanismi molecolari di diverso tipo, un cambiamento della polarizzazione della membrana del neurone postsinaptico che

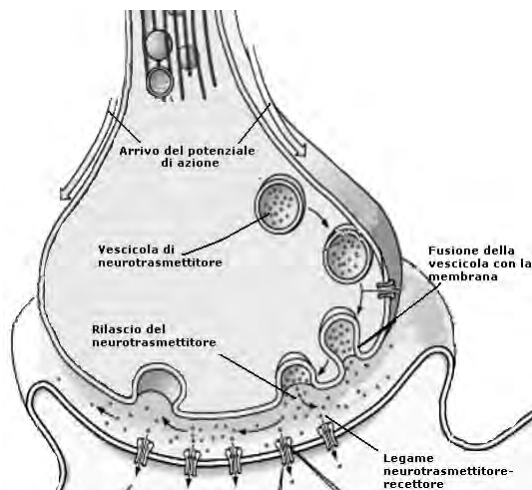


Fig. 4 - Funzionamento delle sinapsi

origina l'impulso nervoso. Subito dopo la liberazione, i neurotrasmettitori vengono rapidamente rimossi e demoliti da enzimi specifici, oppure ricaptati dalla membrana presinaptica per essere riutilizzati. Questa fase è di fondamentale importanza nella regolazione della trasmissione nervosa. I principali neurotrasmettitori del SNP sono l'**acetilcolina** e la **noradrenalina**, mentre quelli del SNC sono la **dopamina** (attività muscolare), la **serotonina** (associata allo stato di veglia) e il **GABA**.

Oltre che dai neurotrasmettitori, la trasmissione sinaptica è regolata anche da **neuromodulatori** (*endorfine, interferoni, interleuchine*), che hanno la funzione di affinare la risposta della cellula al neurotrasmettitore principale. Ogni neurone riceve segnali da centinaia o migliaia di sinapsi e produce una risposta che è il risultato dell'integrazione delle informazioni ricevute dal sistema nervoso e da quello endocrino, elaborazione che avviene all'interno di ogni singolo neurone.

4 Sistema nervoso centrale e organi di senso

a) Encefalo

All'aumentare della complessità degli organismi viventi nel corso dell'evoluzione, le funzioni da controllare e da integrare sono diventate sempre più numerose; ciò ha determinato la progressiva centralizzazione dei punti di elaborazione e di controllo, fino alla comparsa di un centro di elaborazione unico: l'**encefalo**. Nell'uomo esso pesa circa 1,4 kg, racchiude almeno 100 miliardi di neuroni e, analogamente al midollo spinale, è formato da **sostanza bianca** (costituita da tratti mielinizzati) e **sostanza grigia** (formata dai corpi cellulari dei neuroni e da cellule gliali).

È suddiviso in tre sezioni:

1. il **cervelletto**, che funge da centro di regolazione dei movimenti muscolari di precisione;
2. il **tronco cerebrale**, la parte più antica dell'encefalo, che regola funzioni quali il battito cardiaco e la respirazione, e comprende sia i neuroni sensoriali e motori che innervano la testa, sia i tratti che collegano il midollo spinale con i centri del cervello, i quali spesso si incrociano in modo che il lato destro del cervello riceva i segnali dal lato sinistro del corpo e viceversa;
3. il **cervello** (*telencefalo*), che occupa l'80% dell'encefalo ed è diviso in due **emisferi**, collegati tra loro da fibre nervose compatte, le quali costituiscono il corpo calloso (Fig. 5).

La parte esterna del cervello, detta **corteccia cerebrale**, è uno strato dello spessore di pochi millimetri, con circonvoluzioni che ne aumentano la superficie. Nella parte più interna del cervello, tra i due emisferi telencefalici, troviamo il **diencefalo** (di cui le più importanti strutture sono il *talamo*, l'*ipotalamo* e l'*epitalamo*), che contiene i centri associativi di pulsioni ed emozioni, nonché i centri di controllo di fame, sete e sessualità.

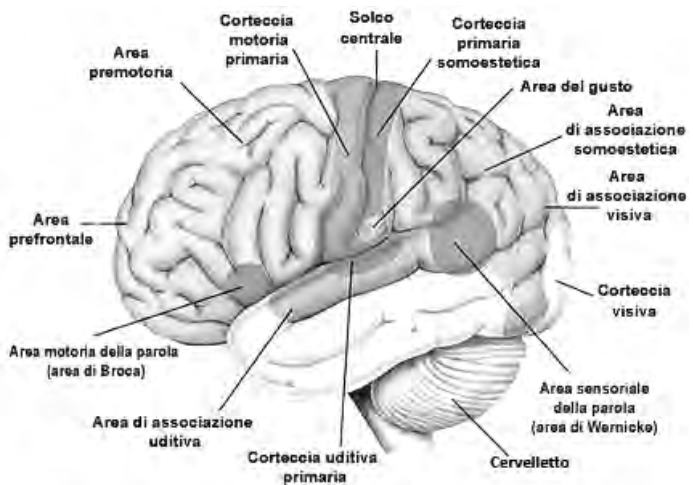


Fig. 5 - Cervello umano

Attraverso la corteccia, è visibile l'ulteriore suddivisione di ciascuno dei due emisferi cerebrali in quattro lobi, **frontale**, **parietale**, **temporale** e **occipitale**, ognuno deputato a funzioni specifiche.

L'area del *lobo frontale* corrisponde alla **corteccia motoria**, quella del *lobo parietale* alla **corteccia sensoriale** (a cui giungono gli stimoli tattili, del gusto, del dolore e della temperatura), quella del *lobo temporale* alla **corteccia uditiva** (dove le diverse regioni sono raggiunte da stimoli di recettori di frequenze diverse) e quella del *lobo occipitale* alla **corteccia visiva**.

Vi sono aree della corteccia cerebrale, raggruppate sotto la denominazione di **corteccia di associazione**, particolarmente ampie nel cervello umano, le quali sono state definite anche «**aree di elaborazione intrinseca**», poiché i neuroni che le costituiscono comunicano con i neuroni di altre aree dell'encefalo, ma non ricevono né inviano segnali al di fuori di esso. Queste aree sono probabilmente responsabili del pensiero, della memoria, delle emozioni, dell'organizzazione delle idee, cioè di tutti gli elementi distintivi della mente umana.

Tra emisfero destro e sinistro esiste una sorta di ripartizione di funzioni, cosa che può aumentare le capacità funzionali senza accrescere le dimensioni (è noto, ad esempio, che le aree preposte al linguaggio si trovano nell'emisfero sinistro). Tuttavia, questa suddivisione di funzioni sembra essere parte dello sviluppo, poiché se uno dei due emisferi subisce un danno, le funzioni delle aree danneggiate possono essere svolte dall'altro, soprattutto se ciò si verifica in età precoce.

b) Orecchio

L'*orecchio* da un lato assolve alla funzione uditiva, trasformando le onde sonore in stimoli nervosi uditivi, dall'altro è deputato al controllo dell'equilibrio, rilevando la posizione della testa rispetto al corpo e allo spazio circostante.

È anatomicamente distinto in orecchio **esterno**, **medio** e **interno**: quello esterno e quello medio hanno la funzione di convogliare e trasmettere l'energia meccanica vibratoria del suono, mentre l'orecchio interno è addetto sia al controllo dell'equilibrio sia alla trasduzione dell'energia meccanica vibratoria in impulso nervoso.

Le onde sonore, raccolte e convogliate dal **padiglione auricolare** sulla **membrana timpanica**, ne provocano la vibrazione, che attraverso tre ossicini, **martello**, **incudine** e **staffa**, viene trasmessa alla membrana della **finestra ovale**. Questa, a sua volta, provoca la vibrazione dei fluidi contenuti nella **coclea**, una struttura a chiocciola costituita da tre canali colmi di sostanza fluida: le onde generate dalle vibrazioni della membrana ovale entrano nella coclea attraverso il canale superiore, e riemergono dal canale inferiore (Fig. 6).

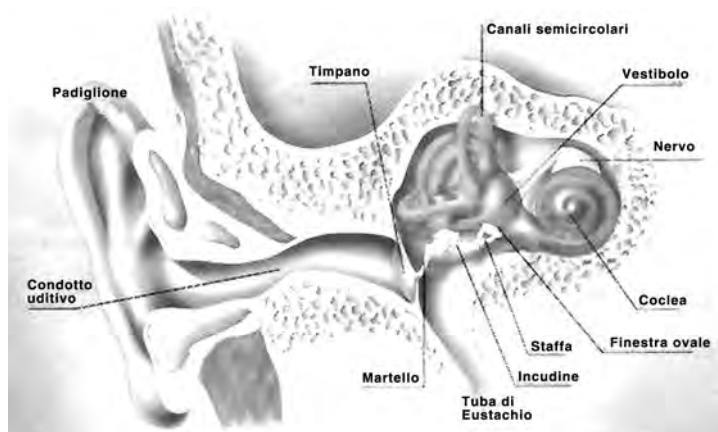


Fig. 6 - Struttura dell'orecchio umano

La membrana basilare del canale centrale della coclea vibra in seguito al passaggio delle onde e tale vibrazione viene colta dalle cellule ciliate dell'**organo del Corti** (Fig. 7), che formano sinapsi con i neuroni sensoriali, i cui assoni costituiscono il **nervo uditivo**. La membrana basilare è meno elastica ad una delle sue estremità e, quindi, non vibra allo stesso modo per tutta la sua lunghezza, il che permette di distinguere suoni di frequenza diversa, i quali producono vibrazioni di energia differente.

L'**apparato vestibolare**, preposto al controllo dell'equilibrio, si trova invece nel *labirinto osseo* dell'orecchio interno ed è composto da tre canali semicircolari posti su piani ortogonali l'uno rispetto all'altro, anch'essi pieni di fluido. I movimenti del capo determinano lo spostamento del fluido nei canali e il movimento delle cellule ciliate, che formano sinapsi con i neuroni sensoriali.

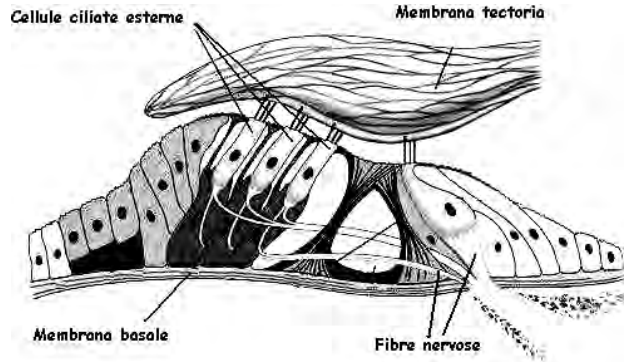


Fig. 7 - Organo del Corti

c) Occhio

L'**occhio** è un organo costituito da tre tonache concentriche (Fig. 8):

1. la **tonaca fibrosa** esterna, distinta in una parte perfettamente trasparente, la **cornea**, e una biancastra e resistente, la **sclera**;
2. la **tonaca vascolare** intermedia, formata dalla **corioidea**, dal **corpo ciliare** e dall'**iride**, all'interno della quale si trova il **cristallino**, elemento che svolge la funzione di una lente;
3. la **tonaca nervosa** o **retina**, la più interna, sede dei fotorecettori.

Vi sono, inoltre, tre camere, che funzionano da mezzi di rifrazione: una **camera anteriore**, tra cornea

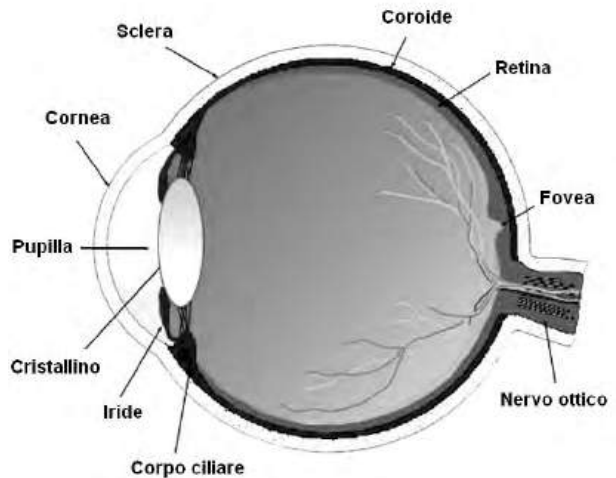


Fig. 8 - Struttura dell'occhio

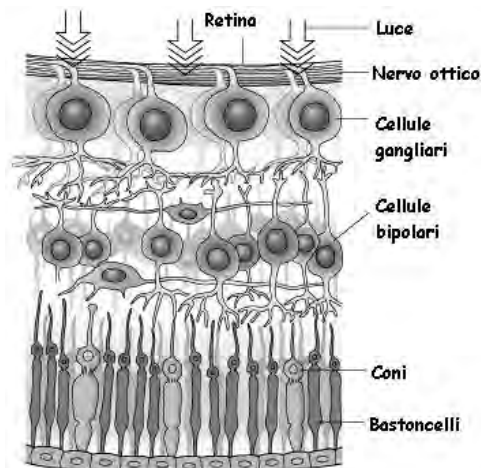


Fig. 9 - Struttura della retina

e iride, e una **camera posteriore**, tra iride e cristallino, contenenti umor acqueo; e una **camera vitreale**, dietro il cristallino, contenente il corpo vitreo.

La luce attraversa la cornea e il cristallino, il quale favorisce la focalizzazione sulla retina di un'immagine capovolta dell'oggetto. La **retina**, a sua volta, possiede due tipi di fotorecettori: i **bastoncelli**, più sensibili alla luce e responsabili della visione in bianco e nero, e i **coni**, meno sensibili alla luce e responsabili della visione a colori. Attraverso le **cellule bipolari**, questi fotorecettori si connettono alle **cellule gangliari**, i cui assoni formano il nervo ottico (Fig. 9).

Mentre la retina contiene circa 125 milioni di fotorecettori, il nervo ottico ospita circa un

milione di assoni: ciò vuol dire che i messaggi di più fotorecettori convergono sulla stessa cellula bipolare e che diverse cellule bipolari convergono su un'unica cellula gangliare. Questa riduzione riguarda principalmente i bastoncelli, le cui informazioni vengono elaborate prima di lasciare la retina. Esiste un'area della retina, la **fovea**, dove si produce l'immagine più nitida; la fovea, infatti, è formata da coni strettamente ammassati, connessi 1:1 con le cellule bipolari.

d) Pelle

Denominata anche **cute**, la *pelle* è un organo addetto allo svolgimento di importanti funzioni di protezione dall'esterno e di percezione, essendo sensibile a stimoli relativi alla temperatura, alla pressione, al dolore. Lo strato esterno della pelle, l'**epidermide**, è composto da cellule continuamente rinnovate, prodotte dalla sua parte più profonda, lo **strato basale** (Fig. 10). Da

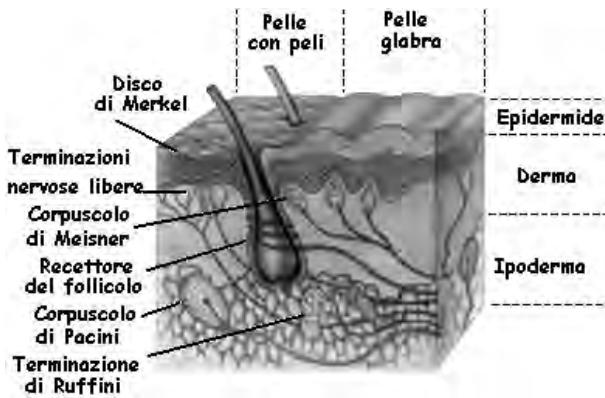


Fig. 10 - Sezione di un tratto di pelle

qui le cellule migrano verso lo strato esterno, **strato corneo**, subendo un processo di differenziazione, al termine del quale la cellula risulta «corneificata», ossia trasformata in una squama cornea nella quale il citoplasma è sostituito da cheratine, che le conferiscono un'elevata resistenza meccanica. Sotto l'epidermide si trova il **derma**, che provvede al nutrimento e all'ossigenazione dell'epidermide non vascolarizzata, con la quale è in contatto tramite la giunzione dermo-epidermica, caratterizzata da estroflessioni che aumentano la

superficie di contatto. Il derma, costituito da una fitta trama di fibre di collagene e di elastina, è il tessuto connettivo della cute, dal quale dipendono l'elasticità e lo spessore della pelle.

Lo strato più profondo del derma, l'**ipoderma**, è formato da tessuto adiposo che, oltre a costituire una riserva energetica, isola dal freddo e fornisce un'efficace protezione meccanica agli organi sottostanti.

La sensibilità tattile della pelle si deve a cellule presenti sia nel derma sia nell'epidermide. Tra quelle contenute nel derma figurano i **bulbi di Krause**, sensibili al freddo, i **corpuscoli di Pacini**, che danno sensibilità alla pressione, i **corpuscoli di Ruffini**, posti nel derma profondo e sensibili al calore, i **dischi di Merkel** e i **corpuscoli di Meissner**, sensibili alle stimolazioni tattili. Vi sono, inoltre, sia terminazioni nervose libere, soprattutto nell'epidermide, che ricevono stimoli dolorifici e tattili, sia terminazioni nervose associate ai bulbi piliferi, poste nel derma profondo e stimulate dai movimenti del pelo. Le fibre nervose sensoriali possono essere molto lunghe: quella che collega la punta del piede con la corteccia sensoriale, ad esempio, arriva a misurare anche due metri. Tuttavia, la risposta agli stimoli dolorifici è di tipo riflesso e parte direttamente dal midollo spinale, producendo la reazione ancor prima che il cervello abbia percepito la sensazione di dolore.

e) Naso e bocca

Sia il *naso* che la *bocca* possiedono recettori dotati della capacità di *tradurre stimoli chimici in sensazioni olfattive e gustative*. L'adulto possiede circa 9.000 *papille gustative*, formate da *gemme gustative* distribuite principalmente sulla superficie superiore della lingua, ma presenti anche su palato, faringe e tonsille. Ogni gemma è legata a più fibre nervose e ciascuna fibra

può essere collegata a più gemme, per cui risulta difficile stabilire da quale parte della lingua si origini un determinato sapore.

Tradizionalmente si distinguono quattro aree, relative ad altrettanti sapori: *dolce* e *salato*, la cui area gustativa sembra essere principalmente la punta della lingua; *acido*, percepito soprattutto nella parte mediana; e *amaro*. La parte anteriore della lingua, però, percepisce anche sensazioni relative alla temperatura, al tatto e al dolore, che vengono integrate con le sensazioni gustative.

Il motivo per il quale sostanze diverse abbiano sapori diversi non è del tutto chiaro. Probabilmente ciò che percepiamo come sapore è un'informazione elaborata dal cervello sulla base di contemporanee sensazioni olfattive. I *recettori olfattivi*, ammassati nella parte superiore delle cavità nasali, sono dotati di una sensibilità molto più alta dei *recettori gustativi*, ma tra loro esiste una connessione funzionale facilmente evidenziabile quando siamo raffreddati: infatti, essendo compromessa la funzione olfattiva, non riusciamo più a distinguere bene il diverso sapore dei cibi.

1 L'apparato riproduttore maschile

Ogni giorno l'uomo produce, all'interno dei testicoli, centinaia di milioni di spermatozoi. Ciascun **testicolo** è diviso in 250 comparti, completamente riempiti di **tubuli seminiferi**. La membrana basale dei tubuli è costituita da cellule spermatogeniche, che producono per differenziamento gli **spermatociti primari**. Questi, in seguito alla meiosi, originano **spermatidi** che a loro volta, dopo una maturazione della durata di 8-9 settimane, diventano **spermatozoi**.

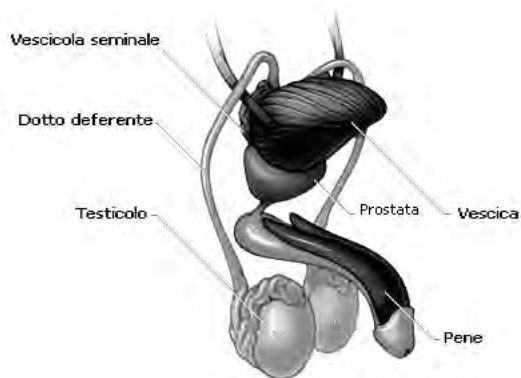


Fig. 1 - Apparato riproduttore nell'uomo

La maturazione degli spermatozoi avviene nell'**epididimo**, un tubulo posto sopra il testicolo, in cui gli spermatidi acquistano la mobilità. Di qui passano poi nel vaso deferente, che, contraendosi, li spinge verso la **prostata** (Fig. 1). In questo percorso vengono aggiunti il liquido prodotto dalle vescicole seminali, contenente fruttosio necessario al nutrimento degli spermatozoi, prostaglandine, che agevolano la risalita degli spermatozoi stessi nel canale riproduttore femminile, inducendone le contrazioni muscolari, e il liquido prodotto dalla prostata, la cui alcalinità serve a neutralizzare l'acidità del canale riproduttore femminile.

Il liquido così prodotto è detto **sperma**

e viene depositato all'interno dell'apparato riproduttore femminile mediante il **pene**, un organo costituito di tessuto spugnoso ed erettile, che si ingrossa e si irrigidisce in seguito all'aumento dell'afflusso di sangue. I tessuti, rigonfiandosi, comprimono le vene e impediscono al sangue di refluire all'esterno, mentre gli impulsi nervosi provenienti dai neuroni sensoriali provocano, attraverso archi riflessi, la stimolazione dei neuroni motori, i quali, inducendo la contrazione della muscolatura dei vasi deferenti, provocano l'**ejaculazione**, cioè l'espulsione del liquido seminale.

Oltre agli spermatozoi, i testicoli producono anche *ormoni maschili* o **androgeni**, i quali, durante lo sviluppo embrionale, inducono lo sviluppo del feto in senso maschile. Successivamente, fino ai dieci anni di età, la produzione di androgeni rimane molto bassa, per poi riprendere nel periodo della pubertà, determinando l'ingrossamento del pene e dei testicoli, nonché l'inizio della produzione di spermatozoi. Parallelamente si sviluppano i cosiddetti «*caratteri sessuali secondari*», come l'ingrossamento della laringe, che comporta l'ispessimento della voce, lo sviluppo dello scheletro e dei muscoli, l'incremento della peluria, la stimolazione delle ghiandole sudoripare, con la comparsa di sudore dall'odore caratteristico.

La produzione degli androgeni, in particolare del **testosterone**, è sotto il controllo dell'*ormone LH*, rilasciato dall'ipofisi e regolato attraverso un sistema a *feedback* negativo, in virtù del quale l'aumento della concentrazione di testosterone rallenta il rilascio di LH. Lo sviluppo degli spermatozoi, invece, è regolato dall'*ormone FSH*.

2 L'apparato riproduttore femminile

Nella donna gli organi produttori dei gameti sono le **ovaie**, connesse attraverso gli ovidotti all'**utero**, l'organo che ospita il feto durante la gravidanza e che normalmente ha la grandezza di un pugno (Fig. 2). Il rivestimento interno dell'utero, l'**endometrio**, si compone di due strati, il più esterno dei quali si sfalda per essere eliminato nel corso della *mestruazione* e viene rigenerato da quello sottostante.

L'utero è dotato di una muscolatura liscia, le cui contrazioni agevolano lo sfaldamento dell'endometrio durante la *mestruazione*, favoriscono la risalita degli spermatozoi verso gli ovidotti e, al momento del parto, determinano l'espulsione del feto. L'organo che riceve il pene e costituisce il canale del parto è la **vagina**, un tubo muscolare della lunghezza di 8-10 cm, che collega la cervice uterina con l'esterno.

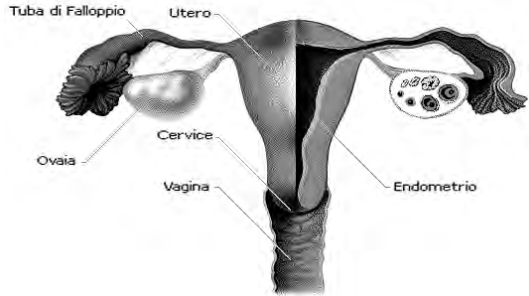


Fig. 2 - Apparato riproduttore nella donna

All'interno del corpo femminile, gli *ovociti* (detti anche *oociti*) *primari* si formano nel terzo mese di vita fetale e alla nascita le ovaie ne contengono circa due milioni, che rimarranno fermi alla profase della prima divisione meiotica fino alla maturazione sessuale della donna. Soltanto 300-400 fra gli ovociti primari raggiungono la maturazione, di norma uno alla volta, ogni 28 giorni circa, dalla pubertà alla menopausa.

La meiosi degli ovociti avviene in modo asimmetrico: dalla prima divisione meiotica, che si completa poche ore prima dell'ovulazione, si originano un corpuscolo polare, contenente la metà dei cromosomi, e un grosso ovocita secondario. A sua volta, la seconda divisione meiotica, che ha luogo solo dopo la fecondazione, genera l'ovulo e un secondo corpuscolo polare. Questo tipo di divisione è necessario per rifornire la cellula uovo di un'elevata quantità di sostanze nutritive, sufficienti durante le prime fasi di sviluppo dell'embrione. Per questo la cellula uovo è molto grande, tanto che nella donna arriva a misurare 100μ.

Nell'organismo femminile la produzione degli ovociti è ciclica e la regolazione del **ciclo mestruale** è competenza dell'ipotalamo. Gli ormoni coinvolti sono gli *estrogeni*, il *progesterone* e le *gonadotropine LH e FSH*. Nelle fasi iniziali del ciclo, dal primo giorno di *mestruazione* in poi, la produzione di ormoni è bassa, ma dopo pochi giorni, sotto l'influenza di LH e di FSH, inizia la maturazione dell'ovocita, che ha luogo vicino alla superficie dell'ovaio, nel **follicolo ovarico**. Questo produce quantità crescenti di estrogeni, che favoriscono la ricostruzione dell'endometrio, preparando l'utero ad accogliere l'uovo. A metà del ciclo la concentrazione degli estrogeni è massima e determina la liberazione dell'ovocita dal follicolo, in modo che possa iniziare il suo viaggio attraverso l'ovidotto, fino a raggiungere l'utero nel giro di 3-4 giorni.

Intanto, sotto l'influenza dell'LH, le cellule del follicolo svuotato si accrescono e generano il **corpo luteo**, che produce quantità sempre maggiori di progesterone ed estrogeni, che inibiscono il rilascio di LH e FSH. Se non c'è stata fecondazione, il corpo luteo si riassorbe, la quantità di ormoni diminuisce e il ridotto afflusso di sangue provoca lo sfaldamento dell'endometrio, ossia la **mestruazione**, con la quale comincia un nuovo ciclo mestruale.

3 Fecondazione e sviluppo embrionale

La vita di un nuovo individuo di qualsiasi specie ha inizio quando lo spermatozoo si fonde con la cellula uovo, evento indicato come **fecondazione**. A partire da quest'unica cellula diploide, l'individuo si sviluppa attraverso un processo divisibile in tre momenti:

- **differenziamento**, cioè la specializzazione delle cellule, dei tessuti e degli organi;
- **morfogenesi**, ossia l'acquisizione dell'aspetto tipico della propria specie;
- **crescita**.

Lo **spermatozoo** è una cellula complessa, composta da una **testa** contenente il nucleo e l'acrosoma (un lisosoma specializzato nel penetrare gli strati della cellula uovo), una **coda** molto potente, che gli conferisce la sua estrema mobilità, e una **terza zona**, posta fra testa e coda, in cui si riscontra un'elevata concentrazione di *mitocondri*, i quali forniscono l'energia necessaria al movimento.

La **cellula uovo** è invece più semplice e molto più grande, circondata da una membrana attaccata alla membrana cellulare — denominata **involucro vitellino**, in cui si collocano recettori proteici degli spermatozoi specifici della specie — e da strati di sostanze gelatinose. Quando lo spermatozoo viene a contatto con l'uovo, gli enzimi dell'acrosoma sciolgono queste sostanze, permettendo alle membrane delle due cellule di venire a contatto, in modo da consentire l'ingresso del nucleo spermatico nella cellula uovo.

In seguito alla fecondazione, sulla superficie dell'uovo si verificano modificazioni che lo rendono inaccessibile ad altri spermatozoi e immediatamente si innesca un'intensa attività metabolica: in seguito all'unione dei materiali genetici dei due gameti, la nuova cellula avvia i processi mitotici.

Lo sviluppo dell'embrione è scandito da tre fasi:

- **segmentazione**, consistente in una serie di divisioni mitotiche che consentono di pompare ioni sodio negli spazi extracellulari, seguiti per osmosi dall'acqua, determinando la graduale formazione di una cavità piena d'acqua, il **blastocoele**. A questo punto, nei mammiferi, l'embrione prende il nome di **blastula** o **blastocisti**;
- **gastrulazione**, che inizia con la formazione, sulla superficie della blastula, di un foro indicato come **blastoporo** (la forma embrionale dell'ano), mentre le cellule ad esso adiacenti si piegano all'interno del blastocoele, dirigendosi verso il polo opposto. Si forma così una cavità interna, l'**archenteron**, che darà origine al canale digerente. Durante la gastrulazione si distinguono inoltre tre strati di tessuto, denominati **foglietti embrionali**, dai quali si origineranno diversi organi. Il più esterno è l'**ectoderma**, da cui prendono origine l'epidermide, l'encefalo e il sistema nervoso; il mediano è il **mesoderma**, che dà origine allo scheletro, ai muscoli e agli apparati escretore, riproduttore e circolatorio; il più interno, infine, è l'**endoderma**, che origina le ghiandole, nonché parte degli apparati digerente e respiratorio;
- **organogenesi**, che rappresenta la terza ed ultima fase dello sviluppo.

4 Gravidanza e parto

Una volta rilasciato dal follicolo, l'ovocita non fecondato ha un tempo di vita di 72 ore, di cui il tempo effettivamente utile alla fecondazione corrisponde a circa 36 ore (la fecondazione, quindi, ha luogo generalmente nell'ovidotto). Dopo 36 ore dalla fecondazione, la cellula uovo si divide in due cellule, dopo 60 ore in quattro, dopo tre giorni in otto, finché, dopo circa cinque giorni, l'embrione risulta formato da 120 cellule ed è ormai giunto allo stadio di blastocisti. In questa prima fase di sviluppo l'uovo è completamente indipendente, in quanto dispone di tutto il nutrimento necessario per far fronte agli iniziali processi di sviluppo.

Dopo circa sei giorni dalla fecondazione inizia l'**impianto** dell'embrione: lo strato di cellule che lo circonda, il **trofoblasto**, entrato in contatto con l'utero, rilascia la gonadotropina corionica, un ormone che stimola il corpo luteo a produrre progesterone ed estrogeni, così da impedire la mestruazione. Penetrando nei tessuti dell'endometrio, il trofoblasto si ispessisce e forma estroflessioni che si insinuano all'interno dell'utero. Intorno all'embrione, intanto, si sviluppano le quattro membrane extraembrionali: il **sacco vitellino**, l'**allantoide** (che forma il cordone ombelicale), l'**amnios** (che racchiude lo spazio pieno di liquido in cui galleggia l'embrione) e il **corion** (che partecipa, insieme all'endometrio, alla costruzione della placenta).

La **placenta** è un organo discoidale che alla fine della terza settimana di gestazione occupa il 20% dell'utero e al termine della gravidanza pesa oltre 500 grammi. Essa provvede al passaggio, dal sangue materno a quello fetale, di tutte le sostanze necessarie all'accrescimento e al nutrimento del feto, nonché all'eliminazione dei prodotti di rifiuto.

Già al ventesimo giorno di gravidanza comincia a pulsare il cuore, mentre al termine del primo mese l'embrione è lungo 5 mm ed ha accresciuto la sua massa di 7.000 volte.

Nel corso del secondo mese si formano mani, piedi, braccia, gambe, ginocchia e, nonostante le dimensioni ridotte (appena 3 cm al termine dei primi sessanta giorni) e la grandezza sproporzionata del capo rispetto al resto del corpo, dovuta al rapido sviluppo del cervello, l'embrione assume sembianze umane, prendendo il nome di **feto**. I primi due mesi di gestazione sono i più delicati per ciò che riguarda l'influenza dell'ambiente sullo sviluppo embrionale: infatti, numerosi farmaci, agenti esterni come i raggi X, ma anche banali infezioni virali possono arrecare gravi danni all'embrione.

Durante il terzo mese si sviluppano praticamente tutti gli apparati e cominciano i primi movimenti fetali. Alla fine del terzo mese il feto è lungo 9 cm e pesa soltanto 15 grammi, ma arriverà a pesarne circa 115 alla fine del quarto mese e 250 alla fine del quinto.

Nei due mesi finali il feto raddoppia le sue dimensioni, registra la formazione di molti nuovi neuroni e acquista dalla madre gli anticorpi, i quali vengono spostati attraverso la placenta con un meccanismo di trasporto attivo altamente selettivo: ciò conferisce al nascituro una temporanea immunità, che lo proteggerà fino a quando il suo sistema immunitario avrà prodotto i propri anticorpi.

Nell'ultimo mese di gravidanza la placenta inizia ad indurirsi e a regredire. Dopo circa 40 settimane di gestazione ha luogo il parto, che si articola in tre momenti:

- **dilatazione**, in cui, a seguito delle contrazioni uterine, il feto viene spinto nel canale del parto. Tale fase termina quando la cervice raggiunge i 10 cm di dilatazione;
- **espulsione**, che inizia con la comparsa del capo;
- **secondamento**, in cui le contrazioni uterine continuano fino all'espulsione della placenta.

Dopo il parto, infine, la prosecuzione delle contrazioni uterine favorisce il ritorno dell'utero alle sue dimensioni normali.

1 L'epidemiologia

L'**Igiene** e la **Microbiologia** occupano un vasto campo di studio sia della Biologia che della Medicina. L'Igiene, in particolare, mira alla conservazione e alla promozione della salute psico-fisica dell'individuo e della collettività, attraverso un'educazione alla **profilassi** e, soprattutto, mediante un attento studio epidemiologico delle malattie, siano esse infettive, cronico-degenerative, professionali etc.

Gli **agenti patogeni** sono *cause dirette* di malattia; vi sono poi *cause indirette* di malattia, definite di *luogo, organiche e fisiche*. Ad esempio: una persona che vive nei Paesi tropicali ha sicuramente più probabilità di contrarre la malaria rispetto a una persona che vive in Europa; un bambino, che ha un organismo più vulnerabile di un adulto, si ammala più facilmente di un adulto in buone condizioni di salute, così come avviene per un adulto denutrito.

Le malattie sono **endemiche** quando sono sempre presenti in una data zona, **epidemiche** quando sono caratterizzate da un'improvvisa comparsa e diffusione, **pandemiche** quando l'epidemia riguarda una zona molto vasta. La trasmissione della malattia, o *contagio*, avviene tramite un **veicolo** (oggetto) o **vettore di trasmissione** (animale o persona infetta), che fa da tramite fra la persona malata e la persona sana.

Si chiamano **portatori sani** gli individui infetti ma non malati, i quali, pur possedendo in sé gli agenti patogeni di una malattia, non ne presentano i sintomi; quindi, tali soggetti sono spesso causa di contagio. Vi sono poi degli stati morbosi, le **parassitosi**, trasmessi da animali pluricellulari che vanno dalle dimensioni di qualche mm fino a raggiungere i 5-8 m nelle *tenie*. Fra i Platelminti ricordiamo la **Taenia solium**, che ha come ospite intermedio il maiale e come ospite definitivo l'uomo. Le tenie hanno il corpo suddiviso in *proglottidi*, ognuna delle quali ha un apparato riproduttore completo; la prima, detta *scolice*, si fissa alle pareti dell'intestino e man mano libera le proglottidi mature, dalle quali fuoriescono le uova che, attraverso le feci, giungono nel terreno e possono, dunque, essere ingerite dal maiale. Nell'intestino del maiale si sviluppa la *larva esacanta*, che raggiunge i muscoli e si incistida come *cisticerco*. A questo punto, se l'uomo ingerisce carne di maiale infetta, il ciclo ricomincia. Pericolosi parassiti sono, inoltre, le *Planarie* e le *Fasciole*. Fra i Nematodi parassiti dell'uomo, ricordiamo: l'**Ascaride**, l'**Ossiuero**, l'**Anchilostoma**, la **Trichinella** e la **Filaria**. Le dimensioni di questi ultimi variano da qualche mm a poche decine di cm.

2 Il sistema immunitario

Nel corso della loro evoluzione, gli organismi hanno sviluppato sistemi di difesa contro l'azione di potenziali invasori. Nei vertebrati, tali sistemi hanno raggiunto un alto grado di complessità, efficienza e specificità, coinvolgendo barriere di tipo anatomico (come la pelle o il rivestimento interno delle mucose), difese di tipo aspecifico e difese altamente specifiche nei confronti dell'invasore, denominate **risposte immunitarie**. Nell'uomo, il sistema immunitario, costituito da un insieme di organi dislocati più o meno in tutto il corpo, comprende *midollo osseo, timo, milza, vasi linfatici, linfonodi e tonsille*.

a) Rete linfatica

A livello dei capillari avvengono gli scambi tra sangue e tessuti; nel contempo, a causa della pressione esercitata dal sangue, una discreta quantità di acqua e soluti passa nel liquido interstiziale. Buona parte di questa viene direttamente riassorbita nei capillari per osmosi, poiché il sangue, divenuto più concentrato, richiama l'acqua all'interno dei capillari stessi.

Il liquido che rimane all'interno dei tessuti viene raccolto dai capillari del **sistema linfatico**, un sistema di vasi che finiscono a fondo cieco nei tessuti, che si riuniscono in vasi progressivamente più grandi, fino ad arrivare a due grandi dotte che riversano la linfa nelle vene succlavie. Analogamente alle vene, anche i vasi linfatici possiedono valvole che impediscono il riflusso della linfa, che penetra al loro interno spinta dalle contrazioni della muscolatura. Lungo il sistema dei vasi linfatici si trovano, inoltre, i **linfonodi**, la cui funzione è quella di produrre *linfociti*, globuli bianchi specializzati nella risposta immunitaria, e di purificare il sangue prima che venga riversato nel torrente ematico.

b) Milza e timo

La *milza* e il *timo* sono i due organi più direttamente coinvolti nella produzione e nella maturazione dei linfociti: in particolare, i *linfociti B* derivano da cellule staminali che si differenziano principalmente nella milza, mentre i *linfociti T* derivano da cellule staminali che si differenziano nel timo. La **milza** possiede anche funzioni di regolazione della crasi cellulare ed umorale del sangue e del volume circolante. In essa, infatti, si possono distinguere nettamente due zone: la **polpa bianca**, costituita da tessuto linfoide dove si trovano i centri germinativi, detti **noduli di Malpighi**, e la **polpa rossa**, deputata alla demolizione delle emazie.

Il **timo**, organo molto sviluppato in età fetale, involge progressivamente nel tempo, aumentando la vulnerabilità dell'organismo. Come detto, in esso le cellule staminali si differenziano definitivamente in linfociti T, i quali completano poi la loro maturazione presso gli organi linfatici periferici.

3 Le difese aspecifiche

Le *difese aspecifiche* costituiscono le «risposte» messe in atto dall'organismo per contrastare l'ingresso di qualcosa riconosciuto come estraneo e vengono indicate genericamente con il nome di **risposta infiammatoria**. In base all'esperienza, è noto che l'infiammazione è caratterizzata da calore, gonfiore, rossore e dolore della zona interessata. Nel caso di una lesione cutanea, ad esempio, le cellule della zona lesionata liberano **istamina**, una sostanza che aumenta il flusso sanguigno nell'area interessata e la permeabilità dei capillari, in modo che i *globuli bianchi* si possano diffondere attraverso le loro pareti e fagocitare gli organismi estranei. Il locale aumento di temperatura ha la funzione di ostacolare la proliferazione dei microrganismi e, in alcuni casi, tale aumento può estendere i propri effetti a livello sistemico, coinvolgendo, cioè, tutto l'organismo e determinando l'insorgere della febbre.

I globuli bianchi, in particolare i **macrofagi**, sono essenziali nelle difese di tipo aspecifico e si trovano dislocati in molti tessuti ed organi, come i linfonodi, il fegato, la milza, i polmoni, i tessuti connettivi, pronti ad eliminare tutte le particelle estranee o i microbi che si siano introdotti nel corpo.

Un'ulteriore categoria di difese aspecifiche è rappresentata dagli **interferoni**, specializzati nel contrastare i virus. Si tratta di piccole molecole proteiche rilasciate dalle cellule infettate, che si legano ai recettori delle cellule vicine, stimolandone la *difesa* nei confronti dell'agente virale. Questa consiste essenzialmente nella produzione di enzimi che bloccano la traduzione dell'RNA

virale. Le molecole di interferone, oltre ad agire sulle cellule circostanti, attivano anche i globuli bianchi, i quali innescano la risposta infiammatoria e immunitaria.

4 Le difese specifiche

Indicate generalmente col nome di **risposta immunitaria**, le *difese specifiche* sono di insorgenza lenta rispetto alle precedenti, ma offrono il vantaggio di essere costruite «su misura» per l'agente estraneo da eliminare. La risposta immunitaria si articola in una **difesa primaria**, a breve termine, che garantisce la protezione dell'organismo contro l'invasione in atto, e una **difesa secondaria**, a lungo termine, che assicura una rapida risposta ai successivi tentativi di «invasione» da parte dello stesso agente e deriva dall'attività dei **linfociti B e T**.

a) Linfociti B

Prodotti e differenziati nel midollo osseo, i *linfociti B* sono costituiti da piccole cellule rotonde metabolicamente inattive, che si trovano accumulate nei tessuti linfatici e portano inserito nella loro membrana un anticorpo specifico per un determinato antigene. Gli **anticorpi** (chiamati anche *immunoglobuline*) sono molecole proteiche in grado di formare complessi stabili con gli **antigeni**, cioè molecole o parti di molecole che l'organismo riconosce come estranee e che stimolano, una volta introdotte nell'organismo, la produzione di «tutele *ad hoc*». L'anticorpo, legandosi stabilmente all'antigene, ne determina l'inattivazione. Antigeni possono essere, ad esempio, le proteine o i polisaccaridi della membrana di cellule batteriche o le proteine dei capsidi virali.

A seguito del contatto con l'antigene, i linfociti B cominciano a proliferare, dando vita a due tipi di cellule:

- le **plasmacellule**, deputate alla produzione di enormi quantità di anticorpi, di cui arrivano a sintetizzare fino a 30.000 molecole al secondo. Hanno un intervallo di maturazione di circa cinque giorni, che rappresenta il tempo minimo necessario al sistema immunitario per contrastare l'infezione, e un periodo di vita limitato;
- le **cellule della memoria**, che invece producono anticorpi e continuano a circolare per lunghi periodi, talvolta anche per tutta la vita dell'individuo, proteggendolo da eventuali infezioni successive da parte dello stesso antigene. I vaccini sfruttano proprio questa capacità del sistema immunitario di mantenere in memoria gli anticorpi di determinate tipologie di antigeni.

Le molecole degli anticorpi (Fig. 1) sono composte da due subunità più lunghe, dette **catene pesanti**, e due subunità più corte, denominate **catene leggere**, le quali presentano sia regioni costanti, caratteristiche della specie che le produce, sia regioni variabili, che determinano la specificità dell'anticorpo nei confronti dell'antigene, originando strutture tridimensionali di forma complementare a questo.

Si è visto che gli organismi sono in grado di generare una varietà enorme di anticorpi, molto più numerosi di quanto si potrebbe prevedere in base al loro *genoma*: l'uomo, ad esempio, è capace di produrre 100 milioni di anticorpi, nonostante il genoma umano non contenga una quantità così elevata di geni strutturali. Questa apparente contraddizione si deve al fatto che le diverse catene proteiche sono codificate in un numero limitato di geni, trascritti e rielaborati attraverso processi di *splicing* alternativi, in modo da originare parecchie possibili combinazioni. Tale

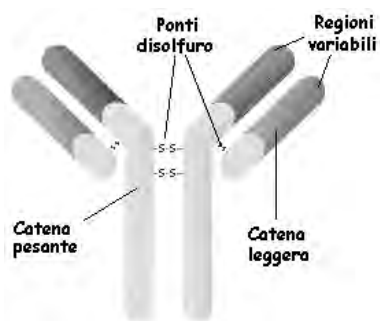


Fig. 1 - Struttura degli anticorpi

processo fa parte del differenziamento del linfocita B, che poi esporrà sulla propria superficie soltanto l'anticorpo per il quale risulta specializzato.

b) Linfociti T

Mentre i linfociti B si occupano di difendere l'organismo da virus, batteri e dalle tossine che questi producono, i *linfociti T* rivolgono la propria azione principalmente contro cellule eucariote come protisti e funghi, nonché contro le cellule dell'organismo infettate dai virus. Questo tipo di linfociti deriva dalle cellule staminali del midollo osseo, le quali, nel corso della vita fetale, migrano nel timo, dove subiscono un complesso processo di differenziazione, selezione e maturazione. Una volta differenziate, sulla loro superficie vengono esposte glicoproteine [1] di tipo diverso, che ne determinano la funzione e la specificità. Inoltre, i linfociti T espongono un recettore attraverso il quale riconoscono le cellule del proprio organismo (*self*); tale recettore ha una struttura simile a quella degli anticorpi esposti dai linfociti B, analogamente ai quali viene assemblato a partire da alcuni geni strutturali. Tuttavia, poiché l'assemblaggio del recettore avviene in modo casuale, i linfociti T devono essere selezionati dopo il differenziamento, al fine di ottenere cellule con una buona capacità di riconoscimento delle cellule *self*.

A seconda del tipo di glicoproteine esposte sulla membrana e del tipo di cellule bersaglio, i linfociti T si distinguono in *citotossici*, *soppressori* ed *helper*.

I **linfociti T citotossici** hanno il compito di aggredire le cellule dell'organismo infettate dai virus, le quali, in seguito all'infezione, risultano riconoscibili perché espongono sulla propria superficie gli antigeni virali. I linfociti si legano a queste cellule attraverso il loro recettore e si differenziano in linfociti attivi che provocano la lisi della cellula, secernendo, nel contempo, **linfocine** (proteine di piccole dimensioni) che richiamano macrofagi.

Sia l'attività dei linfociti T citotossici sia quella dei linfociti B è regolata dai **linfociti T helper** e dai **linfociti T soppressori**: i *primi* riconoscono cellule del sistema immunitario che sono venute in contatto con l'antigene e secernono **interleuchine**, atte a stimolare la proliferazione di altri linfociti B e T in modo da amplificare la risposta immunitaria; i *secondi* sono probabilmente coinvolti nello «spegnimento» della risposta immunitaria.

5 La profilassi

La *profilassi* è il ramo dell'Igiene che si occupa della prevenzione delle malattie attraverso tre vie:

- la **profilassi diretta**, che prescrive le norme per isolare il malato fino alla sua completa guarigione;
- la **profilassi immunitaria**, la quale cerca di ostacolare l'insorgere di un evento morboso mediante l'uso di sieri e vaccini;
- la **profilassi indiretta**, che indica precise norme per la conservazione delle buone condizioni igieniche, affinché sia ostacolato il diffondersi di agenti patogeni.

6 Sieri e vaccini

Un individuo, il quale abbia contratto una malattia, produce anticorpi contro l'antigene responsabile dello stato morboso; questa immunità è quindi *specific*a. I *vaccini* e i *sieri* permettono di far acquisire l'immunità contro determinate malattie che, se contratte, potrebbero essere addirittura mortali.

[1] Si definisce «*glicoproteina*» ogni proteina coniugata (chiamata anche *glicoprotide*) contenente carboidrati legati covalentemente alla struttura polipeptidica; costituiscono un vasto gruppo di composti biologici di considerevole importanza.

I **vaccini** sono costituiti da microrganismi vivi ma attenuati, morti, lisati di batteri o elaborati microbici, che, pur senza essere patogeni, inducono la produzione di anticorpi specifici. Si tratta quindi di un'**immunità attiva** e in genere di lunga durata.

I **sieri**, ottenuti da animali, determinano invece un'**immunità passiva** di breve durata: l'immissione di anticorpi già formati non provoca, infatti, nell'organismo alcuna risposta immunitaria e non vengono prodotti ulteriori anticorpi. La sieroterapia è perciò utilizzata nei casi in cui urge una risposta anticorpale immediata e certamente non duratura. Poiché l'organismo umano può rispondere contemporaneamente all'introduzione di più antigeni, oggi vengono praticate le *vaccinazioni cumulative*. Si possono avere *vaccini misti*, quando vi sono germi morti di più specie batteriche; *associati*, quando accanto a germi morti vi sono sostanze vaccinanti diverse; *polivalenti*, quando sono costituiti da antigeni diversi, ma vicini tra loro (ad esempio, tifico-paratifico). Per ovviare a eventuali reazioni allergiche secondarie, provocate dalla frazione proteica eventualmente presente nel siero, la medicina moderna tende a usare *sieri deproteinati*, nei quali è presente solo la frazione delle **gamma-globuline**.

7 Gli accertamenti delle malattie infettive

Gli *accertamenti delle malattie infettive* si dividono in:

- **microscopici**, vale a dire *a fresco* e mediante colorazioni semplici o particolari;
- **culturali**, ossia con *terreni base* o con *terreni speciali*;
- **sierologici**, cioè attraverso *siero* e *batteriodiagnosi*;
- **biologici**, ovvero *inoculazioni di materiale patologico* in animali e *reazioni allergiche*.

Il terreno di coltura più comune è il **brodo**. Si prepara con carne fatta bollire in acqua, a cui viene aggiunto peptone all'1,5% e cloruro di sodio allo 0,5%. Il pH viene poi aggiustato a 7,2. L'**agar cioccolato** non è altro che brodo al quale si aggiunge *agar* (alga di mare del Giappone) al 2-3% e *sangue*. I metodi di colorazione sono vari; i più comuni e di più facile esecuzione sono lo *Ziehl-Neelsen* e il *Gram*.

Lo **Ziehl-Neelsen** è detto anche *dell'acido-resistenza* ed è legato alla presenza di lipidi, molto abbondanti sugli strati superficiali, di un particolare genere batterico, quello dei *Micobatteri*, al quale appartengono il bacillo tubercolare di alcuni animali, il bacillo della lebbra e alcuni batteri saprofiti. Questi batteri, dopo essere stati colorati a caldo con la fucsina, anche se trattati con alcool o acidi, non si decolorano, per la presenza di speciali sostanze, quali cere e grassi, che rivestono la cellula; vengono per questo definiti «acido-resistenti».

La colorazione di **Gram** è basata, invece, sulla proprietà che hanno alcuni colori, ad esempio il cristallvioletto, il violetto di genziana etc., di combinarsi con lo iodio, dando così origine a composti non dissociabili con l'alcool. I batteri *Gram-positivi* possiedono una parete molto ricca di acido muramico e meno di lipidi, perciò una volta colorati, se trattati con alcool, non perdono il colore e restano in viola; i *Gram-negativi*, invece, si scolorano, per cui possono assumere la colorazione dell'altra sostanza usata come contrasto.

L'esame di Gram è utile per un orientamento riguardante l'identificazione dei germi. Infatti, fra i patogeni più comuni vi sono i:

- **Gram +**: lo stafilococco, lo streptococco, il bacillo del carbonchio, il bacillo della difterite, il bacillo del tetano e quelli della gangrena gassosa;
- **Gram -**: il meningococco, il gonococco, gli enterobatteri, le brucelle, gli emofili e il vibrione del colera.

8 Allergie e disfunzioni del sistema immunitario

Le disfunzioni del sistema immunitario possono provocare gravi danni. Esse si manifestano in tre modi: *allergie*, *malattie autoimmuni*, *immunodeficienze*.

- Le **allergie** sono reazioni anomale ad **antigeni** presenti nell'ambiente, quali polvere, polline, peli, forfora di animali, particolari alimenti come noci, fragole, latte, uova etc., alcuni farmaci ed il veleno di certi insetti, i quali nella maggior parte delle persone non producono alcun effetto. In quelle predisposte, invece, scatenano una notevole produzione di **immunoglobuline E** (IgE) che, a differenza degli altri anticorpi, si fissano sulla membrana dei mastociti, globuli bianchi presenti nella mucosa, sulla pelle e nei connettivi in generale. Se si verifica un altro contatto con l'antigene, le IgE fissate sui mastociti liberano immediatamente una gran quantità di **istamina**, in grado di scatenare una reazione infiammatoria localizzata. Una reazione allergica particolare è lo **shock anafilattico**, scatenato da punture di insetti, da farmaci, da sieri o da vaccini. Tale situazione provoca soffocamento, convulsioni, un'improvvisa dilatazione dei vasi e conseguente collasso; spesso risulta fatale.
- Le **malattie autoimmuni** si manifestano quando un organismo non riconosce più come sue determinate parti del proprio corpo. Questo accade in alcune malattie, come l'artrite reumatoide.
- Le **immunodeficienze** si sviluppano quando il sistema immunitario non svolge più il suo compito come dovrebbe; la persona che ne è affetta contrae facilmente infezioni e sviluppa con altrettanta facilità forme tumorali. L'**AIDS** è una malattia dovuta ad una scarsa attività di risposta immunitaria; la sua sigla sta proprio per «**sindrome da immunodeficienza acquisita**». Portatori di tale malattia sono i **virus HIV 1 e 2**, i quali attaccano i linfociti T-helper, che vengono fortemente ridotti da 1.000 cellule a poche decine per millimetro cubo. L'infezione da HIV si trasmette con rapporti sessuali non protetti, sia etero sia omosessuali, o direttamente con il sangue, tramite siringhe sporche o attraverso trasfusioni. Finché l'individuo manifesti la malattia possono passare diversi anni, ed è questo il periodo in cui, inconsapevolmente, può infettare molte altre persone. Nonostante i progressi nella ricerca, non esiste ancora una cura specifica e la difesa più efficace resta la prevenzione.

9 Cos'è la «Microbiologia» e di cosa si occupa?

La *Microbiologia* è una **branca delle Scienze biologiche che studia i microrganismi (batteri, protozoi, funghi, virus etc.) e le loro interazioni con altri organismi e con l'ambiente**. Questa disciplina ebbe inizio solo nel XVII secolo, con l'invenzione del *microscopio*; rimase tuttavia oggetto di approfondimenti puramente accademici fin verso la metà dell'Ottocento, fino a quando cioè non si comprese l'importanza dei microrganismi nell'«economia della natura».

Un notevole impulso allo sviluppo della Microbiologia fu fornito dall'opera di **L. Pasteur**, il quale con i suoi studi sulla *rabbia* aprì anche la strada alla **vaccinoterapia** e all'**immunoprofilassi**. Contributi fondamentali vennero inoltre dalle osservazioni di **R. Koch**, che riuscì a coltivare *in vivo* il bacillo del carbonchio e identificò i batteri responsabili della tubercolosi e del colera. La Microbiologia nacque, pertanto, principalmente come un *settore della scienza medica, volto all'individuazione e all'analisi dei microrganismi responsabili delle malattie, nonché alla ricerca dei mezzi con cui combatterli*.

Più tardi, però, gli studi microbiologici si estesero ai batteri non patogeni e agli altri microrganismi, che vennero esaminati dal punto di vista morfologico, fisiologico e recentemente anche genetico, con la progressiva differenziazione di rami d'indagine specifici per i vari tipi di microrganismi (**batterologia**, **virologia**, **micologia** etc.) e per le loro possibili applicazioni pratiche (ad esempio, *microbiologia agraria, alimentare e industriale, biotecnologie*).

La **microbiologia agraria** e la **microbiologia industriale**, in particolare, nacquero e si svilupparono parallelamente all'esigenza di sfruttare adeguatamente tutte le principali attuazioni concrete.

La *prima* studia le interrelazioni tra i batteri e il terreno e, inoltre, l'azione svolta dai batteri stessi nei processi naturali sfruttati dalle tecniche agrarie, come la *fermentazione* e la *macerazione*, o, al contrario, nei processi che devono essere evitati, come la *putrefazione*, per la conservazione dei prodotti agricoli.

La *seconda* tende allo sfruttamento delle azioni microbiche economicamente utili, come ad esempio la:

- realizzazione di determinate sostanze chimiche (alcol, acetone, acidi organici, grassi, amminoacidi, vitamine, antibiotici, enzimi);
- trasformazione di prodotti agricoli per la distribuzione su scala industriale (formaggio, pane, birra, bevande alcoliche, conserve alimentari).

1 La vita nelle acque

La vita è nata nell'acqua e, ancora oggi, quest'ultima ospita la maggior parte dei componenti della **biosfera**. Il tipo di vita presente nelle *acque dolci* varia a seconda che si tratti di acque ferme o correnti, mentre la vita nelle *acque marine* cambia essenzialmente in rapporto alla loro profondità.

a) Fiumi

Gli organismi che colonizzano i *fiumi* variano in base alla velocità di scorrimento dell'acqua: nei tratti in cui questa è elevata, come accade alla sorgente, piccoli organismi fotosintetici possono vivere nelle pozze o nelle secche, aderendo alla superficie delle rocce, mentre sotto i ciottoli è possibile che si trovino vari tipi di insetti. Lungo il corso di un fiume si incontrano spesso zone in cui l'acqua è più calma, presso le quali possono accumularsi e decomporre sostanze organiche. Una volta che, ingrossato dagli affluenti, il fiume rallenta il proprio corso, la vita esistente al suo interno comincia ad assumere la fisionomia di quella dei laghi e degli stagni.

b) Acque ferme

Quando si parla di *acque ferme*, s'intende quella di laghi, stagni, fiumi molto lenti, in cui si distinguono zone con forme di vita diverse. La **zona litorale**, la più popolata, comprende piante acquatiche radicate al fondo, che emergono in superficie (tife, giunchi, ninfee etc.), e piante totalmente sommerse. Lumache, larve di zanzara, artropodi, girini si nutrono di piante ed alghe, mentre fra le piante sommerse ci sono piccoli carnivori e sul fondo si trovano bivalvi, vermi e larve di insetti. La zona litorale ospita anche rane, salamandre, tartarughe, serpenti d'acqua. La **zona limnetica**, ossia quella delle acque aperte dove arriva la luce solare, è popolata essenzialmente da pesci (persico, trota) e fitoplancton. La **zona profonda**, che non viene raggiunta dalla luce solare, è priva di vita vegetale ed ospita principalmente detritivori che decompongono i resti organici. Le sostanze riciclate da questi organismi vengono poi ridistribuite durante il rimescolamento delle acque, che avviene due volte all'anno.

c) Oceani

Gli *oceani* sono popolati da organismi fotosintetici solo nello strato più superficiale, poiché l'acqua del mare assorbe rapidamente la luce solare e già ad un metro di profondità giunge meno del 40% della radiazione. Gli organismi che popolano gli oceani si dicono **pelagici**, se liberamente natanti (come il plancton), oppure **bentonici**, se vivono sul fondo (come spugne, anemoni di mare, bivalvi, stelle di mare, pesci, lumache). Nelle zone bentoniche sono più numerosi i funghi e i batteri, i quali si nutrono di resti organici.

d) Litorali

I *litorali* rappresentano le zone in cui le acque sono relativamente poco profonde, come lungo le piattaforme continentali, e corrispondono alle aree più popolate e caratterizzate dalla produttività maggiore, poiché qui giungono sostanze dilavate dai continenti. A seconda del tipo

di litorale, gli organismi hanno sviluppato forme di adattamento diverse. Lungo i litorali *rocciosi*, ad esempio, essi possiedono sistemi di ancoraggio alla roccia, come i pedicelli ambulacrali delle stelle marine o i dischi adesivi delle alghe. Inoltre, la zona **sopracotidale** (bagnata solo dagli spruzzi), la **intercotidale** (periodicamente sommersa o emersa, a seconda delle maree) e la **subcotidale** (costantemente sommersa) ospitano forme di vita differenti. Nei litorali *sabbiosi*, i continui movimenti della sabbia, appunto, limitano il numero degli organismi, al punto che si riscontrano solo piante erbacee con fusti sotterranei, che contribuiscono a stabilizzare le dune, e animali che vivono sotto la superficie della sabbia stessa. I litorali *fangosi*, invece, molto importanti per la deposizione delle uova e per lo sviluppo di forme larvali di esseri viventi marini, ospitano un numero di organismi assai elevato, anche se meno vario rispetto ai litorali rocciosi.

2 La vita sulle terre emerse

Le forme di vita presenti sulle terre emerse sono determinate da fattori fisici, come la temperatura, le precipitazioni, ma anche dalla composizione della crosta. Sulla Terra si possono chiaramente individuare zone caratterizzate da un particolare tipo di vegetazione dominante, indicate con il nome di **biomi**, ognuno dei quali è rappresentato da una comunità di piante e da tutte le forme di vita animale che vivono in associazione con essa. Aree geografiche molto lontane possono ospitare biomi assai simili, poiché se le loro caratteristiche fisiche sono somiglianti, anche gli organismi che le popolano tendono ad essere affini, pur rimanendo distinti sotto il profilo genetico ed evolutivo. I biomi simili forniscono quindi numerosi esempi di **evoluzione convergente**.

a) Foresta decidua

Nelle zone contraddistinte da stagioni calde con precipitazioni modeste e inverni freddi, sfavorevoli alla crescita delle piante, le *foreste* sono *miste* o *decidue*, con piante che perdono le foglie per limitare la perdita d'acqua. Queste sono caratterizzate da **quattro livelli di crescita vegetale**: il *primo* è costituito da alberi le cui fronde formano una volta continua a 10-35 metri di altezza; il *secondo* comprende arbusti che arrivano fino a 5 metri di altezza; il *terzo* è il livello del prato; il *quarto* è il livello del suolo. Nelle regioni settentrionali, gli alberi dominanti sono querce, betulle, faggi, aceri, mentre nelle foreste miste si trovano anche abeti e pini. La vita animale è costituita da arvicole, scoiattoli, procioni, lupi, volpi, linci, cervi. Il suolo, ricco di sostanze, è popolato da batteri, funghi, protisti.

b) Foresta di conifere

Là dove gli inverni sono troppo lunghi e le estati troppo corte, le *conifere* (abeti, pini) prevalgono sulla foresta decidua. La *taiga*, ad esempio, tipo di foresta dominata da alberi dotati di foglie aghiformi, si trova in aree climatiche con inverni molto rigidi, caratterizzati dalla presenza di neve, mentre le piante annuali sono praticamente assenti. La *foresta alpina*, invece, è tipica dell'ambiente montano, umido d'inverno e secco d'estate, dove la scarsa umidità non agevola la crescita delle piante decidue.

c) Tundra artica

Nelle zone estremamente fredde, contraddistinte dalla presenza del *permafrost*, uno strato di terreno gelato posto sotto la superficie dei suoli, la taiga lascia il posto alla *tundra* artica. Durante l'estate, lo strato superficiale disgela parzialmente, per poi gelare nuovamente in inverno: ciò determina la lacerazione delle radici delle piante. La vegetazione della tundra si limita, dunque, a piante erbacee e arbusti legnosi, al di sotto dei quali si trovano muschi e licheni.

d) Prateria

La *prateria* è la zona di transizione tra la foresta temperata e i deserti, caratterizzata da climi caldo-freddi e da una vegetazione principalmente erbosa, sebbene i frequenti incendi ne impediscano la colonizzazione da parte di specie arboree. Nelle praterie si trovano bisonti, gazzelle, zebre e molti altri grossi erbivori, predati da leoni e lupi.

e) Savana

Vere e proprie praterie tropicali sono le *savane*, tipiche di zone con alternanza di stagioni secche ed umide. In queste aree dominano piante come le graminacee, che hanno un'ottima capacità di assorbire acqua nella stagione delle piogge, sopravvivendo in forma sotterranea anche a lunghi periodi di siccità. Gazzelle, impala, antilopi, bufali, elefanti, gnu sono gli animali caratteristici.

f) Macchia mediterranea

Ubicata in aree climatiche con inverni miti ed estati lunghe e asciutte, la *macchia mediterranea* è contraddistinta dalla presenza di piccoli alberi e di arbusti spinosi, un tipo di vegetazione che s'incontra in molte parti del pianeta (Europa, Stati Uniti, Sudafrica), sebbene le specie presenti nelle diverse zone siano tutte filogeneticamente indipendenti. Animali tipici sono i cinghiali e diversi piccoli vertebrati (conigli, fringuelli, lucertole).

g) Deserto

Le *zone desertiche*, tutte concentrate entro i 30° di latitudine nord e sud, sono caratterizzate da aria calda discendente e, di conseguenza, da scarsissime precipitazioni. La poca umidità determina forti escursioni termiche fra il dì e la notte. Molte delle piante desertiche completano il proprio ciclo vitale nel breve periodo di disponibilità d'acqua, quando il deserto si popola di fiori, mentre le piante perenni (come i cactus) si sono adattate ad accumulare grandi riserve idriche. Tutti gli animali delle aree desertiche evidenziano particolari forme di adattamento al clima (come rivestimenti impermeabili ed escrezioni secche), che consentono loro di conservare l'acqua.

h) Foreste tropicali

Le *foreste tropicali* si trovano nelle aree in cui la temperatura e la durata del dì sono praticamente costanti nel corso dell'anno. Le precipitazioni variano di zona in zona e a seconda che siano più o meno abbondanti si distinguono **foreste di arbusti** (con piogge scarse), **foreste miste** (con piogge stagionali), **foreste pluviali** (con precipitazioni copiose durante tutto l'anno).

La *foresta pluviale tropicale*, in particolare, corrisponde all'ecosistema più complesso in assoluto e comprende più specie di tutti gli altri biomi messi insieme. La maggiore competizione in questo tipo di foresta riguarda la luce, e ciò spinge le piante a raggiungere altezze considerevoli, fino a 60 metri per gli alberi più alti, favorendo le specie con foglie grandi e scure. Numerosissimi sono i *rampicanti* e le **epifite** (piante che crescono su altre piante), fra le quali figurano felci, orchidee, muschi. In questa vegetazione così ricca vive una straordinaria varietà di animali e al livello del suolo, dove giunge pochissima luce, si trovano specie adattatesi a resistere in queste condizioni. I cicli dei nutrienti sono estremamente rapidi e non si verifica accumulo di materiale organico all'altezza del suolo, che risulta pertanto povero di principi nutritivi, totalmente e rapidamente assorbiti dalla vegetazione.



Espansione Web

Rispettare e preservare la «biodiversità» nei sistemi ambientali, realizzando esperienze quali, ad esempio, la «costruzione di nidi per uccelli selvatici» oppure l'«adozione di un orto»

IV

Astronomia e Scienze della Terra

Sommario Sezione IV

1 L'Universo	7 La geodinamica esogena
2 Il Sistema Solare	8 Elementi di stratigrafia e di tettonica
3 L'atmosfera	9 Paleogeografia ed evoluzione della vita
4 L'idrosfera	10 Il clima
5 La struttura della Terra solida	11 Elementi di fitogeografia e di zoogeografia
6 La geodinamica endogena	12 L'ecologia

1 I metodi d'indagine in astrofisica

L'**astrofisica** può essere definita come quel ramo dell'astronomia che studia le caratteristiche fisiche dei corpi celesti (natura e struttura).

La capacità dei fisici di compiere nuove scoperte è strettamente legata alla qualità del metodo d'indagine che essi applicano nell'affrontare i problemi che sono chiamati a risolvere.

Il metodo normalmente utilizzato si potrebbe definire *circolare*, in quanto basato su un continuo susseguirsi di fasi di osservazione, di costruzione di modelli teorici e di verifiche sperimentali sui risultati raggiunti. Tali fasi si possono sintetizzare nel seguente modo:

- **osservazione** accurata del fenomeno oggetto di studio;
- **enucleazione degli elementi fondamentali** del fenomeno osservato;
- **costruzione di un modello** del fenomeno, normalmente su basi matematiche;
- **risoluzione formale** del modello;
- **verifica sperimentale** della coerenza fra il fenomeno osservato e il modello costruito.

Il percorso di *modellizzazione* della realtà presuppone la capacità di:

- **utilizzare gli strumenti matematici** adeguati, al fine di ottenere le soluzioni quantitative dei problemi affrontati e di prevedere con buona approssimazione lo sviluppo di fenomeni analoghi;
- **utilizzare gli strumenti informatici** opportuni;
- **gestire le procedure di raccolta e di analisi dei dati** propedeutiche alla costruzione dei modelli;
- **valutare criticamente i modelli utilizzati**;
- **reimpostare i modelli** in seguito all'eventuale discordanza tra gli stessi e il fenomeno osservato.

2 La sfera celeste

Gli antichi definivano *sfera celeste* la parte di spazio che circonda la Terra e che oggi comunemente chiamiamo cielo: in realtà tale spazio non ha né forma né dimensioni definite, anche se sembra incontrare la superficie terrestre lungo una linea circolare che prende il nome di **orizzonte**. Tale evidenza e la necessità di determinare la posizione di qualsiasi corpo celeste rispetto alla Terra spiegano il perché pensiamo al cielo come ad una calotta emisferica ed individuiamo al suo interno dei punti di riferimento, che corrispondono poi alla proiezione nello spazio dei principali riferimenti terrestri.

Il **polo nord** e il **polo sud celesti** sono i punti in cui l'asse terrestre, prolungato (asse mondiale), incontra la sfera celeste, mentre lo **zenit** e il **nadir** rappresentano, rispettivamente, il punto di intersezione tra la sfera celeste, la verticale innalzata sopra la testa di un qualsiasi osservatore e il punto ad esso opposto. Anche la sfera celeste, come quella terrestre, è segnata da una rete di circonferenze definite **meridiani** e **paralleli**: le due circonferenze massime sono il **meridiano celeste fondamentale**, su cui si trovano zenit e nadir, e l'**equatore celeste**, ad esso perpendicolare. Quest'ultimo coincide con l'orizzonte quando l'osservatore si trova al polo (in tal caso lo zenit corrisponderà al polo nord celeste): in tutti gli altri casi le due circonferenze si intersecano

in due punti che sono l'**est** e l'**ovest**, ossia le posizioni da cui sembrano sorgere e tramontare tutti gli astri sull'orizzonte. Il **nord** e il **sud**, invece, sono dati dall'intersezione tra l'orizzonte e il meridiano del luogo. Questi quattro punti insieme formano i **punti cardinali**.

Per indicare la posizione di una stella sulla sfera celeste gli antichi facevano ricorso all'**altezza** (angolo verticale rispetto all'orizzonte) e all'**azimut** (angolo orizzontale tra il sud e il punto in cui il meridiano celeste passante per la stella incontra l'orizzonte). Tale metodo, spesso usato anche oggi perché molto semplice, ha però il grave difetto di far dipendere la posizione della stella dalla località in cui si trova l'osservatore. Oggi gli astronomi utilizzano la **declinazione** (arco di meridiano celeste compreso tra l'equatore ed il piano del parallelo passante per l'astro: corrisponde alla latitudine terrestre e si misura in gradi o frazioni di grado, partendo dall'equatore, con il segno +, se si procede verso nord, e il segno -, se si procede verso sud) e l'**ascensione retta** (arco di parallelo celeste compreso tra la posizione dell'astro e il meridiano celeste fondamentale; corrisponde alla longitudine terrestre e si misura in ore, minuti e secondi): la posizione della stella così individuata è la stessa da qualsiasi punto di osservazione sulla Terra.

Per calcolare le *distanze tra corpi celesti* bisognerebbe utilizzare l'unità di misura prevista dal *Sistema Internazionale* (SI), ossia il *metro*, con i suoi relativi multipli e sottomultipli espressi come potenze di 10; tuttavia, l'entità di tali distanze e l'introduzione piuttosto recente di questo Sistema spiegano perché in Astronomia siano ancora usate *unità di misura alternative*:

- l'**unità astronomica** (U.A.): corrisponde alla distanza media tra la Terra e il Sole, pari a circa 149.600.000 km, e si usa all'interno del Sistema Solare;
- l'**anno luce** (a.l.): è la distanza percorsa in un anno, nel vuoto, dalla luce con un velocità di circa 300.000 km/s. Corrisponde a circa 9.463 miliardi di km e a 63.300 U.A.;
- il **parsec** (pc): è la distanza dalla quale un osservatore vedrebbe il raggio medio dell'orbita terrestre sotto l'angolo di un secondo d'arco. Corrisponde a circa 30.900 miliardi di km e a circa 206.265 U.A.

Per stabilire la *distanza tra la Terra e un corpo celeste* si possono usare tre metodi diversi:

- **metodo della parallasse trigonometrica**: si stabilisce una base, ossia il segmento alle cui estremità sono posti due osservatori, individuata la quale si calcolano i due angoli che essa forma con il corpo celeste. Il terzo angolo di questo triangolo virtuale, osservatore — corpo celeste — osservatore, si ottiene per differenza, ed è possibile ricavare anche la misura della distanza Terra-corpo celeste;
- **metodo delle stelle pulsanti**: le *cefeidi*, stelle pulsanti, sono caratterizzate da un rapporto costante tra periodo di pulsazione e luminosità. Ciò significa che se si riesce a stabilire la distanza di una sola cefeide dal Sole, o da un'altra stella di distanza nota o direttamente dalla Terra, si potrà ricavare quella di qualsiasi stella di cui si conosce il periodo di pulsazione;
- **metodo dello spostamento verso il rosso (*red shift*)**: il calcolo della distanza si ottiene attraverso lo spostamento verso il rosso dello spettro della luce emessa dal corpo celeste. Questo spostamento è direttamente proporzionale non solo alla velocità della sorgente luminosa ma anche alla sua distanza. Il rapporto tra velocità e distanza è la *costante di Hubble*, vale per tutto l'Universo ed è utilissima per calcolare la distanza di corpi estremamente lontani.

3 Stelle e galassie

Le **stelle** sono delle gigantesche masse gassose, ad altissima temperatura, generalmente sferiche e compatte per effetto del proprio campo gravitazionale, in grado di emettere luce propria. Comunemente si usa raggruppare gli astri in **costellazioni**, in base alla loro apparente vicinanza, ma si tratta di un'operazione arbitraria in quanto tali gruppi in realtà non esistono.

Lo scienziato Tolomeo, nel II secolo d.C., ordinò le stelle in base allo splendore, introducendo sei classi di *grandezza*: la prima comprendeva gli astri più luminosi. Oggi, invece, per indicare la luminosità di una stella parliamo di **magnitudine** e la misuriamo con i fotometri fotoelettrici; affinché essa non risenta della distanza a cui si trova la stella dalla Terra e rappresenti quindi un valore oggettivo, si fa riferimento alla *magnitudine assoluta*, immaginando che tutte le stelle si trovino ad una stessa distanza dal nostro pianeta (per convenzione pari a 10 parsec).

Oltre alla *luminosità*, delle stelle possiamo conoscere non solo la *distanza* dalla Terra, ma anche, attraverso gli *esami spettroscopici* [1], la *grandezza*, la *composizione chimica* (che mediamente è: 50-75% di H, 45-20% di He e 5% di altri elementi), la *temperatura* ed il *colore*.

Le caratteristiche principali delle stelle possono essere rappresentate con il **Diagramma Hertzsprung-Russel**: sull'asse delle ordinate è riportata la *luminosità* (o *grandezza*) assoluta, sull'asse delle ascisse il valore della *temperatura*, che oscilla tra i 2.000°C-3.000°C delle stelle rosse ed i 30.000°C delle azzurre. Le stelle non si dispongono a caso, ma la maggior parte di esse si concentra nella parte centrale del grafico, lungo una fascia definita, a formare la cosiddetta *sequenza principale*.

Il diagramma H-R illustra essenzialmente qual è l'evoluzione di una stella, attraverso le tappe principali della sua vita. Una stella ha origine per *collasso gravitazionale* di nubi di polveri e di gas interstellare (soprattutto idrogeno): la contrazione della materia porta ad un aumento della temperatura della *protostella*, che comincia ad emettere radiazioni infrarosse. Quando si raggiungono temperature superiori ai 10 milioni di gradi Kelvin, hanno inizio una serie di reazioni chimiche che corrispondono a diversi stati di equilibrio della stella. L'*idrogeno* si trasforma in *elio*, producendo *energia*; l'*elio*, generando ancora calore, si trasforma in *carbonio* e, quando il combustibile nucleare termina, la stella collassa definitivamente e muore.

Secondo le teorie attuali, ciò che rimane delle stelle dopo questa fase dà luogo alle **stelle a neutroni** e ai **buchi neri**. Le prime (di cui un esempio sono le *pulsar*) hanno origine, nel corso della contrazione, dalla collisione tra elettroni e protoni, che si fondono a formare neutroni: il loro diametro è di soli 10 km, ma la forza di gravità è 100 miliardi di volte più intensa di quella terrestre. I buchi neri sono, invece, corpi celesti che hanno subito un collasso gravitazionale di tale portata che la loro velocità di fuga [2] è uguale a quella della luce: la forza di attrazione è tanto grande che le radiazioni elettromagnetiche, luce compresa, non riescono a sfuggirne. Un buco nero, secondo alcuni astronomi, costituirebbe il nucleo delle *quasar* (*Quasi Stellar Radio Source*), corpi celesti simili a stelle, scoperti nel 1961, la cui origine risulta ancora oggi misteriosa. Rimane insoluto il problema della causa che determina l'*instabilità gravitazionale iniziale*, la quale rappresenta il primo passo per la formazione di una stella. Sembra indiscutibile l'intervento di un fattore esterno: si potrebbe trattare di un'onda d'urto in grado di condensare la materia costituita da *nubi molecolari* (idrogeno allo stato molecolare).

Le **galassie** sono enormi agglomerati di stelle, polvere cosmica e gas, tenuti insieme da forze di attrazione gravitazionale: il numero di quelle visibili è di qualche centinaio di miliardi [3]. Il nostro Sistema Solare si trova in una *galassia spirale*, o a *spirale* (in base alla forma si distinguono galassie *ellittiche*, *spirali* e *irregolari*), con un nucleo a forma di disco, chiamata **Via Lattea**: contiene appros-

[1] Tale metodo, messo a punto da Fraunhofer (fisico e astronomo tedesco) agli inizi dell'Ottocento, permette di risalire alle caratteristiche di un corpo attraverso l'analisi delle radiazioni da esso emesse o assorbite.

[2] Per ogni corpo posto ad una certa distanza da un determinato astro esiste un limite di velocità, detto *velocità di fuga*, superato il quale il corpo non può più essere trattenuto dall'astro, ma sfugge alla sua attrazione gravitazionale e se ne allontana, disperdendosi nello spazio.

[3] Va ricordato che E.P. Hubble, noto astrofisico americano, osservando lo spostamento verso il rosso negli spettri di quasi tutte le galassie, nel 1929 arrivò alla conclusione che esse si stanno allontanando l'una dall'altra con velocità tanto più alta quanto più sono lontane (*legge di Hubble*), con un rapporto costante tra velocità e distanza (*costante di Hubble*).

simativamente 100 miliardi di stelle, ha un diametro di circa 30.000 parsec, uno spessore massimo di 4.000 parsec e fa parte di un ammasso denominato **Gruppo Locale**. Tra le galassie lo spazio non è vuoto, ma contiene gas molto rarefatti, atomi di idrogeno ed elio, particelle elementari e, inoltre, un segnale continuo proveniente da tutte le direzioni: una sorta di **radiazione cosmica di fondo**.

4 L'origine dell'Universo

Le *teorie cosmologiche* cercano di spiegare e di ordinare tutti i fenomeni che avvengono nell'Universo, i meccanismi della sua formazione e le tappe della sua evoluzione. Le teorie principali sono due.

La **teoria dello stato stazionario**, o della *creazione continua*, formulata nel 1948 da Hermann Bondi, Thomas Gold e Fred Hoyle, prevede che l'Universo sia sempre esistito ed esisterà senza soluzione di continuità anche nel futuro: esso cioè non ha avuto inizio e non avrà fine. Ciò non significa che nell'Universo non avvengano cambiamenti ma soltanto che statisticamente essi si compensano. Così per ogni galassia che muore se ne forma una nuova. Se con tale teoria sembra avere soluzione il problema dell'individuazione dell'inizio dell'Universo, restano tuttavia insoluti alcuni quesiti come quelli sulla formazione di nuova materia e sull'origine della radiazione cosmica di fondo.

Le ultime osservazioni e scoperte astronomiche sembrano invece dare credito alla *teorie evolutive dell'Universo*, di cui la più nota resta la **teoria del Big Bang** (il «grande scoppio»), elaborata per la prima volta nel 1927 da G.E. Lemaître, poi ripresa e riveduta nel 1946 da George Gamow. Ne sono state poi proposte numerose varianti. Secondo essa, tutta la materia e l'energia dell'Universo ebbero origine da un agglomerato molto denso che esplose in un determinato momento del passato, l'«istante zero», che secondo alcuni risalirebbe a 15 miliardi di anni fa. Questo «uovo cosmico» improvvisamente si squarciò con un'esplosione immane, il Big Bang appunto, creando una sfera di fuoco in rapidissima espansione, che da allora ha continuato a raffreddarsi e a rallentare. Alla base di tale teoria c'è, quindi, l'ipotesi di un **Universo in espansione**: se quest'ultima proseguirà senza fine, le stelle consumeranno tutto il combustibile, le galassie diverranno corpi oscuri, freddi, inerti ed il Cosmo si trasformerà in un immenso cimitero buio; se la forza di gravità riuscirà a fermare l'espansione, tutti i corpi celesti invertiranno il loro movimento, dando inizio ad una contrazione dell'Universo che porterebbe di nuovo alle condizioni dello stato primordiale, da cui potrebbe innescarsi ancora un Big Bang. Ma il campo d'indagine è estremamente vasto e continuerà ad affascinare gli studiosi a ogni latitudine.

1 Caratteri generali

Il **Sistema Solare** è un insieme di corpi celesti diversi per natura e dimensioni, ma accomunati dalla medesima origine e dal movimento comune in una stessa porzione di spazio (assimilabile ad una sfera con diametro tra i 10.000 e i 20.000 miliardi di km), causato dalle reciproche forze di attrazione. Quasi tutta la massa della materia presente nel Sistema Solare, ben oltre il 99%, è concentrata nella stella chiamata **Sole**: attorno ad essa ruotano otto **pianeti**; più di cinquanta **satelliti** principali; migliaia di **asteroidi** (o «pianetini»); frammenti di varia origine e natura, che attratti dalla forza gravitazionale di un oggetto più grande possono bruciare come **meteore** (le note *stelle cadenti*) o cadere al suolo come **meteoriti**; e infine piccole masse ghiacciate, che se raggiungono il centro del Sistema si trasformano in **comete**. Tra questi diversi corpi celesti si trova la **materia interplanetaria**, ossia un insieme di pulviscolo, gas e frammenti subatomici, estremamente rarefatto.

2 Il Sole

Il Sole è una stella *nana gialla*, situata al centro del Sistema Solare. Le sue caratteristiche principali, in valore assoluto e rispetto alla Terra (rapporto S/T), sono le seguenti:

Caratteristiche	Valore assoluto	Rapporto S/T
Distanza dalla Terra	149.000.000 km	—
Diametro	1.392.000 km	109,2
Massa	$1,99 \times 10^{30}$ kg	332,8
Volume	1.406.000 triloni di km ³	1.302.000
Densità	1,41 (con <i>d</i> acqua = 1)	0,25
Gravità	$2,7 \times 10^4$ cm/s ²	27,9
Temperatura	da 6000°C a > 2.000.000°C	
Composizione	75% H, 25% He, < 1% elementi più pesanti	

Tab. 1

La **struttura** del Sole, per comodità, si immagina suddivisa in una serie di involucri concentrici, ma nella realtà i diversi «strati» non hanno limiti precisi e definiti. Si distinguono un interno opaco, costituito da un *nucleo* avvolto in una *zona radiativa*, e un'atmosfera trasparente, divisa in *cromosfera* e *corona*, separati da una superficie visibile, detta **fotosfera** (sfera di luce). Quest'ultima è uno strato gassoso dello spessore di molte centinaia di chilometri, con una temperatura media di 5.780 K, che costituisce il disco luminoso del Sole, cioè la sorgente del suo spettro continuo. La sua superficie non è liscia, ma presenta una struttura a *granuli brillanti* (masse di gas più calde delle zone circostanti, in continuo movimento) ed è costellata da piccole aree depresse, che appaiono scure, di forme varie, più fredde del resto della superficie di almeno 1.500 K, chiamate *macchie solari*. La **cromosfera** (sfera colorata) è invece lo strato

dell'atmosfera del Sole situato immediatamente al di sopra della fotosfera. Ha uno spessore di circa 10.000 km e una temperatura intorno ai 4.000 K. La parte più esterna dell'atmosfera è la **corona**: si tratta di un involucro di gas ionizzati che, nella parte più esterna, hanno velocità sufficienti per sfuggire all'attrazione gravitazionale del Sole e disperdersi nello spazio come *vento solare*. La corona è interessata da due fenomeni molto caratteristici: le *protuberanze*, grandi nubi filamentose, con temperature tra i 15.000 ed i 20.000 K, che si innalzano dalla cromosfera fino a penetrare nella corona, ed i *brillamenti*, violentissime esplosioni di energia, con le quali viene emesso anche un flusso di elettroni e protoni che raggiunge il nostro pianeta viaggiando con velocità pari a 1.500 km/s. Colpendo le particelle ionizzate dell'atmosfera terrestre, esso dà origine a eventi molto particolari, come le *aurora polari* (boreali ed australi), fenomeni luminosi che si verificano all'aurora alle alte latitudini, o le *tempeste magnetiche*, forti perturbazioni del campo magnetico terrestre.

L'interno del Sole non è direttamente visibile, ma se ne conosce la struttura. Il **nucleo** è la zona in cui avvengono le reazioni termonucleari, con aumento di elio a discapito dell'idrogeno [1], e si ha produzione di energia, che si diffonde verso l'esterno attraverso una zona detta *radiativa*: si calcola un'emissione pari a 100 miliardi di miliardi di Kwh ogni secondo, di cui la Terra riceve soltanto una parte infinitesimale (2 miliardesimi del totale).

Anche il Sole, come tutti gli altri corpi celesti del Sistema, è in continuo movimento. Osservando lo spostamento delle macchie solari, Galileo intuì il movimento principale della nostra stella: si tratta di una **rotazione** attorno ad un asse inclinato di 7°11' rispetto alla perpendicolare al piano dell'eclittica. Il movimento si compie da ovest verso est e dura in media 25 giorni e 8 ore, essendo più rapido all'Equatore (2 km/s) che nella zona dei Poli. Ma, contemporaneamente, il Sole compie anche un movimento di **traslazione** verso un punto della costellazione di Ercole, detto Apice, insieme con gli altri corpi del Sistema Solare, ad una velocità di circa 20 km/s, ed uno di **rotazione generale**, assieme a tutte le stelle della Via Lattea.

3 Pianeti, leggi di Keplero e legge di gravitazione universale di Newton

Intorno al Sole ruotano **8 pianeti**, corpi che, al contrario delle stelle, non hanno luce pulsante ma sono visibili in quanto illuminati dal Sole stesso. Sono, in ordine crescente di distanza dal nostro astro: *Mercurio, Venere, Terra, Marte, Giove, Saturno, Urano, Nettuno*. Essi possiedono delle caratteristiche comuni: compiono un movimento di **rotazione** intorno al proprio asse; descrivono **orbite** pressoché **circolari** che giacciono quasi tutte sullo stesso piano; ruotano attorno al Sole nello stesso senso, con un movimento detto di **rivoluzione**; hanno **forma** quasi **sferica** con schiacciamento polare (non evidente solo in Venere e Mercurio). Inoltre le loro distanze dal Sole non sono casuali, ma vengono rappresentate dalla *Legge di Titius e Bode* (eccetto per l'ultimo). I pianeti più distanti dal Sole, da Giove a Nettuno, hanno numerosi satelliti, lunghi periodi di rivoluzione e brevi periodi di rotazione; mentre i pianeti più vicini al Sole, da Mercurio a Marte, sono piccoli e densi, senza o con un massimo di 2 satelliti, con brevi periodi di rivoluzione ed i primi due con lunghi periodi di rotazione.

Generalmente, i pianeti del Sistema Solare si raggruppano in *due famiglie*, in base alle dimensioni, alla densità, alla distanza dal Sole, alla composizione e alle caratteristiche dell'atmosfera che li circonda. Mercurio, Venere, Terra e Marte sono considerati *di tipo terrestre*: hanno un diametro relativamente piccolo, una densità in media 5 volte superiore a quella dell'acqua,

[1] La quantità di idrogeno ora presente nel nucleo impiegherà circa 5 miliardi di anni per trasformarsi completamente in elio: quando la reazione si arresterà, il nostro Sole sarà destinato a morire.

una distanza dal Sole inferiore ai 300 milioni di km; sono costituiti essenzialmente di rocce e metalli, con un'atmosfera tenue o addirittura assente e pochi o nessun satellite. Giove, Saturno, Urano e Nettuno appartengono alla famiglia di *tipo gioviano o solare*: sono di enormi dimensioni, hanno una densità al massimo una volta e mezzo quella dell'acqua, e la loro distanza dal Sole è elevatissima; inoltre, sono costituiti in prevalenza da idrogeno ed elio, con poco materiale roccioso, una quantità variabile di ghiacci, un'atmosfera molto densa e generalmente un gran numero di satelliti.

Secondo i dati di ricerche pubblicati nell'agosto 2006, **Plutone**, che fino ad allora era considerato un pianeta, il più lontano dal Sole e l'ultimo scoperto dall'uomo (1930), sarebbe da considerarsi un *pianeta nano*, ossia un corpo celeste che ruota attorno al Sole oltre l'orbita di Nettuno con dimensioni non tali da rientrare nella definizione di *pianeta*.

Tutti i pianeti si muovono all'interno del Sistema Solare seguendo delle *leggi specifiche*, individuate da *Keplero* nei primi anni del Seicento, in base alla sola osservazione sperimentale, e poi mirabilmente racchiuse da *Newton*, a distanza di qualche decennio, nella *legge di gravitazione universale*.

La **I legge di Keplero** afferma che *tutti i pianeti descrivono attorno al Sole delle orbite che hanno la forma di un'ellisse; il Sole occupa uno dei due fuochi, comune a tutte le ellissi*. Sull'ellisse che rappresenta l'orbita del pianeta si individuano due punti importanti: l'*afelio*, che rappresenta il punto della curva più distante dal Sole, ed il *perielio*, che è invece il punto più vicino.

La **II legge di Keplero** stabilisce che *il raggio vettore copre aree uguali in tempi uguali*. Per «raggio vettore» si intende il segmento che congiunge il centro del pianeta con il centro del Sole. Ne deriva che il moto del pianeta è più rapido nella zona del perielio piuttosto che in quella dell'afelio.

La **III legge di Keplero** sostiene che *il quadrato dei tempi di rivoluzione dei pianeti è proporzionale ai cubi degli assi maggiori delle orbite*. Ciò significa che il rapporto fra il cubo degli assi maggiori e il quadrato dei tempi è un valore costante. Ne consegue che per i pianeti più lontani, i quali devono percorrere orbite più grandi, i tempi di rivoluzione saranno maggiori.

Le tre leggi di Keplero non sono altro che dei casi particolari della **legge di gravitazione universale di Newton**, secondo cui ogni corpo attira qualsiasi altro corpo con una forza che ha per direzione la retta che congiunge i due corpi e la cui grandezza è direttamente proporzionale al prodotto delle loro masse, nonché inversamente proporzionale al quadrato della loro distanza:

$$F = G \frac{m_1 m_2}{r^2}$$

dove **G** è un coefficiente di proporzionalità chiamato **costante universale**.

4 I pianeti del Sistema Solare

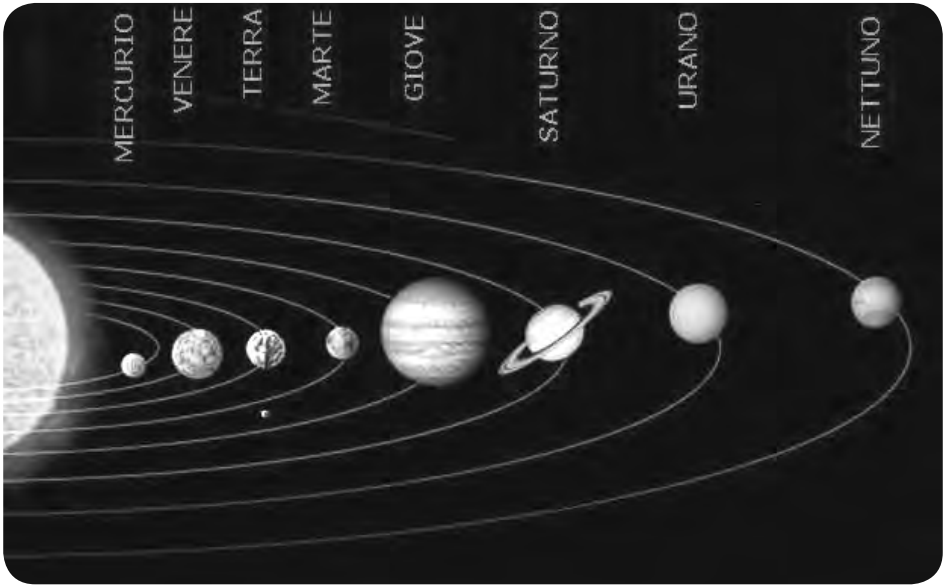


Fig. 1 - Sistema Solare

MERCURIO

Pianeta interno:	primo pianeta in ordine di distanza dal Sole; l'atmosfera è assente
Diametro equatoriale:	4.878 km
Distanza media dal Sole:	58 milioni di km
Rivoluzione completa attorno al Sole:	88 giorni
Rotazione attorno al proprio asse:	58,6 giorni
Satelliti noti:	nessuno

VENERE

Pianeta interno:	secondo pianeta in ordine di distanza dal Sole; l'atmosfera è densa
Diametro equatoriale:	12.104 km
Distanza media dal Sole:	108 milioni di km
Rivoluzione completa attorno al Sole:	225 giorni
Rotazione attorno al proprio asse:	243 giorni (rotazione retrograda)
Satelliti noti:	nessuno

TERRA

Pianeta interno:	terzo pianeta in ordine di distanza dal Sole
Diametro equatoriale:	12.756 km
Distanza media dal Sole:	149 milioni di km
Rivoluzione completa attorno al Sole:	365 giorni, 6 ore e 9 minuti
Rotazione attorno al proprio asse:	23 ore, 56 minuti e 4 secondi
Satelliti noti:	Luna

MARTE

Pianeta esterno:	quarto pianeta in ordine di distanza dal Sole; è il più simile alla Terra; è chiamato <i>pianeta rosso</i> ; l'atmosfera è molto rarefatta
Diametro equatoriale:	6.794 km
Distanza media dal Sole:	228 milioni di km
Rivoluzione completa attorno al Sole:	687 giorni
Rotazione attorno al proprio asse:	24,6 ore
Satelliti noti:	2

GIOVE

Pianeta esterno:	quinto pianeta in ordine di distanza dal Sole; è circondato da fasce di nubi, in cui spicca una grande macchia rossa; è sede di un intenso campo magnetico
Diametro equatoriale:	142.984 km
Distanza media dal Sole:	778 milioni di km
Rivoluzione completa attorno al Sole:	11,9 anni
Rotazione attorno al proprio asse:	9,9 ore
Satelliti noti:	63

SATURNO

Pianeta esterno:	sesto pianeta in ordine di distanza dal Sole; secondo pianeta del Sistema Solare per dimensioni; presenta un esteso sistema di anelli: 7 fasce formate da particelle solide coperte di ghiaccio e rappresentanti un intreccio di più di un centinaio di sottoanelli
Diametro equatoriale:	120.540 km
Distanza media dal Sole:	1.430 milioni di km
Rivoluzione completa attorno al Sole:	29,5 anni
Rotazione attorno al proprio asse:	10,6 ore
Satelliti noti:	61

URANO

Pianeta esterno:	settimo pianeta in ordine di distanza dal Sole; terzo pianeta del Sistema Solare per dimensioni; ha un'atmosfera di colore blu-verde per la presenza di metano
Diametro equatoriale:	51.118 km
Distanza media dal Sole:	2.870 milioni di km
Rivoluzione completa attorno al Sole:	84 anni
Rotazione attorno al proprio asse:	17,2 ore (rotazione retrograda)
Satelliti noti:	27

NETTUNO

Pianeta esterno:	ottavo pianeta in ordine di distanza dal Sole; ha un'atmosfera con bande di nubi e una grande macchia blu
Diametro equatoriale:	49.528 km
Distanza media dal Sole:	4.500 milioni di km
Rivoluzione completa attorno al Sole:	164 anni
Rotazione attorno al proprio asse:	16,1 ore
Satelliti noti:	13

5 L'origine del Sistema Solare

Una volta definite le leggi che regolano tutti i movimenti dei pianeti all'interno del Sistema Solare, resta da spiegare la sua *origine*. Le varie ipotesi si possono ricondurre a *due scuole di pensiero*: una ha alla base l'idea che la nascita del Sistema Solare sia legata ad un *evento catastrofico*, l'altra parte dal presupposto che essa rientri invece in una normale *evoluzione del Cosmo*.

L'evento catastrofico previsto a conferma della prima ipotesi dovrebbe essere una forza in grado di estrarre materia dal Sole e di proiettarla verso l'esterno, formando, per condensazione, i pianeti (Buffon 1794, Chamberlein e Moulton 1900, Jeans 1919, Jeffreys 1929). Tale forza potrebbe essere stata generata dal passaggio di una stella accanto al Sole, con una conseguente onda di marea in grado di provocare il distacco di una grossa nube di gas.

Tuttavia oggi sembrano prevalere, sull'ipotesi dell'evento catastrofico, le teorie evoluzionistiche (Laplace 1796, von Weizsäcker 1944, Kuiper 1951). La loro idea di fondo è che nessun fattore esterno sia intervenuto per la genesi del Sistema Solare, il quale invece deriverebbe da un normale processo di evoluzione di una stella. Con il collasso, la condensazione e il movimento di rotazione della nebulosa solare si sarebbero innescati moti turbolenti nelle parti più esterne (quindi più veloci) della massa stellare, causando il distacco di diversi granuli: raggiunte le densità sufficienti a creare propri campi gravitazionali, essi avrebbero dato origine ai pianeti, evolutisi poi ognuno indipendentemente dagli altri. Per spiegare il perché lo spazio del Sistema Solare, dopo la formazione dei pianeti, non sia rimasto pieno di gas, polveri, frammenti solidi, ma sia al contrario risultato «pulito» alle navicelle spaziali che lo hanno attraversato, gli astronomi hanno immaginato l'intervento di un fortissimo vento stellare, un'enorme corrente elettrizzata, in grado di spazzare via tutti i residui.

6 Il pianeta Terra

La **Terra** è il *terzo pianeta* del Sistema Solare, la sua orbita è compresa tra quelle di Venere e di Marte, e dista dal Sole mediamente 149.600.000 km. Come tutti i corpi solidi, ha forma e dimensioni proprie. Gli antichi ebbero bisogno di ricercare fatti indiscutibili per provarne la **sfericità**: l'orizzonte circolare, la sua ombra proiettata sulla Luna durante l'eclissi, la forza di gravità che agisce lungo i raggi di una sfera, la comparsa e la scomparsa graduale degli oggetti all'orizzonte, i viaggi di circumnavigazione, l'analogia con gli altri pianeti. Oggi bastano le meravigliose fotografie scattate dallo spazio per descrivere perfettamente la forma del nostro pianeta.

Non si tratta di una sfera perfetta, dal momento che risulta «schiacciata» in corrispondenza dei Poli: è assimilabile ad un **ellissoide di rotazione**, ossia a quel solido che si ottiene facendo ruotare un'ellisse intorno al suo asse minore. Più precisamente, la distanza tra i due Poli terrestri rappresenterebbe l'asse minore dell'ellissoide, il diametro dell'Equatore l'asse maggiore. Ma, in effetti, anche tale forma teorica si discosta dalla realtà: la superficie terrestre non è liscia ed

omogenea, a causa delle depressioni e dei rilievi che la caratterizzano. Pertanto, la forma della Terra si identifica con quella del solido costituito dall'insieme dei punti in cui il filo a piombo è perfettamente perpendicolare alla superficie: a tale corpo è stato dato il nome di **geoide**. Tuttavia, lo sfasamento tra ellissoide teorico e geoide reale è soltanto di poche decine di metri ed è, perciò, all'ellissoide internazionale che si fa riferimento per le dimensioni della Terra:

Raggio equatoriale (massimo)	6.378.388 m
Raggio polare (minimo)	6.356.912 m
Schiacciamento polare	1/297
Lunghezza circolo meridiano	40.009.152 m
Lunghezza Equatore	40.076.594 m
Superficie totale	510.100.000 km ²
Massa	5.977×10^{27} g
Volume	1.083.319.780.000 km ³
Densità	5,517 g/cm ³
Accelerazione media di gravità	9,806 m/s ²

a) Moti fondamentali

La Terra compie contemporaneamente diversi movimenti che non soltanto ne fanno mutare la posizione nello spazio ma sono anche causa di fenomeni importanti sulla sua superficie. In primo luogo, il nostro pianeta compie un **moto di rotazione** intorno al proprio asse, da ovest verso est, quindi in senso contrario all'apparente movimento del Sole e degli astri. La durata completa di tale rotazione, ossia il tempo impiegato da un punto qualsiasi della superficie terrestre per compiere un giro di 360°, è di circa 24 ore ed è detta *giorno*. Se si considera il tempo che intercorre tra due passaggi del Sole sul meridiano del luogo di osservazione si parla di *giorno solare*, mentre se si fa riferimento al movimento di una qualsiasi stella si parla di *giorno sidereo*: la sua durata è di 23 ore, 56 minuti e 4 secondi. La velocità angolare di rotazione con cui la Terra compie questo movimento è identica a tutte le latitudini, mentre la velocità lineare è massima all'Equatore (463 m/s) e diminuisce progressivamente verso i Poli, dove diventa nulla.

Sono molteplici i fatti che provano l'esistenza di questo moto di rotazione, in quanto ne costituiscono una diretta conseguenza. Elenchiamo di seguito i principali:

- *l'analogia di comportamento con gli altri corpi celesti* [2];
- *l'esperienza del pendolo di Foucault*: solo ammettendo il moto della Terra si spiega perché il piano di oscillazione di un pendolo, che le leggi della fisica definiscono immutabile nello spazio, ruoti con il passare del tempo;
- *la deviazione della caduta dei gravi*: senza l'effetto della rotazione terrestre i corpi in caduta libera dovrebbero seguire una traiettoria perfettamente perpendicolare alla superficie del pianeta;
- *lo spostamento della direzione dei corpi in moto sulla superficie terrestre*: la legge di Ferrel definisce che tale deviazione avviene verso destra nell'emisfero boreale e verso sinistra in quello australe, come è riscontrabile osservando il moto dell'acqua che defluisce da un tubo di scarico [3];

[2] Si tratta dell'unico tra i fatti elencati che costituisce una «prova» della rotazione terrestre e non una «conseguenza».

[3] La causa di tale deviazione è la ben nota *forza di Coriolis* (chiamata così perché descritta per la prima volta dal fisico Gaspard Gustave de Coriolis), una forza in realtà *apparente*, che nasce dal contrasto di velocità tra un corpo in movimento e la superficie della Terra, anch'essa in movimento a causa della rotazione.

- lo *schacciamento polare*, che fa del globo terrestre un ellissoide, forma di equilibrio di un corpo in rotazione;
- il *movimento notturno della sfera celeste*, quindi di tutti gli astri, *da est verso ovest*, cioè in senso retrogrado rispetto alla rotazione terrestre;
- il *movimento diurno del Sole da est verso ovest*, ovvero anch'esso in senso retrogrado rispetto alla rotazione terrestre;
- la *variazione della gravità con la latitudine*, che non è soltanto l'effetto dello schiacciamento ai Poli ma anche la conseguenza della forza centrifuga dovuta alla rotazione.

Sebbene, come si è detto, tutti i fenomeni elencati siano una conseguenza della rotazione terrestre, certamente l'effetto di tale moto che meglio percepiamo è l'**alternarsi del dì e della notte**. A causa della forma della Terra, i raggi solari che giungono ad essa, paralleli tra loro, illuminano soltanto la parte rivolta verso il Sole, lasciando al buio la parte opposta. Grazie al movimento di rotazione, in un giorno si alternano, dunque, all'esposizione al Sole le varie zone della Terra: in ogni momento c'è un emisfero illuminato e l'altro in ombra, separati da una fascia che prende il nome di **circolo di illuminazione**. Il passaggio, in ogni punto della nostra superficie, dal dì alla notte non è brusco ma avviene gradualmente grazie alla presenza dell'atmosfera che, con i fenomeni di diffusione, rifrazione e riflessione della luce, assicura periodi di chiarore già prima del sorgere del Sole e anche un po' dopo il suo tramonto.

Contemporaneamente al moto di rotazione intorno al proprio asse, la Terra ruota anche intorno al Sole, come tutti gli altri pianeti, percorrendo un'*orbita ellittica*, definita *eclittica*, lunga 940 milioni di km, con un **moto** che è detto **dì rivoluzione**. Tale percorso viene effettuato in senso antiorario, ad una velocità media di 29,8 km/s e in un periodo di tempo pari a 365 giorni, 6 ore, 9 minuti e 10 secondi, definito *anno siderale*. Il piano dell'Equatore terrestre è inclinato di 23°27' sul piano dell'eclittica, con il quale, quindi, l'asse di rotazione terrestre (perpendicolare all'Equatore) forma un angolo di 66°33', che si mantiene costante durante tutto il periodo di rivoluzione. Il Sole occupa uno dei due fuochi dell'eclittica terrestre: in base alla II legge di Keplero, la Terra si muove con una velocità maggiore in perielio (30,3 km/s) e minore in afelio (29,3 km/s).

Anche per dimostrare il moto di rivoluzione della Terra esistono dei fatti osservabili, i quali sono al tempo stesso prove e conseguenze di questo fenomeno. Se ne elencano i principali:

- l'*analogia di comportamento con gli altri pianeti del Sistema Solare*;
- il fatto che quando due corpi celesti sono l'uno nel campo gravitazionale dell'altro si ha *rivoluzione del corpo di massa minore intorno a quello di massa maggiore*;
- la *periodicità degli spostamenti di alcuni gruppi di stelle nel corso dell'anno e di alcune piogge meteoriche*;
- l'*aberrazione della luce proveniente dagli astri*, ossia il piccolo spostamento apparente delle stelle, visto dalla Terra;
- l'*alternarsi delle stagioni e la variazione della loro durata*;
- la *diversa durata del dì e della notte*.

Gli ultimi due fenomeni citati richiedono qualche considerazione in più, se non altro per gli effetti che essi provocano sulla vita del pianeta e sulle attività umane.

Si è detto che l'asse di rotazione terrestre è inclinato rispetto al piano dell'eclittica e che si mantiene parallelo a se stesso durante la rivoluzione. Se invece fosse perpendicolare, i raggi del Sole, nel corso di tutto l'anno, avrebbero sempre, nei confronti di uno stesso punto, il medesimo angolo di incidenza: quindi il giorno avrebbe sempre la stessa durata della notte, cioè ogni punto posto ad una data latitudine riceverebbe sempre la stessa quantità di radiazione luminosa nel corso dell'anno. Dal momento che la premessa non è vera, durante il moto di ri-

voluzione non soltanto cambia la durata dell'insolazione alle varie latitudini, ma anche, per un medesimo luogo, nel corso dell'anno, l'incidenza dei raggi solari. La durata del dì non sarà mai uguale a quella della notte, fatta eccezione per i punti che si trovano sull'Equatore (che hanno notte e dì sempre uguali) e per due giorni nel corso dell'anno, corrispondenti agli **equinozi**. In tali giorni (che cadono approssimativamente il 21 marzo ed il 23 settembre) i raggi solari sono allo zenit sull'**Equatore** (formano cioè un angolo di 90° con l'orizzonte) e tangenti ai Poli: ciò fa sì che il dì e la notte abbiano uguale durata in tutti i punti della superficie terrestre e l'intensità del riscaldamento diminuisca con l'aumentare della latitudine.

Le massime elevazioni a nord e a sud, rispetto al piano equatoriale, il Sole le raggiunge in due posizioni, dette **solstizi**: la prima si verifica intorno al 21 giugno (*solstizio d'estate*), la seconda intorno al 22 dicembre (*solstizio d'inverno*). Nel solstizio d'estate, il Sole è allo zenit sul **Tropico del Cancro**, che corrisponde al parallelo con latitudine $23^\circ 27' N$: l'illuminazione ed il riscaldamento sono maggiori nell'emisfero settentrionale, dove il dì dura più a lungo della notte; all'Equatore si hanno 12 ore di luce e 12 di buio; la calotta artica resta illuminata per 24 ore, mentre quella antartica è al buio. Nel solstizio d'inverno, il Sole è allo zenit sul **Tropico del Capricorno** e la situazione si capovolge: l'illuminazione ed il riscaldamento sono maggiori nell'emisfero meridionale, mentre in quello boreale la notte risulta più lunga del dì; le condizioni dell'Equatore restano invariate; la calotta antartica è completamente illuminata, mentre quella artica è nella piena oscurità.

Per capire, dunque, come variano le condizioni di illuminazione e di riscaldamento nei diversi periodi dell'anno, nell'ambito dei due emisferi, occorre fare riferimento alle *quattro posizioni fondamentali* che la Terra assume rispetto al Sole nei giorni *21 marzo (equinozio di primavera)*, *21 giugno (solstizio d'estate)*, *23 settembre (equinozio d'autunno)* e *22 dicembre (solstizio d'inverno)*, in cui presenta le situazioni estreme. Nei periodi di tempo che intercorrono fra queste quattro posizioni, le condizioni saranno intermedie rispetto a quelle descritte, variando progressivamente dall'una all'altra. E siccome la quantità di calore ricevuta da qualsiasi punto sulla superficie terrestre dipende, appunto, dalla durata del dì e della notte e dall'inclinazione dei raggi solari, ne conseguirà che durante la rivoluzione, cioè nel corso di un anno, nei diversi punti della Terra si avrà l'**alternarsi delle stagioni** (perfettamente invertite nei due emisferi).

A causa della differente velocità della Terra durante l'orbita (maggiore in perielio quando l'inverno interessa l'emisfero boreale, minore in afelio quando nello stesso emisfero è invece estate), le stagioni non risulteranno della medesima durata: nell'emisfero settentrionale il semestre caldo dura circa 7 giorni e 6 ore in più rispetto a quello freddo, laddove nell'emisfero meridionale si verifica esattamente il contrario.

I due nuovi importanti *paralleli* introdotti per illustrare meglio il movimento di rivoluzione della Terra, il *Tropico del Cancro* e il *Tropico del Capricorno* (che rappresentano le massime latitudini alle quali il Sole può raggiungere lo zenit), individuano, insieme con l'*Equatore* e il *Circolo Polare Artico* ed *Antartico*, le 5 **zone astronomiche** del nostro pianeta:

- la *zona torrida*, compresa tra i due Tropici e tagliata in due dall'Equatore, in cui il Sole è sempre alto e a mezzogiorno raggiunge lo zenit, e dove il dì più lungo (circa 13 ore) si ha in estate (solo in corrispondenza dell'Equatore il dì e la notte sono sempre di eguale durata);
- le *zone temperate*, *boreale* ed *australe*, comprese tra i rispettivi Tropici e Circoli Polari, in cui il Sole non raggiunge mai lo zenit, e si ha una marcata differenza stagionale;
- le *zone polari*, *artica* ed *antartica*, comprese tra i rispettivi Circoli Polari ed i Poli, in cui il Sole non è mai alto sull'orizzonte e la notte e il dì durano ciascuno 6 mesi.

b) Moti secondari

Accanto ai moti di *rotazione* e di *rivoluzione*, la Terra compie anche altri **movimenti secondari** [4]. Questi i più significativi:

- *moto di precessione*: l'asse terrestre, come si è detto, si mantiene parallelo a se stesso durante il moto di rivoluzione, ma ciò è vero solo per tempi non troppo lunghi. In realtà, infatti, la sua direzione va lentamente mutando a causa dell'attrazione che Sole e Luna esercitano maggiormente sul rigonfiamento equatoriale, provocando un raddrizzamento dell'asse stesso. Con la forza di reazione dovuta alla rotazione terrestre, si ottiene un movimento che porta l'asse a descrivere due coni con i vertici verso il centro della Terra. Tale movimento avviene in senso contrario alla rotazione, in un periodo pari a 26.000 anni. Le conseguenze principali sono due:
 1. i Poli con il tempo si spostano e con essi anche il Nord, oggi indicato dalla Stella Polare;
 2. gli equinozi si anticipano e questo fenomeno di precessione, alla lunga, potrà portare all'inversione delle stagioni.

I coni descritti dall'asse terrestre, in realtà, risultano ondulati: l'azione attrattiva della Luna provoca, infatti, delle perturbazioni periodiche nel moto di precessione, dette *nutazioni*;

- *variazione dell'inclinazione dell'asse terrestre*: il valore dell'angolo che l'asse terrestre forma con la perpendicolare al piano dell'eclittica non è in realtà costante ma varia, entro i 2°, in un periodo pari a 40.000 anni. Tale variazione modifica l'incidenza dei raggi del Sole nel corso dell'anno e, quindi, influenza il clima;
- *variazione dell'eccentricità dell'orbita terrestre*: l'eccentricità, che è il rapporto tra la distanza centro dell'ellissi-Sole ed il semiasse maggiore dell'orbita, varia nel tempo con un periodo medio di 92.000 anni, influenzando le posizioni dell'afelio e del perielio. Di conseguenza si ha una variazione della quantità di radiazione solare negli equinozi e nei solstizi, che senza dubbio comporta delle modificazioni climatiche.

7 La Luna

Attorno ad alcuni pianeti del Sistema Solare ruotano dei corpi celesti relativamente piccoli, definiti **satelliti**: sono generalmente sferici, non hanno luce propria, sono costituiti da materiale roccioso e, oltre ad orbitare intorno ai pianeti (moto di rivoluzione), ruotano anche su sé stessi (moto di rotazione). La **Luna** è l'unico *satellite naturale* della Terra [5]. La sua massa e dimensioni, modeste se rapportate a quelle di tanti corpi esistenti nell'Universo, sono invece considerevoli rispetto a quelle degli altri satelliti.

[4] Si ricorda che, secondo l'ipotesi di Milanković, formulata nella prima metà del Novecento, sarebbero proprio questi movimenti secondari le «cause astronomiche» delle variazioni dell'insolazione media, le quali spiegherebbero il susseguirsi di epoche glaciali ed interglaciali nella storia del nostro pianeta.

[5] La definizione di *satellite naturale* si usa oggi per distinguere tali corpi celesti dagli oggetti *artificiali* inviati dall'uomo nello spazio, in orbita attorno alla Terra o ad altri pianeti. Vengono utilizzati sia per raccogliere e trasmettere a terra informazioni dallo spazio (*satelliti astronomici*), sia per ripetere segnali telefonici, radiofonici e televisivi intorno alla superficie curva del nostro pianeta (*satelliti per telecomunicazioni*).

Schematizziamo qui di seguito le caratteristiche principali del nostro satellite, in valore assoluto e rispetto alla Terra (rapporto L/T):

Caratteristiche	Valore assoluto	Rapporto L/T
Distanza dalla Terra	384.400 km	—
Raggio medio	1.738 km	0,273
Massa	$7,35 \times 10^{22}$ kg	0,012
Volume	$22,0 \times 10^9$ km ³	0,020
Densità	3,34 (con d acqua = 1)	0,606
Gravità	1,62 m/s ²	0,165

Tab. 2

Anche la Luna, come tutti i satelliti del Sistema Solare (eccetto Titano, legato a Saturno) non possiede né atmosfera né acqua. L'assenza è dovuta al basso valore della gravità e della velocità di fuga: se anche ammettessimo che in passato la Luna, con condizioni di temperatura diverse, abbia avuto qualche forma di atmosfera, i gas e le molecole d'acqua evaporata non avrebbero potuto essere trattenute all'interno del suo campo gravitazionale, ma si sarebbero disperse nello spazio.

Proprio l'assenza di un involucro gassoso come quello che costituisce l'atmosfera terrestre fa sì che sulla Luna non si abbiano fenomeni crepuscolari, pertanto il passaggio dall'illuminazione all'oscurità (e viceversa) è molto brusco. Inoltre i periodi di illuminazione e di oscurità hanno una lunga durata (circa 15 giorni ciascuno) perché la rotazione lunare è molto lenta. Questi fattori, combinati con le caratteristiche del suolo lunare (assenza di vegetazione, materiali lapidei che trattengono poco l'energia ricevuta dal Sole), causano inoltre i fortissimi sbalzi di temperatura tra il dì e la notte: da 150°C (a mezzogiorno, all'equatore) a -140°C (minimo notturno). Per gli stessi motivi il potere riflettente della Luna (*albedo*) è molto basso, pari circa a 0,07: cioè solo il 7% della luce ricevuta dal Sole viene rinviata verso la Terra, mentre il restante 93% viene assorbito dal suolo lunare, trasformato in calore ed immediatamente disperso per irraggiamento [6].

La superficie lunare è caratterizzata da aree depresse, molto ampie, piane, coperte da una polvere soffice costituita da frammenti rocciosi e vetrosi (*regolite*), che ci appaiono come delle macchie scure e sono definite **mari**, e da aree a morfologia complessa ed accidentata, che invece appaiono chiare e vengono chiamate **continenti**. Mari e continenti sono costellati da **rilievi** che raggiungono anche i 9.000 metri di altitudine [7]. Si tratta di *catene montuose* (cui spesso è stato dato il nome dei più noti rilievi terrestri: ad esempio, Alpi, Appennini etc.), *crateri* e *circhi*. Questi ultimi due tipi morfologici sono delle grandi aree circolari a fondo piatto e bordi rilevati. La loro origine dovrebbe essere legata, secondo alcuni, all'impatto, con esplosione, di meteoriti (di cui oggi non si ritrovano resti in quanto sarebbero stati immediatamente vaporizzati a causa dell'assenza di atmosfera) sul suolo lunare; secondo altri, all'intensa attività vulcanica avvenuta durante il consolidamento della Luna, testimoniata dall'origine ignea delle rocce che costituiscono sia i mari che i continenti [8]. L'età di queste rocce, calcolata utilizzando il metodo della radioattività, che consente di stabilire una cronologia assoluta, risalirebbe a 3-3,3 miliardi di anni per le più recenti e a circa 4,5 miliardi per le più antiche.

[6] L'irraggiamento è uno dei meccanismi di conduzione del calore.

[7] Va precisato che si tratta di valori assoluti, in quanto, sulla Luna, manca un opportuno livello di riferimento, quale è la superficie del mare per tutte le altezze e le profondità misurabili sul nostro pianeta.

[8] In particolare, per i primi si tratterebbe di *rocce effusive* simili ai nostri basalti, per i secondi di *rocce intrusive* assimilabili alle anortositi terrestri.

Altre forme caratteristiche della superficie lunare sono i **solchi**, trincee lunghe o corte, dritte o sinuose, che tagliano pendii e pianure. Potrebbero essere delle fessure formatesi per il raffreddamento delle lave, linee di fuoriuscita di masse gassose, canali di flusso delle lave o fratture da cui esse hanno avuto origine, o ancora faglie tettoniche. Infine, sono da ricordare quelle aree definite dagli astronomi **mascons** (dall'inglese *mass concentrations*, ossia «concentrazioni di massa»), che, con il loro campo gravitazionale particolarmente elevato, sarebbero responsabili delle perturbazioni provocate nell'orbita dei satelliti artificiali.

a) Moti della Luna

La Luna, come gli altri satelliti, è soggetta ad una serie di movimenti contemporanei, il primo dei quali è rappresentato dal **moto di rotazione** intorno al suo asse, che si compie, come per la Terra, da ovest verso est. La velocità angolare di rotazione è pari a 13° al giorno, quindi la durata di una rotazione completa è di 27 giorni, 7 ore, 43 minuti e 12 secondi. Tale periodo corrisponde esattamente al tempo impiegato dalla Luna per compiere il suo **moto di rivoluzione** intorno al nostro pianeta (*mese sidereo*). È questo il motivo per cui essa rivolge alla Terra sempre la stessa «faccia». Se per questo movimento si fa riferimento alla Terra, anziché ad una stella, il nostro satellite presenta un tempo di rivoluzione più lungo. Quando la Luna avrà terminato di compiere il suo giro di rivoluzione siderea, il nostro pianeta non si troverà più nel medesimo punto, in quanto si sarà spostato intorno al Sole di circa 27° : la Luna, dunque, per raggiungere la stessa posizione di partenza rispetto all'allineamento Terra-Sole dovrà percorrere ancora un tratto, il che allungherà il periodo di rivoluzione di circa due giorni (*mese sinodico*).

Il movimento di rivoluzione avviene con una velocità di circa 1 km/s, in senso antiorario, lungo un'orbita ellittica di cui la Terra occupa uno dei fuochi: il punto più vicino Luna-Terra è detto **perigeo**, quello più lontano **apogeo**. Il piano dell'orbita lunare forma un angolo di circa 5° con il piano dell'eclittica terrestre e quindi la interseca in due punti detti *nodi*, lungo una linea che è chiamata, appunto, **linea dei nodi**.

Contemporaneamente, la Luna si troverà a ruotare insieme con la Terra intorno al Sole, con un **movimento** denominato **di traslazione**: l'orbita lunare, che risulta dalla combinazione di questi tre movimenti principali rispetto al Sole, sarà una curva sinusoidale particolarmente complessa, definita *epicicloide*. L'interazione con due corpi celesti di massa considerevole, il Sole e la Terra, fa sì che la Luna sia in realtà soggetta anche ad altri movimenti più lenti e di minore entità.

Ricordiamo unicamente il *moto di regressione della linea dei nodi*, in quanto è responsabile sia dei fenomeni delle nutazioni terrestri, citate a proposito dei movimenti secondari della Terra, sia della ciclicità con cui si verificano le eclissi.

b) Eclissi

Il fenomeno delle **eclissi** è collegato alla posizione assunta dalla Luna rispetto al sistema Terra-Sole durante il suo moto di rivoluzione. Quando la Luna si trova in **congiunzione**, cioè dalla stessa parte del Sole rispetto alla Terra, l'emisfero che essa rivolge verso di noi non è colpito dai raggi solari e, quindi, ci risulta oscuro: si è nella fase di *Luna nuova* (o *novilunio*). Quando, invece, la Luna si trova in **opposizione**, cioè dalla parte opposta del Sole, sempre rispetto alla Terra, vediamo l'emisfero rivolto verso di noi illuminato: è la fase di *Luna piena* (o *plenilunio*). Queste due posizioni sono definite **sizigie**. Quando invece la Luna forma con il Sole e la Terra un triangolo rettangolo, mentre il nostro pianeta occupa il vertice dell'angolo retto, si parla di **quadrature**, e dalla Terra vediamo il *primo* e l'*ultimo quarto* di Luna. Tra queste quattro fasi principali (2 sizigie e 2 quadrature) si avranno tutte le condizioni di illuminazione intermedie, con la Luna in *fase crescente* dal novilunio fino al plenilunio e in *fase calante* dal plenilunio al novilunio seguente.

Se l'orbita terrestre e quella lunare si trovassero sullo stesso piano, ogni mese, durante la fase di plenilunio, il cono d'ombra della Terra proiettato sulla Luna la oscurerebbe completamente e avremmo quindi un'**eclissi di Luna**; allo stesso modo, ad ogni fase di novilunio si verificherebbe un'**eclissi di Sole**. Invece le eclissi sono fenomeni rari: questo perché l'orbita lunare e quella terrestre non giacciono sullo stesso piano e quindi l'allineamento Sole-Terra-Luna, responsabile dell'eclissi di Luna, e l'allineamento Sole-Luna-Terra, che provoca l'eclissi di Sole, si verificano soltanto quando c'è l'allineamento (anche non perfetto) lungo la linea dei nodi.

Il fenomeno dell'eclissi, sia essa di Luna o di Sole, può essere *totale* o *parziale* a seconda se ci sia o meno l'oscuramento completo. Un tipo particolarmente interessante di eclissi solare è l'**eclissi anulare**. Essa si verifica quando la Luna è in apogeo (ferma restando la condizione di allineamento lungo la linea dei nodi): il cono d'ombra non riesce ad eclissare completamente il disco solare ma ne lascia illuminato il bordo, che appare come un anello splendente e risulta dunque facilmente osservabile ed analizzabile per gli astronomi.

c) Teorie sull'origine della Luna

Definite la struttura e le caratteristiche della Luna, e descritti i suoi movimenti fondamentali, occorre rispondere al quesito di come essa abbia avuto origine. Esistono diverse ipotesi, riconducibili a tre meccanismi di formazione principali, ai quali di seguito si farà un rapido accenno.

La **teoria della fissione**, avanzata da George Darwin nel XIX secolo, prevede la *separazione* di una grossa goccia di materiale, dal cui raffreddamento avrebbe poi avuto origine la Luna, da una Terra allo stato fuso in rapida rotazione su se stessa e quindi soggetta, per effetto dell'attrazione gravitazionale solare, a continue contrazioni e dilatazioni: le cosiddette *maree*. Una versione più moderna di tale teoria identifica la causa della «separazione» non nelle maree bensì nell'instabilità dovuta alla concentrazione, verso il centro della Terra, del materiale più pesante, che avrebbe portato ad un aumento di velocità verso la parte più esterna e al distacco di una goccia della massa fusa.

Secondo un'altra ipotesi, nota come **teoria della cattura**, la Luna sarebbe invece nata come un corpo indipendente nel Sistema Solare ma, ad un certo punto, si sarebbe talmente avvicinata alla Terra da rimanere attratta nel suo campo gravitazionale.

La **teoria dell'accrescimento**, infine, vuole che la Luna sia il risultato dell'unione di materiali diversi già in orbita intorno alla Terra.

Malgrado il dibattito sull'origine del nostro satellite sia ancora aperto, attualmente, grazie anche alle numerose missioni lunari, siamo in grado di delineare, in linea di massima, le *fasi principali della «vita» della Luna* secondo la seguente linea evolutiva: una volta formatasi, essa subì una *separazione della crosta*, cui fecero seguito una prima epoca di *vulcanismo*, quindi un periodo di *bombardamento meteorico*, una nuova era di vulcanismo, poi una fase di declino dell'attività e, infine, uno stato di *quiescenza* che perdura ancora oggi.

8 L'esplorazione dello spazio: l'ultima frontiera della scoperta

Esplorare fisicamente lo spazio esterno al nostro pianeta comporta un insieme di attività tecnico-scientifiche, ma anche economiche. Resa possibile dalle ricerche effettuate prima e durante il secondo conflitto mondiale, la «corsa allo spazio» è iniziata nel 1957, anno in cui viene messo in orbita il **primo satellite artificiale**. I successivi trent'anni sono caratterizzati dalla *guerra fredda* della competizione fra **USA** e **URSS**. La fine dell'Unione Sovietica ha segnato, anche in campo spaziale, una svolta; si è passati dalla «competizione» alla «cooperazione», e la diminuita importanza dell'«uso dello spazio» per fini militari ha aperto la strada alle applicazioni civili della ricerca: dai **satelliti meteorologici** a quelli per le **telecomunicazioni**.

a) Agli albori

I primi tentativi di costruire un mezzo con propulsione a razzo capace di staccarsi dalla Terra risalgono agli anni Trenta del secolo scorso. Durante la seconda guerra mondiale, queste ricerche ricevettero un grande impulso nella Germania nazista, che ne aveva intuito l'importanza nelle applicazioni belliche. Il gruppo di scienziati e studiosi creato da **Werner von Braun**, nella base militare tedesca di Peenemunde, riuscì a sviluppare le prime armi basate su razzi, le famose **V-2**, che vennero utilizzate per bombardare Londra dal territorio tedesco nel 1944. Alla fine del conflitto, molti ricercatori di quel gruppo entrarono a far parte dei progetti spaziali degli Stati Uniti e dell'Unione Sovietica, entrambi particolarmente interessati alle applicazioni belliche di quella tecnologia.

La vera e propria corsa allo spazio, in un mondo allora diviso dalla celeberrima «cortina di ferro» in due zone contrapposte e ostili, iniziò il 4 ottobre 1957 con il lancio, da parte dei Sovietici, del *primo satellite artificiale* terrestre, lo **Sputnik 1**, il cui nome vuol dire «compagno di viaggio». Costituito da una sfera di metallo di 50 cm di diametro e dotato di 4 antenne lineari di circa 3 m di lunghezza ciascuna, lo Sputnik era equipaggiato con due trasmettenti, grazie alle quali inviò, durante le tre settimane di durata della sua missione, un segnale radio rimasto storico. Con una massa di poco più di 83 kg, percorse un'orbita compresa fra i 215 e i 939 km di altezza dal suolo, con un periodo di 96,2 minuti. Lo *Sputnik 1* ricadde sulla Terra nel gennaio 1958, disintegrandosi durante il rientro.

Pochissimo tempo dopo fu la volta degli Americani, che accettarono la sfida: la loro prima missione fu un successo, con il lancio dell'**Explorer 1**.

b) Animali e uomini in orbita: l'iniziale primato sovietico

Il **primo essere vivente** a entrare in orbita attorno alla Terra è stata una **cagnetta russa** chiamata **Laika**. Il suo nome significa letteralmente «quella che abbaia» ed è comune nei Paesi dell'Europa orientale per indicare una razza simile agli husky. Il 3 novembre 1957 Laika fu condotta in orbita a bordo di un satellite sovietico, lo **Sputnik 2**. La missione venne realizzata in modo parecchio affrettato, e quindi necessariamente approssimativo, per ordine del governo, il cui scopo era quello di poter vantare un «successo» nella corsa spaziale entro la data del 7 novembre 1957, quarantesimo anniversario della Rivoluzione bolscevica. La missione di Laika, che doveva durare diversi giorni, finì anzitempo e l'animale morì per vari malfunzionamenti all'interno della capsula spaziale, principalmente a causa dell'aumento della temperatura nell'abitacolo. La vicenda della cagnolina russa ebbe forti ripercussioni nel mondo e molte associazioni per i diritti degli animali protestarono in modo energico. Il sacrificio di Laika, che commosse tutti a quel tempo, servì comunque per aprire le porte dell'osservazione diretta dello spazio anche all'uomo.

L'era dell'*esplorazione lunare* si aprì nel 1959, quando tre sonde sovietiche — **Lunik 1, 2 e 3** — raccolsero le prime informazioni dirette sul nostro satellite e finalmente ne fotografarono la mitica «faccia nascosta».

Pochi anni dopo si ebbe l'attesissimo primo volo di un essere umano al di fuori dell'atmosfera terrestre: fu il sovietico **Yuri A. Gagarin**, in assoluto, il **primo cosmonauta**. Il 12 aprile 1961, a bordo del **Vostok 1**, raggiunse le condizioni di volo orbitale attorno al Globo e rimase nello spazio, in quelle condizioni estreme, per 1 ora e 48 minuti. Rispetto ai voli umani successivi, quello di Gagarin fu molto breve e pericoloso: si disse in seguito che la maggioranza dei tecnici, anche russi, erano convinti che l'astronauta non sarebbe riuscito a ritornare vivo sulla Terra. Invece andò tutto nel migliore dei modi, Gagarin atterrò perfettamente e divenne così il «primo eroe dello spazio». L'impressione e l'emozione furono fortissime a tutte le latitudini e l'opinione pubblica internazionale percepì, forse anche in maniera esagerata, che l'umanità era entrata in

una nuova epoca. Gagarin, dopo un periodo di straordinaria notorietà, tornò a fare il collaudatore aeronautico. La sua vita, purtroppo, finì tragicamente a soli 34 anni, in un incidente di volo, mentre pilotava un nuovo aeromobile militare: il *Mig 15*.

Nel 1962 toccò a **John Glenn** [9], il primo americano in orbita attorno alla Terra. Nel 1967, poi, si verificò il primo tragico incidente, con tre vittime, legato alla conquista dello spazio, che avvenne però a terra durante un test sull'**Apollo 1**.

c) La rivincita degli Stati Uniti

Il sorpasso americano arrivò nell'anno 1969, quando gli Stati Uniti riuscirono a far scendere sul suolo lunare i primi uomini con la famosa missione **Apollo 11**.

La navicella, che faceva parte del **Programma Apollo** (fortemente voluto dal Presidente Kennedy), il quale condusse più volte astronauti statunitensi sulla Luna, partì il **16 luglio 1969** da **Cape Kennedy** (l'attuale **Cape Canaveral**), in Florida, e fu portata in orbita dal potente razzo *Saturno V*. Sebbene durante le procedure di avvicinamento alla superficie del nostro satellite i computer di bordo della capsula di allunaggio *Eagle* segnalassero diversi allarmi, al centro di controllo di Houston si decise di proseguire; e infatti, la discesa vera e propria, nella regione del *Mare della Tranquillità*, si concluse nel migliore dei modi. Il 21 luglio 1969 **Neil Armstrong**, il comandante della missione, ed **Edwin Aldrin**, il pilota, misero piede per la prima volta nella storia dell'umanità sul suolo lunare. Il terzo membro dell'equipaggio, **Michael Collins**, restò a bordo del modulo di comando *Columbia*, da cui la navicella *Eagle* si era staccata. Alle 2,56, ora di Greenwich, Armstrong posò il piede sulla Luna, pronunciando una frase che sarebbe rimasta famosa: «Un piccolo passo per un uomo, un grande passo per l'umanità». Gli astronauti piantarono sulla Luna la bandiera degli Stati Uniti, effettuarono una «passeggiata» di due ore e mezzo e il 24 luglio ritornarono, da eroi, sulla «madre Terra», con il «prezioso» carico di 22 kg di rocce lunari. L'ultimo allunaggio umano, con l'**Apollo 17**, risale al 1972.

Gli USA si erano presi una bella «rivincita» simbolica, in un campo che nel frattempo dominavano sempre più, inaugurando, nel ventesimo anniversario del primo volo di Gagarin, il 12 aprile 1981, quello che doveva essere il progetto rivoluzionario dell'esplorazione spaziale, il programma di navette spaziali riutilizzabili **Shuttle**. In realtà, lo *Shuttle* non mantenne del tutto le promesse principali di «riutilizzabilità», «economicità» e «sicurezza», creando notevoli problemi all'*Agenzia Spaziale Statunitense*, la **NASA**, che aveva incentrato l'intero suo programma, fino al 2010, proprio attorno a esso, e nel 1986, in seguito a un evento drammatico — lo *Shuttle Challenger* esplose pochi secondi dopo il decollo, causando la morte di tutti i componenti dell'equipaggio — le missioni relative a quel programma vennero immediatamente sospese.

d) Dalla «competizione» alla «cooperazione»

Verso la fine degli anni Ottanta del XX secolo, mentre la «guerra fredda» allentava la sua morsa e l'Unione Sovietica iniziava a sgretolarsi come confederazione di Stati comunisti, si fece strada l'idea di «collaborare» invece che «competere». Nacquero così le prime **imprese congiunte** russo-americane, soprattutto nella costruzione e nell'impiego delle **stazioni spaziali**, costruite pezzo dopo pezzo in orbita. Anche l'Europa si unì a questo progetto che portò, verso la fine del Novecento, alla creazione della *ISS*, la **Stazione Spaziale Internazionale**, nella quale astronauti di varie nazionalità vivono assieme per periodi più o meno lunghi (da pochi giorni ad

[9] La missione *STS-95* dell'astronave **Discovery** portò nuovamente nello spazio John Glenn, trentasei anni dopo il suo primo e storico viaggio: nel 1998, a 77 anni, il leggendario astronauta tornò a bordo di una navicella per compiere 144 giri intorno al nostro pianeta in poco meno di 9 giorni.

alcuni mesi), per condurre ricerche ed esperimenti nei campi dell'astronomia, della fisica, della biologia e della chimica.

Nei primi anni dell'attuale secolo, l'esplorazione spaziale, che fino a quel momento era stata un'impresa sostenuta «direttamente e soltanto» dagli Stati nazionali, per gli enormi investimenti di risorse richiesti, ha dato inizio a due fenomeni completamente nuovi:

- il **turismo spaziale**;
- lo sviluppo e la messa in orbita dei primi **vettori da parte di privati**.

e) Perché andare nello spazio

In tanti, soprattutto all'inizio della corsa allo spazio, si sono interrogati sulla necessità di questo tipo di missioni, considerati i molteplici problemi concreti da risolvere qui sulla Terra.

Tali discussioni partivano dalla constatazione del fatto che l'esplorazione spaziale richiede:

- lo stanziamento d'ingenti somme di denaro;
- lo sforzo di migliaia di tecnici specializzati e scienziati;
- il dispiegamento di grandi mezzi.

Oggi, discussioni del genere sembrano superate anche perché, a differenza di allora, l'aspetto militare connesso a quello scientifico — con lo scontro fra le due superpotenze geopolitiche, USA e URSS, sempre sullo sfondo — è, almeno in parte, venuto meno. Inoltre, i benefici che derivano dall'utilizzo di informazioni e dati scaturiti dalle attività spaziali in campo civile sono oramai sotto gli occhi di tutti.

Infinite, infatti, sono le *applicazioni* nella nostra vita quotidiana:

- abbiamo, ogni giorno, *previsioni del tempo* sempre più precise grazie ai satelliti meteorologici;
- i satelliti per le telecomunicazioni ci permettono di *telefonare*, d'inviare *trasmissioni audio-video* in tutte le aree del mondo e con estrema precisione;
- le nostre automobili e gli *smartphone* di ultima generazione sono dotati di un *sistema di navigazione* che si basa sugli apparati satellitari per il rilevamento della posizione, come ad esempio il **GPS** (*Global Positioning System*, ovvero «Sistema di Posizionamento Globale»).

Tutte sono state sviluppate a partire dalle tecnologie messe a punto per la realizzazione delle varie imprese spaziali.

Non dimentichiamo, poi, che il desiderio di conoscere ed esplorare costituisce, probabilmente, sin dalla notte dei tempi, una delle pulsioni più positive e insopprimibili dell'umanità.

f) Dove siamo stati finora

La **Luna** ha rappresentato il primo obiettivo dell'esplorazione spaziale; posta a poco più di 300.000 km da noi è stato quasi naturale considerarla la prima meta.

Ma l'attenzione degli scienziati si è rivolta, nel corso degli anni, all'intero Sistema Solare e a tutti i pianeti che lo compongono.

Nel 1961, agli albori di questa straordinaria «avventura», dopo il successo della mitica impresa di Gagarin, i Sovietici partirono alla conquista di **Venere**, su cui nel 1966 scese **Venera 3**, mentre, attorno alla prima metà degli anni Settanta, gli Americani raggiunsero **Mercurio** con la sonda **Mariner 10**, ad oggi l'unico mezzo spaziale a giungere sul più piccolo pianeta (quello più vicino al Sole) del nostro sistema planetario. Mercurio è anche l'obiettivo di una delle missioni europee più interessanti del prossimo futuro, la quale porta il nome di «Giuseppe «Bepi» Colombo», un grande ingegnere aerospaziale italiano, morto nel 1984, che lavorò a lungo per la NASA.

Le prime foto di **Marte** sono americane e risalgono al 1965, ma già nel 1976 l'esplorazione del «pianeta rosso» sembrava essere giunta al culmine, con lo sbarco delle **Viking 1 e 2**: si trattava di

sonde costituite da un *orbiter*, destinato a restare, appunto, in orbita attorno al pianeta, e da un *lander*, preposto all'atterraggio. La missione, estremamente fruttuosa, terminò nel 1981. Questo è anche l'anno in cui gli USA lanciarono il primo *Space Shuttle* guidato da un equipaggio umano.

Intanto, nel 1979, cominciarono a partire le prime spedizioni verso i *pianeti giganti* più lontani: le sonde americane **Voyager 1** e **2** prima sorvolarono **Giove**, poi vennero indirizzate alla volta di **Saturno**. Successivamente, la *Voyager 1* fu spinta verso l'esterno del Sistema Solare e si prevede che tra circa 40.000 anni si troverà in prossimità dell'Orsa Minore; la *Voyager 2*, invece, proseguì in direzione di Urano, fino a incontrare, nel 1989, anche Nettuno e continuare poi il suo incredibile viaggio oltre l'orbita di Plutone.

Nel 1986, i Sovietici lanciarono la stazione orbitante **Mir** e, quasi in contemporanea, la sonda **Giotto** dell'**Agenzia Spaziale Europea** conquistò un obiettivo prestigioso, fotografando da vicino, come mai era accaduto prima, il nucleo della **Cometa di Halley**.

Sempre negli anni Ottanta, *Giove* e *Saturno* furono raggiunti dalle sonde **Pioneer**, ma saranno oggetto di missioni ripetute. Nel 1995, infatti, un'altra importante sonda americana, la **Galileo**, arrivò nelle vicinanze del «pianeta gigante».

Nel 1997, gli Stati Uniti conquistarono Marte: la navicella **Pathfinder** scese sulla superficie marziana, regalandoci immagini straordinarie e indimenticabili.

La sonda **Cassini-Huygens** raggiunse l'orbita di Saturno nel 2004 e l'anno successivo si ebbe il primo «atterraggio», effettuato dal modulo *Huygens*, su un satellite naturale diverso dalla Luna, ovvero **Titano**, il più grande tra quelli che gravitano attorno al meraviglioso «pianeta degli anelli».

Altre missioni spaziali studiarono da vicino alcune **comete**, riportando a terra campioni di materiale della coda, e la superficie dell'**asteroide Eros** venne ripresa in dettaglio dalla sonda **NEAR Shoemaker** della NASA, la quale ha orbitato attorno al piccolo corpo celeste, situato tra Marte e Giove, per circa un anno.

Nel futuro prossimo si prevede, fra l'altro, l'arrivo di una missione europea (**Rosetta**) su una cometa, per analizzarne il suolo, e di una statunitense su **Plutone**, il lontanissimo «pianeta nano», l'unico del Sistema Solare finora mai esplorato. Quest'ultimo veicolo spaziale, con i suoi sofisticati strumenti, potrà fornire ulteriori e più aggiornate informazioni relative a Urano e Nettuno, di cui possediamo finora solo i dati ottenuti dalle sonde *Voyager* e *Pioneer* nella loro corsa verso i confini del Sistema Solare.

g) Ma come si fa ad andare in orbita?

Uno dei più importanti problemi del volo spaziale è rappresentato dalla difficoltà di staccarsi dalla Terra, ossia di «vincere» l'*attrazione gravitazionale*, applicando al mezzo cosmico una forza contraria e maggiore. Per «scappare» dall'attrazione gravitazionale terrestre occorre raggiungere la velocità di almeno 11,18 km/s, un valore chiamato, proprio per questo, «*velocità di fuga*». Ovvio, pertanto, che il peso «a terra» del velivolo da portare in orbita sia un aspetto fondamentale nella fase di lancio: più grande è il vettore, più elevata deve essere la spinta per raggiungere la «velocità di fuga» e, dunque, più potenti i razzi vettori. Purtroppo, più questi ultimi sono potenti, maggiormente pesano al momento del lancio, a causa della superiore quantità di carburante che devono trasportare. Fortunatamente, però, a mano a mano che ci si allontana dal nostro pianeta, l'attrazione di questo, che dipende molto dalla distanza, diminuisce rapidamente: è essenziale, quindi, una vigorosa spinta all'inizio del volo per «staccare» il mezzo dal suolo e ottenere la velocità di fuga, ma in seguito la forza propulsiva necessaria per continuare la missione diventa minore perché l'astronave pesa sempre meno, dato che la gravità via via si riduce. Per questo motivo i satelliti artificiali, e in genere tutti i mezzi spaziali, hanno motori piccolissimi, capaci di spinte modeste, ma sono condotti in orbita da possenti razzi vettori, capaci di spinte particolarmente energiche, che vengono utilizzati solo nelle primissime fasi del decollo.

h) L'Italia, terzo Paese del mondo a lanciare un satellite

Unione Sovietica e Stati Uniti sono stati i primi due Paesi a lanciare con successo un satellite nello spazio, l'Italia è stata il terzo. Nel 1964, dalla **Base San Marco**, in Kenya, partiva, su un razzo vettore statunitense, il primo satellite italiano, il **San Marco 1**. Forse oggi quest'ottimo risultato ottenuto dalla nostra Nazione è poco noto, ma vale la pena di ricordare che, grazie all'illuminata ed energica opera di **Luigi Broglio**, il padre delle attività spaziali italiane, dai primi anni Sessanta del secolo scorso al 1988, data dell'ultimo lancio, la nostra piattaforma in Kenya ha giocato un ruolo considerevole a livello internazionale, con 23 lanci complessivi, tutti andati a buon fine, di cui 6 riguardanti satelliti della serie *San Marco*.

Il *Centro Spaziale Italiano* è situato nei pressi di Malindi, sull'Oceano Indiano; realizzato dall'Università di Roma «La Sapienza» e gestito fin dagli inizi mediante il Centro Ricerche del Progetto San Marco, esso è composto da un'unità marina per il lancio e da una terrestre per la ricezione dati. Ancora oggi, questa base, intitolata proprio a Luigi Broglio e gestita ora dall'**Agenzia Spaziale Italiana** (ASI), è fra le meglio attrezzate al mondo e conserva una notevole importanza — per la sua posizione geografica e strategica — nel garantire un semplice ed efficace controllo di volo sui satelliti artificiali. Proprio come stazione di rilevamento e verifica viene utilizzata anche nell'ambito di varie collaborazioni internazionali, ad esempio con gli Stati Uniti e la Cina.

9 I propellenti dei vettori spaziali

Il **propellente** è un *materiale usato per ottenere la propulsione di un oggetto*. Prevalentemente, si tratta di sostanze chimiche a elevato potere calorifico, ovvero in grado di sviluppare rapidamente grandi quantità di calore nella combustione e conseguentemente di gas ad alta pressione, i quali, a loro volta espulsi, attraverso un ugello, provocano una spinta in avanti secondo il *principio di azione e reazione*.

In riferimento allo stato fisico, un *propellente* può essere:

- **solido**;
- **liquido**;
- **gassoso**;
- **ibrido** (in quest'ultimo caso è composto da *un solido e un liquido* o da *un liquido e un gas*).

I comuni propellenti chimici consistono in un *combustibile* che brucia in presenza di un *ossidante*. Essi trovano largo impiego nei settori balistici e pirotecnici, nonché in quelli dell'*aeronautica* e dell'*astronautica*. Negli ultimi due, il termine «propellente» si riferisce ai prodotti chimici che vengono immessi nei serbatoi di un veicolo/veivolo prima dell'utilizzo, con esclusione dei gas atmosferici. Negli **aerei** i propellenti sono *liquidi*, mentre nei **motori a razzo** possono essere *solidi* o *liquidi*.

Per avviare la *combustione* i propellenti hanno bisogno di un sistema di accensione, che non è necessario nel caso dei cosiddetti «**propellenti ipergolici**», per i quali l'accensione stessa avviene spontaneamente nel momento in cui il *combustibile* e il *comburente* (in questo caso si deve parlare più propriamente di *ossidante*) vengono a contatto. Il propellente ipergolico più adoperato attualmente è dato dalla combinazione di idrazina e tetraossido di diazoto. I motori a razzo di taglia più piccola, usati per il controllo d'assetto delle sonde spaziali, possono anche sfruttare solamente l'espansione di gas inerti, come l'azoto o l'elio.

I **propellenti solidi** vengono adoperati in genere nei *booster* (propulsori ausiliari per il decollo di veicoli spaziali) dei primi stadi di un razzo. Il *vantaggio* offerto è la grande semplicità costruttiva del motore che li utilizza, la quale si traduce in economicità e affidabilità, a differenza di quanto accade per i motori a *propellente liquido*. Inoltre, essi possono essere stivati per lungo

tempo, senza che sia necessaria una manutenzione particolare, e hanno un'elevata «prontezza di funzionamento» (il «transitorio di accensione», rispetto ai motori a propellente liquido, risulta decisamente più rapido ed è pari a circa 150ms). Gli *svantaggi* sono, invece, collegati all'impossibilità di spegnere e poi riaccendere, di modulare la spinta e di riadoperare un motore che si serva di questo tipo di carburante.

I **propellenti liquidi** (spesso impiegati in razzi «ibridi», dove sono combinati con propellenti solidi o gassosi) hanno il vantaggio di raggiungere una densità relativamente elevata, cosa che permette di mantenere il volume dei serbatoi dei razzi piuttosto contenuto, con un corrispondente aumento del rapporto di massa.

I razzi che ne fanno uso sono suddivisi, a seconda del numero di propellenti impiegato, in razzi:

- a *monopropellente*;
- a *bipropellente*;
- a *tripropellente*.

I secondi utilizzano, generalmente, un carburante e un ossidante liquidi, come l'*idrogeno liquido* e l'*ossigeno liquido*. A differenza dei gas, un tipico *propellente liquido* ha una densità simile all'acqua, pari a circa 0,7-1,4 g/cm³ (tranne l'idrogeno liquido, che ha una densità molto inferiore), e richiede una modesta pressione per evitare la «vaporizzazione». Questa combinazione di densità e bassa pressione consente l'impiego di serbatoi leggeri: approssimativamente, con un peso dell'1% del contenuto per i propellenti densi e del 10% per l'idrogeno liquido (a causa della bassa densità e della massa per l'isolamento richiesto).

Il loro utilizzo, però, può essere associato a taluni inconvenienti, che possiamo qui di seguito elencare:

- poiché il propellente costituisce una parte molto grande della massa del veicolo, il baricentro si sposta in modo significativo verso il retro mentre il propellente medesimo viene consumato, con la conseguente possibilità di perdere il controllo;
- sono soggetti a «sciabordare» all'interno dei serbatoi e necessitano spesso di propulsori per evitare il risucchio dei gas nel motore durante l'accensione. Inoltre, essi possono fuoriuscire, formando miscele esplosive;
- i propellenti denominati «*criogenici*» (ossia che hanno bisogno di essere conservati a temperature estremamente basse per mantenere lo stato liquido), come l'*ossigeno liquido*, raffreddano il vapore acqueo atmosferico in cristalli di ghiaccio molto duri, che possono danneggiare o bloccare i sigilli e le valvole, provocando perdite e altri fastidiosi contrattempi. Per evitare tali problematiche sono previste elaborate procedure di rimozione del vapore dal sistema;
- i razzi tendono a essere molto complessi (con maggiori possibilità di malfunzionamento) e richiedono meticolosi preparativi prima del lancio (il che li rende poco pratici rispetto a quelli a propellente solido);
- quando sono adoperati all'interno dell'atmosfera, deve essere mantenuta una «pressurizzazione» positiva dei serbatoi, per evitare il loro collasso.

1 La composizione

L'atmosfera è la massa gassosa che circonda il globo terrestre. Essa svolge un'azione fondamentale per il nostro pianeta, in quanto contiene degli elementi indispensabili alla vita ed è interessata da fenomeni fisico-chimici che influiscono sull'ambiente naturale terrestre. Mediamente è costituita (in volume) dal 78,08% di **azoto**, dal 20,95% di **ossigeno**, dallo 0,93% di **argon**, dallo 0,03% di **anidride carbonica** e da ridottissimi quantitativi di altri gas, quali **idrogeno**, **ozono**, **ammoniaca**, **ossido di carbonio**, **elio**, **neon**, **kripton**, **xenon**, **anidride solforosa**. Contiene inoltre polveri, sali e prodotti di combustione (che insieme formano il **pulviscolo atmosferico**), e **vapor d'acqua** in quantità fortemente variabile (ma sempre inferiore al 5%) che diminuisce con il procedere verso l'alto e verso i Poli.

2 Le suddivisioni

Generalmente l'atmosfera viene distinta, in base alla temperatura, in **strati** verticali separati da fasce di discontinuità definite **pause**.

Lo strato a diretto contatto con la superficie terrestre prende il nome di **troposfera** (*sfera dei cambiamenti*): è la parte più densa dell'atmosfera, in cui si concentrano il 90% dell'intera massa gassosa e il 75% di vapor d'acqua ed in cui si verificano tutti i fenomeni meteorologici più vistosi che interessano la superficie terrestre. La sua temperatura diminuisce in ragione dell'altezza con una media di 0,6°C ogni 100 metri, che rappresenta il *gradiente termico verticale* della troposfera.

Ad un'altezza media di 10-12 km (in realtà a circa 8 km, in corrispondenza dei Poli, e a circa 17 km, sulla fascia tropicale) la temperatura [1] raggiunge un minimo di circa -70°C: si ha il passaggio alla **tropopausa**, uno strato particolarmente turbolento con moti verticali ed orizzontali che raggiungono velocità intorno ai 400-500 km/h, dove soffiano i venti più forti e si trovano le nubi più alte, che danno origine a precipitazioni e tempeste.

Segue la **stratosfera**, che si estende fino a 40-50 km d'altezza: la sua composizione è simile a quella dello strato sottostante, ma i gas si presentano più rarefatti. Diminuiscono pulviscolo atmosferico e vapor d'acqua, la temperatura si mantiene costante fino agli ultimi chilometri per poi aumentare fino a salire sopra lo zero in corrispondenza della **stratopausa**.

Tale incremento corrisponde ad un livello ricco di O₃, definito **ozonosfera**, prodotto dall'azione dei raggi solari che scindono le molecole di ossigeno (O - O) in atomi singoli che si uniscono a molecole rimaste intatte (O + O₂). Tale strato gioca un ruolo fondamentale nei fenomeni di irraggiamento del nostro pianeta: l'ozono assorbe gran parte delle radiazioni solari ultraviolette, che se arrivassero in forti dosi creerebbero gravi danni per la vita degli organismi. Anche la stratosfera è caratterizzata da una forte circolazione d'aria, con venti di intensità notevole che a volte si trasformano in violente *correnti a getto*.

A partire dalla stratopausa inizia la **mesosfera**, strato relativamente caldo con temperatura in diminuzione verso l'alto, in cui si verifica il fenomeno della combustione delle meteoriti. I gas risultano estremamente rarefatti e abbondano quelli leggeri rispetto a quelli pesanti.

[1] Si tenga presente che le temperature cui si è fatto e si farà riferimento sono una misura dell'attività cinetica delle particelle.

La **mesopausa** segna un nuovo minimo di temperatura (intorno a -80°C) e il passaggio alla **termosfera**. Quest'ultimo strato dell'atmosfera parte da un'altezza di circa 80 km ma non ha un limite superiore (la **termopausa** è, quindi, un riferimento relativo): verso l'alto diminuisce la densità mentre aumenta considerevolmente la temperatura, fino a raggiungere le migliaia di gradi. La presenza di una grande quantità di particelle con cariche elettriche (ioni), prodotte dall'assorbimento di radiazioni solari da parte dei gas, spiega perché tale strato venga anche detto **ionosfera**. Questo fenomeno di ionizzazione è l'origine del potere riflettente della zona più alta dell'atmosfera rispetto alle onde radio, che emesse dalla Terra tornano poi indietro ad essa e vengono captate da apparecchi riceventi. Sempre un processo di ionizzazione è alla base del fenomeno della *luminescenza*, riscontrabile al crepuscolo, e delle *aurora polari*, tipiche delle alte latitudini.

La suddivisione dell'atmosfera finora illustrata, come si è detto, segue le variazioni di temperatura. Un'ulteriore distinzione può essere, invece, fatta in base alla composizione chimica: in tal caso l'atmosfera risulta divisa in **ommosfera**, che si estende fino alla mesopausa ed ha composizione omogenea e peso molecolare costante, e in **eterosfera**, che si estende fino alla fine dell'atmosfera ed ha composizione e peso molecolare estremamente variabili.

Per **esosfera** (*sfera esterna*), invece, si intende la regione nella quale la fuga delle molecole costituenti l'atmosfera terrestre è particolarmente significativa: il suo limite inferiore è ancora impreciso, ma probabilmente si aggira tra i 300 e i 1.000 km di altezza, mentre il suo limite superiore (intorno ai 1.500 km) segna il passaggio tra l'atmosfera e lo spazio interplanetario.

Anche la **magnetosfera** rappresenta uno strato che circonda la Terra: essa costituisce l'estensione verso l'esterno, nello spazio, del campo magnetico terrestre e comprende tutta l'atmosfera, spingendosi fino ad un'altezza compresa tra i 64.000 e i 34.000 km. Nel 1958 i satelliti Explorer I e III individuarono la presenza, all'interno della magnetosfera, di due fasce di particelle atomiche, ad intensa radioattività, definite poi *fasce di radiazione di Van Allen*, dal nome del fisico che per primo le descrisse: la prima è localizzata ad un'altezza di 2.600 km e contiene soprattutto protoni; la seconda, posizionata tra i 13.000 e i 19.000 km, è composta da elettroni e protoni provenienti dal Sole. La magnetosfera funge per la Terra da schermo protettivo contro il vento solare, ad altissimo potere radioattivo, mentre le fasce di Van Allen svolgono un'azione riflettente nei confronti dei segnali radio emessi dai trasmettitori terrestri.

Di tutta la radiazione emessa dal Sole, pari a circa 5.2×10^{24} kilocalorie al minuto, soltanto mezzo miliardesimo arriverebbe al limite superiore dell'atmosfera, cioè un flusso di 2 calorie al minuto per cm^2 (*radiazione solare*). Tale valore, meglio definito come 2 langley al minuto ($1 \text{ langley} = 1 \text{ cal/cm}^2$), corrisponde alla **costante solare**. Di questa radiazione soltanto il 51% arriva poi effettivamente alla superficie terrestre (*radiazione globale o insolazione*). Tuttavia il riscaldamento dell'atmosfera è imputabile solo in misura ridotta alla radiazione solare diretta, dipendendo in realtà dall'irraggiamento della Terra stessa. L'atmosfera dunque si comporta come una serra, lasciando passare in entrata l'energia solare e intercettando in uscita le radiazioni terrestri: in questo modo può mantenersi calda (*effetto serra*).

3 La temperatura

La *temperatura dell'aria* (ci si riferisce chiaramente alla troposfera, in quanto è lo strato dell'atmosfera che influenza direttamente il clima sulla superficie terrestre) dipende in primo luogo dall'**altitudine**, decrescendo con l'altezza secondo il già richiamato gradiente termico di $0,6^{\circ}\text{C}/100 \text{ m}$, a meno di fenomeni di inversioni termiche, legati generalmente ad un più rapido irraggiamento terrestre, che porta ad un raffreddamento notturno più intenso degli strati d'aria a diretto contatto con il suolo. Ma l'andamento della temperatura è legato anche ad altri fattori: l'**angolo d'incidenza dei raggi solari** con la superficie terrestre, che dipende a sua volta

dalla latitudine, dalle stagioni, dall'ora del giorno e dall'esposizione ed inclinazione del terreno (quanto maggiore è l'angolo tanto maggiore è la quantità di calore per unità di superficie); la **presenza di terre o di acque** (le prime hanno un minore calore specifico, raggiungono presto alte temperature, ma solo negli strati più superficiali, e rapidamente si raffreddano; le seconde invece hanno un maggiore calore specifico, si riscaldano lentamente, ma fino in profondità, e altrettanto lentamente si raffreddano); la **copertura vegetale** (quanto più abbondante e rigogliosa è la vegetazione tanto più si ridurranno i valori estremi di temperatura).

Questi fattori meteorologici principali si combinano in diversi modi, con il risultato che la temperatura dell'aria varia non solo da un luogo all'altro della Terra ma anche nel tempo. Così in uno stesso posto, nel corso di una singola giornata, si registreranno temperature diverse che oscilleranno tra un massimo ed un minimo: la loro differenza è l'*escursione termica giornaliera*. Essa tende a diminuire con il crescere della latitudine (ai tropici raggiunge i 40°C) ed è generalmente maggiore d'estate che d'inverno. La differenza invece tra la temperatura media del mese più caldo e quella del mese più freddo in un anno si definisce *escursione termica media annua*. È minima all'Equatore (2°-4°C) ed aumenta verso i Poli (40°-50°C), mentre a parità di latitudine è maggiore per i continenti rispetto agli oceani e minore nelle zone piovose rispetto a quelle aride. Come l'escursione termica giornaliera, essa tende a diminuire con l'altitudine.

Per rappresentare sulle carte la distribuzione delle temperature si usano delle linee curve, dette **isoterme**, che uniscono tutti i punti in cui è stata registrata la stessa temperatura.

4 Pressione e venti

L'altra proprietà fondamentale dell'aria è la *pressione*. Anche l'atmosfera ha un peso e il suo rapporto con la superficie su cui agisce è la **pressione atmosferica**. Essa al livello del mare, alla latitudine di 45° e alla temperatura di 0°C è in equilibrio con una colonnina di mercurio di sezione pari a 1 cm² ed altezza di 760 mm. Questo valore rappresenta l'unità di pressione, è definito *pressione normale* o *atmosfera* e corrisponde a circa 1033 g/cm². La pressione atmosferica diminuisce con l'**altitudine**, con l'aumento della **temperatura** e con l'incremento della quantità di **vapore acqueo** nell'aria. La distribuzione della pressione atmosferica viene rappresentata su carte utilizzando le **isobare**, linee che uniscono i punti a uguale pressione.

Le **aree anticicloniche** sono zone di alta pressione, con aria più fredda e secca che si muove verso il basso e verso le zone a bassa pressione, definite **aree cicloniche**, in cui l'aria più calda e umida converge verso l'alto e verso il centro. Le differenze di pressione (*gradiente barico*) generano dunque dei movimenti verticali d'aria ma anche dei flussi orizzontali, paralleli cioè alla superficie terrestre, che si spostano dalle zone anticicloniche a quelle cicloniche e che prendono il nome di **venti** [2].

La *circolazione nella bassa troposfera* non avviene semplicemente lungo la direzione del gradiente barico: le differenti velocità di rotazione della Terra alle diverse latitudini, l'azione della *forza di Coriolis* e l'attrito dell'aria con il terreno modificano sensibilmente la direzione dei venti in superficie.

La circolazione di questi **venti planetari** rappresenta un elemento fondamentale del clima nelle diverse zone del pianeta e può essere descritta, in maniera molto sintetica e schematica, come l'insieme di tre sistemi principali di venti: gli **alisei**, che hanno direzione da est verso ovest (in quanto sono più lenti della velocità di rotazione terrestre) e convergono verso l'Equatore, area di bassa pressione; i **venti occidentali**, che partono dalle zone di alte pressioni tropicali (da cui

[2] I venti vengono individuati e classificati in base alla *direzione* di provenienza e in base alla *velocità*, misurata con l'*anemometro*.

hanno origine anche gli alisei) e muovono appunto verso est, ossia verso le zone di basse pressioni polari (in queste zone la velocità dell'aria eguaglia quella di rotazione terrestre); i **venti orientali polari**, che provengono dalle zone di alte pressioni polari e convergono anch'essi verso le zone di basse pressioni polari, ma spirando da est. Venti planetari sono considerati anche i **monsoni**: venti periodici a carattere stagionale, localizzati soprattutto in Asia e in Nord America, che esercitano un'azione fortissima sul clima e sulle attività agricole di queste terre. I **monsoni estivi** spirano dall'oceano verso il continente, causando intense precipitazioni, mentre i **monsoni invernali** si muovono dalla terraferma verso il mare, apportando condizioni di tempo sereno ed asciutto.

Venti locali invece sono le **brezze**, che nelle zone costiere soffiano di giorno dal mare verso la costa e di notte dalla costa verso il mare, mentre nelle zone montuose spirano verso monte di giorno e verso valle di notte [3].

La *circolazione nell'alta troposfera* risulta diversa a causa delle variate condizioni di pressione con l'altitudine: già a 5.000 m la pressione equatoriale è più alta rispetto a quella polare e i venti muoverebbero quindi verso il nord. L'azione della rotazione terrestre li devia e fa sì che essi spirino da ovest verso est (**correnti occidentali**), seguendo la direzione dei paralleli. In corrispondenza delle fasce intertropicali il movimento è inverso e si originano dunque **correnti orientali**.

5 Umidità e precipitazioni

La terza e ultima caratteristica fondamentale dell'aria è l'**umidità**, ossia la quantità di vapore acqueo prodotta dalla continua **evaporazione del mare** e, in minima parte, dalla **traspirazione delle piante**. L'evaporazione dipende prevalentemente dalla temperatura, aumentando al crescere di questa, ma è legata anche alla velocità del vento e alla secchezza dell'aria sovrastante; la traspirazione è influenzata dall'insolazione e varia a seconda delle specie vegetali.

L'**umidità assoluta** è la quantità di vapore acqueo (in grammi) presente in un'unità di volume di aria (1 m^3), in un preciso momento, in un determinato punto dell'atmosfera. Nelle regioni equatoriali è intorno ai 20-25 g/m^3 , in quelle polari è inferiore ai 2 g/m^3 . L'**umidità relativa**, espressa in percentuale, è invece il rapporto tra la quantità di vapore acqueo effettivamente presente in uno specifico volume d'aria e la massima quantità che potrebbe esservi contenuta alla stessa temperatura (ossia il rapporto tra l'umidità assoluta e quello che viene definito *punto di saturazione*).

Quando un volume d'aria è saturo di vapor acqueo, si può assistere ad un passaggio o allo stato liquido, con un processo detto di **condensazione**, o allo stato solido, con il fenomeno della **sublimazione**. Questi due fenomeni sono molto frequenti nell'atmosfera e spiegano l'origine delle *nubi*, insieme di goccioline liquide di diametro intorno a 1/100 di mm che «condensano» attorno a dei nuclei (particelle di sali, polveri, ioni etc.) e galleggiano nell'aria. I tipi di nubi sono raggruppati in primo luogo secondo la forma (nubi stratiformi e nubi cumuliformi), quindi in base all'altitudine.

Le **nubi stratiformi** ricoprono aree ampie e risultano sottili rispetto allo sviluppo orizzontale. A seconda dell'altitudine a cui si trovano, si distinguono in: *strati* (nubi stratiformi, dense, di colore grigio scuro, che preannunciano pioggia), *nembostrati* (strati con precipitazioni in corso) e *stratocumuli* (cumuli di nubi grigiastre che generalmente preludono a tempo buono o a schiarite), entro i 2.000 m dal suolo; *altostrati* (coperture nuvolose grigiastre e continue, associate generalmente a cattive condizioni atmosferiche) e *altocumuli* (masse nuvolose distinte, di

[3] Accanto a questi *venti locali periodici*, esistono anche *venti locali variabili*, che risultano fortemente condizionati da fattori locali. Tra i tanti che interessano il territorio italiano ricordiamo: la *bora*, freddissima e asciutta, che spira dalle Alpi Dinariche verso l'Adriatico; il *maestrale*, freddo e secco, che giunge nel Tirreno settentrionale dall'interno della Francia; il *libeccio* e lo *sciocco*, caldi e umidi, che arrivano dall'Africa.

diverse forme, caratteristiche di tempo buono) tra i 2.000 e i 6.000 m di altezza; *cirri*, nubi sottili, a ciuffi, composte da cristalli di ghiaccio, con forma continua ed omogenea (*cirrostrati*) o a globuli (*cirrocumuli*), per altezze tra i 6.000 e i 12.000 m. Anche la **nebbia** appartiene alla famiglia delle nubi stratiformi e si caratterizza per la forte vicinanza al suolo.

Le **nubi cumuliformi** mostrano un'altezza considerevole, pari, se non maggiore, rispetto allo sviluppo orizzontale. Si distinguono i *cumuli*, masse nuvolose bianche, con la classica forma a «cavolfiore», tipiche di bel tempo, e i *cumulonembi*, nubi scure di enormi dimensioni, da cui hanno origine i temporali.

Quando le goccioline delle nubi che sono più spesse verticalmente ed hanno la base a basse quote si fanno troppo grandi e pesanti, hanno luogo le **precipitazioni**. Esse consistono in **pioggia**, in **neve** (fiocchi di cristalli di ghiaccio formatisi direttamente dal vapor d'acqua) e in **grandine** (grani di ghiaccio a struttura concentrica, con diametri compresi tra 0,5 e 5 cm). La loro distribuzione sulla superficie terrestre è estremamente diseguale e variabile: generalmente per uno studio scientifico e valido di tali fenomeni si prendono in considerazione osservazioni e misure relative a periodi di tempo sufficientemente ampi, almeno di 30-50 anni [4]. La rappresentazione sulle carte avviene usando le **isoiete**, linee curve che uniscono tutti i punti in cui è stata misurata la stessa quantità di pioggia media in un anno (*isoiete annue*) o in un mese (*isoiete mensili*).

Le variazioni stagionali e mensili costituiscono il **regime pluviometrico** di una località. A scala mondiale si distinguono: il *regime equatoriale*, con piogge tutto l'anno e due massimi di precipitazione equinoziali; il *regime subequatoriale*, con due stagioni umide alternate a due secche; il *regime tropicale*, con un solo periodo piovoso solstiziale ed un periodo asciutto; il *regime monsonico*, con la stagione estiva piovosa e l'invernale asciutta; il *regime mediterraneo*, con l'estate secca e l'inverno piovoso; il *regime continentale*, con piovosità tutto l'anno, ma soprattutto d'estate; il *regime polare*, con precipitazioni prevalentemente nevose, massimamente concentrate in estate e in autunno.

L'insieme delle condizioni atmosferiche (quali la temperatura, la pressione, i venti, l'umidità, la nuvolosità, le precipitazioni) in un certo luogo e in un determinato momento costituisce il **tempo atmosferico** [5]. La scienza che si occupa dello sviluppo e dell'evoluzione delle perturbazioni atmosferiche (che nascono dagli spostamenti e dalle interazioni di cicloni ed anticicloni) e dei movimenti delle masse d'aria (l'incontro tra masse d'aria fredda e masse d'aria calda lungo i ben noti **fronti** è anch'esso causa di perturbazioni) è la **meteorologia**, che negli ultimi decenni ha registrato dei progressi straordinari: grazie ai satelliti, infatti, le previsioni meteorologiche sono sempre più precise. La conoscenza del tempo atmosferico costituisce per l'uomo un elemento fondamentale per poter difendere dalle intemperie se stesso e le proprie attività. Eventi atmosferici estremi, come le alluvioni, le trombe d'aria, i cicloni tropicali, il gelo, le intense precipitazioni nevose, spesso rappresentano fonti di pericolo per l'incolumità dell'uomo e quasi sempre apportano danni enormi alle attività agricole. Per converso, l'irrigazione artificiale, l'uso di fasce boschive come frangivento e le colture nelle serre rappresentano esempi emblematici del tentativo di difendere l'ambiente antropizzato dalle intemperie.

[4] Lo strumento di misurazione della quantità di pioggia è il *pluviometro*.

[5] È forse opportuno precisare che il tempo atmosferico (o meteorologico) è cosa ben diversa dal *clima*, che, invece, come si vedrà, indica il complesso di condizioni meteorologiche caratteristiche di una certa area per un certo periodo di tempo, condizioni che si succedono di anno in anno con modalità più o meno analoghe.

1 Il ciclo idrologico

L'**idrosfera** è la parte della crosta terrestre costituita da oceani, mari, laghi, fiumi, nevi, ghiacciai e acque sotterranee. Essa è in stretta relazione con l'atmosfera, in quanto tutte le acque del nostro pianeta sono in circuitazione continua da un ambiente all'altro, attraverso un insieme di fenomeni definiti complessivamente *ciclo idrologico*.

Il motore è l'energia solare. Il calore infatti provoca l'**evaporazione** delle acque degli oceani e delle terre emerse verso l'atmosfera dove il vapore acqueo, attraverso la **condensazione**, origina le nubi e quindi cade sulla superficie terrestre sotto forma di pioggia o neve. Una parte dell'acqua delle **precipitazioni** penetra nel sottosuolo (*deflusso profondo*) attraverso l'**infiltrazione**, una parte scorre sulla superficie (*deflusso superficiale*), un'altra parte ancora può tornare direttamente all'atmosfera attraverso l'**evaporazione** e la **traspirazione** da parte delle piante. L'acqua del sottosuolo, in continuo movimento, con il tempo si rimette in circolo, riaffiorando in superficie, sfociando negli oceani o alimentando le piante. Grazie al **bilancio idrologico** [1] è possibile quantificare questo ciclo, calcolando la quantità di acqua che passa da un ambiente all'altro attraverso i diversi processi elencati.

2 Gli oceani

Circa il 97% dell'idrosfera è rappresentato dalle *acque oceaniche*, che, con la loro estensione, occupano il 78% della superficie terrestre. Gli **oceani**, al contrario dei continenti, sono una scoperta alquanto recente dell'uomo: risalgono all'inizio degli anni Sessanta, infatti, le prime conoscenze scientifiche in questo campo. Oggi sappiamo che l'ambiente oceanico è costituito da una serie di regioni con caratteristiche molto diverse; partendo dalla riva dei continenti troviamo: la **piattaforma continentale**, che con una debole pendenza (circa il 2‰) raggiunge profondità di circa 200 m; la **scarpata**, che con un'inclinazione media di 4° raccorda, a circa 2.000 m di profondità, la piattaforma con i **fondi oceanici** o **piane abissali**, i quali si estendono fino a profondità intorno ai 6.000 m; infine le **fosse**, depressioni profonde fino ad 11.000 m (è il caso della Fossa delle Marianne, nell'Oceano Pacifico). Ognuna di queste regioni è ricoperta da **sedimenti** [2], più grossolani vicino alla costa, sottilissimi (*fanghi*) alle alte profondità, ed è caratterizzata da strutture tipiche: gli **atolli** (isolotti corallini dei mari tropicali costruiti sugli orli di crateri di vulcani sottomarini estinti) e le **barriere coralline** (formazioni coralligene molto estese, caratteristiche dei mari dell'Australia) della piattaforma continentale, nonché i rilievi delle **dorsali** oceaniche, che si stagliano sulla superficie convessa dei fondi marini.

Componente principale degli oceani è senz'altro l'acqua, la cui caratteristica fondamentale è la **salinità**. Il contenuto medio dei sali è di 35 grammi per litro, ossia del 35‰, ma si va da punte massime del 68‰ nel Mar Rosso o addirittura del 250‰ nel Mar Morto, a valori minimi del

[1] La relazione di riferimento è la seguente: $P = D + ET + I$, dove P = precipitazioni, D = deflusso superficiale, ET = evapotraspirazione, I = infiltrazione.

[2] Si tratta di depositi che si formano dalla disgregazione fisico-chimica delle rocce e che poi, a loro volta, danno origine a rocce sedimentarie.

3,5‰ nel Golfo di Finlandia. I sali principali sono: cloruro di sodio (77,76%), cloruro e solfato di magnesio (10,87% e 4,73%), solfato di calcio e di potassio (3,60% e 2,45%), carbonato di calcio (0,33%) e bromuro di magnesio (0,26%). Insieme ad essi sono presenti alcuni **componenti secondari** in quantità variabili nello spazio e nel tempo, come l'azoto (sotto forma di nitrati, nitriti e sali ammoniacali) e il fosforo (dei fosfati), che vengono utilizzati dagli organismi marini, e **gas disciolti**, tra cui ancora l'azoto, l'ossigeno e l'anidride carbonica.

Anche la **temperatura** ha una sua valenza: mediamente si aggira intorno ai 3,8°C, ma si passa da valori molto diversi anche in zone ugualmente profonde e poste alla stessa latitudine (dai 2°-3°C sul fondo dell'Atlantico ai 13°C del fondo del Mediterraneo); il punto di congelamento è generalmente intorno a -2°C. Le acque marine, con la loro capacità di incamerare calore fino alle alte profondità e di conservarlo a lungo, esplicano un'importantissima *funzione termo-regolatrice* sul clima della Terra.

Salinità e temperatura hanno inoltre un ruolo fondamentale per l'ecosistema marino (uno dei componenti fondamentali della biosfera), che comprende un numero svariato di specie.

Sebbene le classificazioni e le distinzioni nell'ambiente marino presentino molti limiti ed eccezioni, è possibile affermare che tutti gli organismi di questo ecosistema appartengono a una delle seguenti categorie:

- *plancton*;
- *benthos*;
- *necton*.

Con il termine **plancton** si indica l'insieme degli organismi che vivono in sospensione nel mezzo acquoso incapaci di vincere, con movimenti propri, i moti del mare (correnti, onde etc.) e che, pertanto, vengono da esso trasportati passivamente. All'interno del *plancton* si distinguono, tradizionalmente, gli organismi del **fitoplancton** (di genere vegetale) e dello **zooplancton** (di genere animale).

Gli organismi del *plancton* si distinguono da quelli *dotati di movimenti propri*, in grado, dunque, di opporsi ai moti dell'acqua e che vivono in *acque libere* (**necton**) o a *contatto con il fondo* (**benthos**).

a) Onde

Il mare è caratterizzato da una serie di **movimenti** che hanno una notevole importanza sia sull'ambiente naturale sia su alcune attività dell'uomo. Il **moto ondoso** è legato all'azione del vento: la pressione e l'attrito che esso esercita sulla superficie dell'acqua genera delle increspature, che si ingrandiscono fino a formare le onde vere e proprie e che si muovono seguendo orbite circolari.

In un'onda si distinguono una **cresta** (la parte più rilevata) ed un **ventre** (la parte più depressa); i suoi parametri principali sono: la **lunghezza**, ossia la distanza tra due creste successive; l'**altezza**, cioè il dislivello tra la cresta e il ventre; il **periodo**, ovvero il tempo impiegato da due creste successive a passare per lo stesso punto di riferimento; e infine la **velocità**, vale a dire lo spazio percorso nell'unità di tempo dalla cresta (generalmente viene espressa in km/h).

Mentre in mare aperto il moto ondoso non risente della presenza dei fondali e si manifesta quindi unicamente come movimento circolare, senza trasporto di acqua (*onde di oscillazione*), vicino alla costa l'interazione con il fondo fa sì che il moto ondoso risulti più rallentato verso il basso (*onde di traslazione*) e di conseguenza l'onda, che stavolta consisterà in un reale trasporto di acqua, si rovescerà sulla riva (*frangente di spiaggia*), mentre un flutto di ritorno (*risacca*) si muoverà dalla sua parte più bassa diretto verso il largo. Le onde che battono su un ostacolo (si pensi ad un promontorio o ad un isolotto) subiscono quello della **riflessione**, mentre in acque

basse, a causa dell'influenza del fondale, subiscono il fenomeno della **rifrazione**, che le porta ad essere parallele rispetto all'andamento della costa.

b) Maree

Un altro fenomeno tipico del mare, che ha importanti ripercussioni sulla attività di navigazione, ma che, al contrario del moto ondoso, è prevedibile e periodico, è quello delle **maree**: si tratta di oscillazioni regolari del livello delle acque degli oceani, dovute all'azione delle forze gravitazionali della Luna e del Sole, combinate con la forza centrifuga della Terra (causata dal moto di rivoluzione Terra-Luna). Il massimo sollevamento è detto **alta marea**, il massimo abbassamento, invece, **bassa marea**. La differenza tra i due è l'*ampiezza della marea*: in media, per i mari mediterranei essa oscilla intorno a qualche decina di centimetri, mentre per gli altri mari sono frequenti valori più alti (si ricordi la punta clamorosa dei 19,6 m raggiunta nella Baia di Fundy, nella Nuova Scozia).

Durante le *sigizie*, quando Sole, Terra e Luna sono allineati, le azioni gravitazionali si sommano, dando i massimi valori delle maree; al contrario, nelle *quadrature* si verificano le oscillazioni di maree minime. Generalmente, in un giorno lunare si manifestano due oscillazioni complete del mare, ossia due alte e due basse maree (*maree semidiurne*), ma in alcune zone l'alternanza è tra un'alta e una bassa marea (*maree diurne*) o due alte e due basse maree, ma con ampiezza diverse (*maree miste*).

c) Correnti

Altro fenomeno di particolare interesse è quello delle **correnti**. Consiste nello spostamento orizzontale di masse d'acqua, sia superficiale sia profondo, che avviene secondo celle chiuse e separate, in senso orario nell'emisfero boreale e in senso antiorario in quello australe. L'origine è da ricercare nella *forza di Coriolis*, ma non è da sottovalutare l'azione svolta dai venti. La velocità di questi spostamenti è dell'ordine di 1 m/s, mentre le portate si aggirano intorno alle decine di milioni di m³/sec.

La conoscenza dell'andamento delle correnti è fondamentale non soltanto per la loro influenza su alcune attività dell'uomo, quali la navigazione e la pesca, ma anche per l'azione importantissima che questi movimenti del mare esercitano sul clima del nostro pianeta e sulla vita negli oceani. Si ricorda, solo per proporre un esempio classico, la *Corrente del Golfo*: essa ha origine dalle correnti nordequatoriali nell'Oceano Atlantico e si dirige verso l'Europa nordoccidentale, mitigandone il clima grazie all'apporto di acque tiepide.

3 I ghiacciai

Nell'ambito dell'idrosfera, costituita, come si è detto, per il 97% dagli oceani, i **ghiacciai** rappresentano la seconda risorsa idrica della Terra (2%). Per ghiacciai si intendono le masse di ghiaccio, derivate dalla compressione e dalla trasformazione della neve, che scorrono lentamente sulla superficie terrestre. Essi hanno origine nelle zone fredde del globo (zone polari e di alta montagna), ossia al di sopra di una soglia definita **limite delle nevi persistenti**, dove il calore estivo è insufficiente a far fondere le nevi accumulate nell'inverno precedente. Tale soglia si trova a quote molto elevate in prossimità dell'Equatore (4.000-5.000 m) e nelle zone tropicali (6.000 m), mentre si abbassa fortemente già nelle nostre aree alpine (2.400-3.200 m), fino a raggiungere i 300 m in Groenlandia e il livello del mare in Antartide.

Il regime dei ghiacciai dipende da due fattori principali: l'**alimentazione** nevosa e l'**ablazione** dovuta alla fusione. Ghiacciai come gli *inlandsis* groenlandesi si sviluppano nonostante un'alimentazione scarsa, grazie ad un'ablazione molto bassa, mentre nei *ghiacciai di tipo alpi-*

no l'ablazione molto elevata è compensata da un'abbondante alimentazione. Il rapporto fra l'*accumulo* (alimentazione) e la *fusione* (ablazione) rappresenta il **bilancio di un ghiacciaio** e permette di stabilire se esso è in *pregresso* o in *regresso*. Il movimento di queste masse dipende, in primo luogo, dalla gravità e poi, essenzialmente, dalla plasticità del ghiaccio e dai piccoli moti intergranulari; la velocità con cui si compie è generalmente di qualche centinaio di metri all'anno e anche i rapidi ghiacciai groenlandesi non superano gli 8 km/a.

Durante l'ultima glaciazione pleistocenica, la cui acme si manifestò intorno a 20.000 anni fa, i ghiacciai raggiunsero un'estensione di circa 40.000 km², contro i 15.000 attuali. L'inizio della grande fase di riscaldamento si verificò intorno all'8000 a.C., cui fecero seguito alternanze di intervalli freddi e caldi. Il periodo compreso tra il 1500 d.C. e il 1895 d.C. è conosciuto come *piccola epoca glaciale*, caratterizzata da ghiacciai di montagna ben sviluppati. Attualmente ci troviamo, invece, in una fase di forte ritiro dei ghiacciai, che si spiega con la spiccata *tendenza al riscaldamento* del nostro pianeta nel secolo scorso, cui contribuiscono una serie di fattori concomitanti, quali l'intensa *urbanizzazione* (le città costituiscono delle vere e proprie isole di calore), lo *sviluppo industriale* (con il conseguente aumento di produzione di calore), l'uso indiscriminato di *sostanze chimiche* in grado di «bucare» lo strato atmosferico di ozono e l'*effetto serra*.

4 Le acque continentali

Il resto dell'idrosfera è formato dalle **acque continentali** (così dette in contrapposizione alle *acque marine*), che comprendono sia quelle che si trovano ad una certa profondità nel sottosuolo (*acque sotterranee*) sia quelle che scorrono o stagnano sulla superficie emersa della Terra (*acque superficiali*). Le prime si accumulano nelle rocce e/o nei sedimenti permeabili del sottosuolo: percolano in profondità fino a quando non incontrano un terreno impermeabile, che impedisce loro il passaggio. Si forma così un vero e proprio strato d'acqua, che prende il nome di **falda freatica** (o **acquifera**).

La falda è in continua circolazione all'interno delle rocce (tra gli interstizi o le fratture), con movimenti sia verticali che orizzontali (sotto la spinta della gravità, per raggiungere i punti più depressi, in corrispondenza dei quali fuoriesce in superficie). Le oscillazioni verticali del livello di superficie della falda (*livello piezometrico*) sono dovute alle variazioni di alimentazione di acqua nel sottosuolo e alle eventuali perdite. Questi due fenomeni possono essere sia di origine naturale che antropica: si pensi al regime delle precipitazioni, alla presenza di pozzi di drenaggio o alla connessione tra la falda e un fiume o il mare.

Quando la falda taglia la superficie del terreno si forma una **sorgente**. La sua origine è legata alla natura, alla morfologia e alla struttura dei terreni. Le sorgenti hanno un'importanza straordinaria per l'uomo in quanto rappresentano la fonte più accessibile di acqua potabile, che in alcuni casi risulta tuttavia insufficiente a soddisfare tutti i bisogni: spesso l'uomo deve ricorrere ad *opere di drenaggio* profonde (come i pozzi o le gallerie) per captare direttamente la falda acquifera nel sottosuolo, senza contare il caso della *dissalazione* dell'acqua marina.

a) Fiumi

Tra le acque continentali superficiali le più abbondanti sono quelle dei **fiumi**. Un fiume è un corso d'acqua permanente, con regime relativamente costante, che scorre in un alveo, compreso fra argini naturali, con pendenza regolare e non troppo forte. Per **bacino idrografico** si intende, invece, la regione le cui acque (piovane o di fusione di nevi o ghiacciai) affluiscono a un medesimo fiume; è delimitato dalla **linea di spartiacque**, che lo separa da bacini adiacenti.

Le caratteristiche proprie di un fiume sono: la **lunghezza**, che dipende essenzialmente dai caratteri morfologici e geologici dei terreni su cui scorre; la **pendenza**, ossia il rapporto tra il di-

slivello esistente fra la sorgente (origine del corso d'acqua), la foce (sbocco del fiume in un lago, in mare o in un altro corso d'acqua) e la lunghezza del fiume stesso; la **velocità**, estremamente variabile anche all'interno di uno stesso fiume, in quanto influenzata da più fattori (natura del terreno, pendenza, profondità dell'acqua); la **portata**, cioè il volume d'acqua che passa nell'unità di tempo attraverso una sezione trasversale al fiume, espressa in m^3/s ; il **coefficiente di deflusso**, ovvero il rapporto tra la quantità delle precipitazioni e la quantità di acqua incanalata; il **regime**, che rappresenta la variazione di portata in un'unità di tempo che generalmente è un anno.

I fattori sopra elencati rappresentano la «carta d'identità» di un fiume, la cui conoscenza approfondita, seguita dal controllo costante, permette che questa risorsa straordinaria non si trasformi in un elemento a rischio per l'uomo, generando terribili calamità come le alluvioni e le esondazioni che anche recentemente hanno interessato il nostro Paese, con danni ingenti per le popolazioni e per l'ambiente.

b) Laghi

I **laghi** sono accumuli d'acqua generalmente dolce (ma può essere anche salata o salmastra) che occupano cavità o depressioni del terreno alquanto estese e profonde (in base a ciò si distinguono da *paludi* e *stagni*).

La loro **origine** può essere di diversa natura: *glaciale*, quando occupano le conche generalmente circolari scavate dai ghiacciai oppure tratti di valli glaciali (ad esempio, i laghi Maggiore, di Como, di Garda tra la Lombardia e il Veneto); *vulcanica*, se situati in crateri o in caldere di vulcani ormai spenti (i laghi di Bracciano, di Vico e di Bolsena nel Lazio, il lago d'Averno in Campania); *carsica*, se si formano in cavità carsiche come le doline (il lago del Matese in Campania); *tettonica*, in seguito allo sprofondamento di tratti della superficie terrestre (i laghi Vittoria e Tanganika in Africa). Esistono, inoltre, anche laghi di *origine antropica*: si tratta dei *laghi artificiali*, creati, attraverso opere di sbarramento di fiumi (le dighe), per fungere da riserve d'acqua dolce destinata all'agricoltura, all'industria, alle centrali idroelettriche.

Le **caratteristiche** delle acque lacustri sono: la **trasparenza**, inferiore a quella dei mari; la **temperatura**, che in linea di massima diminuisce con la profondità, con la densità massima che si raggiunge a 4°C ; la **salinità**, la quale dipende dalla presenza o meno di un *emissario* (fiume che ne allontana le acque, in contrapposizione con l'*immissario*, che invece le porta), e che comunque è diversa da quella marina, molto più ricca in cloruri.

Le **dimensioni** dei laghi sono estremamente varie: si passa dai piccoli laghetti di montagna al Lago Superiore dell'America del Nord, il quale si estende per più di 82.000 km^2 .

Il **regime** di un lago dipende da molti fattori, fra cui il regime dei suoi fiumi immissari e delle precipitazioni; inoltre la sua presenza influenza fortemente il clima delle regioni limitrofe, addolcendone generalmente gli inverni più rigidi, e regola il regime dei fiumi immissari, accogliendone le acque di piena. Le acque dei laghi sono soggette a **movimenti ondosi**, a **correnti lacustri** e a oscillazioni a bilancia, dette **sesse**, dovute all'azione del vento e alle variazioni locali di pressione atmosferica.

La struttura della Terra solida

1 La litosfera

In passato si indicava con il termine **litosfera** la parte solida della Terra, in contrapposizione all'*atmosfera* e all'*idrosfera* che, come si è visto, ne costituiscono rispettivamente la parte gassosa e la parte liquida. Oggi, invece, si attribuisce al termine *litosfera* un significato più restrittivo rispetto all'uso generale nel quale si comprendeva l'intera componente solida della Terra. In una nuova e più moderna visione della struttura del nostro pianeta, la litosfera rappresenta soltanto l'involucro rigido più esterno [1] di quella che possiamo definire più correttamente, ma pur sempre in modo un po' semplicistico, «**Terra solida**», ossia formata da rocce.

2 Le rocce

Per **roccia** si intende un aggregato di minerali che forma masse più o meno potenti, aventi la stessa composizione, struttura e origine, e costituenti parte integrante della litosfera. Generalmente le rocce sono **eterogenee**, cioè composte da minerali diversi, ma è possibile trovarne anche di **omogenee**, ossia formate da un solo minerale: è il caso delle masse di salgemma o di gesso, che possono essere appunto considerate rocce, data la loro estensione molto ampia in alcune zone.

Lo studio di una roccia comincia presso il sito in cui affiora: se ne analizzano l'*aspetto*, la *struttura*, i *rapporti* con le rocce circostanti e, con l'ausilio di adeguati strumenti, se ne identificano *durezza*, *colore*, *composizione* macroscopica (attraverso il riconoscimento dei minerali visibili ad occhio nudo). Quindi la roccia viene campionata e sottoposta ad analisi di laboratorio: con il microscopio si studiano le caratteristiche non visibili ad occhio nudo e con sofisticate apparecchiature (ad esempio, il diffrattometro e la microsonda) si calcolano i parametri quantitativi, come la composizione chimica.

Le rocce si classificano, in base alla loro genesi, in tre gruppi: **igne**, **sedimentarie** e **metamorfiche**. Le *prime* si formano per solidificazione di una massa fusa, detta *magma*, proveniente dalle zone profonde della litosfera; le *rocce sedimentarie* sono il risultato di processi di erosione, trasporto e accumulo di rocce preesistenti; quelle *metamorfiche* traggono origine in profondità, in condizioni di pressione e/o temperatura particolari, dalla trasformazione di rocce sedimentarie o ignee o già metamorfiche.

Tali processi hanno tempi di sviluppo molto lunghi, in nessun modo comparabili al «tempo dell'uomo»: l'ambiente naturale, così come ci appare oggi, rappresenta il risultato dell'azione di tali fenomeni verificatisi nel passato, nel corso dei «tempi geologici». Si tratta, però, di eventi che hanno luogo anche attualmente, sebbene l'essere umano nella maggior parte dei casi non sia in grado di seguirne l'evoluzione né possa analizzarne gli effetti sul pianeta, che risulteranno visibili soltanto in futuro (si pensi, a tal proposito, a fenomeni quali la formazione di magmi all'interno della Terra o la genesi delle rocce metamorfiche). Alcuni processi, invece, come l'*erosione*, che prelude alla formazione delle rocce sedimentarie, o le *eruzioni vulcaniche*, che danno origine alle rocce ignee effusive, evolvono, almeno

[1] Tale involucro, come si vedrà, comprende la crosta terrestre e lo strato più esterno del mantello.

per quanto riguarda gli stadi conclusivi, in maniera estremamente rapida, così da essere percepiti e compresi dall'uomo.

a) Rocce ignee

Le **rocce ignee**, come si è detto, hanno origine dal raffreddamento e dalla cristallizzazione di una miscela di silicati e gas ad alta temperatura, definita **magma**. Tali processi possono avere luogo all'interno della crosta terrestre, formando le rocce ignee **intrusive**, oppure in superficie, in seguito alla risalita del magma, e in tal caso avremo rocce ignee **effusive** (o *vulcaniche*). Un'ulteriore categoria è quella delle rocce **filoniane**, che si formano in condizioni intermedie di profondità.

Le diverse profondità di formazione di queste rocce, che si traducono poi in differenti condizioni di pressione e temperatura, ne condizionano fortemente i processi di solidificazione e, di conseguenza, la struttura.

Per le *rocce intrusive* il raffreddamento avviene sotto pressione e lentamente, in quanto il magma è protetto dalle rocce circostanti: il fuso ha il tempo di cristallizzare completamente, comportando una *struttura olocristallina granulare* (la roccia è cioè formata interamente da cristalli di dimensioni visibili ad occhio nudo).

Le *rocce effusive* si formano, invece, in superficie: il magma risale, attraverso un condotto vulcanico, dalla camera magmatica (volume della litosfera in cui risiede) fino in superficie, dove si raffredda rapidamente in condizioni di pressione e temperatura quasi normali. Sol tanto una parte del fuso riesce a formare cristalli di dimensioni apprezzabili (fenocristalli), mentre il resto si consolida sotto forma di piccoli cristalli visibili solo al microscopio (*struttura porfirica*), o come pasta vetrosa (*struttura ialina*).

Le *rocce filoniane*, pur non formandosi direttamente in superficie, hanno una struttura simile alle rocce effusive, con cristalli di dimensioni però inferiori, che viene perciò definita *struttura olocristallina porfirica*.

In base alla loro composizione chimica, e quindi mineralogica, le rocce ignee possono essere distinte in **acide** (o *sialiche*), **basiche** (o *femiche*) e **neutre**. Le prime contengono una quantità di silice maggiore del 65% e sono ricche di **minerali sialici** (cioè provvisti di silice e allumina), che presentano una colorazione bianca o comunque chiara (di qui anche il nome di leucocrate per queste rocce): **quarzo**, **feldspati**, **feldspatoidi**, **muscovite**. Le seconde presentano una quantità di silice tra il 45% e il 52%, sono quindi più ricche in **minerali femici** (cioè a base di ferro e magnesio) dai caratteristici colori scuri, dal verde, al bruno, al nero (di qui il nome di melanocrate): **biotite**, **anfibioli**, **pirosseni**, **olivina**. Le rocce neutre contengono silice tra il 52% e il 65% e risultano intermedie rispetto ai due gruppi precedentemente descritti. Un ulteriore gruppo è quello delle **rocce ultrabasiche**, che contengono una quantità di silice inferiore al 45% e sono formate essenzialmente da silicati di ferro e di magnesio (vengono perciò anche dette *rocce ultrafemiche*).

Un altro tipo di classificazione tiene conto della percentuale in ossidi degli elementi alcalini (sodio e potassio) e degli elementi alcalino-terrosi (calcio e magnesio): si individuano due associazioni principali di rocce, la **serie alcalina** e la **serie alcali-calcica**, che occupano *province petrografiche* distinte [2]. Per ciascuna composizione chimica esiste una roccia ignea di tipo intrusivo ed una di tipo effusivo: elencheremo di seguito le più note, con qualche breve cenno alle rispettive caratteristiche principali.

[2] Per *provincia petrografica* si intende una zona della Terra che presenta caratteristiche simili di composizione chimica e mineralogica, e quindi di origine, delle rocce ignee in essa presenti, le quali costituiscono, appunto, una serie magmatica.

I **graniti** costituiscono le rocce intrusive più diffuse sulla Terra, occupandone dal 5 al 10% della superficie, sotto forma di *batoliti* (ammassi rocciosi a forma di cupola) o di *filoni* (corpi rocciosi sviluppati in altezza, che ancora fusi penetrano tra le fratture delle rocce incassanti e solidificano). Si tratta di rocce acide che contengono quarzo, feldspati alcalini e alcali-calcici e pochi minerali femici (generalmente biotite e muscovite). I graniti sono molto usati nell'edilizia sia come pietre ornamentali che per impieghi esterni (facciate di edifici, colonne, monumenti), grazie alla loro notevole resistenza a compressione, rottura e gelività. In Italia si trovano soprattutto nel Novarese (granito di Baveno e di Montorfano) e nel Biellese in Piemonte, nella zona di Sondrio (Lombardia), sull'isola d'Elba e nelle province di Grosseto e Livorno (Toscana), sull'isola di Capraia e nel Nuorese (Sardegna).

Le corrispettive rocce effusive dei graniti sono le **rioliti**, note anche come *lipariti*, in quanto sono caratteristiche di quest'isola delle Eolie, ma in Italia si trovano anche sull'isola di Ponza e nella Maremma Toscana. Presentano aspetto bollosi e spugnoso (*pomici*) o vetroso (*ossidiane*), e vengono generalmente utilizzate come pietre di rivestimento.

Le **dioriti** sono rocce intrusive che presentano, rispetto ai graniti, più alto contenuto in plagioclasio (di tipo sodico), in anfiboli (orneblenda) e pirosseni. Molto comuni nell'arco alpino lombardo (massiccio dell'Adamello, Val Masino, Val Biandino), vengono usate in edilizia per soddisfare richieste di materiali per esterni ad alto grado di durezza. Le rocce effusive aventi la stessa composizione chimica delle dioriti sono denominate **andesiti** (tra cui si ricordano il *porfido rosso antico* e il *porfido verde antico*, utilizzati dai nostri antenati per l'ornamento di chiese e monumenti). In Italia si ritrovano presso i Colli Euganei (Veneto), su alcune isole delle Eolie (Sicilia) e nella zona di Alghero (Sardegna).

I **gabbri** si presentano, come le dioriti, ricchi in plagioclasio, ma di tipo calcico (labradorite). In Italia risultano diffusi nelle Alpi, in particolare nella zona del Monviso e della Val di Susa (Piemonte), e nell'Appennino settentrionale (Liguria, Toscana, Emilia). Le corrispondenti rocce effusive prendono il nome di **basalti**: diffusissimi su tutto il pianeta, costituiscono anche il «pavimento» degli oceani. Si tratta di rocce molto utilizzate per la pavimentazione stradale, fin dall'antichità, e per le massicciate ferroviarie, in quanto resistono bene alle vibrazioni. Nella nostra penisola possiamo trovarle nella zona etnea (Sicilia), in gran parte della Sardegna e nella Liguria orientale.

Alle principali rocce ignee fin qui elencate se ne possono aggiungere, chiaramente, molte altre [3]. Questa grande varietà di composizioni chimiche e mineralogiche pone il problema complesso dell'origine dei magmi: è immaginabile che per ognuna di queste *famiglie* di rocce esista un diverso magma originario? La risposta è chiaramente no. Il magma si ottiene generalmente per la *fusione* di una porzione del *mantello terrestre* [4], che ha in linea di massima composizione omogenea: il risultato è un fuso di tipo basaltico. Sono infatti più rari i casi di *fusione crostale* (fenomeno conosciuto con il nome di *anatessi*), in cui vengono prodotti magmi di tipo granitico. Per spiegare come da un magma unico si ottenga la vasta gamma di rocce ignee esistenti bisogna ricordare che esso può subire una serie di **processi di evoluzione** (ad esempio, differenziazione magmatica, differenziazione gravitativa, cristallizzazione frazionata), nel corso dei quali muta la sua composizione, o può interagire

[3] Tra le rocce ultrabasiche intrusive ricordiamo le *peridotiti* e le *duniti*, formate essenzialmente da olivina e pirosseni, che hanno scarsa diffusione sui continenti ma rappresentano giacimenti minerali di grande valore (cromo, cobalto, nichel, platino, diamanti); tra le rocce alcaline vanno menzionate le *sieniti*, intrusive, e le *trachiti*, corrispondenti effusive, povere in quarzo e ricche in ortoclasio, utilizzate sia per motivi ornamentali che per l'edilizia; tra le rocce effusive a feldspatoidi citiamo le *leucititi*, tipiche dei vulcani laziali e campani, usate dagli antichi Romani per la pavimentazione a basoli (cioè a blocchi) delle strade.

[4] Si tratta di uno degli involucri che costituiscono l'interno della Terra.

con un magma diverso o con elementi esterni (fenomeni di *mixing* tra un fuso mantellico ed uno crostale, processi di assimilazione), dando come risultato un nuovo *magma ibrido* (generalmente a composizione andesitica).

b) Rocce sedimentarie

Le **rocce sedimentarie** si formano per accumulo di materiali di tipo diverso (*sedimenti*), che gli agenti esogeni (acque, vento, ghiacciai) disgregano da rocce preesistenti e trasportano. Il passaggio dal sedimento alla roccia avviene attraverso un processo, definito **diagenesi**, in cui si ha compattazione e cementazione. Il processo completo di sedimentazione prevede, dunque, quattro fasi fondamentali: l'*erosione*, il *trasporto*, l'*accumulo* e la *diagenesi*. L'accumulo può avvenire in ambiente subaereo o subacqueo: di qui la prima distinzione delle rocce sedimentarie in **subaeree** e **subacquee**, appunto. La classificazione più comune viene però fatta sulla base della loro genesi e si individuano tre gruppi principali di rocce: **clastiche** (o *detritiche*), **organogene** e **di origine chimica**.

Le **rocce clastiche** si formano per la diagenesi di sedimenti di origine detritica, cioè di frammenti di rocce o di minerali preesistenti, disgregati per azione degli agenti esogeni. In esse si distinguono elementi visibili ad occhio nudo, i *clasti*, ed elementi di dimensioni minori che appaiono come una pasta omogenea, definita *matrice*. Il *cemento* che litifica i sedimenti può avere composizione estremamente varia (calcarea, silicea, argillosa, ferrifera etc.). La classificazione di queste rocce viene effettuata in base alle dimensioni dei clasti [5].

Le rocce clastiche più grossolane e poco cementate vengono generalmente utilizzate per la produzione di *calcestruzzo*; le sabbie impastate con la calce formano la *malta* per le costruzioni; alcune argille, cotte, danno *laterizi* (mattoni e tegole), altre, mescolate anche con calcare, danno *cemento*; infine, le arenarie più compatte e le altre rocce ben cementate vengono lavorate per ottenere lastre ornamentali o blocchi per la pavimentazione delle strade o i muri degli edifici. La loro diffusione risulta molto ampia su tutto il nostro territorio.

Le **rocce non clastiche** derivano da materiale in soluzione che si deposita, o per precipitazione chimica (*rocce di origine chimica*) o per precipitazione indiretta, attraverso l'azione di un organismo (*rocce organogene*).

Le **rocce organogene** si formano per accumulo dei resti di organismi animali o vegetali, generalmente nelle acque marine. La composizione può essere *calcarea* (ad esempio, calcari corallini, calcari a rudiste etc.), *silicea* (le radiolariti e le diatomiti sono rocce molto diffuse sui fondi oceanici), *fosfatica* (fosforiti), *carboniosa* (i carboni fossili che derivano dalla «fossilizzazione», appunto, di grandi masse di vegetali), *bituminosa* (petrolio e bitumi, rispettivamente idrocarburi liquidi e solidi che derivano dalla decomposizione di sostanze organiche vegetali e animali ad opera di batteri). Gli impieghi delle rocce organogene si rivelano estremamente vari: i carboni e gli idrocarburi vengono notoriamente sfruttati per produrre energia e nell'industria chimica (fertilizzanti, materie plastiche etc.); la selce è stata largamente utilizzata dagli uomini preistorici per fabbricare arnesi e utensili; da alcuni calcari e dagli scisti bituminosi si ricava asfalto; la farina fossile delle diatomiti costituisce la materia inerte per la dinamite; alcuni calcari fossiliferi, infine, vengono lavorati ed usati come lastre ornamentali. Le rocce organogene hanno ampia diffusione in gran parte della nostra penisola. In particolare: i calcari compatti, di svariati colori, abbondano dalle Prealpi Bresciane (Lombardia) fino a Trapani (Sicilia), passando attraverso le zone del Carso

[5] La classificazione corrisponde a: *ruditi* (o *sefiti* o *conglomerati*), se gli elementi hanno il diametro superiore ai 2 mm (**ciottoli**); *areniti* (o *psammiti*), se il diametro degli elementi è compreso tra 2 mm e 1/16 di mm (**sabbia**); *siltiti*, se il diametro degli elementi è compreso tra 1/16 di mm e 1/25 di mm (**silt**); *lutiti*, se il diametro è inferiore a 1/250 di mm (**argilla**).

(Friuli-Venezia Giulia), di Verona (Veneto), di Trani (Puglia); i calcari fossiliferi si ritrovano dalle Prealpi Lombarde fino ai monti della Calabria; le rocce silicee caratterizzano il Monte Amiata (Toscana) e alcune zone della Romagna; i carboni invece scarseggiano, se si eccettua la zona del Sulcis (Sardegna), con i giacimenti di litantrace e lignite; modesti giacimenti di idrocarburi liquidi sono stati individuati nelle province di Caltanissetta e di Ragusa (Sicilia), mentre consistenti giacimenti di gas naturale vengono sfruttati nel Mezzogiorno.

Le **rocce di origine chimica** si formano per precipitazione di sostanze sciolte nelle acque (quando la loro quantità supera il limite di solubilità), per evaporazione delle acque (*rocce evaporitiche*) o per trasformazione in sostanze insolubili [6].

Per quanto riguarda lo sfruttamento da parte dell'uomo, le rocce chimiche residuali vengono utilizzate per estrarre ferro (lateriti) o alluminio (bauxiti), ma anche per fabbricare ceramiche e porcellane (argille caoliniche e caolino); l'alabastro e l'opale sono considerate pietre ornamentali, mentre il travertino, diffusissimo al tempo degli antichi Romani come pietra da costruzione per gli edifici, è tuttora utilizzato a tal fine. Grandi depositi di bauxite si trovano in Istria (penisola situata tra Italia, Croazia e Slovenia) e nelle Murge (Puglia). In Italia, il lago Omodeo (Sardegna) risulta ricco di masse contenenti opale; le aree del Carso (Friuli-Venezia Giulia), del Bergamasco (Lombardia) e del Cuneese (Piemonte) rivelano ampie concentrazioni di alabastro; molte località della Toscana, dell'Umbria e delle Marche, infine, presentano estesi depositi di travertino.

c) Rocce metamorfiche

Le **rocce metamorfiche** si formano da rocce preesistenti, ignee, sedimentarie o anche già metamorfiche (in tal caso si parla di **polimetamorfismo**) attraverso trasformazioni che si hanno in condizioni particolari di temperatura e/o pressione (*fenomeni di metamorfismo*). Il *metamorfismo* è, dunque, un riequilibrio della roccia che si viene a trovare in un ambiente diverso da quello di origine.

Se i fattori che determinano le condizioni di metamorfismo sono le *variazioni di pressione e di temperatura*, gli effetti che esso ha sulle rocce sono modificazioni della struttura, della costituzione mineralogica e della composizione chimica. Quando il fattore predominante è la *temperatura*, si parla di **metamorfismo di contatto**: ad esempio, il fenomeno che interessa le rocce che circondano una massa magmatica in risalita. Ne scaturiscono i marmi, gli scisti macchiettati, le cornubianiti, le quarziti. Quando invece il fattore determinante è la *pressione*, il fenomeno viene detto **dinamometamorfismo** ed è generalmente associato ai movimenti orogenetici. Le rocce tipiche che si formano sono le breccie di frizione e le miloniti. In situazioni di profondità molto elevata, invece, il processo di trasformazione delle rocce diviene più complesso e si definisce col termine di **ultrametamorfismo** [7].

[6] A seconda della natura chimica, si distinguono: *rocce saline* (formate da salgemma e altri cloruri, da gesso e anidrite, da solfati potassici e magnesiaci) di origine evaporitica; *rocce calcaree e dolomitiche*, che si formano prevalentemente per precipitazione di CaCO_3 , carbonato di calcio, e di $\text{CaMg}(\text{CO}_3)$, carbonato di calcio e magnesio, nell'acqua di mare senza l'intervento di organismi costruttori. In acque poco profonde e calde si può avere deposizione di carbonato attorno a piccoli granuli di diversa origine (calcarei oolitici e pisolitici), mentre in acque continentali (soprattutto di fiume) si possono formare calcari concrezionati, come gli alabastrini e i travertini; *rocce selcifere*, che derivano dalla precipitazione diretta di SiO_2 (selce) o dalla deposizione di silice nelle acque sorgive o nelle sorgenti termali vulcaniche (opale); *rocce chimiche residuali*, che hanno origine dall'accumulo dei materiali di alterazione di altre rocce.

[7] Si raggiungono condizioni prossime al punto di fusione e si formano *migmatiti* (dalla litificazione del migma, miscela silicatica semifusa) e *anatessiti* (rocce metamorfiche molto simili ai graniti ignei, che si formano per fusione crostale in un processo definito di «anatessi»).

Il metamorfismo più diffuso e imponente si verifica in seguito ad un *aumento contemporaneo di pressione e temperatura*, e prende il nome di **metamorfismo regionale**: interessa ampie zone del pianeta ed è legato ai grandi movimenti tettonici delle *fasce attive* della litosfera. A seconda delle condizioni di pressione (e cioè di profondità) e di temperatura in cui avvengono le trasformazioni, esse risultano più o meno intense: si hanno cioè diversi **gradi di metamorfismo** e differenti **facies metamorfiche** [8]. La caratteristica delle rocce sottoposte al metamorfismo regionale (quando non si raggiungono temperature e pressioni troppo elevate) consiste nella **scistosità**, ossia l'orientazione preferenziale dei cristalli in piani paralleli, che conferisce loro una divisibilità in lastre più o meno sottili. Se la roccia originaria è *igne*, la roccia metamorfica che si ottiene è un **ortoscisto**, se è *sedimentaria* è allora un **parascisto**. Tra le principali rocce formatesi per metamorfismo regionale si ricordano le filladi (che generalmente derivano da argille), i micascisti, gli gneiss (risultanti di solito dai graniti), le serpentiniti (originate da metamorfismo di basso grado di peridotiti), i marmi (calcarei metamorfosati), gli scisti a glaucofane e le eclogiti (risultanti da diversi gradi di metamorfismo di basalti).

Le rocce metamorfiche vengono generalmente utilizzate come materiale da costruzione, come pietre ornamentali e come lastre di rivestimento per le pareti. In Italia, filladi e micascisti abbondano nelle Alpi Centrali (tra il Piemonte e il Trentino Alto-Adige) e nell'Appennino Meridionale (tra Calabria e Sicilia); gli gneiss sono presenti principalmente sui massicci del Gran Paradiso, del Monte Rosa e del Monte Bianco (tra Val d'Aosta e Piemonte) e in Valtellina (Lombardia); i marmi più noti si trovano nelle Alpi Apuane (Toscana) e vengono lavorati nell'area di Massa-Carrara, ma sono diffusi anche nelle zone del Novarese (Piemonte) e in molte località delle Alpi Centrali.

Le rocce ignee, sedimentarie e metamorfiche, pur avendo origine da processi differenti, sono in realtà legate in un unico **ciclo litologico**: attraverso i processi magmatici, sedimentari e metamorfici, e con l'azione concomitante di fattori esterni (interazione tra la Terra solida, l'idrosfera e/o l'atmosfera; scambi continui di calore e materia tra i vari strati della Terra solida) c'è il passaggio continuo (anche se molto lento rispetto alla concezione umana del tempo) da una roccia all'altra.

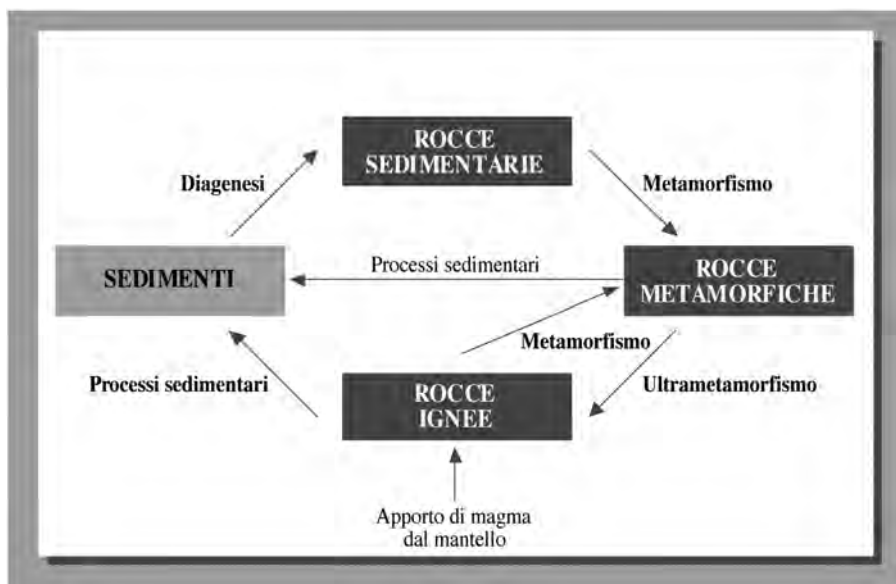


Fig. 1 - Ciclo litologico

[8] Una *facies metamorfica* è l'insieme delle rocce che si sono formate nelle stesse condizioni di temperatura e di pressione. Le *facies metamorfiche* sono: quella delle zeoliti, degli scisti verdi, degli scisti blu, anfibolitica, granulitica.

3 L'interno della Terra e i diversi involucri che lo compongono

Lo sviluppo della *geologia* e, quindi, le prime importanti conoscenze sulla struttura e sulla composizione della superficie della Terra risalgono alla metà del Settecento e sono legate al lavoro nelle miniere, che caratterizza gli anni della Rivoluzione industriale.

Le indagini sulle aree interne del globo sono invece più recenti e hanno portato a nuovi e spesso sconcertanti risultati, grazie soprattutto agli studi relativi alla propagazione delle onde sismiche nel sottosuolo.

Che la Terra non fosse di composizione e struttura omogenee al suo interno era già noto agli scienziati dei secoli scorsi: la densità delle rocce in superficie era mediamente intorno a $2,67 \text{ g/cm}^3$, mentre la densità media del pianeta (calcolata a partire dalla *Legge di Newton*, che finalmente consentiva di conoscere il valore della massa terrestre) era pari a $5,5 \text{ g/cm}^3$. La discrepanza di valori poteva essere giustificata soltanto ammettendo un *interno* della Terra molto più denso della parte superficiale.

Già a fine Ottocento si ipotizzò un modello di Terra ad **involucri**, mentre ai primi del Novecento risale la distinzione in **SIAL**, **SIMA** e **NIFE** ad indicare, rispettivamente, la *zona esterna*, costituita da rocce con silice ed allumina; la *zona intermedia*, formata da rocce con silicati di ferro e di magnesio; e la *zona interna*, ricca di nichel e ferro. Un contributo straordinario alla conoscenza dell'interno terrestre è venuto dalla geofisica: sfruttando la proprietà delle **onde sismiche** di essere rifratte o riflesse al passaggio tra materiali con densità, rigidità e/o composizione terrestre, i geologi, provocando terremoti artificiali, hanno potuto elaborare un *nuovo modello ad involucri* per il nostro pianeta.

Due *discontinuità sismiche* principali (la discontinuità di *Mohorovičić* e la discontinuità di *Gutenberg*) individuano il contatto fra tre involucri primari, che, dall'esterno verso l'interno del globo, sono: *crosta*, *mantello* e *nucleo*.

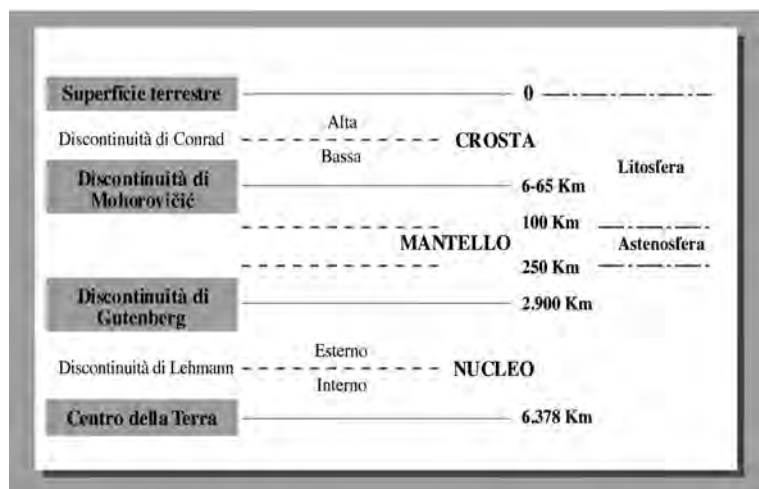


Fig. 2 - Schema della composizione a strati dell'interno terrestre

La **crosta** è un involucro rigido, con uno spessore che varia dai 30 km sotto i continenti (ma in corrispondenza di grandi catene montuose raggiunge fino ai 70 km di profondità) ai 6 km sotto gli oceani. La **crosta oceanica** e la **crosta continentale** hanno dunque spessore diverso e presentano, inoltre, una composizione litologica differente: la prima è formata essenzialmente da rocce basaltiche e peridotitiche, la seconda è costituita da rocce di tipo granitico-granodioritico. All'interno della crosta continentale è stata individuata un'ulteriore

discontinuità, detta di *Conrad*, che distingue una **crosta alta** da una **crosta bassa**: la differenza tra i due involucri è ancora nella composizione, più basica per la seconda.

Il **mantello** si spinge fino a 2.900 km circa di profondità, ha una densità superiore rispetto alla crosta sovrastante (come testimoniato dal forte incremento di velocità che subiscono le onde sismiche in corrispondenza della discontinuità di Mohorovičić, o semplicemente *Moho*) ed una composizione di tipo peridotitico (più basica, però, rispetto alla crosta oceanica). L'improvvisa diminuzione delle onde sismiche tra i 100 e i 250 km si spiega, secondo gli studiosi, con la presenza all'interno del mantello solido di uno strato a uguale composizione ma di natura plastica, definito **astenosfera**. Per contrasto, tutto lo spessore superiore all'astenosfera (ossia crosta continentale/oceanica e mantello solido) prende il nome di **litosfera**.

Il **nucleo** è l'involucro terrestre più interno ed ha uno spessore di 3.600 km. La sua densità raggiunge i 13 g/cm³ al centro della Terra, mentre la composizione chimica, assimilabile a quella delle meteoriti ferrifere (*sideroliti*), dovrebbe prevedere ferro e nichel, con la presenza di silicio. All'interno del nucleo, ad una profondità intorno ai 5.000 km, si rileva la presenza di una quarta zona di *discontinuità*, quella di *Lehmann*, la quale separa un involucro allo stato fuso, il **nucleo esterno**, da un involucro solido, il **nucleo interno**.

La geodinamica endogena

1 Il calore interno

La **geodinamica** è il ramo della geologia che studia le forze che agiscono sul nostro pianeta, sia **esogene** (azione delle acque, del vento, dei ghiacciai etc.), sia **endogene** (vulcanismo, sismi, tettonica, orogenesi). Il motore che è alla base di tutti i fenomeni geologici endogeni e che, insieme con l'energia solare, spiega anche i processi esogeni, è il *calore interno* della Terra.

Le osservazioni effettuate nei pozzi minerari rivelano che la temperatura della Terra aumenta, in media, di 1°C ogni 33 metri di profondità: 30°C/km è, dunque, il **gradiente geotermico medio**. Se esso si mantenesse costante lungo tutto il raggio terrestre, in corrispondenza del nucleo si raggiungerebbero temperature stellari. Le cose, ovviamente, non stanno così. Gli studi geofisici, basati sull'analisi della propagazione delle onde sismiche nell'interno della Terra, dimostrano che alla profondità dell'astenosfera «plastica», cioè intorno ai 200 km, laddove hanno inizio i processi di fusione, la temperatura è intorno ai 1.200°C, mentre in corrispondenza del mantello solido più profondo supera i 3.000°C. Entrando nel nucleo esterno fuso, la **geoterma** (ossia la curva che descrive l'andamento della temperatura con la profondità) raggiunge i 3.700°C, superando la curva di fusione delle rocce in quelle specifiche condizioni di pressione. Con il passaggio al nucleo interno solido, la geoterma si riavvicina alla curva di fusione e le temperature sono dell'ordine dei 4.300°C.

Il calore interno della Terra è il residuo dei fenomeni di condensazione delle polveri cosmiche che hanno dato origine al nostro pianeta. Gli studiosi sostengono che questo calore si rinnovi continuamente grazie a processi come la precipitazione del ferro, alla trasformazione di porzioni di nucleo in mantello e soprattutto al decadimento radioattivo di isotopi di alcuni elementi, ad esempio l'uranio, il torio e il potassio, che si trovano concentrati prevalentemente in rocce acide come i graniti.

La propagazione di questo calore in tutto il pianeta avviene attraverso un meccanismo di convezione e si è calcolato che il flusso sia mediamente pari a 1/1.500.000 cal/cm²s, sia per le aree oceaniche che per quelle continentali. Tuttavia queste ultime, granitiche, hanno un contributo rilevante dal decadimento radioattivo che invece interessa pochissimo le aree oceaniche basaltiche: i valori confrontabili del flusso di calore si spiegano ipotizzando un trasporto più efficace al di sotto degli oceani grazie alla risalita di materiale caldo dal mantello (**movimenti convettivi** e punti caldi, di cui si parlerà diffusamente più avanti).

2 Il magnetismo

Già nel Seicento gli scienziati ipotizzavano che il **campo magnetico terrestre** fosse legato a cause agenti dall'interno del pianeta. I principi della fisica stabiliscono che tutte le proprietà magnetiche si perdono al di sopra di una temperatura pari al *punto di Curie* (che varia da sostanza a sostanza), e questo significa che nessun elemento all'interno della Terra può comportarsi da magnete.

Il **magnetismo**, tuttavia, può essere generato anche da un **campo elettrico**. Il modello più attuale prevede che il centro del fenomeno sia il nucleo: una dinamo ad auto-eccitazione avrebbe preso il via da un debole campo magnetico iniziale, generato dall'elettricità prodotta

da alcuni dipoli irregolari. Una volta innescato il meccanismo, la dinamo ha continuato ad agire ininterrottamente: al momento della formazione del nucleo il processo ha avuto luogo la prima volta e da lì in poi non si è più arrestato.

Una roccia che contiene elementi magnetici può rivelare informazioni sul campo magnetico esistente nel momento in cui si è formata: tramite l'**inclinazione magnetica** calcoliamo la paleo-latitudine della roccia; con la **declinazione magnetica** si ha l'inclinazione del polo magnetico e quindi la paleo-direzione del polo, attraverso il *nord* e il *sud magnetici* è possibile risalire ad eventuali fenomeni di **inversione magnetica** [1].

Si è calcolato che queste inversioni si verificano con un tempo di ritorno di circa 200.000 anni: in tal modo è possibile distinguere la storia della Terra in **epoche magnetiche**. Negli ultimi 4,5 milioni di anni se ne identificano 4 principali, con i nomi dei geofisici che le hanno individuate: *Gilbert*, *Gauss*, *Matuyama* e *Brunhes* (l'attuale).

3 La tettonica delle placche

La teoria secondo la quale i continenti possono muoversi orizzontalmente e assumere nei diversi momenti geologici posizioni relative differenti nasceva agli inizi del Novecento ad opera di Wegener, geofisico e meteorologo tedesco. Questi intuì che l'attuale configurazione delle terre emerse potesse essere il risultato della lacerazione di un unico grosso continente, che egli definì **Pangea**, e del successivo allontanamento dei vari blocchi frammentati in diverse direzioni (**teoria della deriva dei continenti**).

La straordinaria corrispondenza «geometrica» tra gli attuali margini dei continenti (ad esempio, la costa orientale dell'America Meridionale combacia con la costa occidentale dell'Africa, così come l'India si incastra perfettamente con il margine orientale dell'Africa stessa), nonché la loro eccezionale conformità di strutture geologiche e di *facies* paleontologiche (ossia la sorprendente corrispondenza, lungo margini di continenti oggi lontani e separati da oceani, di terreni della stessa età, con uguali caratteristiche geologiche e con fossili del medesimo tipo) costituivano per Wegener prove inconfutabili della sua teoria. Il motore del movimento dei continenti era attribuito all'energia gravitazionale e la deriva avveniva per galleggiamento delle terre sul mantello (gli studiosi adoperano l'espressione: «SIAL SU SIMA») [2].

Le ricerche oceanografiche e sismiche in mare, condotte a partire dagli anni Cinquanta del Novecento, produssero una serie straordinaria di dati, dimostrando che i fondali marini sono percorsi da un sistema complesso di rilievi di natura vulcanica (*dorsali oceaniche*) e zone depresse (*fosse oceaniche*). Le dorsali si estendono per una lunghezza superiore ai 60.000 km e sono larghe fino ai 1.000 km, raggiungendo altezze comprese tra i 1.000 e i 3.000 metri. Le loro creste sono percorse da larghi solchi longitudinali conosciuti come **rift valleys**, tagliati orizzontalmente da faglie (fratture) trasversali. Le dorsali sono caratterizzate da elevati flussi di calore e sono formate da rocce effusive basaltiche o da rocce intrusive di tipo gabbrico-peridotitico. Localizzate in prossimità dei continenti, le fosse raggiungono, come si è già ricordato, profondità medie intorno ai 4.000-6.000 m (con punte fino a 11.000 m), sono molto allungate, con fondi stretti poche decine di chilometri e versanti ripidi.

I dati geologici, paleontologici e geofisici raccolti permisero allo studioso Hess, e poi a Dietz, di elaborare la **teoria dell'espansione dei fondi oceanici**, secondo la quale le dorsali oceaniche rappresentano in superficie i rami ascendenti delle correnti convettive del mantello: in tal modo

[1] Il campo magnetico terrestre subisce periodicamente fenomeni di inversione di polarità, passando alternativamente da *normale*, cioè orientato con il Polo Nord (com'è oggi), a *inverso*.

[2] In geologia il termine SIAL (**S**ilicio + **A**lluminio) indica lo strato superiore della crosta terrestre, mentre SIMA (**S**ilicio + **M**agnesio) identifica lo strato più profondo.

si crea continuamente crosta oceanica che si espande, spingendo lateralmente la crosta già formata. Le datazioni dei fossili e la registrazione del paleomagnetismo delle varie sezioni di crosta oceanica, a partire dalle *rift-valleys* delle dorsali fino al bordo dei continenti, fornivano, già allora, prove certe di questo movimento: si riscontravano infatti, simmetricamente rispetto alle *rift-valleys*, alternanze di bande magnetiche con polarità dirette ed inverse, a testimoniare che fasce di crosta ugualmente distanti dalla rift-valley hanno la stessa età, così come i fossili risultavano via via più antichi allontanandosi dalla dorsale, dove quindi si trova materiale più recente.

Gli oceani sono dunque tuttora in formazione e la recente scoperta che nessuno di essi è più antico del Giurassico (secondo periodo dell'Era Mesozoica, compreso tra 195 e 135 milioni di anni fa), mentre le rocce più antiche dei continenti sono datate intorno ai 4 miliardi di anni fa, è un'ulteriore testimonianza di come siano in costante formazione e si rinnovino continuamente.

Le intuizioni di Wegener e le straordinarie scoperte sui fondi oceani hanno poi portato, negli anni Settanta del Novecento, alla formulazione definitiva di un nuovo modello dinamico della Terra, denominato «**tettonica delle placche**» o «**tettonica a zolle**». Tale teoria prevede che la litosfera terrestre sia formata da una serie di **zolle litosferiche**, spesse circa 100 km, formate da crosta oceanica o crosta continentale (con le relative coperture sedimentarie) e dalla parte superiore del mantello, che poggiano e si muovono sull'astenosfera. Non si tratta dunque semplicemente di masse continentali alla deriva, come sosteneva Wegener, ma di grosse placche che trasportano anche continenti. Messa da parte l'idea dell'energia gravitazionale come «motore», il movimento delle zolle è invece giustificato con moti convettivi del mantello terrestre.

L'estensione delle zolle è molto variabile: le più grandi, come quella euroasiatica e quella africana, comprendono uno o più continenti e una parte degli oceani circostanti; le più piccole, come quelle di Nazca o delle Filippine, soltanto una frazione di oceano, casomai con qualche isola. Sono tutte legate tra loro e ricoprono interamente la superficie terrestre, senza alcuno «spazio vuoto».

I **margini** che le delimitano sono di tipo attivo o passivo, a seconda se ci sia o meno produzione/distruzione di crosta. In particolare:

- i **margini attivi** corrispondono a:
 - **dorsali oceaniche**, dove si produce continuamente nuova crosta oceanica (*margini divergenti costruttivi*),
 - **fosse di subduzione**, dove le zolle si consumano per sottoscorrimento di una zolla rispetto all'altra (processo di *subduzione*) e per fusione del materiale crostale nell'astenosfera (*margini convergenti distruttivi*);
- i **margini passivi** si identificano con le **faglie trasformi** e **trascorrenti**, in corrispondenza delle quali le placche scorrono lateralmente l'una all'altra senza produzione né distruzione di crosta (*margini conservativi*).

Alcune placche, come quella africana, sono dunque attualmente in accrescimento, altre sono in fase stazionaria, essendo delimitate da una parte da una dorsale e dall'altra da una fossa di subduzione, altre ancora sono in riduzione: il bilancio è sempre in equilibrio, dal momento che complessivamente la Terra ha superficie e volume stabili. Questo modello è in grado di spiegare i principali lineamenti strutturali del nostro pianeta, processi complessi come la nascita delle catene montuose (*orogenesi*) e l'esistenza di «fasce attive» in cui si rileva la massima concentrazione dell'attività vulcanica e sismica.

4 I terremoti

Un **terremoto**, o *sisma*, è un movimento brusco e rapido che si verifica sulla superficie terrestre con una serie di scosse: la causa risiede nelle onde sismiche che hanno origine nella profondità

della crosta o del mantello per un'improvvisa rottura di equilibrio. Il punto in cui si libera l'energia e da cui partono le onde elastiche viene definito **ipocentro**: quanto più esso è profondo, tanto più si dilata l'area in cui si risentiranno gli effetti del terremoto. L'**epicentro** rappresenta, invece, il punto della superficie terrestre che si trova perpendicolarmente sopra all'ipocentro e indica la zona di massima intensità di un sisma. L'**intensità** viene stabilita valutando gli effetti prodotti dal terremoto su persone e strutture e si determina, generalmente, facendo riferimento alla **scala Mercalli-Cancani-Sieberg** (MCS), divisa in 12 gradi. Tale valutazione risente di troppi fattori per essere realmente oggettiva: è più corretto dunque fare riferimento alla **magnitudo** (grandezza), che dipende dall'ampiezza delle onde sismiche registrate dagli idonei strumenti di misurazione (sismografi), ed è misurata con la **scala Richter** (per i terremoti più violenti degli ultimi decenni è stata registrata una magnitudo pari quasi a 9 gradi Richter). Dalla magnitudo è possibile risalire all'energia meccanica del sisma.

Le **onde sismiche** sono di tre tipi:

- **onde longitudinali**, dette anche **prime (P)**, giacché arrivano per prime avendo velocità mediamente maggiori (circa 5,5 km/s). Sono onde di compressione, in quanto producono nella roccia sollecitazioni con la loro stessa direzione di propagazione, dal che conseguono variazioni di volume;
- **onde trasversali**, o **seconde (S)**, che arrivano dopo le prime perché più lente (circa 4,4 km/s). Sono onde di taglio: le oscillazioni nella roccia si compiono perpendicolarmente alla direzione del moto di propagazione delle onde;
- **onde superficiali**, o **di Rayleigh**, che si generano in superficie a partire dall'epicentro, quando arrivano le onde P ed S, e viaggiano con velocità più moderate (3,3 km/s).

La velocità e la direzione di propagazione delle onde interne (P e S) cambiano al passaggio tra materiali fisicamente diversi, subendo fenomeni di rifrazione e riflessione. È proprio sfruttando tali caratteristiche delle onde che, con la sismica, è stato possibile avere informazioni preziose sulla composizione e sulle caratteristiche degli involucri terrestri.

Se il sisma provoca in superficie vibrazioni verticali si parla di **sosse sussultorie**, se le vibrazioni sono orizzontali le **sosse** saranno definite **ondulatorie**. Grazie allo studio dei **sismogrammi**, che rappresentano la traccia grafica del terremoto registrato dalle reti di **sismografi**, ben diffuse nelle aree maggiormente esposte al **rischio sismico** (inteso come la probabilità che in una certa area e in un certo intervallo di tempo si risentano gli effetti di un terremoto), è possibile conoscere tutte le peculiarità dell'evento: istante di innesco, durata, periodo e ampiezza delle oscillazioni, quindi magnitudo ed energia. Utilizzando le **dromocrone**, curve che indicano i tempi di propagazione di ogni tipo d'onda in funzione della distanza dall'epicentro, si risale, per ogni sisma, alla posizione dell'epicentro e alla profondità dell'ipocentro. Si distinguono *terremoti superficiali*, con profondità ipocentrali tra 0 e 70 km; *terremoti intermedi*, per profondità tra i 70 e i 300 km; *terremoti profondi*, con ipocentro oltre i 300 km.

In base all'origine, i terremoti si classificano invece essenzialmente in *vulcanici* e *tettonici*, a seconda che siano legati ad eruzioni o siano invece il frutto di movimenti della crosta o del mantello in conseguenza di situazioni di disequilibrio. A tal proposito, si è già visto come la distribuzione dei terremoti sulla superficie terrestre sia intimamente legata alla tettonica delle placche.

I terremoti rappresentano per l'uomo un gravissimo pericolo: a tutt'oggi sono considerati eventi non prevedibili e, sebbene le recenti normative si siano fatte più attente a tali problematiche e l'ingegneria stia facendo passi da gigante nel campo della progettazione di strutture antisismiche, il rischio resta elevatissimo in moltissimi Paesi del mondo, compresa l'Italia. Accanto all'applicazione dei criteri di sicurezza per le costruzioni occorre, a scopo di prevenzione, raccogliere quanti più dati possibili sulla sismicità delle aree (cataloghi degli eventi), dai quali elaborare carte di zonazione sismica (aree a diversa sismicità).

5 I fenomeni endogeni: vulcanismo, sismi e orogenesi

Le dorsali e le fosse non rappresentano soltanto fattori morfologici, ma vanno interpretate come centri intensi di **attività endogena**: è per questo che tali zone vengono definite **fasce attive**. La distribuzione dei **vulcani** sulla superficie terrestre è in corrispondenza quasi perfetta con la distribuzione dei **terremoti**: entrambe coincidono straordinariamente con le fasce attive.

In particolare le **zone di dorsale** sono caratterizzate sia da un elevato flusso di calore legato alla risalita di magma dal mantello che genera un vulcanismo molto intenso, esteso e di grossi volumi, sia da frequenti terremoti, con ipocentri relativamente superficiali.

Le **zone di fossa** sono invece caratterizzate da vulcanismo di arco (ossia da una serie di isole, con attività vulcanica assai concentrata, allineate sulla placca che non subduce) e da un'intensa attività sismica, con gli ipocentri dei terremoti localizzati a profondità crescenti (fino ai 750 km) mano a mano che ci si allontana dalla fossa verso il continente. Tale fenomeno si spiega con il meccanismo di subduzione che interessa proprio i margini distruttivi tra zolle, cioè l'area di convergenza tra due placche. L'immersione di una zolla sotto l'altra avviene lungo il *piano di Benioff* (dal nome del geofisico russo che lo individuò), una superficie inclinata rispetto alla superficie terrestre di 15°-75°. Le immancabili fratture generate lungo questo piano durante il *processo di subduzione* sono la causa dei terremoti e dei loro diversi e assai profondi ipocentri. Lungo le **zone delle faglie trasformi e trascorrenti**, in corrispondenza del piano di scorrimento laterale tra placche, sia il vulcanismo sia l'attività sismica risultano estremamente ridotti.

Esistono delle eccezioni a questo modello di distribuzione così «regolare» dei fenomeni endogeni: il caso più noto è quello del complesso delle isole vulcaniche Hawaii, in mezzo all'Oceano, al centro della *placca pacifica* e quindi al di fuori delle fasce attive. È possibile, dunque, un meccanismo alternativo per il vulcanismo, che gli scienziati hanno definito fenomeno da **hot spot**. I «punti caldi» rappresentano la risalita di pennacchi di magma dal mantello (*mantle plumes*), ossia rami ascendenti di correnti convettive che formano sulla superficie terrestre apparati vulcanici. L'allineamento dei vulcani sulle diverse isole e la circostanza apparentemente strana che solo una delle isole manifesti attualmente attività vulcanica, si spiega con il movimento della placca sui pennacchi fissi: il punto caldo, quando la placca vi passa sopra, lascia tutta una successione di vulcani che via via si estinguono e che, quindi, appaiono sempre più antichi man mano che ci si allontana da quello ancora in eruzione.

Nella fascia attiva delle zone di fossa c'è la possibilità che si verifichi il processo di **orogenesi**: se, infatti, una delle due zolle che convergono è formata da crosta continentale, la subduzione dà origine al sollevamento delle catene montuose (*orogeni*). In particolare, se soltanto una delle due zolle presenta crosta continentale, quando avviene la collisione, la zolla costituita da crosta oceanica, più leggera, subduce, mentre sul continente si hanno delle deformazioni che portano al «raggrinzimento» e alla nascita della catena montuosa. Le Ande costituiscono l'esempio classico di questo tipo di orogenesi: la zolla oceanica di Nazca è in subduzione da almeno 200 milioni di anni sotto il margine continentale sudamericano; di qui il nome di *orogenesi andina* per questo tipo di convergenza.

Quando la convergenza porta alla completa subduzione della crosta oceanica si può arrivare alla collisione tra due placche con crosta continentale: si verifica, insomma, un vero e proprio scontro. Soltanto la parte più profonda della crosta di una delle due zolle riesce a subdurre, mentre la superficie viene soltanto «scorticata» e si forma la catena montuosa. È con tale meccanismo che hanno avuto origine le nostre Alpi (di qui il nome di *orogenesi alpina*), nate dallo scontro tra la placca africana e quella eurasiatica, e la catena Himalayana, sollevatasi in seguito alla collisione tra il margine meridionale dell'Euroasia e il continente indiano.

La convergenza tra placche oceaniche, con subduzione di una delle due, non porta all'orogenesi, ma unicamente alla risalita di magmi, dovuti alla fusione parziale del materiale subdotto, da cui hanno origine un vulcanismo sottomarino e, in superficie, **archi di isole vulcaniche**.

6 I vulcani

Il **vulcanismo** rappresenta l'effetto sulla superficie terrestre dell'attività magmatica endogena, che ben si inquadra, come si è visto, nel modello generale della tettonica delle placche. Un **vulcano** è una fessura della crosta terrestre attraverso la quale risalgono e vengono emessi magmi e gas provenienti dall'interno della Terra. Il magma risiede in una o più **camere magmatiche** ed è collegato con l'esterno attraverso uno o più **condotti vulcanici**, che terminano in superficie con un **cratere**.

I prodotti eruttati dai vulcani, nel solidificarsi, danno luogo alle **rocce effusive**. La forma e le dimensioni degli *apparati* o *edifici vulcanici* (che corrispondono alla parte dei vulcani che affiora in superficie) dipendono dal tipo di attività e dalle caratteristiche del magma emesso (viscosità, contenuto in gas, composizione etc.).

In un *vulcano a condotto aperto* il magma arriva in superficie già degassato e viene emesso sotto forma di **lave**, masse magmatiche allo stato liquido-pastoso, ad alta temperatura, che scorrono lentamente sul suolo (pochi km/h) e, una volta solidificate, danno origine a rocce effusive come i basalti. Questa **attività effusiva** è tipica dei magmi basici ed è caratteristica delle zone di dorsale (si pensi all'Islanda, costruita proprio dalle lave emesse dalla dorsale medio-atlantica su cui è posizionata), dove dà origine a eruzioni sottomarine con produzione di nuova crosta oceanica, di alcune zone di intraplacca (esempi sono il *plateaux del Deccan*, enorme espandimento lavico dell'India emesso da fratture, ma anche vulcani centrali come l'Etna) e dei vulcani da *hot spot* (come le già citate isole Hawaii, che, formate unicamente da lave, assumono la forma di vulcani a scudo).

In un *vulcano a condotto chiuso* il magma arriva in superficie ancora ricco di gas ed esplode violentemente, innalzando una nube di ceneri e di gas, da cui cadono **prodotti piroclastici** come i lapilli e le bombe vulcaniche. Si tratta di magmi generalmente acidi, molto viscosi, che vengono emessi sotto forma di pomici e scorie di composizione riolitica, tefritica, fonolitica. Ne risulta dunque un'**attività esplosiva** durante la quale possono generarsi, per il collasso della colonna eruttiva o per altri processi più complicati, terribili flussi piroclastici, ossia miscele di gas e prodotti solidi, spesso ad alta temperatura, che viaggiano radenti al suolo con velocità fino ai 250-300 km/h ed hanno altissima forza distruttiva e letale per gli organismi viventi. Il vulcanismo esplosivo è tipico dei vulcani degli archi insulari o dei margini dei continenti vicini alle fosse oceaniche. Le forme tipiche degli apparati con attività esplosiva sono i coni vulcanici (di cui il Vesuvio è uno splendido esemplare, conosciuto in tutto il mondo per la sua eruzione del 79 d.C. che distrusse le città romane di Pompei, Ercolano e Stabia) e le caldere, crateri vulcanici di grosse dimensioni, formati per esplosione o sprofondamento di coni precedenti (un bellissimo esempio è la caldera del Krakatoa, localizzata tra Giava e Sumatra, che eruttò violentemente nel 1883 causando decine di migliaia di vittime).

La convivenza dell'uomo con i vulcani non è semplice: le eruzioni sono fenomeni naturali di straordinaria potenza che è impossibile fermare. L'unico modo per difendersi è prevederne con anticipo l'inizio per mettersi in salvo, mentre non c'è scampo per edifici e strutture che sorgono nei pressi dei vulcani. Eppure, nonostante si sappia che alcune zone sono esposte ad altissimo **rischio vulcanico** (inteso come il prodotto tra la probabilità che si verifichi un determinato evento e i danni ad esso conseguenti), esse risultano oggi densamente popolate ed edificate, a motivo della eccezionale fertilità dei suoli agrari (ricchi di potassio e altri minerali) e della loro alta permeabilità e facile lavorabilità. È il caso dell'area flegrea o di quella vesuviana in Campania. Il controllo dei vulcani attraverso il costante monitoraggio di alcuni parametri, quali la deformazione del suolo, le anomalie termiche e gravimetriche, la composizione delle fumarole, la sismicità (fenomeni precursori), permette oggi di prevedere comunque le eruzioni. Con una corretta informazione alle popolazioni a rischio e con la predisposizione di adeguati piani di

emergenza per le eventuali evacuazioni, è possibile credere che in futuro diminuiscano sempre più le vittime dei vulcani.

Tra gli effetti indiretti che le eruzioni hanno sulla vita dell'uomo va ricordata l'**influenza** che esse esercitano **sul clima**. In seguito a forti eventi esplosivi le ceneri disperse nell'atmosfera non soltanto causano fortissime piogge sulla regione interessata (in quanto fungono da nuclei addensativi per il vapore), che a loro volta possono dare origine a grossi eventi franosi, ma ostacolando il passaggio della radiazione solare possono portare ad un raffreddamento su scala planetaria.

E, infine, un breve cenno meritano i fenomeni di vulcanismo secondario che, se ben sfruttati, risultano economicamente vantaggiosi per l'uomo. È il caso delle **acque termali**, che sgorgano dal sottosuolo a temperatura elevata, in prossimità di focolai magmatici (in Italia sono noti i centri di Salsomaggiore e di Ischia); delle **acque minerali**, che, attraversando terreni vulcanici (come avviene per le acque nella zona del vulcano campano di Roccamonfina), si arricchiscono di elementi benefici all'uomo; dei **geysers** o **soffioni**, utilizzabili per ricavare energia geotermica (come accade nell'area di Lardarello, in Toscana) ma anche prodotti chimici, come acido borico e sali ammoniacali.

1 Il modellamento del rilievo terrestre

Il ramo della geografia fisica e della geologia che ha per oggetto lo studio del rilievo terrestre, delle sue forme e modificazioni, della sua evoluzione e delle cause che le determinano è definito **geomorfologia**. Le **forme del rilievo** si distinguono in **primarie**, se prodotte dalle forze endogene attraverso le attività tettoniche e vulcaniche, e in **secondarie** (o *derivate*), se risultano «rielaborate» per azione degli agenti esogeni. Questi ultimi, intervenendo dall'esterno sulla superficie terrestre, provocano il disfacimento delle rocce e la loro trasformazione in sedimenti; nella maggior parte dei casi provvedono, inoltre, con il contributo della forza di gravità, al loro trasporto e al successivo deposito.

Gli **agenti esogeni** (che corrispondono alle acque meteoriche, al vento, ai fiumi, ai ghiacciai e alle acque del mare) da una parte, quindi, demoliscono i rilievi, dando origine alle **forme di erosione**, dall'altra trasportano e accumulano altrove il materiale disgregato, producendo **forme di deposito**.

Il processo di erosione produce effetti diversi sui differenti tipi di rocce che, proprio in base alla loro reazione a tale fenomeno (proprietà definita *erodibilità*), si classificano in tenere e dure: in un paesaggio costituito da varie litologie si verifica, quindi, un'**erosione selettiva**, che lascia in rilievo i materiali duri, meno erodibili, e crea depressioni nelle rocce tenere, più erodibili. Spesso, poi, anche due rocce ad eguale litologia subiscono l'erosione in maniera diversa: se coesistono in un'unica zona, la causa va ricercata nelle differenze di struttura o di età delle rocce stesse; se invece sono localizzate in regioni diverse, il fenomeno è chiaramente imputabile a differenti situazioni climatiche, in quanto, come si vedrà meglio più avanti, le condizioni di temperatura e di umidità costituiscono dei fattori assai rilevanti nei processi di erosione.

2 I processi di degradazione delle rocce

Per *degradazione* si intende il complesso dei fenomeni di erosione, siano essi di tipo fisico e meccanico (disgregazione fisica) o di natura chimica (alterazione chimica), che producono il disfacimento delle rocce costituenti il rilievo terrestre e la formazione di **materiale detritico**.

La **disgregazione fisica** si esplica in diversi modi, ed ha come effetto la frammentazione della roccia. Nelle regioni a clima freddo, il processo più diffuso è il **crioclastismo**: l'azione del gelo è in grado di frantumare anche terreni estremamente compatti. L'acqua, infatti, si infiltra nei pori o nelle spaccature delle rocce e, con le basse temperature, forma cristalli di ghiaccio: l'aumento di volume origina una pressione verso l'esterno in grado di separare blocchi comunque già individuati da zone di debolezza.

Nelle regioni a clima arido, in cui è chiaramente assente il crioclastismo, può invece accadere che, all'interno delle cavità presenti nelle arenarie, si formino, per evaporazione dell'acqua, grosse quantità di cristalli di sale, la cui azione (*haloclastismo*) è paragonabile a quella del gelo: la pressione sulla roccia ne provoca la frammentazione.

Un fattore fondamentale della degradazione meteorica è costituito dalle variazioni di temperatura: il passaggio brusco e continuo da alte a basse temperature causa un'alternanza di

tensioni per dilatazione e contrazione all'interno della roccia, che, alla lunga, possono portare alla fratturazione, e quindi alla frantumazione (*termoclastismo*) [1].

Laddove invece si succedono stagioni piovose e stagioni secche, risulta diffusa la disgregazione delle rocce per **imbibizione** e **disseccamento**: tale processo è tipico delle argille, che mostrano una spiccata tendenza al rigonfiamento e alla contrazione in seguito all'assorbimento e alla perdita d'acqua. È infine da ricordare l'azione svolta dalle **radici delle piante**, in grado di allargare (ma, in alcuni casi, anche di creare) le fratture nelle rocce, fino a causarne la separazione in blocchi.

L'**alterazione chimica** delle rocce ad opera degli agenti esogeni si compie attraverso diversi processi (legati alla presenza di ossigeno, anidride carbonica e acqua) che ne modificano la struttura e la composizione.

L'**ossidazione** è dovuta all'interazione dell'ossigeno presente nell'aria e/o nell'acqua con rocce che contengono sostanze carboniose, composti ferrosi, zolfo e solfuri. Questo processo di alterazione si manifesta con una variazione del colore superficiale della roccia: così le rocce carboniose o bituminose assumono una tipica colorazione biancastra, mentre quelle ferrose si tingono di rosso e di giallo.

Il processo di **idrolisi** interessa le rocce composte da silicati: l'acqua, dissociata in ioni H^+ e OH^- , attacca le molecole degli allumosilicati e le decompone, facendo passare in soluzione alcuni ioni metallici e lasciando sul posto prodotti residuali [2]. Nelle regioni con clima caldo-umido, la decomposizione si rivela più spinta e porta alla separazione tra silice idrata, idrossidi di alluminio e idrossidi di ferro: la prima viene allontanata sotto forma di soluzione colloidale, mentre gli idrossidi restano come residui, a formare, rispettivamente, rocce bauxitiche e rocce lateritiche.

L'anidride carbonica presente nell'acqua piovana e nell'acqua del suolo esercita una forte azione corrosiva sulle rocce di origine calcarea o dolomitica. Il carbonato di calcio, insolubile, reagendo con la CO_2 , si trasforma in bicarbonato di calcio, solubile, che viene portato via dall'acqua: con questo meccanismo, conosciuto come **carsismo**, si riescono a formare, nel corso dei tempi geologici, anche enormi sistemi di cavità sotterranee.

Le acque meteoriche sono, infine, responsabili del processo di **idratazione** di alcuni minerali (come l'anidrite, che si trasforma in gesso, e come l'ematite, che diventa limonite): l'aumento di volume della roccia che li contiene provoca il fenomeno di degradazione.

I processi di degradazione delle rocce appena descritti interessano sì tutta la superficie del pianeta, ma risultano poco appariscenti all'occhio dell'uomo. Eppure si tratta di fenomeni fondamentali per l'evoluzione del paesaggio, in quanto preparano le rocce sia alla formazione del suolo sia all'azione dell'erosione da parte degli agenti esogeni.

3 Formazione e classificazione dei suoli

I **suoli** rappresentano l'espressione più superficiale della degradazione dei rilievi. La presenza, al loro interno, di frazione organica, originata dalla decomposizione di materia animale o vegetale (*humus*), è la caratteristica che li distingue dal detrito. Il suolo è frutto dell'azione di fattori fisici, chimici e biologici: quest'ultima componente manca, invece, nei processi di formazione del detrito. Un suolo formatosi solo grazie a fattori naturali viene definito **terreno vegetale** (o *terreno vergine*, o *terreno naturale*); quando interviene l'azione dell'uomo prende il nome di **terreno agrario** (o terreno coltivato).

[1] Spesso la frantumazione della roccia per termoclastismo avviene secondo superfici parallele a quella esterna, secondo un processo conosciuto come *esfoliazione*.

[2] Si tratta di nuovi silicati appartenenti alla famiglia delle argille (processo di *argillificazione*).

Gli elementi solidi che compongono il suolo sono il frammento roccioso (*frazione inorganica*) e i resti decomposti (*frazione organica*); la componente liquida è rappresentata da soluzioni complesse (soluzioni humifiche acide); il costituente gassoso è dato dalla combinazione di gas atmosferici e gas liberati dalle reazioni chimiche.

Tra le **proprietà fisico-chimiche** che caratterizzano i suoli, il **colore** è un parametro fondamentale per la loro classificazione. La causa della colorazione è nella presenza di *humus* (quanto più abbondante è, tanto più il suolo risulta scuro), di sali (carbonati, gessi, cloruri, che danno una tipica colorazione bianco-giallastra) e di composti di ferro e di alluminio (responsabili dei colori rossastri). L'abbondanza di **basi** allo stato libero, come Mg^{2+} , Ca^{2+} , Na^+ , K^+ , indica un terreno poco dilavato e, soprattutto, molto fertile (i colloidi provvedono a scambiare basi tra il suolo e le piante, che ne hanno bisogno per lo sviluppo); l'**acidità** rappresenta, invece, la concentrazione di ioni idrogeno (H^+) rispetto agli ossidrili (OH^-), ed è nota come pH del suolo. Un parametro estremamente importante è la **tessitura**, ossia la granulometria delle particelle che costituiscono il suolo. Un suolo molto fine (di tipo argilloso) è pessimo, perché difficilmente lavorabile e poco permeabile; ma un suolo troppo grossolano, e quindi altamente permeabile, è a sua volta poco produttivo in quanto non permette reazioni abbastanza lente da essere fruttuose. In generale, le tessiture di tipo misto rappresentano l'ideale per lo sviluppo delle piante. La **struttura** indica, invece, il tipo di aggregazione delle particelle che compongono il suolo: essa influenza la capacità di assorbimento dell'acqua, la predisposizione all'erosione e la sua lavorabilità. Infine, ciò che più di ogni altro fattore caratterizza un suolo è il suo **profilo**, cioè la suddivisione in strati, definiti **orizzonti**, a differente tessitura e colore. Dal basso verso l'alto troviamo [3] l'**orizzonte D**, che corrisponde al substrato roccioso non alterato; l'**orizzonte C**, costituito da roccia madre alterata, ossia da detrito pronto alla trasformazione in suolo; gli **orizzonti B**, detti anche di illuviazione, molto compatti, ricchi in colloidi, più colorati rispetto agli altri strati; gli **orizzonti A** che si distinguono in: **A₀₀**, (il livello più superficiale, dato dall'accumulo di foglie), **A₀** (con detriti organici in via di decomposizione), **A₁** (ricchissimo di *humus* e quindi molto scuro), **A₂** (povero di sostanza organica e, di conseguenza, molto chiaro, noto anche come orizzonte di eluviazione).

I fattori principali che condizionano lo sviluppo di un suolo (**fattori pedogenetici**) sono:

- la **roccia madre**. Il substrato roccioso, e di conseguenza il detrito, costituisce il materiale-base di un suolo. Tale fattore è determinante per la natura del suolo soltanto al momento di inizio della *pedogenesi* [4], mentre ha scarsa importanza per suoli già maturi;
- la **topografia**. Questo fattore influenza essenzialmente la velocità dei fenomeni di erosione e di circolazione dell'acqua. Ad esempio: laddove il detrito presenta spessori maggiori, si formano suoli più potenti; in prossimità di un fiume il terreno risulta più ricco d'acqua; su un versante poco inclinato l'evoluzione del suolo si compie più rapidamente che su un versante acclive e così via;
- il **tempo**. La formazione di un suolo non è un processo immediato, ma implica un tempo di evoluzione in un periodo di stabilità geomorfologica. L'inizio della *pedogenesi* corrisponde al tempo-zero: occorrono millenni per giungere ad una configurazione stabile, con gli orizzonti ben definiti (*suolo maturo*). Risulta chiaro, quindi, che l'età di un suolo non corrisponde a quella della roccia madre, da cui ha avuto origine il detrito;
- il **clima**. Rappresenta il fattore più influente sulla *pedogenesi*, tanto che si parla di **suoli zonali**. Il regime della temperatura e delle precipitazioni condiziona fortemente i processi di formazione e di evoluzione: il trasporto e l'accumulo dei colloidi e degli ioni, infatti, avvengono attraverso le acque che circolano nel terreno, così come l'attività chimica e quella dei batteri sono favorite dalle alte temperature;

[3] Secondo il *Department of Agriculture* degli Stati Uniti, 1938.

[4] Per *pedogenesi* si intende la genesi del suolo e la sua evoluzione.

- i **fattori biologici**. Gli animali esercitano sul suolo un'azione essenzialmente meccanica di compattazione e di rielaborazione, sia dall'esterno (animali selvatici e di allevamento) sia dall'interno (vermi, formiche, animali scavatori). L'azione delle piante è dovuta non soltanto ad alberi, arbusti ed erbe (*macroflora*), che scambiano continuamente con il terreno elementi minerali e forniscono materia organica per l'*humus*, ma anche a batteri e funghi (*microflora*), che invece distruggono l'*humus* e fissano l'azoto dell'aria.

I fattori pedogenetici appena descritti, siano essi di natura passiva o di natura attiva, agiscono contemporaneamente e ciascuno interferisce con l'azione dell'altro: il risultato è una straordinaria varietà di suoli, **difficilmente classificabili** in un ristretto numero di gruppi. Per semplificare il discorso si può fare riferimento alla distinzione fondamentale tra **suoli zonali**, legati cioè a zone climatiche e vegetazionali ben definite, e **suoli azonali**, determinati, invece, da condizioni strettamente locali. Nel primo caso si parla di **regimi pedogenetici**: elencheremo di seguito i principali.

Il regime della **podsolizzazione** è tipico dei climi con precipitazioni abbondanti, che garantiscono acqua sufficiente per lo sviluppo delle conifere, e temperature basse, che non permettono l'azione dei batteri. Il risultato è un **podsol** (o *podzol*), suolo particolarmente acido, caratterizzato da abbondante *humus*, da orizzonti A di colore chiaro e con intensa eluviazione, e da orizzonti B ricchi, invece, di ferro e di alluminio e di colorazione scura. Il regime della **laterizzazione** contraddistingue le zone a clima umido e caldo: l'*humus* manca perché si decompone molto velocemente a causa dell'intensa azione batterica, gli orizzonti B risultano assenti perché sali e colloidali vengono rapidamente allontanati dalle acque, mentre è ben sviluppato l'insieme degli orizzonti A, in cui si concentrano gli ossidi e gli idrati ferrosi e ferrici (*lateriti*), responsabili del colore rosso scuro del suolo. Nelle regioni in cui l'evaporazione in media supera le precipitazioni è comune il processo di **calcificazione**: le basi vengono dilavate dalle acque e nel suolo si concentrano ioni di calcio e di magnesio. L'*humus* risulta abbondante e negli orizzonti B si formano noduli, lamine e straterelli di bicarbonato di calcio (*caliche*). Il regime della **salificazione** (o *salinizzazione*) è tipico dei climi molto caldi e secchi: si formano suoli ricchi in solfati e cloruri di calcio e di sodio. Laddove non c'è presenza di sali, le temperature sono basse e le precipitazioni scarse, si imposta il regime della **gleizzazione**, con abbondante *humus* che forma un potente strato di torba, sotto il quale si individua un livello argilloso, plastico, compatto, di colore grigio-blu, noto come *orizzonte a gley*.

Un suolo rappresenta dunque un importantissimo elemento del paesaggio, in quanto non soltanto testimonia un periodo di stabilità geomorfologica, ma indica chiaramente le condizioni climatiche e, più in generale, ambientali, in cui esso stesso si è formato e si sta evolvendo. In ugual modo, suoli formati in passato su antiche superfici topografiche (**paleosuoli**) forniscono utilissime indicazioni sui paleoclimi e sui paleoambienti.

4 Modellamento dei versanti e dilavamento

Le rocce disgregate e alterate dai processi di degradazione possono dare origine ai suoli, come si è appena descritto, ma possono anche essere rimosse, soprattutto se si trovano su superfici inclinate (*pendii*). Tale rimozione si compie ad opera della forza di gravità (*trasporto gravitativo*) o per mezzo dell'acqua, con **processi** che complessivamente vengono definiti di **denudazione**; può interessare un intero versante (*erosione areale*) oppure soltanto delle fasce (*erosione lineare*). Il trasporto gravitativo raramente avviene con uno spostamento «granulo a granulo» del detrito [5]: generalmente si verifica un movimento di massa, che può rivelarsi lento o rapido.

[5] I *coni* e le *falde detritiche* rappresentano due esempi di trasporto ed accumulo «granulo a granulo». I clasti si staccano singolarmente ed indipendentemente dalla parete di un versante e si accumulano, trasportati dalla gravità, ai suoi piedi. Generalmente i frammenti più grossolani arrivano più in basso, mentre quelli più fini si arrestano prima: in ogni caso il movimento si blocca quando tutti i clasti raggiungono un profilo di equilibrio.

Il **soil creep** è un movimento di massa lento, che avviene per strisciamento (*reptazione*) del mantello detritico di un pendio. Le cause di tale spostamento possono essere varie e combinarsi assieme: la circolazione d'acqua nel terreno, le variazioni di volume dovute ai fenomeni di imbibizione e disseccamento o al gelo-disgelo, l'azione degli animali o delle piante. Il *creep* interessa materiale a granulometria alquanto grossolana, privo quindi di matrice fine limosa-argillosa: in tal caso, infatti, prevale il movimento detto **soliflusso**. Quest'ultimo consiste in una sorta di colamento lento di materiale molto viscoso ed è tipico delle rocce argillose che subiscono dilatazione in seguito all'assorbimento di acqua. Una morfologia del terreno a lobi e alberi e/o paletti tiltati verso valle indicano chiaramente un fenomeno di *creep*, o di soliflusso, in atto. In regioni con climi freddi, caratterizzate dalla presenza di ghiaccio sul suolo, il soliflusso prende il nome di **geliflusso**. La differenza tra i due fenomeni non risiede nel tipo di movimento, ma nella causa di innesco, che, nel primo caso, come si è detto, è l'acqua di infiltrazione, nel secondo è invece l'acqua ottenuta dallo scioglimento del ghiaccio. È utile ricordare che tutti i movimenti di massa lenti sono di tipo areale, coinvolgono cioè vaste porzioni di pendio, si svolgono tutti con ritmi uniformi, costanti, e necessitano di pendenze anche minime ($<10^\circ$).

I movimenti gravitativi di massa rapidi sono conosciuti con il nome di **frane**. Si tratta, in realtà, di un termine generico che accomuna fenomeni anche molto diversi tra loro per genesi, caratteristiche ed impatto sul territorio. In una frana si distinguono una zona di distacco (**nicchia**), di forma varia, una zona di movimento lungo il pendio (**alveo di frana**), e una zona, generalmente ai piedi del pendio, di accumulo del detrito o della roccia (**accumulo di frana**). Un fenomeno franoso si innesca quando in un versante si raggiunge la rottura di equilibrio; ciò generalmente si verifica a causa dell'intervento di un fattore esterno, che fa diminuire la resistenza delle rocce. Il movimento vero e proprio avviene, invece, sotto l'azione della forza di gravità.

La denudazione dei versanti ha luogo, come si è detto, anche per mezzo delle acque, con un processo che si definisce di **dilavamento**. L'acqua piovana che raggiunge il suolo, e non si infiltra in profondità a raggiungere la falda freatica, scorre superficialmente, seguendo la direzione di massima pendenza. Questo fenomeno di **ruscellamento** esplica, anche su pendii non eccessivamente inclinati, un'azione di erosione, trasporto e deposito, legata proprio all'energia dell'acqua che si trasmette sul terreno. Già il semplice impatto delle gocce di pioggia sul terreno ha potere erosivo, riuscendo a staccare granuli dal suolo (*splash erosion*); con il ruscellamento diffuso si arriva all'erosione areale (*sheet erosion*). Quando l'acqua comincia a concentrarsi, si formano prima piccoli rivi, che erodono secondo linee di scorrimento preferenziale (*rill erosion*), poi flussi più intensi che scavano solchi e fossi, producendo un'erosione lineare più spinta (*gully erosion*). È in questo modo che pian piano nasce un bacino imbrifero, il quale raccoglie le acque di alimentazione di un fiume principale. La pendenza del rilievo, la natura del suolo e delle rocce, la copertura vegetazionale, l'intensità e la durata delle precipitazioni, costituiscono tutti fattori che controllano ed influenzano lo scorrimento superficiale dell'acqua e la sua infiltrazione. Ad essi vanno aggiunti anche fattori antropici che intervengono, in alcuni casi, ad aggravare i fenomeni erosivi (attraverso azioni quali il disboscamento o le colture intensive), in altri a diminuire il dilavamento (come con la sistemazione dei pendii o con l'incanalamento delle acque).

Tra le forme più caratteristiche generate dai processi di dilavamento ricordiamo i **calanchi**, incisioni lunghe, profonde e ramificate che solcano i pendii argillosi di molti versanti italiani [6], soprattutto dell'Appennino Toscano e di quello Lucano, e le **piramidi di terra**, colonne di roccia tenera sormontate da massi tabulari di roccia più resistente, di cui si possono ammirare degli splendidi esemplari a Renón, presso Bolzano, e a Segonzano, vicino Trento.

[6] Generalmente i *calanchi* si ritrovano sui versanti esposti a sud dei rilievi e mancano invece su quelli rivolti a nord. Secondo i geomorfologi, probabilmente, il fenomeno si spiega con la maggiore insolazione dei primi rispetto ai secondi.

Sotto gli effetti di un processo di denudazione, un pendio tende a raggiungere il suo profilo più stabile, la forma di equilibrio in grado di contrastare l'ulteriore azione di quel determinato processo. Se non mutano le condizioni climatiche, tettoniche o ambientali, e se non intervengono fattori esterni, di cui spesso è artefice anche l'uomo, l'evoluzione del pendio si compie con un ritmo regolare e costante, e si conclude con la formazione di un **versante regolarizzato**, quasi perfettamente rettilineo, inclinato intorno ai 45°, estremamente stabile. In Italia versanti di questo tipo sono diffusi nelle Prealpi e lungo tutto l'Appennino.

L'evoluzione di un versante regolarizzato prosegue a seconda del tipo di roccia che lo costituisce, delle condizioni climatiche e vegetazionali, dei movimenti tettonici, dell'intervento di agenti di erosione e di trasporto. Nelle regioni a clima semiarido si riscontra spesso un **arretramento parallelo** (*retreat*) del versante, sempre che alla base del pendio operi un agente di trasporto per allontanare continuamente il detrito prodotto dalla denudazione. Se ciò non avviene, il detrito accumulato ostacola l'erosione al piede: prevale allora un fenomeno di **addolcimento** (*decline*), che porta ad una complessiva diminuzione di pendenza, con la formazione di un versante convesso-concavo.

5 Le frane

Una **frana** consiste in un qualsiasi movimento di terra o di roccia da un versante, che avvenga con distacco improvviso o progressivo, con moto rapido o lento, e sotto l'effetto della forza di gravità. In una frana si distinguono una **zona di distacco**, una **zona di movimento** ed una **zona di accumulo**. La zona di distacco, una volta che si è verificata la frana, si presenta come un'incavatura sul versante, una nicchia di distacco circondata da una scarpata, cioè da una parete rocciosa quasi verticale. Eventuali fessure e crepe che contornano la zona stanno ad indicare un ulteriore imminente franamento. La zona di movimento, definita alveo di frana, consiste in un corridoio, più o meno lungo e più o meno largo, anch'esso incavato, all'interno del quale discende il materiale franato. In alcuni casi esso può mancare e si registra, dunque, un passaggio diretto dalla nicchia alla zona di accumulo. In quest'ultima si ammassa, caoticamente, il materiale franato: la sua posizione dipende dalla velocità, cioè dall'energia cinetica, del movimento. Generalmente il cumulo si arresta lungo il versante o ai suoi piedi, ma in alcuni casi è possibile che risalga addirittura sul versante opposto (così è accaduto per il cumulo della frana di Scanno, in Abruzzo, che ha saturato completamente la valle ai piedi del versante, formando un lago di sbarramento).

Sono stati fatti vari tentativi di **classificazione delle frane**, basandosi su diversi criteri (effetti indotti, cause, pericolosità, fattori morfologico-strutturali etc.). In realtà per ogni movimento franoso esiste un gran numero di variabili da considerare e ciò irrimediabilmente rende molto complicato un raggruppamento in classi. Una descrizione esauriente di una frana dovrebbe quantomeno indicarne la posizione (di versante o di piede), lo stato di attività (inattiva, quiescente, attiva), l'età (antica, recente, attuale), i materiali coinvolti (roccia, detrito grossolano, terra fine) e il tipo di movimento. Riguardo a quest'ultimo punto si riporta di seguito la classificazione assai semplice e chiara del Castiglioni (1991), che si ispira ad alcuni studi del Desio.

Le **frane di crollo** rappresentano un distacco improvviso di roccia o terra, da un versante, con caduta libera di materiali in verticale. La velocità risulta estremamente elevata e il movimento termina con il raggiungimento di una superficie pianeggiante. Questo tipo di frana spesso caratterizza le rocce fortemente fratturate e i versanti molto acclivi; le falesie di tufo della Penisola Sorrentina (Campania) sono frequentemente soggette a fenomeni di crollo, esponendo ad un elevato rischio residenti e turisti, che soprattutto d'estate affollano queste zone.

Nelle **frane di scorrimento rotazionale** il distacco avviene lungo una superficie curva di neoformazione: la rotazione si verifica in contropendenza, descrivendo un arco di cerchio rivolto verso l'alto. Generalmente i materiali maggiormente soggetti a tale movimento sono le terre fini, soprattutto se costituite da alternanze di strati.

Quando il movimento si compie lungo superfici di strato, o strutturali, già esistenti, le **frane** si dicono **di scivolamento** (o *di scorrimento traslativo*). Il distacco può essere improvviso e la velocità del corpo di frana risulta tanto maggiore quanto più è inclinato il versante.

Nelle **frane di colata** (o *colamento*) non c'è l'individuazione di una vera e propria superficie di rottura: l'acqua piovana o di falda imbeve i terreni (generalmente argille, ma in alcuni casi anche sabbie o piroclastiti), ne riduce la coesione e li fa muovere come un fluido plastico-viscoso. Lungo i versanti dell'Appennino, caratterizzati da una forte presenza di terreni argillosi, le frane di colata hanno un'ampia diffusione. Quando il movimento è improvviso e rapido, il fenomeno può risultare estremamente pericoloso.

Qualunque sia il tipo di frana, con il passare del tempo, gli agenti esogeni ne modellano sia la nicchia di distacco sia il cumulo, impedendo così, spesso, il riconoscimento immediato del fenomeno avvenuto, che tuttavia non può e non deve sfuggire ad un'analisi accurata del territorio. La conoscenza della **franosità avvenuta** si rivela, infatti, fondamentale per stabilire la **franosità temuta**, che tiene appunto conto delle frane che si possono riattivare e di quelle che si possono verificare ex novo sui versanti. Le moderne tecniche di fotointerpretazione e di telerilevamento si sono dimostrate estremamente utili nel campo della geomorfologia e in particolare per l'individuazione e lo studio delle frane. L'uso di Sistemi Informativi Territoriali consente inoltre di elaborare carte di franosità di un'area, periodicamente aggiornabili con estrema facilità, in cui siano mappati tutti i dissesti in atto e quelli potenziali e le zone in erosione accelerata, con le relative cause di natura geologica ed idrogeologica.

Alle cause naturali delle frane, legate alla litologia, alla morfologia e alla struttura dei terreni, all'azione dell'acqua e a fenomeni tettonici, come i sismi, in grado di rompere l'equilibrio delle rocce, va aggiunto l'intervento dell'uomo sul territorio che spesso è determinante per l'innescio di tali fenomeni. L'apertura di cave, la costruzione di strade, gallerie, ponti, gli interventi sui fiumi, la distruzione della vegetazione, l'appesantimento del terreno con costruzioni di diverso tipo, rappresentano intense modificazioni dei versanti, in grado di alterarne profondamente e irrimediabilmente gli equilibri. Negli ultimi anni una maggiore attenzione da parte di Autorità e cittadini verso le problematiche ambientali e i temi di protezione civile hanno limitato i disastri di origine antropica e hanno posto rimedio alle situazioni potenzialmente pericolose. La sistemazione delle frane, attraverso gli interventi sulle possibili cause che potrebbero rimetterle in movimento, il controllo costante delle aree maggiormente esposte a questo tipo di rischio (attraverso reti di monitoraggio dotate di pluviometri, estensimetri, inclinometri etc.), il rispetto dei vincoli idrogeologici nell'edilizia, la predisposizione di opportuni sistemi di allarme per la popolazione in caso di emergenza, costituiscono degli importanti strumenti predisposti dall'uomo per difendere se stesso e il territorio da questo tipo di calamità.

6 Morfologia fluviale ed erosione

Le acque costituiscono l'agente atmosferico principale per il modellamento della superficie terrestre. All'azione di dilavamento prodotta sui versanti dalle acque meteoriche, si aggiungono i processi di erosione, trasporto e sedimentazione da parte delle acque incanalate che formano un **reticolo** (o sistema) **idrografico**.

L'azione morfologica esercitata da un fiume si deve all'**energia** dell'acqua che scorre nell'alveo fluviale, in parte spesa per vincere gli attriti interni (viscosità) ed esterni (rugosità del letto, resistenza dell'aria), in parte per trasportare i sedimenti, in parte per erodere ulteriore materiale dal fondo [7]. Laddove il fiume ha elevati valori di pendenza e di velocità, l'energia risulta sufficiente

[7] Oltre all'erosione di tipo «meccanico», le cui forme tipiche sono le *marmitte dei giganti*, cavità cilindriche scavate nel letto, le acque fluviali sono anche in grado di degradare le rocce «chimicamente», con processi quali l'idrolisi.

per l'azione di erosione e di trasporto; laddove invece pendenza e velocità sono ridotte, l'energia non basta nemmeno per il trasporto e i sedimenti vengono, quindi, depositati.

Il **trasporto** avviene in **sospensione** o in **soluzione**, per i materiali fini e finissimi (che sono causa della torbidità delle acque), oppure per **trascinamento** (rotolamento, slittamento e salvezza), per i clasti più grossolani, che risultano così arrotondati e levigati. Si spiega in questo modo la forte selezione granulometrica dei depositi fluviali. La **portata solida** corrisponde alla quantità di materiale trasportato che passa per una sezione del fiume nell'unità di tempo, mentre la **capacità di trasporto** (o *carico limite*) è la massima quantità di materiale che un corso d'acqua può trasportare sul fondo.

Per capire come agiscono i processi di **erosione fluviale** vanno fatte alcune premesse. In linea generale va detto che l'azione erosiva tende a regolarizzare i corsi d'acqua, eliminando col tempo tutti i gradini presenti sul fondo, dovuti alla coesistenza di rocce a diversa litologia ed erodibilità. Il fiume raggiunge il suo **profilo di equilibrio** (o *regolarizzato*) quando le pendenze di ogni suo tratto garantiscono un'energia tale da vincere unicamente le forze di resistenza e di attrito, senza cioè che vi sia erosione, ma solamente scorrimento dell'acqua. Non si tratta tuttavia di una situazione definitiva né immutabile, a causa degli innumerevoli fattori che possono influenzare la dinamica fluviale, come le variazioni del livello di base [8], i movimenti tettonici, gli sbarramenti del corso d'acqua, le variazioni di portata o di trasporto solido, i cambiamenti del reticolo idrografico, gli interventi dell'uomo, l'evoluzione generale del rilievo. Tutti questi fattori incidono essenzialmente sulla pendenza del fiume e quindi sull'energia dell'acqua: di conseguenza influenzano i processi di erosione, di trasporto e di sedimentazione.

Anche il fenomeno dell'**erosione regressiva** è finalizzato al raggiungimento dell'equilibrio. L'approfondimento del letto in un tratto di fiume fortemente inclinato provoca una diminuzione di dislivello rispetto al tratto più a valle ed una accentuazione rispetto a quello più a monte: l'erosione, in tal caso, si sposta verso monte, con un processo appunto regressivo. Tale fenomeno è osservabile attualmente nelle Cascade del Niagara, dove il salto, pur mantenendo la sua altezza, risulta in continuo arretramento.

L'**erosione fluviale** si esercita sia in profondità, sia lateralmente. L'erosione laterale provoca lo scalzamento delle sponde del corso d'acqua e la formazione delle scarpate d'erosione fluviale; l'erosione in profondità incide verticalmente. Le forme di erosione fluviale generalmente nascono dall'interazione tra i due fenomeni, con la prevalenza ora dell'uno, ora dell'altro.

Nel caso dell'evoluzione dei **meandri**, anse regolari che si susseguono regolarmente lungo il corso di un fiume, prevale l'erosione laterale: lo scalzamento della riva concava e il conseguente accumulo sulla riva convessa provocano la migrazione, generale e lenta, di tutte le anse verso valle. Quando due anse si avvicinano particolarmente, può avvenire che il fiume abbrevi il suo percorso, «saltando» un lobo (*salto del meandro*): in tal caso nel tratto abbandonato (*meandro morto*) l'acqua ristagna e può dare origine ad una palude (*lago di meandro*). Sebbene tali forme siano caratteristiche di zone subpianeggianti (*meandri liberi*), dove il fiume ha tutto lo spazio per divagare dal suo corso rettilineo, cercando il percorso più semplice (attraverso cioè le rocce più tenere all'erosione), non è raro trovarle anche in strette valli, in cui anche i versanti curvano in anse concave e convesse (*meandri incassati*) [9].

[8] Il *livello di base* di un fiume corrisponde all'altitudine della sua foce. Esso generalmente coincide con il livello medio del mare. Ma, a livello locale, può essere costituito anche da un lago. In ogni caso è ad esso che si raccorda il profilo dell'acqua, ed è in funzione di esso che si effettua l'erosione a monte.

[9] Un esempio spettacolare di *meandri incassati* si può ammirare lungo il fiume San Juan, nello Utah (Stati Uniti). La loro origine è spiegata con un fenomeno di sollevamento progressivo della regione e di conseguente abbassamento relativo del livello di base. Per adattarsi ad esso il fiume ha continuato ad approfondirsi, senza però abbandonare il suo percorso sinuoso a meandri.

Nel caso delle **valli fluviali** prevale invece l'incisione verticale del fiume, che agisce contemporaneamente ai processi di denudazione dei versanti. La caratteristica di queste valli è la sezione trasversale a V, prodotta da un fondo molto stretto e pareti assai acclivi. Nel corso del tempo, per un abbassamento del livello di base, si può avere un approfondimento dell'alveo del fiume che va ad incidere il suo antico letto. Ai lati della nuova incisione restano due superfici pianeggianti, alla stessa altezza, che sono definite **terrazzi fluviali orografici**, se costituiti da roccia, oppure **alluvionali**, se costituiti da materiali detritici (sedimenti del fiume, detti alluvioni). In una stessa valle si possono individuare vari ordini di terrazzi: ognuno di essi, dal più recente, al fondo, al più antico, posto più in alto sul versante, rappresenta un diverso livello di base.

Anche la **sedimentazione** da parte di un corso d'acqua produce forme caratteristiche. Il deposito avviene generalmente quando diminuisce la pendenza del fiume. Vengono abbandonati prima i clasti di dimensioni maggiori (ciottoli), poi il detrito più fine (sabbia, limo); gli spessori dei depositi diminuiscono con la distanza dal fiume stesso e generalmente hanno un aspetto lenticolare, che segue l'andamento della corrente. L'insieme dei sedimenti depositati dal fiume (alluvioni) costituisce una **pianura alluvionale**. Immediatamente allo sbocco del corso d'acqua in una zona pianeggiante si formano invece le **conoidi di deiezione** (o **alluvionali**), ventagli convessi di sedimenti con l'apice verso monte. Se lo sbocco avviene in mare ha origine un **delta**, la cui forma viene però spesso modificata dalle correnti, dalle maree o dal moto ondoso. Ve ne sono di grandissimi: basti ricordare il delta del Fiume Giallo e quello famosissimo del Nilo.

7 Le alluvioni

Quando la portata di un fiume non può essere contenuta all'interno del suo alveo naturale, e l'acqua è costretta a straripare dagli argini, si verifica un evento di **piena**. La causa scatenante corrisponde ad un periodo di intense precipitazioni, mentre la conseguenza immediata consiste nell'allagamento dei terreni circostanti il corso d'acqua, con un fenomeno che è definito **alluvione**.

Nelle regioni a clima umido, caratterizzate da una stagione fortemente piovosa, le alluvioni risultano abbastanza frequenti, facilmente prevedibili e, tutto sommato, innocue, in quanto inondano le pianure alluvionali, destinate a foreste o a coltivazioni adattate a sopportare grandi carichi d'acqua. Invece, nelle regioni caratterizzate da climi instabili, come quello mediterraneo o tropicale, le alluvioni sono fenomeni più rari ma estremamente pericolosi: in primo luogo perché possono allagare anche terreni che si trovano a quote elevate, poi perché in molti casi si rivelano eventi non prevedibili. Nascono in seguito a precipitazioni definite «eccezionali», ossia ben al di sopra dei valori registrati, che possono aumentare la portata di un fiume anche di 20-50 volte rispetto a quella media. Le conseguenze sono spesso disastrose: coltivazioni prima allagate poi ricoperte da sabbie e limo, abitati spazzati via dalla forza dell'acqua, decine, centinaia, a volte migliaia di vittime. Basti ricordare la terribile alluvione del Fiume Giallo, in Cina, nel 1887, che causò oltre un milione di morti.

Se la dinamica di ogni fiume e torrente potesse essere continuamente monitorata, così come avviene per i corsi d'acqua principali, le piene e le alluvioni potrebbero sempre essere previste. Con i dati forniti dagli **idrometri**, strumenti che misurano il livello dell'acqua all'interno dell'alveo, è possibile, infatti, individuare il **livello di esondazione**, cioè il livello critico al di sopra del quale si ha tracimazione, e calcolare la propagazione dell'**onda di piena**; una volta previsti l'inizio e il successivo sviluppo dell'evento, si potrebbe procedere all'allertamento della popolazione ed evitare almeno la perdita delle vite umane.

Sebbene la causa di questi fenomeni sia essenzialmente di origine naturale (sono le intense precipitazioni a causare le piene), c'è da dire che spesso l'uomo aggrava la situazione: gli interventi sugli alvei, che variano il naturale regime dei corsi d'acqua, e il disboscamento, che rende più vulnerabile e meno assorbente il suolo, costituiscono due esempi a proposito. L'uomo ha

però inventato anche dei sistemi artificiali per evitare lo sviluppo della piena: i più riusciti sono i **serbatoi di piena** e gli **scolmatori**. I primi consistono in dighe che trattengono a monte una parte delle acque piovane per evitare che vadano ad accrescere la portata fluviale; i secondi sono canali o gallerie che, in caso di ingrossamento dei fiumi, prelevano parte delle acque e le smaltiscono lontano dall'alveo.

8 Il carsismo

Per **carsismo** si intende l'insieme dei processi di corrosione chimica provocati dalle acque meteoriche su rocce solubili come i calcari, le dolomie, i gessi e i depositi salini, che, di conseguenza, assumono forme di erosione caratteristiche. Tale fenomeno è particolarmente evidente nei calcari, i cui massicci sono più estesi e più ampi rispetto a quelli delle altre rocce solubili.

La **reazione chimica** che è alla base del processo consiste nella trasformazione del carbonato di calcio, insolubile, in bicarbonato di calcio, solubile, per azione dell'acqua acida (ossia combinata con l'anidride carbonica proveniente dall'atmosfera):



Il bicarbonato disciolto viene portato via dall'acqua, mentre i minerali insolubili, contenuti nelle rocce carbonatiche, si accumulano nelle depressioni, formando le tipiche terre rosse, di natura argillosa, che caratterizzano i paesaggi carsici. La reazione può avvenire anche in senso inverso, sebbene più lentamente e solo in ambienti specifici: il bicarbonato di calcio in soluzione può, cioè, precipitare come carbonato di calcio, producendo concrezioni calcaree.

L'intensità dell'azione carsica dipende essenzialmente dall'acqua (quantità, velocità, turbolenza, temperatura), dalla percentuale di anidride carbonica, dalla composizione e dalla struttura (discontinuità, fratture etc. che favoriscono l'infiltrazione dell'acqua) della roccia, dalla temperatura. A parità delle altre condizioni, il carsismo risulta più sviluppato nelle regioni a clima tropicale, in quanto l'umidità e il calore favoriscono la reazione chimica.

Le forme prodotte dal carsismo si raggruppano in due classi principali, le *ipogee* (forme carsiche sotterranee) e le *epigee* (forme carsiche di superficie), sebbene in realtà entrambe derivino spesso da un unico fenomeno.

La **morfologia epigea** è rappresentata da **microforme** e macroforme. Alle prime appartengono: i **karren** e le **scannellature carsiche**, rispettivamente solchi più grandi e più piccoli, simili a valli fluviali in miniatura, che tagliano il terreno roccioso, e le **vaschette di corrosione** (*kamenitza*), piccole cavità, poco profonde, più larghe sul fondo che in superficie. Le **macroforme** epigee più tipiche sono le **doline**, depressioni chiuse, di forma più o meno circolare, con dimensioni comprese tra alcuni decimetri e parecchie centinaia di metri, di profondità variabile. Si formano per dissoluzione ma anche per crollo, a causa dello sprofondamento del suolo sospeso su cavità carsica sotterranee. Non contengono acqua, perché essa, attraverso le pareti altamente permeabili, o attraverso un solo condotto principale (**inghiottitoio**), si infila in profondità. Con il tempo, tuttavia, è possibile che le argille di decalcificazione ostruiscano i condotti carsici al loro interno: in questo caso le acque riempiono le doline e hanno origine stagni o paludi. Spesso le doline si dispongono lungo i sistemi di frattura delle rocce e risultano quindi tutte allineate; quando invece crescono l'una accanto all'altra, si possono unire a dare forme particolari, definite **uvale**.

I **polije** (voce serbo-croata che significa *pianura*) sono invece depressioni più grandi e più irregolari, vere e proprie conche carsiche, con fondo piatto, costellato di doline, e versanti ripidi. Generalmente si impostano su fosse di origine tettonica, poi modellate dall'erosione. Hanno un'ampia diffusione sulle Alpi Dinariche, dove raggiungono lunghezze e larghezze di chilometri, ma di più piccole se ne trovano anche in Italia (si pensi a Campo Imperatore sul Gran Sasso, in Abruzzo).

Sebbene le valli non siano considerabili come forme carsiche, in quanto sono sempre il risultato dell'erosione di un corso d'acqua superficiale, assumono tuttavia delle caratteristiche particolari in ambiente carsico che giustificano in qualche modo l'appellativo di **valli carsiche**. In alcuni casi può mancare completamente il corso d'acqua sul fondo, che appare cosparso unicamente di punti di assorbimento (**valle morta**); oppure la valle può risultare occlusa in un senso, mancando della testata, perché il fiume nasce bruscamente da una sorgente carsica (**valle cieca**); o ancora può accadere che il fiume venga inghiottito in profondità prima di raggiungere la foce: mentre la valle a monte dell'inghiottitoio si approfondisce (*valle cieca*), il tratto oltre l'inghiottitoio resta stabile e senza acqua (*valle morta*) [10].

Le forme epigee appena descritte sono tutte imputabili al fenomeno dell'erosione carsica, ma esistono anche forme che nascono dalla precipitazione e dalla ricristallizzazione di carbonato di calcio. Il travertino, una tra le rocce sedimentarie più diffuse, ad esempio, consiste in concrezioni calcaree, che inglobano anche resti di organismi vegetali.

Le **grotte** rappresentano le forme tipiche del carsismo ipogeo. Si tratta di cavità orizzontali (*gallerie*) o verticali (*pozzi*) scavate dall'azione chimica e meccanica delle acque sotterranee. L'infiltrazione si spinge in profondità attraverso le fratture, fino a raggiungere gli strati di roccia non erodibili (**livello di base carsico**). Il primo risultato del carsismo consiste nella formazione dei fusi, piccole grotte allo stato iniziale, sviluppate prevalentemente in verticale. Il perpetrarsi dell'azione causa il crollo delle pareti che separano i vari fusi e dà quindi origine a sistemi complessi di gallerie. Se si verifica il crollo del tetto si genera un **pozzo carsico**. Il forte sviluppo delle gallerie in profondità e, di contro, le dimensioni estremamente ridotte delle aperture che mettono in comunicazione tali cavità con l'esterno, si spiega con il meccanismo stesso del carsismo ipogeo: il modellamento avviene infatti dal basso verso l'alto, con un processo definito di **erosione inversa**.

Le grotte assumono forme e dimensioni assai varie, in relazione alla modalità di circolazione delle acque. Ricordiamo tra le più profonde, la grotta della Pierre-Saint Martin (>1.200 m) in Francia, e, tra le più lunghe, la grotta del Mammut (circa 250 km) negli Stati Uniti. In Italia restano celebri le grotte di Postumia, ai confini tra Italia e Slovenia, che si allungano per circa 22 km, e la grotta di Monte Cucco, in Umbria, che raggiunge profondità superiori ai 900 m [11].

La straordinarietà di queste grotte è accentuata dalla presenza di spettacolari concrezioni che sono definite **stalattiti**, se pendono dal soffitto, e **stalagmiti** se si innalzano dal pavimento. Esse rappresentano le forme ipogee di deposito carsico: l'acqua che cola dalla volta delle grotte perde anidride carbonica e deposita, sul soffitto, piccoli anelli di carbonato di calcio, che col tempo formano una specie di tubicino che si ingrossa e si allunga fino ad assumere una struttura a tronco d'albero (*stalattite*); se la goccia d'acqua raggiunge il pavimento, la concrezione che si forma risulta più irregolare (*stalagmite*). All'interno delle grotte possono scorrere anche veri e propri fiumi: generalmente si tratta di corsi d'acqua catturati dagli inghiottitoi superficiali, che percorrono alcuni chilometri in sotterraneo per poi risorgere, improvvisamente, in superficie. Un esempio di questo fenomeno di risorgenza riguarda il fiume Timavo, in Venezia Giulia, che si inabissa nella Grotta di San Canziano per poi riaffiorare presso Duino dopo circa 40 km.

La scienza che studia le grotte, dedicandosi avventurosamente alla loro esplorazione, è la **speleologia**: i risultati di queste ricerche si rivelano estremamente utili non soltanto per la geologia, ma anche per numerose altre scienze, dalla biologia, all'archeologia, all'antropologia.

[10] La caratteristica principale della morfologia carsica consiste proprio nell'assenza di un reticolo idrografico superficiale ben sviluppato, in quanto le acque meteoriche si infiltrano per la massima parte nelle rocce e nel sottosuolo.

[11] Altre celebri zone carsiche sono l'area prealpina veneta, larghi tratti dell'Appennino centrale, in particolare di quello marchigiano, le Alpi Apuane, alcune zone della Sardegna, l'altopiano delle Murge in Puglia.

9 Morfologia glaciale e periglaciale

I **ghiacciai**, che coprono attualmente il 10% circa delle terre emerse, in un passato non molto lontano, circa 20.000 anni fa, risultavano almeno tre volte più estesi. La loro importanza non è dunque legata soltanto alla notevole azione morfogenetica che essi svolgono oggi in alcune zone del pianeta, ma anche al modellamento che hanno operato in passato su aree ben più vaste, e di cui oggi restano soltanto alcune forme relitte [12].

Si è già avuto modo di sottolineare che i ghiacciai rappresentano masse di ghiaccio in movimento lungo i pendii, con velocità differenti, che dipendono da diversi fattori (la topografia, l'attrito con il terreno, la spinta da parte di altro ghiaccio in discesa, la sezione di scorrimento etc.). In virtù di questo movimento, i ghiacciai esercitano un'azione di **erosione**, di **trasporto** e di **accumulo**.

a) Azione erosiva dei ghiacciai

L'**erosione glaciale**, indicata con il termine di **esarazione**, avviene attraverso processi di *rimozione* del detrito già presente sul terreno, di *abrasione* della roccia su cui il ghiaccio si muove e di *sradicamento* di blocchi e scaglie di roccia. L'esarazione, e in particolare l'abrasione, risulta tanto maggiore quanto più abbondante è la quantità di detrito trasportato dal ghiaccio. Le superfici rocciose levigate e arrotondate, o solcate da strie e scanalature, rappresentano la prova più evidente del passaggio del ghiaccio. La sua **azione erosiva** tende essenzialmente ad ammorbidire la morfologia preesistente, dando origine a forme prevalentemente dolci, arrotondate, di diverso tipo (**forme dell'esarazione**): qui di seguito verranno elencate e brevemente descritte le principali.

Le **rocce montonate** corrispondono a gobbe di roccia più resistente che il ghiaccio non è riuscito a sradicare, ma che ha soltanto modellato, arrontondandone la superficie e il lato rivolto a monte (il lato a valle resta scabro). Si ritrovano spesso in gruppi irregolari e risultano particolarmente utili in quanto la loro presenza sui versanti di una valle indica con precisione la massima altezza raggiunta dal ghiacciaio.

Le **valli glaciali** (o *truogoli*), ossia le incisioni lungo le quali ha avuto luogo il movimento del ghiacciaio, presentano una classica sezione trasversale ad U (ben riconoscibile dai profili a V delle valli fluviali, assai più strette e dai fianchi maggiormente inclinati), prodotta da un'erosione molto spinta sul fondo e sui fianchi da parte del ghiaccio. Il fondo non è piatto e regolarizzato come quello delle valli fluviali, ma risulta costellato di forme concavo-convexe (**morfologia ad ombelichi e selle**): la loro origine dipende da vari fattori, tra i quali la diversa durezza delle rocce e la variabilità dell'intensità erosiva del ghiacciaio. Sui fianchi ripidi delle valli si possono trovare tratti di versanti meno inclinati (**spalle glaciali**), che costituiscono relitti di un'incisione più antica, o anche **valli sospese**, create in passato da ghiacciai secondari meno potenti, e oggi percorse da fiumiciattoli che raggiungono la valle principale con caratteristiche cascate (spesso utilizzate dall'uomo per la produzione di energia elettrica). Immediatamente al di sotto delle dorsali, in corrispondenza della «testa» delle valli glaciali, si individuano i **circhi**, nicchie semicircolari, da cui un tempo muovevano le lingue di ghiaccio e che oggi risultano spesso riempiti da piccoli laghi.

Infine, tra le forme d'esarazione vanno menzionati i **fiordi**, profonde insenature formatesi in seguito all'invasione di una valle glaciale da parte del mare, dopo il ritiro dei ghiacciai. Il loro fondo può raggiungere profondità superiori ai 1.000 metri e possono allungarsi, dalla costa verso l'interno, per decine e decine di chilometri (il *Sognefjord*, in Norvegia, è lungo quasi 200 km). Hanno un'ampia diffusione sulle coste alte formate da rocce dure, dove i ghiacciai scendevano con forte pendenza dall'interno (Norvegia, Groenlandia, Labrador, Cile, Nuova Zelanda).

[12] In Italia le tracce delle glaciazioni quaternarie sono conservate sulle Alpi, ma anche in alcuni gruppi montuosi particolarmente elevati dell'Appennino.

b) Funzione di trasporto e di deposito dei ghiacciai

Accanto all'*azione erosiva*, i ghiacciai esplicano anche un'importante **funzione di trasporto**, sia dei clasti caduti dalle pareti rocciose sovrastanti il ghiacciaio, sia del materiale strappato dal letto per esarazione, sia dei frammenti inglobati lungo il percorso. Complessivamente, tale materiale trasportato viene definito **morena**, così come il deposito che dal suo accumulo ha origine. Il trasporto avviene senza alcun processo di selezione granulometrica da parte del ghiacciaio che, data la sua elevata compattezza, è in grado di sostenere e far muovere, insieme con sabbie e fango, anche massi di dimensioni enormi.

Tali prodotti, una volta abbandonati, si accumulano caoticamente, e, in base alla posizione rispetto alla lingua glaciale in cui avviene il deposito, vengono classificati in: **morene di fondo**, alla base; **morene laterali**, ai bordi; **morene frontali**, immediatamente davanti; **morene mediane**, nel mezzo, per la confluenza di due lingue. Una successione di morene frontali, disposte ad arco a diverse altezze (definita con il termine **anfiteatro morenico**), testimonia gli stadi di arretramento di un ghiacciaio: esempi di tali forme si trovano anche in Italia, davanti ai laghi Maggiore, di Como e di Garda. Accanto ai depositi morenici accumulati direttamente dai ghiacciai, esistono anche **depositi fluvioglaciali, glaciomarini e glaciolacustri**. Tra i primi, che hanno origine dall'azione delle acque dei torrenti glaciali, ricordiamo gli *esker*, colline allungate, a sommità piatta, costituiti da ghiaie e sabbie; i *kame*, terrazzi ai margini delle masse glaciali in scioglimento; le *sandur*, pianure alluvionali formate dalle acque di fusione dei ghiacciai. Le caratteristiche di questi depositi fluvioglaciali (riconoscibili per la stratificazione e gli andamenti lenticolari assenti nei depositi glaciali, in quanto tipici dell'azione delle acque correnti) dipendono chiaramente dall'ambiente di sedimentazione, dal comportamento idrologico dei torrenti e dal tipo di materiale che essi trasportano.

Quando i ghiacciai trasportano e accumulano materiale solido in mare o in un lago si originano i depositi glaciomarini e glaciolacustri. Tra questi ultimi hanno particolare interesse le **varve**, banchi di argille stratificate in lamine chiare e scure, alternate. Ogni coppia di lamine corrisponde ad un ciclo di deposizione annuale [13], e questa indicazione si dimostra utilissima per stabilire correlazioni tra prodotti simili in ambienti diversi e definire una cronologia relativa di eventi.

Lo studio dei depositi glaciali si dimostra estremamente prezioso per le ricostruzioni paleoclimatiche e paleoambientali: l'alternanza di epoche glaciali e interglaciali ha caratterizzato tutta la storia del nostro pianeta, e nel passato geologico più recente ha inevitabilmente influenzato l'evoluzione dell'uomo e delle sue attività.

c) Ambiente periglaciale

L'**ambiente periglaciale** comprende le regioni con clima freddo, in cui non c'è l'azione diretta dei ghiacciai ma i processi geomorfologici sono dovuti all'azione gelo-disgelo e, in minima parte, alla nivazione. In queste zone la struttura del terreno risente fortemente del clima: se il livello più superficiale (**strato attivo**) oscilla tra temperature immediatamente sopra e sotto lo zero, a seconda della latitudine, dell'escursione termica giornaliera, stagionale ed annuale, lo strato sottostante, invece, risulta costantemente gelato (**permafrost**). A partire dalla base del *permafrost*, la temperatura invece risale, seguendo il ben noto gradiente geotermico.

Il *permafrost*, dunque, si trova, durante tutto l'anno, in una condizione di stabilità [14], mentre lo strato attivo, sottoposto a continue variazioni termiche, risulta costantemente in

[13] La lamina chiara, più spessa, si sedimenta in estate, in periodo di ablazione, mentre quella scura, più sottile, ricca di materia organica, si deposita d'inverno.

[14] Si calcola che attualmente il *permafrost* ricopra circa 1/5 delle terre emerse, comprendendo gran parte della Russia e del Canada, e quasi la totalità della Groenlandia e dell'Antartide.

trasformazione: il passaggio dell'acqua dallo stato liquido allo stato solido è responsabile di numerosi processi «periglaciali».

Hanno così origine i **suoli poligonali**, superfici fratturate dall'azione del gelo-disgelo che contrae e dilata continuamente il terreno; i **ghiacciai di pietre**, lobi di detrito roccioso trasportato, lungo i versanti, dall'acqua disciolta dal ghiaccio; i **pingo** e i **palsa**, rispettivamente collinette di sabbie e di torba, sollevate ad opera dell'acqua del sottosuolo che, gelando, esercita una pressione verso l'alto.

10 La morfologia eolica

Il **vento** rappresenta, dopo l'acqua meteorica, il principale agente atmosferico dell'erosione. La sua azione sui corpi si manifesta come una pressione, che dipende dall'angolo di incidenza dell'aria con la superficie colpita, dalla velocità del movimento e dalla densità dell'aria stessa. In particolare sono le particelle solide trasportate dal vento ad esercitare un'azione erosiva sulle rocce, definita come **corrasione** [15]. Si tratta di un processo lento e costante che, se protratto nel tempo, può però produrre morfologie davvero particolari. Si formano in questo modo tutti i ciottoli sfaccettati o levigati che ricoprono i suoli dei deserti, ma anche gli splendidi archi sospesi, le sculture alveolari e le cavità che bucano anche le rocce più dure (come i *tafoni* nei graniti della Sardegna). In alcuni casi il vento si limita invece ad asportare detrito e polveri da superfici di rocce o da cavità con un processo di denudamento, detto **deflazione** [16]: oltre, infatti, ad essere un agente erosivo assai efficace, in grado di attaccare da ogni direzione, il vento si dimostra anche un importante agente di **trasporto**, che, rispetto a tutti gli altri, ha in più la capacità di agire contro la forza di gravità. Il trasporto, particolarmente selettivo, avviene in tre diversi modi: per **sospensione**, cioè all'interno del flusso d'aria, nel caso delle particelle più fini; per **saltazione**, ossia attraverso balzi successivi, quando si tratta di sabbie fini o medie; per **rotolamento**, o reptazione, ossia per la spinta in avanti dovuta all'urto di altre particelle, nel caso dei clasti più grossolani.

Il trasporto termina quando diminuisce l'energia del vento o quando il flusso d'aria incontra un ostacolo sul suo cammino: in tali situazioni hanno origine le forme di **accumulo** eolico. Le più comuni e conosciute sono le **dune**, collinette costituite da sabbia trasportata dal vento e poi depositata, che si ritrovano sia in zone interne che lungo le spiagge (**dune costiere**), generalmente raggruppate a formare campi di dune [17]. Esse variano per altezza (da qualche decimetro a molte decine di metri), per lunghezza (da 1 metro a decine di chilometri) e per forma. Le dune **longitudinali** sono allungate secondo la direzione del vento dominante, le dune **trasversali** sono ad essa perpendicolari, le dune **ad arco** (o barcane) e le dune **paraboliche** mostrano invece una caratteristica forma a semiluna, rispettivamente con la convessità e la concavità rivolte al vento. Il loro profilo trasversale non si presenta regolare: la pendenza dei fianchi sopravento (cioè esposti al vento) è minore rispetto ai pendii sottovento (cioè riparati) e la cresta risulta quindi asimmetrica.

Al loro interno i granuli di sabbia, costituiti perlopiù da minerali leggeri e duri, sono disposti a straterelli con inclinazioni divergenti, in una classica **stratificazione incrociata**. È alquanto raro che le dune risultino fisse nella loro posizione, perché trattenute dalla vegetazione o da interventi antropici: generalmente esse sono dotate di una forte mobilità, sotto l'azione continua del vento.

L'azione geomorfologica del vento, come agente di erosione e di trasporto, risulta particolarmente evidente nelle zone di alta montagna, in cui la sua energia risulta maggiore, e nelle regioni a clima arido o semiarido (aree desertiche), dove le acque svolgono un ruolo secondario

[15] In particolare, i granuli di quarzo presentano un'elevata capacità abrasiva.

[16] Gli *hamada*, deserti rocciosi, e i *serir*, deserti ciottolosi, sono forme assai caratteristiche e molto comuni nel Sahara, generate proprio per deflazione da parte del vento, che trasporta via i detriti più fini e lascia sul posto il materiale più grossolano.

[17] Le distese più ampie di dune, che caratterizzano deserti come il Sahara, sono definite *erg* (mari di sabbia).

e la vegetazione, capace di trattenere il suolo, è quasi assente. Tuttavia è possibile riconoscere morfologie eoliche anche in alcune zone delle terre emerse che non siano gli attuali deserti e le odierne aree costiere: si tratta delle aree che, in un passato non troppo lontano, ricadevano in una fascia climatica diversa. Un esempio è dato dalle regioni africane a sud del Sahara, verso il Sudan, coperte da vegetazione a steppa, a boscaglia o a savana, tra cui è facile individuare ampie estensioni di dune di sabbia, ormai fissate e in parte rimaneggiate.

Un cenno a parte, infine, merita un deposito eolico molto particolare definito **loess**: si forma per accumulo di polveri trasportate dal vento in regioni più o meno vaste, ricoperte da vegetazione erbacea tipo steppa, e si presenta sotto forma di banchi di polveri giallastre compatte, porose, in cui si individuano tracce di radici e fossili di molluschi terrestri. I depositi più estesi e potenti si ritrovano oggi in Cina e nelle *pampas* dell'Argentina; il *loess* presente in Europa (Russia meridionale, Svizzera, Belgio, Francia settentrionale e anche Italia, nell'anfiteatro morenico di Rivoli, in Piemonte) si è formato sicuramente nel Quaternario, nei periodi glaciali secondo alcuni studiosi, o in quelli interglaciali secondo altri. La sua importanza è notevole dovunque, dal momento che rende i terreni particolarmente fertili.

11 La morfologia costiera

La **costa** è, per definizione, la zona di contatto tra la terraferma e il mare, che si estende dalla linea di bassa marea fino al limite di influenza delle onde e dell'alta marea, verso l'interno. La classificazione più semplice prevede una distinzione tra le *coste alte*, o **falesie**, che consistono in pareti di roccia verticali, o molto inclinate, alte decine di metri, e le *coste basse*, o **spiagge**, che corrispondono a fasce orizzontali, o a debole pendenza, costituite da sabbie e ciottoli. Tale distinzione, estremamente pratica, risulta valida soltanto se non si incorre nell'errore di identificare le falesie con coste di erosione e le spiagge con coste di deposizione.

Una classificazione più completa e corretta, dal punto di vista scientifico, è stata proposta nel 1952 dal tedesco Valentin ed è oggi comunemente utilizzata da tutti i geomorfologi. Essa si basa sull'andamento morfodinamico delle coste, prendendo in considerazione quattro fattori: l'*erosione marina*, la *deposizione marina e fluviale*, l'*innalzamento della costa (emersione)* e l'*abbassamento della costa (sommersione)*. In tal modo, tenendo conto degli *spostamenti orizzontali* causati dai due primi fattori e degli *spostamenti verticali* [18], rispetto al livello del mare, prodotti dagli altri due, è possibile distinguere *coste in avanzamento*, *coste in arretramento* e *coste in equilibrio*. L'azione del **mare** sulle fasce costiere si esplica anche attraverso le correnti e le maree ma, soprattutto, per mezzo del moto ondoso; essa dà luogo sia a fenomeni distruttivi (*erosione*) sia a fenomeni costruttivi (*deposito*).

L'**erosione marina** consiste essenzialmente in un processo di disgregazione meccanica della costa (**abrasione**) ad opera della sabbia e dei ciottoli trasportati dalle onde. Le onde, battendo su un versante roccioso che degrada, con varia inclinazione, verso il mare, sono in grado di scavarlo [19], creando un intaglio orizzontale che prende il nome di **solco di battigia** [20]. Il perpetrarsi dell'abrasione porta all'approfondimento del solco, che si traduce in un vero

[18] Gli *spostamenti verticali* della costa consistono in variazioni relative tra mare e terra. La causa può essere sia un innalzamento/abbassamento del livello del mare (*movimenti eustatici*), sia un innalzamento/abbassamento tettonico. Se il mare avanza rispetto al continente, il fenomeno prende il nome di *trasgressione*; se invece esso si ritira da una regione che aveva precedentemente occupato, si parla di *regressione*.

[19] All'azione di disgregazione meccanica, si associa, nel caso si tratti di rocce carbonatiche, anche un processo chimico, simile al carsismo.

[20] La *battigia* corrisponde al tratto inclinato della spiaggia su cui avviene il movimento alternato del flusso montante e della risacca.

e proprio fenomeno di scalzamento al piede del versante: in tal modo si può innescare il crollo della roccia sovrastante, con la conseguente formazione di una falesia. Sulle pareti di alcune falesie, a qualche metro dall'attuale livello del mare, è oggi possibile osservare antichi solchi di battaglia che stanno a testimoniare fasi interglaciali del recente passato.

Il detrito prodotto dall'azione marina durante la formazione di una falesia si raccoglie sulla **piattaforma di abrasione marina**, una sorta di scogliera inclinata, che, col tempo, allargandosi ed ispessendosi, può proteggere la falesia da un ulteriore fenomeno di arretramento. Quando cambia l'altezza relativa tra terra e mare, a causa di movimenti eustatici o fenomeni tettonici, le piattaforme possono emergere, ed appaiono allora come ripiani delimitati da una scarpata a monte (l'antica falesia) ed una a valle, che vengono definiti **terrazzi marini**.

L'erosione che agisce sulle coste alte produce inoltre una serie di forme assai caratteristiche: si pensi alle **grotte marine** scavate nelle falesie (rinomate per bellezza, come la Grotta del Bue marino nella zona dell'Ogliastro, in Sardegna, o la Grotta Azzurra di Capri, in Campania; oppure per interesse scientifico, come la Grotta Romanelli in Puglia o la Grotta Guattari al Circeo, in Lazio, che ospitarono gli uomini preistorici), agli **archi naturali**, che tagliano i promontori, ai **faraglioni**, scogli staccatisi dalla costa.

Sulle coste basse l'abrasione marina produce risultati meno appariscenti: l'azione delle onde provoca unicamente un rimaneggiamento dei clasti che costituiscono la spiaggia, arrotondandoli e frammentandoli, e, in alcuni casi, provvede ad un loro allontanamento verso il largo. Accanto all'erosione il mare svolge, infatti, un'importante azione di **trasporto**, che avviene per **sospensione** delle particelle più fini, e per **trasporto longitudinale**, ossia parallelo alla riva, di sabbie e ghiaie [21]. I detriti prodotti dall'abrasione o erosi dai fondali, insieme con i sedimenti trasportati dai fiumi (e, in misura minore, da altri agenti di trasporto) e con i resti di organismi marini, depositandosi costituiscono la materia prima per la formazione di una spiaggia. Il deposito avviene soprattutto nelle insenature più protette, dove si formano spiaggette a forma di falce, ma, nel caso di acque basse, si può avere anche a largo della costa: nascono così lunghe ed ampie distese di sabbia. Le coste assumono poi una morfologia particolare in corrispondenza delle foci dei fiumi: i sedimenti trasportati dai corsi d'acqua vengono depositati all'ingresso in mare, formando **delta** (conoidi di sedimenti) o **estuari** (profonde insenature), a seconda della quantità di materiale presente nel fiume e dell'azione di dispersione dei sedimenti da parte di maree, moto ondoso e, soprattutto, di correnti. Si tratta di forme poco stabili, che mutano rapidamente al variare dei rapporti tra terraferma e mare.

Come si è detto, anche gli organismi marini partecipano alla costruzione e al modellamento della morfologia costiera. Le forme più straordinarie sono costituite dalle **scogliere coralline**, costruzioni prodotte dalla sovrapposizione di organismi con guscio calcareo (coralli, madrepore, alghe, briozoi, molluschi echinodermi etc.), che si ergono più o meno a largo rispetto alla riva [22]. Quando si dispongono ad anello intorno ad una laguna centrale le isole coralline prendono il nome di atolli: la loro origine, ancora in parte dibattuta, andrebbe collegata a isole vulcaniche in corso di sommersione. Le condizioni ottimali di sviluppo delle scogliere sono rappresentate da acque mosse, pulite, poco profonde, con temperature non inferiori ai 18°C.

L'azione combinata di erosione e di deposito da parte del mare tende complessivamente a rendere rettilinea la costa. Il raggiungimento di questo stato di equilibrio potrebbe rappresentare la fase finale di un processo generale di evoluzione costiera. Immaginando un **ciclo di erosione** paragonabile a quello fluviale si potrebbe individuare una fase di giovinezza, in cui la costa si

[21] Questo meccanismo di trasporto ben spiega l'origine delle *barre* (o scanni) di sabbia orientati parallelamente alla costa. Quando esse si accrescono possono formare dei *cordoni litoranei* emersi, in grado di chiudere le insenature, producendo *lagune*, *stagni* e *laghi costieri*, o di unire isole alla terraferma (in tal caso prendono il nome di *tomboli*).

[22] Le *scogliere* più prossime alla riva si definiscono *frange coralline*, quelle più distanti *barriere coralline*.

presenterebbe fortemente articolata, con profonde insenature, lunghi promontori, falesie in arretramento e piattaforme di abrasione in accrescimento. Con il passaggio alla successiva fase di maturità, la costa comincerebbe ad assumere un andamento più rettilineo, grazie all'erosione dei promontori e al riempimento delle insenature. Soltanto nella fase di vecchiaia si raggiungerebbe l'equilibrio, con lo sviluppo di ampie piattaforme di abrasione a frenare l'ulteriore arretramento delle falesie. Tale evoluzione rappresenta in realtà un processo assolutamente ideale, in quanto le variazioni del livello del mare e i movimenti tettonici costituiscono elementi di instabilità che interrompono continuamente le diverse fasi del processo.

Un ulteriore fattore di disturbo per il processo evolutivo naturale delle coste è costituito dall'**attività antropica**. Gli interventi umani in grado di modificare pesantemente la dinamica dei litorali, accelerandone spaventosamente l'erosione sono molteplici: la costruzione di laghi artificiali e di opere di regimazione di torrenti e fiumi (briglie, arginature etc.) impoverisce di sedimenti i corsi d'acqua che, con il carico ridotto, non riescono a bilanciare l'azione erosiva da parte del moto ondoso [23]; l'emungimento eccessivo di acqua di falda o di gas in aree retrocostiere fa abbassare il suolo e di conseguenza causa l'ingresso dell'acqua marina in zone di terraferma [24]; i moli, le scogliere frangiflutti, i pennelli, se da un lato producono l'ampliamento di alcuni tratti di spiaggia, dall'altro limitano gli apporti di sedimenti nelle zone contigue, che subiscono quindi il processo inverso; il prelievo indiscriminato di sabbie e ghiaie dai letti fluviali o direttamente dalle spiagge costituisce una vera e propria azione di erosione antropica.

12 Ciclo d'erosione e superfici di spianamento

Il **paesaggio** rappresenta, come si è visto, l'effetto di una duplice azione: quella dei processi endogeni, che innalzano i rilievi, e quella dei contemporanei o successivi fenomeni esogeni, che tendono invece a livellare le forme, attraverso lo smantellamento dei rilievi stessi e il riempimento delle depressioni. La situazione di equilibrio corrisponderebbe ad un completo appiattimento del rilievo, ossia al raggiungimento, da parte della superficie terrestre, della forma di geoide. Si tratta in realtà di un equilibrio irraggiungibile, dati i tempi molto più rapidi della tettonica rispetto a quelli dell'erosione.

Il geologo e geografo americano W.M. Davis per primo propose, a fine Ottocento, una teoria in grado di spiegare la tendenza evolutiva del paesaggio attraverso la ciclicità del processo di erosione. Tale teoria, ritenuta valida ancora oggi, ben si applica, però, unicamente alle regioni umide, in cui i corsi d'acqua esercitano un'azione predominante rispetto agli altri agenti esogeni. La situazione iniziale è data da un territorio sollevato da fenomeni tettonici fino ad una certa altitudine su cui cominciano ad agire i processi esogeni, guidati dall'erosione fluviale. Versanti ripidi, valli profondamente incise, laghi tettonici, tracce dell'antico paesaggio piatto, fenomeni franosi, caratterizzano lo **stadio di giovinezza** del rilievo, che corrisponde alla prima fase del **ciclo di erosione**. Le Alpi, l'Himalaya e le Ande, tutte catene di recente sollevamento, ben rappresentano oggi tale stadio. L'addolcimento delle forme, lo sviluppo di una rete fluviale più ramificata, l'allargamento di alcuni fondovalle e l'erosione più lenta dei versanti contraddistinguono lo **stadio di maturità** del rilievo. È così che si presentano i Carpazi (tra Polonia, Slovacchia, Ucraina e Romania), la catena franco-svizzera del Giura, la regione del Delfinato nella Francia sud-orientale. Lo **stadio di senilità**, che segna la fine del ciclo di erosione, corrisponde allo stadio iniziale: il paesaggio presenta rilievi a debole pendenza, valli fluviali larghe e con i fianchi poco inclinati e forme generalmente appiattite, definite **penepiani** (di qui il nome di **modello di**

[23] L'impovertimento dei fiumi Agri e Sinni, in Basilicata, e il conseguente arretramento della costa di circa un metro all'anno, sono ascrivibili proprio alla costruzione di alcuni invasi artificiali.

[24] A tale fenomeno si deve l'arretramento del litorale di Ravenna di circa 100 metri nel giro di pochi anni.

peneplanazione per questa teoria). Questo stadio è tipico di regioni geologicamente antiche e stabili da lungo tempo, come le Ardenne (tra Belgio, Lussemburgo e Francia), il Massiccio Boemo (Repubblica Ceca), lo Scudo Canadese.

Il punto debole di tale teoria, che ha comunque avuto il merito di individuare un'origine alternativa all'erosione marina per le superfici pianeggianti così diffuse in un certo tipo di paesaggio, è nella presupposta stabilità tettonica in cui si verificherebbe il ciclo di erosione. Tale situazione risulta, invece, rarissima: i fenomeni tettonici sono, infatti, generalmente contemporanei ai processi erosivi, e questo fattore irrimediabilmente complica il normale sviluppo ciclico. Risulta quindi molto difficile trovare in natura paesaggi che abbiano raggiunto lo stadio di maturità, ossia uno sviluppo completo dei penepiani: generalmente il ciclo di erosione si interrompe prima, per il sopraggiungere di un evento tettonico o anche, ad esempio, di una variazione climatica, e accade così che sul rilievo si vada ad impostare un nuovo ciclo (fenomeno di **ringiovanimento**), spesso testimoniato dalla presenza di terrazzi fluviali (un classico esempio è costituito dai *canyons* americani).

Negli stessi anni in cui Davis proponeva il suo modello di peneplanazione, Penck, geografo tedesco, elaborava la **teoria della pedimentazione** per spiegare l'origine delle superfici d'erosione in roccia, a debolissima pendenza, caratteristiche delle regioni a clima arido e semi-arido. La situazione di partenza prevede un rilievo su cui agisce l'erosione che ne causa un arretramento parallelo; l'acqua pluviale, che scende dal rilievo stesso, provvede, con un'azione di dilavamento diffuso, ad allontanare i detriti dal piede del versante, in modo da favorire il processo di erosione successivo. Ha così origine il **pediment**, che con i suoi pochi gradi di pendenza (generalmente $<5^\circ$) correla ciò che resta del versante con il fondovalle. Il progressivo arretramento del versante e lo sviluppo dei *pediment* attorno ad esso può portare alla formazione di rilievi isolati, detti *Inselberg* (termine tedesco che significa *monte-isola*), e di un unico **pedipiano** per fusione dei diversi pediment (evoluzione per **pediplanazione**).

Una teoria più recente, risalente agli anni Cinquanta-Sessanta del Novecento, si deve allo studioso tedesco Büdel ed è conosciuta come **etchplanation**: tale modello prevede l'esistenza, in zone caldo-umide tropicali, di superfici di spianamento dovute essenzialmente all'azione di fenomeni di alterazione e disfacimento chimico. L'acqua agisce sia in superficie sia in profondità, creando spessori di decine e decine di metri di alteriti, che vanno a costituire suolo e sottosuolo. In tali zone, i fattori fisici che tendono a portar via l'alterite risultano molto più lenti dei processi chimici responsabili della sua origine.

Elementi di stratigrafia e di tettonica

1 La stratigrafia

La ricerca geologica non è finalizzata unicamente all'interpretazione del paesaggio attuale, ma anche alla ricostruzione dei paesaggi del passato, ossia degli antichi ambienti in cui si sono formate e poi modificate rocce e strutture. In tal senso risulta particolarmente utile lo studio delle rocce sedimentarie, che più delle altre riflettono le condizioni paleoambientali in cui hanno avuto origine e si sono evolute. Per indicare l'insieme dei caratteri litologici e paleontologici di una roccia sedimentaria, si usa il termine **facies** (dal latino, *aspetto*) **sedimentaria**: il suo riconoscimento permette di risalire all'ambiente di sedimentazione. Tuttavia non basta studiare una singola roccia ma occorre considerare l'insieme cui essa appartiene, cioè descrivere una successione verticale di rocce sedimentarie diverse. Il metodo consiste nel suddividere tali successioni in unità contraddistinte in base ai caratteri di *facies*; di ciò si occupa la **stratigrafia** (il nome nasce dalla caratteristica propria delle rocce sedimentarie di essere per lo più stratificate), una branca estremamente interessante della geologia.

I principi fondamentali su cui si basa la stratigrafia (che ammettono, comunque, delle eccezioni) sono tre:

- il **principio di sovrapposizione**, secondo cui, in una successione di strati, ciascuno di essi è più antico del sovrastante e più recente del sottostante [1];
- il **principio di orizzontalità** originaria, in base al quale, in linea generale, gli strati si dispongono orizzontalmente [2];
- il **principio di continuità**, secondo il quale la sedimentazione è un processo continuo, se non altro all'interno di uno stesso ambiente. Pertanto, all'origine di una discontinuità va ricercato l'intervento di un fattore esterno.

Le unità stratigrafiche sono di tre diversi tipi: **unità litostratigrafiche** (basate sulle caratteristiche petrografiche delle rocce), **biostratigrafiche** (individuate in base al contenuto in fossili), **cronostratigrafiche** (definite in relazione al tempo). Per ognuna di esse esiste un'unità fondamentale: si tratta, rispettivamente, della **formazione**, della **zona** e del **piano**. L'analisi dei rapporti verticali e orizzontali tra le diverse unità consente di interpretare la serie stratigrafica. Ogni passaggio tra un'unità e la successiva segna un cambiamento significativo nelle condizioni di sedimentazione. I fenomeni tettonici e i movimenti eustatici del mare sono i maggiori responsabili delle forti variazioni che interessano le serie stratigrafiche.

Per **eustatismo** si intende la variazione del livello medio del mare, la cui origine può essere legata a movimenti tettonici che causano sollevamenti o abbassamenti relativi del fondo marino, oppure ai processi di glaciazione e «deglaciazione» che provocano una diminuzione o un aumento di volume delle acque.

[1] In caso di *terreni piegati o fagliati* (tali termini saranno spiegati tra non molto, in questo stesso capitolo) è l'intervento di eventi tettonici successivi alla sedimentazione la causa dell'inapplicabilità di tale principio. Vere eccezioni sono invece costituite dall'intrusione di un filone recente in un terreno più antico, dalla morfologia a terrazzi fluviali, dai depositi in grotta e dalle scogliere coralline, che si innalzano al di sopra di terreni più antichi.

[2] È chiaro che si intende *orizzontalità* rispetto al piano di sedimentazione: se esso è inclinato, risulteranno inclinati anche gli strati. Costituisce, invece, un'eccezione, la stratificazione incrociata delle dune.

Al fenomeno di avanzamento del mare su una terra precedentemente emersa si dà il nome di **trasgressione**, mentre il fenomeno inverso (ritiro del mare) è definito **regressione**. L'intervallo di sedimentazione (inteso come serie di formazioni) tra una trasgressione e una regressione è definito **ciclo sedimentario** [3], mentre per **lacuna stratigrafica** si intende l'assenza di sedimentazione tra due formazioni [4].

Il riconoscimento in una successione di rocce di diversi cicli sedimentari e di lacune stratigrafiche, grazie alle caratteristiche litologiche e paleontologiche, risulta di fondamentale importanza per ricostruire la distribuzione di terre e mari nel passato geologico (paleogeografia) e l'evoluzione degli organismi viventi.

2 La tettonica

I **movimenti tettonici** comprendono tutte le deformazioni e gli spostamenti prodotti sulle rocce e sulle strutture da parte delle forze endogene. In senso ampio, in questa categoria rientrano anche i movimenti della litosfera e l'orogenesi. Più precisamente, però, l'espressione «movimenti tettonici» si riferisce agli effetti che hanno sulle diverse rocce, forze di natura distensiva e di natura compressiva. La deformazione prodotta risente non soltanto del tipo, dell'intensità e della durata della forza, ma anche, e soprattutto, dal grado di rigidità della roccia, e da altri fattori, quali la pressione e la temperatura.

Le **rocce** si distinguono in **fragili** e **duttili**: le prime reagiscono alle sollecitazioni prima con la frattura e poi con la rottura, le seconde invece con il piegamento. A seconda di quale fenomeno prevalga, si hanno **regioni a stile rigido** e **regioni a stile plastico**. Nel primo caso la struttura tipica è la faglia, nel secondo la piega.

Per **faglia** si intende una frattura all'interno di una massa rocciosa, lungo la quale si verifica uno spostamento orizzontale, verticale o obliquo delle parti. La superficie in corrispondenza della quale si verificano il taglio e lo spostamento è detta **piano di faglia**. Se il piano di faglia risulta inclinato, la parte al di sopra di esso è detta **tetto**, quella al di sotto **letto**; se il tetto si innalza rispetto al letto, la faglia è **inversa** (tale fenomeno si verifica quando la forza che agisce è di tipo compressivo e il risultato consiste in un accorciamento della struttura); se invece il tetto si abbassa rispetto al letto la faglia è **diretta** (in tal caso la forza agente è distensiva e l'effetto è un allungamento della struttura).

L'entità dello spostamento, variabile da pochi metri a qualche chilometro, viene definita **rigetto**. Quando il piano di faglia è verticale e il movimento avviene orizzontalmente, la faglia viene definita **trascorrente**.

Raramente le faglie si presentano isolate. Generalmente sono associate a formare sistemi crostali anche molto complessi, come le **strutture ad horst e graben**: si tratta di associazioni di faglie dirette, in fasci paralleli, che determinano ampie fosse tettoniche a gradini (*graben*), delimitate da pilastri di crosta (*horst*). La più celebre struttura di questo tipo è la Great Rift Valley dell'Africa Orientale, lunga oltre 6.000 km e larga tra 40 e 60 km; in Italia ne abbiamo un esempio in Sardegna, con il Campidano, che si estende da Cagliari fino al Golfo di Oristano. Fosse tettoniche sono anche le depressioni che solcano le dorsali oceaniche (*rift valleys*).

Una **piega** consiste in una deformazione, simile ad una flessione, di una massa rocciosa. Tali strutture risultano particolarmente evidenti quando interessano le rocce sedimentarie: gli strati,

[3] Secondo una definizione più recente, per *ciclo sedimentario* si intende l'intervallo di sedimentazione tra due eventi importanti e generalizzati, che non siano esclusivamente in relazione ai movimenti relativi fra terra e mare.

[4] La causa del mancato deposito è imputabile generalmente al passaggio da un ambiente di sedimentazione, quale è quello marino, ad uno di erosione, quale è quello continentale. In questo senso, una *lacuna stratigrafica* individua il succedersi di due cicli sedimentari diversi.

prima orizzontali, divengono ondulati. In una piega si distinguono: un **nucleo**, parte più interna; una **cerniera**, linea lungo la quale si ha la massima curvatura; due **fianchi**, lati che convergono verso la cerniera; un **piano assiale**, superficie che passa per le cerniere di tutti gli strati piegati e che dà la direzione della piega; un **asse**, linea di intersezione tra il piano assiale e un piano verticale che dà l'inclinazione della piega.

Quando gli strati curvano con la convessità verso l'alto e presentano nel nucleo i terreni più antichi, la piega è detta **anticlinale**; quando la convessità è verso il basso e nel nucleo ci sono le rocce più recenti, la piega si definisce **sinclinale**.

Se l'asse è verticale la piega è **diritta**; se è inclinato, si dice **obliqua**; se è fortemente inclinato, si definisce **rovesciata**; se l'asse è praticamente orizzontale, la piega è **coricata**. A seconda poi dell'inclinazione dei due fianchi, le pieghe assumono una serie di altri appellativi.

L'accavallamento di terreni diversi, a partire da una piega o da una faglia inversa, sotto l'azione di spinte tettoniche molto forti, è definito **sovrascorrimento**. Quando il fenomeno è ampiamente esteso fino ad assumere dimensioni regionali, prende il nome di **falda** (o coltre) **di ricoprimento**: il terreno che sovrascorre viene definito **alloctono**, in quanto proviene da una zona diversa, mentre il terreno in posto si dice **autoctono** [5]. Tale processo risulta più diffuso di quanto si immagini: sia sulle nostre Alpi sia sull'Appennino se ne possono ammirare splendidi casi.

Le strutture generate dai movimenti tettonici hanno un'importanza assai rilevante, in quanto possono influenzare fortemente i processi di dinamica esogena. Soltanto comprendendo a fondo i processi che le determinano è possibile, di conseguenza, procedere ad una corretta lettura e interpretazione del paesaggio.

[5] In corrispondenza di profonde incisioni (*finestre tettoniche*) è possibile osservare il substrato roccioso alloctono. Quello che resta di una falda di ricoprimento interessata dall'erosione è invece definito *klippen* (scoglio).

Paleogeografia ed evoluzione della vita

1 Le ere geologiche

Le stime moderne, basate su più metodi di decadimento radioattivo, calcolano per le più antiche rocce della crosta terrestre un'età di 3,8 miliardi di anni. Tuttavia oggi si pensa che l'**età della Terra** sia di 4,6 miliardi di anni, in base soprattutto a tre considerazioni principali: l'età radiometrica U/Pb e Rb/Sr delle meteoriti è di 4,6 miliardi di anni; le più antiche rocce lunari hanno questa stessa età; i rapporti Pb^{206}/Pb^{204} e Pb^{207}/Pb^{204} nei minerali di piombo terrestri sono gli stessi di quelli delle meteoriti datate 4,6 miliardi di anni. Di questo lunghissimo tempo geologico solamente gli ultimi 570 milioni di anni (indicati come **Eon Fanerozoico**) sono ben conosciuti, grazie ai numerosi **fossili** [1] ritrovati in rocce sedimentarie, mentre sui primi quattro miliardi di anni (**Eon Archeozoico** o **Precambriano**) è possibile unicamente avanzare delle ipotesi. Il Fanerozoico è distinto in tre **ere**, ognuna di esse è divisa in **periodi** (o *sistemi*), a loro volta distinti in **epoche** (o *serie*) suddivise in **età** (o *piani*).

Come si evince dalla tabella riportata qui a seguire, la più recente «suddivisione» del tempo geologico prevede che il Quaternario non sia un'era a parte ma, a causa della sua breve durata, risulti un periodo accorpato al Cenozoico [2]. Quest'ultimo sarebbe, quindi, distinto in tre periodi e sette epoche.

EON FANEROZOICO (dati della Società Geologica Americana)			
Era	Periodo	Epoca*	Milioni di anni**
Cenozoica o Terziaria (attuale- 66,4 m.a.)	Quaternario	<i>Olocene</i>	0,01- attuale
		<i>Pleistocene</i>	1,8-0,01
	Neogene	<i>Pliocene</i>	5,3-1,8
		<i>Miocene</i>	23,7-5,3
	Paleogene	<i>Oligocene</i>	36-23,7
		<i>Eocene</i>	57,8-36
		<i>Paleocene</i>	66,4-57,8
Mesozoica o Secondaria (66,4- 245 m.a.)	Cretacico o <i>Cretaceo</i>	<i>(inf.-sup.)</i>	144-66,4
	Giurassico	<i>(inf.-med.-sup.)</i>	208-144
	Triassico	<i>(inf.-med.-sup.)</i>	245-208
Paleozoica o Primaria (245-570 m.a.)	Permiano	<i>(inf.-sup.)</i>	286-245
	Carbonifero	<i>(inf.-sup.)</i>	360-286
	Devonico	<i>(inf.-med.-sup.)</i>	408-360
	Silurico	<i>(inf.-sup.)</i>	438-408
	Ordoviciano	<i>(inf.-med.-sup.)</i>	505-438
	Cambriano	<i>(inf.-med.-sup.)</i>	570-505
* Mentre alle Epoche dei Periodi del Cenozoico sono stati attribuiti dei nomi, i Periodi del Paleozoico e del Mesozoico vengono unicamente distinti in Inferiore, Medio e Superiore.			
** Si intendono milioni di anni dall'attuale.			

Tab. 1 - Suddivisioni dell'Eon Fanerozoico

[1] Un *fossile* è l'impronta di un organismo vegetale o animale, vissuto anteriormente all'epoca attuale, conservatosi in una roccia.

[2] Altri studiosi, invece, preferiscono suddividere il tempo geologico in *quattro ere*, anziché *tre*; in questo modo, l'ultima era, quella attuale, è chiamata *Neozoica Quaternaria*, a sua volta distinta in *due periodi*: *Pleistocene* e *Olocene*.

Il discorso della cronologia relativa è strettamente legato alla stratigrafia e alla necessità di individuare degli intervalli di tempo definiti nei loro limiti inferiori e superiori da associazioni di fossili. È in questo modo che nasce nel 1759, ad opera di Giovanni Arduino, l'ordinamento dei terreni in quattro gruppi (primario, secondario, terziario e quaternario) corrispondenti a quattro ere. Un'ulteriore era, anteriore alla Primaria, e definita *Arcaica*, fu introdotta dall'americano Dana nel 1876. L'intenzione dei padri della stratigrafia era di suddividere il tempo in intervalli, in base a fatti naturali riconoscibili su tutto il pianeta.

In origine si pensò a cataclismi o a grossi eventi, poi con lo sviluppo della paleontologia si individuarono nell'evoluzione degli organismi i riferimenti cronologici adatti per elaborare una valida scala dei tempi (di qui i termini *Paleozoico*, *Mesozoico* e *Cenozoico*, che significano, rispettivamente, *organismi di un periodo antico*, *intermedio* e *recente*). Ma in realtà anche i processi di origine, evoluzione ed estinzione degli organismi non possono essere considerati né come fenomeni globali né tantomeno come processi immediati, in grado quindi di segnare in modo univoco il passaggio tra un'era e un'altra o tra un periodo e il successivo. Qualsiasi classificazione o suddivisione rappresenta, insomma, una convenzione ed ha validità soltanto fino a quando nuove scoperte o intuizioni non ne propongano di più valide e moderne.

2 La paleogeografia

La *paleogeografia* è la scienza che studia la distribuzione delle terre e dei mari nel «passato geologico», basandosi su indagini paleomagnetiche (calcolo delle paleolatitudini), litologiche (analisi delle *facies* delle rocce sedimentarie per definire gli ambienti di sedimentazione, le trasgressioni e le regressioni marine, i fenomeni glaciali) e paleontologiche (studio della flora e della fauna fossili).

Della situazione paleogeografica del **Precambriano** ancora oggi si sa molto poco. Probabilmente è intorno a 2 miliardi di anni che la crosta granitica si separa in zolle che cominciano a muoversi rapidamente, dando origine a fenomeni di convergenza e di divergenza. Inizia così a formarsi una crosta oceanica. L'Africa e l'America del Sud appartengono ad una sola placca, ma non si è certi che siano già legate alle altre terre dell'emisfero australe, come accadrà nel Paleozoico. La zolla nord-americana sembra sia stata indipendente fino a 1,4 miliardi di anni fa, quando invece si unisce alle altre placche. Per quanto riguarda il clima dell'Eon Precambriano è difficile fare un discorso complessivo, a causa della sua lunghissima durata [3]. Molto poco sappiamo anche della vita nel Precambriano: tra gli organismi più antichi ci sono filamenti cilindrici, simili a batteri, che convivono con alghe blu, datate tre miliardi di anni, cui seguono i primi metazoi (organismi pluricellulari), che compaiono già circa 1,6 miliardi di anni fa, ma si svilupperanno poi nel Paleozoico. Anche i modelli paleogeografici proposti per il **Paleozoico** sono considerati ancora delle ipotesi da verificare ulteriormente.

Le aree continentali sono in espansione: la convergenza tra le zolle porta alla formazione nel periodo *Cambro-Ordoviciano* del **Continente di Gondwana** nell'emisfero meridionale, a comprendere l'India, l'Africa (con annessa l'Europa meridionale, secondo alcuni), il Madagascar, l'Australia e l'America Meridionale intorno all'Antartide. L'America Settentrionale e l'Europa sono ancora separate dall'esistenza di un Proto-Atlantico.

Durante il *Silurico-Devonico* la convergenza della zolla americana con quella europea porta alla scomparsa del Proto-Atlantico, alla formazione di un unico continente chiamato **Laurasia** e all'orogenesi caledoniana. L'Asia risulta ancora indipendente dalle altre terre (**Continente Siberiano**).

[3] Essenzialmente si alternano intervalli caldi e glaciazioni, queste ultime concentrate maggiormente tra 2,4 e 2,3 miliardi di anni e nel Precambriano superiore, il quale si chiude, appunto, con un fortissimo raffreddamento che, in base alle prove di cui disponiamo, interessò la maggior parte della Terra.

Il *Carbonifero* si apre con un'enorme ingressione marina a livello globale, che porta ad un'espansione dei mari epicontinentali, e con un clima fortemente umido, responsabile della straordinaria abbondanza di carbone (risultato della fossilizzazione in profondità di enormi foreste pluviali) fino a latitudini relativamente basse. Inoltre, con il sollevamento della catena dell'Europa centrale, prende il via il ciclo dell'orogenesi ercinica che si compirà completamente nel *Permiano*. In quest'ultimo periodo dell'era Paleozoica, infatti, per la convergenza tra le zolle, si formano gli Appalachi e gli Urali: la Laurasia si unisce al Continente Siberiano e insieme si fondono con Gondwana a formare un unico immenso continente emerso, **Pangea**. Il Permiano, caratterizzato da localizzate ingressioni marine, si chiude con una fase di intenso riscaldamento (dopo le frequenti glaciazioni verificatesi tra il Carbonifero medio e il Permiano medio) che spiega l'enorme diffusione di evaporiti, *red beds* e scogliere.

In ognuno dei sei periodi del Paleozoico si verificano glaciazioni, molte delle quali geograficamente ristrette, ma complessivamente la tendenza dei climi della prima *era* è verso temperature relativamente alte, che seguono alla straordinaria glaciazione del Precambriano superiore. Già dall'inizio del Paleozoico esistono quasi tutti i gruppi di Invertebrati (ricordiamo, tra i più noti, i gruppi dei Foraminiferi, dei Brachiopodi, degli Artropodi e degli Echinodermi) e pian piano si diversificano anche i Vertebrati: nel Silurico, i Pesci; nel Devonico, gli Anfibi; nel Carbonifero, i Rettili. Alle alghe cambriane si aggiungono anche le piante terrestri: nel Devonico compaiono le Pteridosperme; nel Carbonifero, le Cicadofite; nel Permiano, le Conifere.

La fine del Paleozoico è segnata da un'immane crisi biologica, causata, probabilmente, dalle importanti modificazioni paleogeografiche e paleoclimatiche, che vede l'estinzione definitiva di alcuni organismi invertebrati.

All'inizio del **Mesozoico** il quadro paleogeografico è identico a quello permiano: persiste ancora la Tetide. Ma già nel *Trias* medio, intorno a 200 milioni di anni fa, la parte meridionale della Pangea (la ex Gondwana) comincia a smembrarsi, sottoposta com'è a forze distensive che provocano la formazione dei primi bracci oceanici (tra Africa e India, e tra l'India e l'Antartide-Australia), pur senza separare definitivamente le terre. Intanto i continenti settentrionali della Pangea cominciano a separarsi dal blocco America meridionale-Africa, segnando l'inizio dell'apertura dell'Atlantico e del Mar dei Caraibi. Si hanno, inoltre, i primi fenomeni distensivi della Tetide, grande area di sedimentazione marina che separa gli scudi dell'emisfero settentrionale dal Continente di Gondwana, con la formazione di nuova crosta oceanica.

Questo movimento si accentua nel *Giurassico*, mentre il Proto-Atlantico si allarga verso nord, l'Oceano Indiano si espande con il procedere della migrazione della zolla antartico-australiana verso est, e la zolla africano-sudamericana, con quella indiana, spingendo verso nord e cominciando a distendersi, dà inizio alla compressione della Tetide. Si forma un *rift* tra Africa e America del Sud che nel *Cretacico* porta all'apertura dell'Atlantico meridionale (che raggiunge la larghezza di 3.000 km). Sempre nel Cretacico, lo spostamento verso ovest delle Americhe causa il sollevamento della cordigliera delle Ande e delle Montagne Rocciose. Lo smembramento della Pangea, che si verifica nell'ultimo periodo del Mesozoico, è all'origine della quasi scomparsa della Tetide e delle variazioni delle correnti oceaniche, che spiegano bene il forte raffreddamento con cui si chiude la seconda *era* e probabilmente la crisi biologica (di cui furono vittime, secondo una delle ipotesi sulla loro estinzione, anche i *dinosauri*) che segna il passaggio al Cenozoico.

Nonostante alcune fluttuazioni, il clima del Mesozoico, caldo e secco (se si eccettua quello del *Trias* inferiore, freddo e umido a seguito della glaciazione permiana), è caratterizzato da una straordinaria stabilità, mai raggiunta in altri momenti geologici. La vita ha un enorme sviluppo: *fitoplancton* e *zooplancton* abitano i mari, insieme ad alghe e svariati Invertebrati (ricordiamo, fra i tanti, i Molluschi, tra cui le famose Ammoniti, e i Lamellibranchi, tra cui le Rudiste), mentre i Vertebrati raggiungono una straordinaria diversificazione. In particolare evolvono i Mammiferi

e gli Uccelli, mentre tra i Rettili compaiono i Dinosauri, che popolano terre emerse, mari e cieli. Tra le piante dominano le Gymnosperme e le Angiosperme.

Il **Cenozoico** vede nel *Paleocene* la separazione tra Antartide ed Australia e l'inizio dell'apertura dell'Atlantico settentrionale, che si compie definitivamente nell'Oligocene, quando contemporaneamente si prosciuga il Mare degli Urali. Cambiano così in maniera sostanziale le vie di migrazione dei Mammiferi: tra Paleocene ed *Eocene* esisteva il collegamento tra l'America Settentrionale e l'Europa, ma entrambe risultavano isolate dall'Asia; nell'*Oligocene* l'Europa è invece isolata verso ovest a causa dell'Atlantico ma è collegata all'Asia, a sua volta in comunicazione con l'America attraverso lo Stretto di Bering, che saltuariamente diviene terraferma.

L'apertura dell'Atlantico, momentaneamente bloccatasi nell'Oligocene per riprendere poi nel Miocene, causa, per compressione, la definitiva chiusura della Tetide, di cui il Mediterraneo rappresenta un relitto e, con lo spostamento dell'Eurasia verso sud-est e la rotazione antioraria dell'Africa, induce l'orogenesi alpina. La zolla indiana si salda a quella asiatica e si solleva la catena dell'Himalaya. Il Mediterraneo diviene un bacino evaporitico e comunicherà di nuovo con l'Atlantico solo nel Pliocene, quando l'Istmo di Panama unirà le due Americhe e tra la zolla arabica e quella africana si aprirà il Mar Rosso.

Alla fine del Cenozoico la situazione paleogeografica si presenta identica alla distribuzione odierna di terre e mari.

La situazione climatica della terza *era* vede un'alternanza di intervalli caldi e forti raffreddamenti, che si fa più fitta a partire dalla fine del Neogene, fino ad assumere, durante il Quaternario, un vero e proprio carattere ciclico [4]. In Europa si individuano, a partire dal Pliocene, quattro stadi glaciali principali, conosciuti come *Günz*, *Mindel*, *Riss* e *Würm*, verificatisi tra 1,6 m.a. e 18.000 anni fa, intervallati da tre periodi interglaciali in cui le temperature si innalzano di 10°-15°C. Attualmente stiamo vivendo in uno stadio interglaciale iniziato 18.000 anni fa.

All'inizio del Paleogene sono già rappresentati tutti i grandi gruppi oggi viventi: il Cenozoico è dunque l'*era* dell'evoluzione e della diversificazione. I Mammiferi, nel Neogene, raggiungono il loro apogeo e tra i Carnivori compaiono generi assai diffusi oggi, come il *Canis* e il *Felis*. Nel Miocene appaiono tra i Primati le prime Scimmie, mentre a fine Pliocene si impongono i generi *Equus* ed *Elephas* ed evolve rapidamente il ramo degli Ominidi, già preannunciato nel Miocene Superiore dall'*Oreopithecus* (parecchi esemplari del quale sono stati ritrovati in Toscana) e dal *Rhamapithecus*, sviluppatosi nel Pliocene con l'*Australopithecus*, i cui resti, vecchi 4 milioni di anni, sono stati scoperti presso il lago Rodolfo in Kenya e nella Valle dell'Omo in Etiopia. Gli *Australopithecus* erano bipedi, di piccola statura, avevano un cranio appiattito, fortemente prognato, con arcate sopraorbitali sporgenti; la capacità cerebrale del cranio era meno della metà di quella dell'uomo attuale, nonostante essi avessero già conquistato la stazione eretta e l'uso delle mani [5]. Solo a fine Pliocene, intorno a 2 milioni di anni fa, si ha l'evoluzione nell'*Homo Habilis*, il primo ominide che sperimenta una vera «industria litica» (sassi e ciottoli scheggiati in modo ancora rudimentale, i cosiddetti *choppers*, utilizzati per uccidere e sezionare le prede), e nell'*Homo Erectus*, con il quale il genere ominide perfeziona le tecniche di taglio della pietra e giunge alla scoperta del fuoco. Il Pleistocene Superiore (tra i 100.000 e i 35.000 anni) vede la diffusione del *Neanderthal*, soprattutto in Europa (celebre il ritrovamento italiano, del 1939, di un teschio nella grotta Guattari del Circeo): associate a questo ominide si rinvencono le prime sepolture, datate 80.000 anni fa, che necessariamente attestano la nascita di elementari forme di culto.

Contemporaneamente si registra la comparsa dell'*Homo Sapiens* «arcaico», che presenta ormai molte affinità con l'uomo moderno: capacità e conformazione cranica, impronta facciale,

[4] La ciclicità delle glaciazioni viene spiegata sulla base di fenomeni astronomici.

[5] Nel 1974, in una regione dell'Etiopia, una spedizione archeologica inglese riportò alla luce lo scheletro di *Australopithecus* più completo finora ritrovato per un'epoca così arcaica, cui fu dato il nome di *Lucy*.

corporatura. A lui si attribuiscono le pitture con scene di caccia, di combattimento e di raccolta, scoperte sulle pareti di numerose grotte in Francia, Spagna e Italia. Paleontologi, paleontologi e antropologi discutono ancora oggi da quale dei due rami, indipendenti e coesistenti, sia disceso poi l'*Homo Cromagnon*, nostro più diretto antenato, risalente a 30.000 anni fa.

3 L'origine della vita

La Terra ebbe origine circa 4,6 miliardi di anni fa: dai gas liberati dal suo interno si formò un'**atmosfera primordiale**, costituita da gas nobili, idrogeno, azoto, carbonio e vapore acqueo, che le frequenti eruzioni vulcaniche incrementavano continuamente. Con il progressivo raffreddamento del pianeta, parte del vapore cominciò a condensare e dalle precipitazioni iniziò a formarsi l'idrosfera. Si ritiene che, al momento della solidificazione della parte più esterna del pianeta, l'atmosfera doveva essere composta prevalentemente da **idrogeno, elio, metano** (CH_4), **ammoniaca** (NH_3) e **acqua**, componenti che oggi sappiamo essere indispensabili per la formazione delle sostanze organiche, caratterizzate appunto da carbonio, idrogeno, ossigeno e azoto.

Nonostante l'atmosfera si trovasse in equilibrio termodinamico, la radiazione solare, la radioattività superficiale, l'attività vulcanica, la caduta di meteoriti, nonché l'effetto di scariche elettriche sicuramente producevano cambiamenti fisici e chimici. Avevano così origine composti complessi con alta energia di legame, in grado di contribuire alla formazione di composti ancora più complessi. Se tali prodotti risultavano instabili nell'atmosfera, certamente avevano modo di accumularsi nei più stabili oceani, in cui erano trasportati dalle piogge: qui le reazioni tra composti tipo HCN, acido cianitrico, favorite probabilmente dalla presenza di **catalizzatori**, poterono portare alla sintesi di **aminoacidi, basi puriniche e porfirine**, come dimostrato dalla riproduzione di tali reazioni in laboratorio.

L'ulteriore evoluzione chimica a **zuccheri, lipidi, proteine ed acidi nucleici** è purtroppo ancora oscura. A partire da questi quattro gruppi biologici fondamentali e dai catalizzatori si sarebbero formate le prime macromolecole e, con lo sviluppo in alcune della membrana cellulare, si sarebbero costituite le **protocellule** (di tipo *procariota*, senza cioè nucleo definito) e avrebbe dunque avuto inizio l'**evoluzione biologica**.

Le cellule primitive erano probabilmente *eterotrofe* e si procuravano l'energia per l'accrescimento e la duplicazione dalla fermentazione, come fanno i batteri anaerobi. Ma pian piano cominciarono ad originarsi cellule *autotrofe*, in grado di scomporre l'acqua attraverso l'energia solare e di trasferire l'idrogeno all'anidride carbonica, formando carboidrati e liberando ossigeno secondo una reazione di fotosintesi. È così che si registrò il passaggio da un'atmosfera riducente ad una ossidante, ben testimoniato dai *red beds*, depositi continentali contenenti ferro altamente ossidato, che risalgono a 1,9 miliardi di anni fa. Con l'ossigeno nell'atmosfera si svilupparono le prime **cellule eucariote** (a nucleo definito) e poi, soltanto circa 700 milioni di anni fa, gli organismi pluricellulari eucarioti, sia *autotrofi* (piante) sia *eterotrofi* (animali).

1 Caratteri generali

Il **clima** indica il complesso di condizioni meteorologiche caratteristiche di una certa area per un certo periodo di tempo, condizioni che si succedono di anno in anno con modalità più o meno analoghe. Esso rappresenta il più importante fattore ambientale, da cui dipendono non soltanto tutti i fenomeni della geodinamica esogena ma anche tutti i processi di origine e sviluppo della vita. L'influenza più diretta da parte del clima viene esercitata sul mondo vegetale ma, in realtà, gli effetti si ripercuotono sull'intero processo di evoluzione della **biosfera**. Le piante (**produttori primari**) sintetizzano infatti la sostanza organica che costituisce la fonte principale d'alimentazione per gli animali (**consumatori**). Questi si organizzano in catene alimentari, in modo da creare un circuito continuo di materia e di energia. Quando poi piante e animali muoiono, i loro tessuti, per azione degli organismi **decompositori**, vengono nuovamente trasformati in componenti primari per la vita, come l'acqua e l'anidride carbonica.

Il clima di una regione risulta dalla combinazione dei vari elementi atmosferici (temperatura, pressione, umidità, venti) con i diversi fattori locali (latitudine, altitudine, presenza di mari e laghi, direzione delle catene montuose etc.). Elaborare una classificazione dei climi che tenga conto di tante variabili è un tentativo praticamente impossibile. L'operazione più diffusa consiste, invece, nello scegliere come base uno soltanto tra i fattori essenziali: una volta ottenute le classi principali, le si differenziano ulteriormente sulla base degli altri elementi.

Scegliendo la **temperatura** (intendendo le medie mensili di temperatura dell'aria), si identificano tre gruppi principali:

- **climi delle basse latitudini**, senza inverno [1];
- **climi delle medie latitudini**, con una stagione invernale ed una estiva;
- **climi delle alte latitudini**, senza estate [2].

Se alla base della classificazione si pone il valore medio annuo delle **precipitazioni**, si individuano sette regioni differenti.

REGIONI TERRESTRI IN RAPPORTO ALLE PRECIPITAZIONI			
Denominazione	Intervallo di latitudine	Posizione rispetto ai continenti	Precipitazioni annue (mm)
Zona equatoriale umida	0°-10° N e S	interno, coste	> 2.000
Costa degli alisei	5°-30° N e S	coste	> 1.500
Deserti tropicali	10°-35° N e S	interno, coste occ.	< 250
Deserti e steppe delle medie latitudini	30°-50° N e S	interno	100-500
Regioni subtropicali umide	25°-45° N e S	interno, coste	1.000-1.500
Coste occidentali delle medie latitudini	35°-65° N e S	coste occ.	> 1.000
Deserti artici e polari	60°-90° N e S	interno, coste	< 300

Tab. 1

[1] In nessun mese dell'anno si rilevano temperature medie inferiori a 18°C.

[2] Nessun mese dell'anno presenta temperature medie superiori a 10°C.

Ulteriori tentativi di questo tipo sono stati fatti utilizzando il bilancio dell'**acqua nel suolo** [3], la **radiazione netta**, la **vegetazione** [4].

Tutte queste classificazioni, basate su uno soltanto dei fattori determinanti del clima, sebbene valide di per sé, risultano estremamente limitative e, quindi, poco utilizzabili ai fini di un discorso globale nel quale possano rientrare anche gli organismi vegetali, gli animali e l'uomo.

Per rispondere a tale esigenza l'austriaco **Wladimir Köppen**, nel 1918, individua, sulla base della temperatura e delle precipitazioni, una serie di zone climatiche, i cui limiti di demarcazione risultano adatti anche alla distribuzione della vegetazione e dei suoli. Il suo sistema prevede **cinque gruppi principali di climi** (contraddistinti da lettere maiuscole) con le seguenti caratteristiche [5]:

A	<i>climi tropicali</i> : senza stagione invernale, con precipitazioni annue decisamente abbondanti, che superano l'evaporazione
B	<i>climi aridi</i> : l'evaporazione supera le precipitazioni e non si originano corsi d'acqua a carattere permanente
C	<i>climi temperato-caldi</i> (mesotermici): con una stagione invernale e una estiva
D	<i>climi nivali</i> (microtermici): con la temperatura del mese più freddo inferiore a -3°C e quella del mese più caldo superiore a 10°C
E	<i>climi glaciali</i> : senza stagione estiva

A partire da questi cinque gruppi principali, si individuano *undici sottogruppi*, definiti con una seconda lettera associata alla prima, il cui significato è il seguente:

S	<i>clima della steppa</i> : semiarido, con un valore medio di precipitazioni annue di 380-760 mm (*)
W	<i>clima desertico</i> : arido, con piovosità media annua <250 mm
f	<i>clima umido</i> : senza stagione asciutta (**)
w	<i>clima con l'inverno asciutto</i>
s	<i>clima con l'estate asciutta</i>
m	<i>clima della foresta pluviale</i> : con precipitazioni di tipo monsonico e una breve stagione asciutta (***)
(*) Le lettere S e W si applicano unicamente ai climi aridi B.	
(**) La lettera f si applica soltanto ai gruppi A, C e D.	
(***) La lettera m si associa unicamente al clima tropicale A.	

Gli **undici tipi di climi** che si ottengono sono:

Af-Am	<i>clima tropicale della foresta pluviale</i>	Cs	<i>clima temperato piovoso, con estate asciutta</i>
Aw	<i>clima tropicale della savana</i>	Df	<i>clima nivale, umido in tutte le stagioni</i>
BS	<i>clima della steppa</i>	Dw	<i>clima nivale, con inverno asciutto</i>
BW	<i>clima desertico</i>	ET	<i>clima della tundra</i>
Cw	<i>clima temperato piovoso, con inverno asciutto</i>	EF	<i>clima del gelo perenne</i>
Cf	<i>clima temperato piovoso, umido in tutte le stagioni</i>		

[3] Si tratta del sistema del climatologo Thornthwaite, che utilizza come variabili le precipitazioni e l'evapotraspirazione.

[4] In realtà la *vegetazione* non costituisce un fattore che determina il clima, al contrario rappresenta una risposta alle condizioni climatiche.

[5] Si noti che i gruppi A, C, D ed E sono definiti in base alla temperatura, mentre il clima di tipo B è individuato in base al rapporto precipitazioni-evapotraspirazione.

Per differenziare maggiormente le classi, in funzione della temperatura, è previsto l'uso anche di una terza lettera [6].

L'approccio scelto da Köppen per la classificazione dei climi è essenzialmente di tipo empirico-quantitativo e per questo inevitabilmente trascura le problematiche relative all'origine dei fenomeni descritti. I sistemi di classificazione più moderni, utilizzati nel campo delle scienze naturali, si basano invece su metodi di tipo esplicativo-descrittivi, in grado da una parte di illustrare le caratteristiche dei fenomeni e dall'altra di indagare sulla loro possibile origine. Nelle pagine che seguono viene riportata, in modo schematico e semplificato, la classificazione proposta da Strahler (1984), che prevede comunque il riferimento alle classi introdotte da Köppen.

Strahler individua **tre gruppi principali di climi**, in base alla localizzazione delle regioni di origine delle masse d'aria, nonché alla natura e al movimento di queste ultime. I **climi delle basse latitudini** (gruppo I) sono regolati dalle masse d'aria equatoriale e tropicale; i **climi delle medie latitudini** (gruppo II) subiscono l'influenza delle masse d'aria tropicale, che muovono verso i Poli, e polare, dirette verso l'Equatore; i **climi delle alte latitudini** (gruppo III) sono sotto l'influsso delle masse d'aria polare ed artica. Per ognuno di questi gruppi esistono una serie di climi differenti: nei paragrafi seguenti se ne descriveranno brevemente le caratteristiche (tra parentesi saranno riportate le sigle del sistema di Köppen).

2 I climi delle basse latitudini

Le regioni comprese nell'intervallo di 5° di latitudine N e S, lungo la zona di bassa pressione equatoriale, rappresentano aree di convergenza di masse d'aria calda e instabile. Tale situazione ben spiega l'abbondanza di precipitazioni (superiori, generalmente, a 2.000 mm) che caratterizza il **clima equatoriale** (*Af*); la loro differente distribuzione nei diversi mesi è dovuta allo spostamento stagionale della zona di convergenza equatoriale. Nella maggior parte dell'anno si registra una notevole eccedenza idrica, che dà vita a corsi d'acqua ricchi e costanti (si pensi al Rio delle Amazzoni, nell'America del Sud). L'escursione di temperatura annua risulta estremamente ridotta, perché il Sole è alto nel corso di tutto l'anno e i valori medi si aggirano intorno ai 27°C; leggermente più alta l'escursione giornaliera. Le piogge e le alte temperature favoriscono l'azione dei processi chimici sulle rocce: si formano *latosuoli*, poveri di *humus* ma ricchi di idrossidi di ferro, di manganese e di alluminio (lateriti e bauxiti), e giacimenti di minerali di diverso tipo. Il clima equatoriale caratterizza terre come l'Amazzonia, l'America Centrale, il bacino del Congo, il Golfo di Guinea, l'Indonesia; anche in alcune zone poste a latitudini più alte (comprese, in ogni caso, nei 25° N-S), come l'Australia settentrionale, l'India, la Thailandia, il Giappone, si riscontrano condizioni climatiche simili. L'origine delle masse d'aria è tuttavia diversa, in quanto prevale il regime dei monsoni (*Am*): si ha dunque un'alternanza di una breve stagione asciutta (monsone invernale, con l'aria che muove dall'interno verso l'oceano), con una più lunga stagione piovosa (monsone estivo, con le masse d'aria dirette dal mare verso le terre).

Anche lungo le coste orientali del Centro e del Sud America, dell'Indocina, delle Filippine e dell'Australia nord-orientale, a latitudini comprese tra 10° e 25°, si individuano fasce ad elevata piovosità. L'origine del fenomeno è, in questo caso, ascrivibile alla forte umidità del lato occidentale delle masse d'aria tropicale (alisei provenienti dalle celle di alta pressione), che si scontrano

[6] Con la lettera *a* si indica un clima con estate molto calda e temperatura massima superiore ai 22°C; con la *b*, invece, un clima con estate calda e temperatura massima inferiore a 22°C; la lettera *c* indica la presenza di un'estate fresca e breve (meno di 4 mesi con temperature maggiori di 10°C); la *d* un clima con inverno molto freddo, con temperatura minima superiore ai -38°C; la lettera *h* contraddistingue un clima caldo-asciutto, con temperatura media annua sopra i 18°C; la *k* un clima freddo-asciutto, con temperatura media annua inferiore ai -18°C. Le lettere *a*, *b* e *c* si riferiscono unicamente alle classi *C* e *D*; la lettera *d* è relativa soltanto al clima di tipo *D*; le lettere *h* e *k* definiscono il clima *B*.

con i versanti montuosi, provocando precipitazioni cosiddette *orografiche*, con valori intorno ai 2.000 mm annui. Il **clima costiero degli alisei** (variazione di *Af* e *Am*), rispetto a quello equatoriale, mostra, oltre ad un diverso meccanismo di genesi delle precipitazioni, anche escursioni termiche più pronunciate e valori di temperatura leggermente più elevati.

Le terre emerse comprese tra i 15° e i 35° di latitudine sono per la massima parte caratterizzate da climi asciutti, dovuti all'azione delle masse d'aria tropicale continentale: i moti d'aria discendenti provocano un riscaldamento adiabatico, che riduce estremamente il valore dell'umidità (solitamente intorno al 20%). Complessivamente si parla di **climi dei deserti e delle steppe tropicali**, caratterizzati da estrema aridità: in certe zone la siccità può durare anche diversi anni. Più in particolare si distinguono *zone veramente aride*, i deserti, con precipitazioni annue inferiori ai 250 mm (*BWh*), e *zone semiaride*, le steppe, nelle quali le precipitazioni variano tra 250 e 760 mm all'anno (*BSh*). L'evaporazione è enorme e l'idrografia risulta praticamente assente, a causa del perdurante deficit idrico, se non nei brevissimi periodi di piogge intense (quando c'è infiltrazione di correnti marittime), durante i quali si possono verificare anche fenomeni di piena. Le temperature sono altissime nel periodo estivo (fino ai 50°C), l'escursione annua è moderatamente forte (17°-22°C), mentre assai intensa (spesso intorno ai 22°C) risulta quella diurna. Nelle zone interessate da questo tipo di clima (Nord Africa, Arabia, Iran, Australia occidentale, Cile settentrionale, Sud Africa), si rivela particolarmente intensa l'azione modellatrice del vento; i suoli sono privi di *humus* (o quasi privi, nelle aree steppiche), contengono ferro, che li rende rossastri o giallastri, risultano ricchi in sali e praticamente sterili.

Le coste occidentali delle regioni localizzate tra i 15° e i 30° di latitudine, al contrario di quelle orientali, appaiono estremamente aride, con valori delle precipitazioni inferiori ai 250 mm annui. Il fenomeno, all'apparenza strano, di deserti posti in prossimità degli oceani, si spiega facilmente ricordando che le cellule di alta pressione subtropicale sono per natura aride sul loro lato orientale. Le masse d'aria sono quindi asciutte e stabili, ma le temperature, addolcite dalla vicinanza di masse d'acqua, risultano relativamente fresche (intorno ai 18°C) e l'escursione termica annua appare assai ridotta. Le condizioni generali favoriscono la formazione di suoli molto simili a quelli dei deserti interni. Tale **clima dei deserti delle coste occidentali** (*BWh*, *BWk*) caratterizza il Deserto di Atacama, in Cile, quello della Namibia, nell'Africa sud-occidentale, e alcuni tratti di costa del Marocco, della California e dell'Australia occidentale. Le caratteristiche dei suoli di queste zone sono sostanzialmente simili a quelle dei deserti interni.

Esistono delle zone in cui le condizioni climatiche si rivelano intermedie rispetto al clima equatoriale e a quello desertico: per esse si parla di **clima tropicale secco-umido** (*Aw*, *Cwa*). Tali regioni sono posizionate tra i 5° e i 25° di latitudine N e S (si pensi al Centro e al Sud America, all'Africa centrale, all'Australia settentrionale, all'Asia sud-orientale) e sono caratterizzate da una forte alternanza stagionale. Ad un'estate umida, causata dall'azione delle masse d'aria equatoriale e tropicale marittima (le precipitazioni possono raggiungere i 2.500 mm annui), segue un inverno secco, influenzato dalle masse d'aria tropicale continentale. La piovosità non è così regolare come per il clima equatoriale, ma lo è di più rispetto al clima desertico. Di conseguenza, i corsi d'acqua presentano un regime stagionale molto intenso, che produce frequenti episodi di piena. Le temperature più alte (intorno ai 30°C) si registrano nei mesi primaverili. I suoli sono di tipo residuale, assimilabili a quelli delle regioni equatoriali, generalmente poco fertili, e quindi poco sfruttati, se non in corrispondenza di locali pianure alluvionali.

3 I climi delle medie latitudini

Il **clima subtropicale umido** (*Cfa*) rappresenta un'estensione a latitudini più alte del già descritto *clima costiero degli alisei*. Dal lato occidentale umido delle celle di alta pressione tropicale

si muovono masse d'aria umida e instabile, che invadono prima le coste orientali dei continenti, posti a latitudini comprese tra i 25° e i 35°, quindi si spingono al loro interno. È quello che accade in Argentina, nel Brasile meridionale, nella Cina orientale, nel Giappone meridionale, lungo la costa orientale dell'Australia. Tale clima non ha una stagione veramente secca (infatti anche nel mese più asciutto cadono 30 mm di pioggia): le precipitazioni abbondano durante tutto l'anno (in inverno anche sotto forma di neve e gelo), anche se i temporali sono particolarmente frequenti in estate (il totale annuo si aggira intorno ai 1.000 mm), quando prevale l'azione delle masse d'aria tropicale. Si ha quindi un'eccedenza idrica nel corso di tutto l'anno, appena mitigata da un ridotto deficit estivo. L'estate si presenta molto calda (temperature medie superiori ai 22°C); l'escursione termica si rivela moderatamente forte, paragonabile a quella dei deserti tropicali. Si formano, per azione delle intense e frequenti piogge, suoli di tipo lateritico, poveri di elementi nutritivi e, quindi, poco idonei allo sfruttamento agricolo.

Le coste occidentali delle regioni poste tra i 40°-45° e i 60° di latitudine N e S presentano anch'esse un clima umido, definito **clima temperato oceanico** (*Cfb*, *Cfc*), la cui origine è, però, legata alle masse d'aria polare marittima. Si tratta di un clima piovoso ad estati calde: le precipitazioni sono ben distribuite in tutto l'arco dell'anno [7] (si registra una riduzione evidente soltanto nei mesi estivi), con un valore medio che oscilla tra 750-1.000 mm nelle zone a costa bassa (ad esempio, nella Francia settentrionale e nell'Inghilterra meridionale) e tra 1.500-2.000 mm, lungo le coste montuose [8] (come in Norvegia e in Cile). Il rapporto precipitazioni/evapotraspirazione fa sì che vi sia sempre un'eccedenza d'acqua notevole. Le estati sono da calde a fresche; la vicinanza con l'oceano spiega l'escursione termica annua ridotta e gli inverni prevalentemente miti. I suoli caratteristici delle regioni a clima oceanico sono i *podsol*, ricchi di *humus*, particolarmente adatti a forme di agricoltura diversificata.

Nella zona di transizione tra i deserti tropicali e le regioni a clima oceanico, tra i 30° e i 45° di latitudine, è localizzata una fascia costiera in cui si alternano una stagione asciutta ed una piovosa. Un tipico esempio è costituito dalle terre che si affacciano sul Mediterraneo, da cui il nome di **clima mediterraneo**. Le estati sono asciutte e calde, gli inverni miti e umidi [9]; l'escursione termica risulta ridotta (*Csa*). Spostandosi verso le fasce costiere delimitate dagli oceani (si pensi alla California o ad alcuni tratti di coste africane e australiane), l'estate diviene più fresca (*Csb*) e il valore medio annuo delle precipitazioni diminuisce. I suoli associati a questo clima risultano estremamente variabili: tra i più caratteristici si ricorda la *terra rossa*, formata sui calcari.

Le aree continentali, poste tra i 30° e i 50° di latitudine, come le zone interne del Nord America e dell'Asia, presentano climi di genesi alquanto complessa, definiti **climi dei deserti e delle steppe delle medie latitudini**. Si tratta di una zona d'influenza di masse d'aria diverse: d'estate prevalgono le masse d'aria tropicale continentale, che portano aridità e temperature molto elevate; d'inverno predominano le masse d'aria polare continentale (con lo sviluppo degli anticicloni siberiano e canadese) asciutte e fredde; l'influenza delle masse d'aria marittima polare e tropicale non si manifesta a causa dei versanti montuosi, i quali ne provocano un innalzamento e un riscaldamento «adiabatico», che le priva del contenuto di umidità. Si spiegano in questo modo le modeste o scarse precipitazioni di queste aree (mediamente comprese tra i 250 mm e i 500 mm), che si distinguono in *semiaride*, o *steppe* (*BSk*), e in *propriamente desertiche* (*BSk*). Il deficit estivo di acqua si protrae a lungo nel corso dell'anno e neanche la ricarica invernale serve a bilanciarlo. I valori delle temperature oscillano parecchio a seconda delle zone: rispetto

[7] Soltanto durante i mesi estivi si registra una riduzione evidente. La causa va ricercata nel particolare sviluppo, in tale periodo, delle alte pressioni subtropicali oceaniche, che si spostano più a nord, portando aridità in un'area più vasta.

[8] Alle medie latitudini, l'influenza delle *catene montuose* costiere sulla piovosità risulta fortemente marcata.

[9] Tale situazione di piovosità, associata alla stagione fredda e di siccità nella stagione calda, rappresenta un caso davvero unico tra i vari tipi di clima.

ai deserti delle latitudini più basse, mostrano, però, inverni più freddi ed escursioni termiche annue molto maggiori. I suoli, classificati come suoli bruni e suoli del deserto, risultano poveri di *humus* ma ricchi di calcio, che li rende fertili: tuttavia la scarsità d'acqua rende difficile il loro impiego per l'agricoltura.

Il **clima continentale umido** interessa le regioni situate tra i 40° e i 55° di latitudine N, che si estendono dall'interno dei continenti fino alle coste orientali, come vaste aree del Nord America e dell'Asia. Si registrano grandi contrasti stagionali: in inverno prevalgono le masse d'aria polare continentale fredda (con temperature anche sotto lo zero), d'estate predominano le masse d'aria tropicale calda (con valori fino ai 25°-28°C). L'escursione termica annua oscilla intorno ai 30°C, ma va detto che la situazione cambia da zona a zona, a seconda della latitudine e della vicinanza agli oceani. Le precipitazioni, poiché si tratta di zone disposte lungo il fronte polare, abbondano tutto l'anno, anche durante l'estate, che può essere molto calda (*Dfa*) o calda (*DFb*). Tuttavia, in alcune zone, come la Siberia orientale o la Corea settentrionale, l'inverno è generalmente più asciutto (*Dwa*, *Dwb*). Anche i suoli risentono della forte variabilità del clima: si formano suoli bruni, suoli della prateria e *podsol*. In ogni caso, sono ricchi di *humus* e possono essere utilizzati per un'agricoltura diversificata.

4 I climi delle alte latitudini

Le terre localizzate tra i 50° e i 70° di latitudine N, come le regioni interne più settentrionali dell'America e dell'Asia, subiscono essenzialmente l'influenza delle masse d'aria polare continentale. Il **clima subartico continentale** che ne deriva è caratterizzato da inverni molto rigidi, con temperature minime mediamente di 20°-30°C sotto zero [10], estati brevi, miti (la media del mese più caldo si aggira intorno ai 10°C, sebbene quotidianamente si raggiungano anche i 20°C) e piovose. Le precipitazioni, infatti, sono concentrate nei mesi caldi e il valore annuo totale oscilla intorno ai 300 mm. Complessivamente si tratta di un clima umido (classificato da Köppen come *Dfc* o *Dfd*, a seconda delle temperature estive e invernali), in quanto l'evapotraspirazione risulta praticamente inesistente, se non in brevissimi periodi in estate, in cui, clamorosamente, si può verificare anche una situazione di deficit idrico. In alcune regioni dell'Asia nord-orientale, tuttavia, è possibile trovare una variante di questo clima che mostra un inverno assolutamente asciutto (*Dwc*, *Dwd*). Mediamente l'escursione termica annua del clima subartico continentale raggiunge valori di circa 40°C (con punte di 60°C in Siberia), rivelandosi maggiore di qualunque altra sulla Terra. Dal momento che per sei o sette mesi le temperature si mantengono al di sotto del punto di congelamento (l'inverno rappresenta senz'altro la stagione dominante in questo tipo di clima), il suolo che prevale in queste zone è il *permafrost*: *continuo* in regioni come la Siberia, *discontinuo* in altre poste a latitudini minori. Tuttavia durante i 3-4 mesi estivi, da maggio ad agosto, quando il Sole si mantiene al di sopra dell'orizzonte per 16-18 ore, le condizioni del suolo sono tali da permettere alcune forme di coltivazione.

Lungo le coste comprese tra i 50° e i 70° di latitudine, le condizioni climatiche si mostrano profondamente diverse rispetto a quelle dell'interno dei continenti. Il **clima subartico oceanico** (indicato dai climatologi come *EM*, sebbene in realtà non abbia riscontro nella classificazione di Köppen) è contraddistinto da un totale annuo di precipitazioni eccezionalmente alto (tra 700 e 1.300 mm circa) e da un'escursione termica ridottissima per latitudini così elevate. La spiegazione va ricercata nell'influenza esercitata dalle masse d'aria polare marittima. La temperatura media annua si aggira intorno a 1°-2°C, quella del mese più freddo non è inferiore ai -7°C. Le regioni

[10] In alcune località i valori risultano anche molto inferiori: a Verkhoyansk, in Russia, la temperatura media per il mese di gennaio è di -51°C, ma si è registrato anche un picco negativo pari a -69°C.

interessate da questo clima sono poche e concentrate nell'Emisfero Artico: la costa meridionale della Groenlandia, l'Islanda settentrionale e il nord della Norvegia.

Oltre i 55° di latitudine N e S, lungo il cosiddetto fronte polare, predomina il **clima della tundra (ET)** legato all'azione delle masse d'aria polare continentale e polare marittima. Ne sono interessate le frange settentrionali del Nord America e dell'Eurasia. Le perturbazioni di tipo ciclonico, numerose e intense, provocano precipitazioni concentrate nella breve estate fresca, che raggiungono mediamente i 250 mm annui. L'escursione termica risulta elevata, ma non al pari delle zone a clima subartico continentale: le temperature medie estive sono appena superiori ai 4°C, mentre la media del mese più freddo è intorno ai -20°C. L'evapotraspirazione è pressoché nulla, di modo che si registra un deficit idrico estivo. I suoli sono estremamente caratteristici: si tratta di *permafrost*, che nascono unicamente da disgregazione fisica e sono essenzialmente formati da torba e argilla con un colore grigio-blu.

La Groenlandia, l'Antartide e i ghiacci galleggianti sull'Oceano Artico sono dominati dal **clima delle calotte glaciali (EF)**. Si registrano le temperature più basse di tutto il pianeta (in Antartide è stata misurata una minima record di -88,3°C): le medie annue oscillano tra i -30°C e i -35°C sui continenti e tra i -20°C e i -25°C sugli oceani, dove non esistono rilievi; in nessun giorno dell'anno si registrano valori superiori a 0°C. Si alternano sei mesi d'insolazione continua e sei mesi di notte continua. Le precipitazioni hanno origine da perturbazioni cicloniche di passaggio e assai frequenti sono le *blizzard*, violente bufere di neve molto basse sul suolo ghiacciato.

Elementi di fitogeografia e di zoogeografia

1 Caratteri generali

La **vegetazione** costituisce uno degli elementi fondamentali del paesaggio ed è per questo che il suo studio rientra nella sfera di interessi del geografo. La **fitogeografia**, in generale, è la scienza che studia la distribuzione dei vegetali sulla Terra, dalle loro origini [1] ai giorni nostri, tentando le previsioni sul loro dinamismo. Nel capitolo 9 di questa Sezione, in un discorso più ampio di *paleogeografia*, si è fatto cenno alla nascita, allo sviluppo e alla distribuzione degli organismi vegetali nel corso della storia geologica del pianeta; in questo capitolo ci occuperemo, invece, degli aspetti attuali, correlando in particolare il mondo vegetale agli ambienti climatici precedentemente individuati.

Nel biociclo terrestre, ambiente della biosfera che comprende le terre emerse [2], si distinguono quattro grandi **biocore** (dal greco: *bios*=vita e *chora*=regione): la *foresta*, la *savana*, la *prateria* e il *deserto*. Nell'ambito di ogni *biocora* ci sono una decina di **classi di formazioni**, ossia diverse varietà di associazioni vegetali che, in base al regime climatico, individuano una regione (ad esempio, foresta pluviale, foresta decidua, foresta ad aghifoglie etc.). Ciascuna di queste regioni, o classi di formazioni, rappresenta un insieme di più **habitat** differenti, determinati da condizioni locali di pendenza, altitudine e tipo di suolo [3]. All'interno di ogni habitat si trova il più piccolo ecosistema, la **comunità vegetale**, costituita da un numero limitato di organismi più o meno interdipendenti.

Gli elementi essenziali che occorre conoscere per poter descrivere correttamente un **organismo vegetale** (*descrizione strutturale*) sono i seguenti:

- la **forma biologica**: le piante si classificano in *alberi*, caratterizzati da un unico tronco e da una ramificazione superiore ricoperta da foglie; in *arbusti*, composti da più fusti con le foglie già al livello del suolo; e in *erbe*, senza fusti legnosi. A stretto contatto con il terreno, o sui tronchi, vivono le *brioidi* (licheni e muschi). Forme vegetali inferiori, denominate *tallofite*, sono i batteri, le alghe, le muffe e i funghi;
- le **dimensioni** e la **stratificazione**: in base a degli standard di riferimento, le piante si classificano in *alte*, *medie* e *basse*, e si distinguono in *strati*;
- la **copertura**: si definisce *mancante* o *molto sparsa*, *discontinua*, in *ciuffi* o *gruppi*, *continua*;
- la **funzione** e la **periodicità**: le piante *decidue* perdono le foglie durante la stagione a loro sfavorevole; le *sempreverdi* mantengono le foglie durante tutto l'anno; le *semidecidue* perdono le foglie ad intervalli che non corrispondono per forza alle stagioni; le *sempreverdi succulente*

[1] La *paleobotanica*, o *paleontologia vegetale*, costituisce uno dei settori più importanti della *fitogeografia*: attraverso lo studio dei fossili ricostruisce le caratteristiche della vegetazione nel passato geologico, tentando anche di spiegarne l'evoluzione.

[2] Gli altri ambienti, o *biocicli*, della biosfera sono l'*acqua salata* (oceani) e l'*acqua dolce* (corsi d'acqua, laghi, stagni). Nel descrivere le caratteristiche di una classe di formazioni, si considera generalmente l'habitat di *altopiano*, che rappresenta lo standard di riferimento.

[3] Alcune piante possono vivere anche in ambienti climatici diversi da quelli originariamente ad esse più consoni. In tal caso si parla di piante *acclimatate*: gli agrumi, provenienti dall'Asia, oggi perfettamente ambientati nelle regioni a clima mediterraneo, costituiscono un esempio emblematico di questo fenomeno.

e le *sempreverdi senza foglie*, che, rispettivamente, mantengono tutto l'anno le loro foglie polpose e spesse, non hanno proprio foglie ma solo fusti carnosì, come i cactus;

- la **forma** e le **dimensioni della foglia**: alla forma a *lamina espansa*, tipica di alberi come l'acero o il faggio, si contrappone l'*aghifoglia*, caratteristica dell'abete o del pino. Le *spine* rappresentano la trasformazione di un'aghifoglia;
- la **struttura della foglia**: lo spessore varia da *sottile* a *grosso* e dipende fortemente dall'habitat. Le foglie dure, spesse e coriacee vengono definite *sclerofille*; quelle spesse e spugnose, *succulente*.

La struttura, l'associazione e la distribuzione della vegetazione risultano determinate da più fattori ambientali, di natura climatica, geomorfologica, edafica (ossia in rapporto al suolo) e biologica, che agiscono contemporaneamente e s'influenzano a vicenda.

Tra i **fattori climatici** un ruolo primario spetta all'acqua, al calore e alla luce. Le piante assorbono la maggior parte dell'**acqua**, di cui hanno bisogno per l'accrescimento e per il processo di fotosintesi, dal suolo, attraverso le radici. A causa della traspirazione ne perdono una certa quantità: quanto più è ampia la superficie delle foglie, tanta più alta risulta la quantità d'acqua che traspira. Si spiega così il fatto che, in condizioni di clima caldo e asciutto, le piante siano prevalentemente ad aghifoglie o a spine, o siano addirittura prive di foglie, come i cactus, mentre le strutture a lamina espansa siano diffuse negli ambienti umidi. In base al bisogno d'acqua, le piante si distinguono in **igrofile**, se necessitano di un ambiente umido (come le ninfee); **xerofile**, se sono capaci di vivere in climi aridi (come i cactus); e **mesofile**, se si adattano a zone con valori di umidità intermedi. Si dicono **tropofile** quelle piante capaci, attraverso un mutamento dell'apparato fogliare, di adattarsi a climi con alternanza di periodi secchi ed umidi. Il regime delle precipitazioni rappresenta, dunque, un elemento fortemente significativo nello sviluppo della vegetazione: va detto a tal proposito che il fattore determinante non è rappresentato dall'entità delle piogge, quanto dalla loro intensità e dalla loro distribuzione nel corso dell'anno. È chiaro, inoltre, che il bisogno d'acqua è anche proporzionale alla temperatura: quanto più essa è elevata, tanta più acqua risulterà necessaria.

La **temperatura**, o ancor meglio, il calore, costituisce infatti un altro fattore che condiziona enormemente la vegetazione. Per ogni pianta esiste una temperatura ottimale associata a ciascuna delle sue funzioni (fotosintesi, fiorazione, fruttificazione etc.) ed un valore minimo e massimo al di là dei quali è impossibile la sopravvivenza. In tal senso, le piante si definiscono *megaterme*, se hanno bisogno di una temperatura media superiore ai 20°C; *mesoterme*, se vivono bene in un intervallo termico di 15°-20°C; *microterme*, se necessitano di temperature superiori a 0°C e inferiori a 15°C; e infine *echistoterme*, se sopportano anche temperature sotto zero. I differenti regimi delle temperature, come si è precedentemente visto, sono associati a diverse regioni del pianeta: ciò significa che, nella maggior parte dei casi, i limiti di distribuzione dei vari tipi vegetali corrispondono anche a dei confini geografici (si parla di *frontiere*) [4].

Un altro influsso determinante è quello della **luce**, anch'essa funzione della latitudine; non tutte le piante ne hanno bisogno allo stesso modo, tanto che si distinguono le piante **ombrofile**, che necessitano di poca luce, e le piante **eliofile**, che mostrano un'esigenza opposta.

Accanto a questi elementi di natura climatica, esistono una serie di **fattori geomorfologici**, quali l'inclinazione del pendio, l'esposizione, la forma del rilievo, che, indirettamente, influenzano l'aspetto e la distribuzione delle piante, in quanto modificano le quantità d'acqua, di luce e di calore che ad esse dovrebbero pervenire.

Il rapporto **suolo-vegetazione** (*fattori edafici*) è invece più diretto, ma, in qualche modo, anche più complesso: se è infatti vero che le caratteristiche del suolo influenzano notevolmen-

[4] La paleontologia, attraverso lo studio dei fossili, studia le caratteristiche della fauna nel passato geologico, interessandosi anche di spiegarne l'evoluzione.

te la vegetazione, è anche vero il contrario, ossia che le piante contribuiscono a formare e ad alterare il suolo.

Infine, vanno presi in considerazione i **fattori biologici**: dall'attività dei batteri a quella degli insetti, dall'azione degli animali erbivori all'intervento dell'uomo, tutto influisce, in maniera maggiore o minore, sullo sviluppo della vegetazione, mutandone in alcuni casi anche la naturale evoluzione.

La distribuzione della vegetazione rappresenta, a sua volta, un elemento discriminante e determinante per lo sviluppo della **fauna**. La **zoogeografia**, che si interessa di ricostruire la distribuzione degli organismi animali sulla Terra dalle origini della vita fino ad oggi, risulta dunque intimamente legata alla *fitogeografia* e alla *climatologia*.

La vegetazione costituisce la fonte primaria di vita per gli animali erbivori, che scelgono il proprio habitat in funzione del cibo disponibile. Gli animali carnivori, che si nutrono degli erbivori, risultano a loro volta influenzati dalla distribuzione della vegetazione, anche se indirettamente [5]. Il rapporto fauna-clima è sia indiretto sia diretto (nel senso che il clima influenza le caratteristiche della vegetazione, che a sua volta condiziona lo sviluppo degli animali): la luce, il calore, l'umidità e la pressione atmosferica sono dei fattori estremamente importanti anche per gli organismi animali. Esistono gruppi di animali che ben si adattano a situazioni anche molto diverse, ma ve ne sono altri fortemente legati ad habitat specifici. Così, ad esempio, le civette, i pipistrelli e tutte le specie definite «notturne», non tollerano la luce in quanto il loro organismo si è adattato a vivere nell'oscurità. Ve ne sono alcuni che riescono a compiere il proprio ciclo biologico soltanto in ambienti caldi, altri che prediligono gli ambienti freddi, altri ancora che si adattano anche a forti sbalzi di temperatura. Sono invece pochi gli animali che sopravvivono in condizioni di alta pressione, sia atmosferica (i rapaci) sia sottomarina (alcuni pesci abissali); generalmente la maggior parte della fauna terrestre ha bisogno di notevoli quantità di ossigeno, reperibile esclusivamente in ambienti con pressione normale. Altro elemento determinante è l'acqua: nessun essere vivente riesce a sopravvivere senza; tuttavia, mentre alcune specie necessitano di grosse quantità, disponibili soltanto in ambienti umidi, altre si possono accontentare di piccoli quantitativi e riescono quindi ad abitare anche le regioni più secche.

Anche per la fauna è dunque possibile fare un discorso di distribuzione geografica alquanto definita, sebbene gli animali abbiano, rispetto alle piante, il grande vantaggio di potersi spostare, in caso di condizioni avverse (siano esse variazioni climatiche, eventi geologici o mutamenti che coinvolgono il regno vegetale), alla ricerca di un ambiente di vita più favorevole. I raggruppamenti di specie che hanno esigenze ambientali alquanto simili vengono definiti **associazioni animali** (o **biomi**): essi, insieme con le classi di formazioni vegetali, popolano le diverse regioni della Terra precedentemente distinte in base al clima.

2 Classi di formazioni vegetali e associazioni animali

Di seguito vengono sintetizzate le caratteristiche più significative delle principali **classi di formazioni vegetali** [6] delle *quattro biocore terrestri* (foresta, savana, prateria, deserto), con riferimenti al clima e al regime pedogenetico. Nelle brevi introduzioni generali sulle biocore, sono state inserite anche alcune indicazioni fondamentali sugli organismi animali che è possibile trovare associati ad esse.

[5] Sono pochissimi gli animali che fanno a meno degli organismi vegetali per la nutrizione: si tratta dei *bentonici* di ambiente marino profondo e della fauna che popola le grotte.

[6] Si segue la classificazione proposta da Strahler (1984), ispirata al lavoro di Dansereau, dal titolo *Biogeography, an ecological perspective* (1957).

a) Foresta

La *foresta* è una formazione vegetale costituita da alberi ravvicinati, stratificati in altezza, che, con la copertura pressoché continua del loro abbondante fogliame, ombreggiano il terreno, creando anche locali microclimi; questa biocora richiede precipitazioni abbondanti, anche non necessariamente ben distribuite nel corso dell'anno, e risulta dunque associata a più climi diversi (dall'equatoriale, al mediterraneo, al tropicale etc.).

I biomi forestali risultano estremamente variabili in funzione del clima: ogni tipo di foresta possiede una sua fauna, strettamente legata al patrimonio floristico. Le foreste dei climi caldi ospitano innumerevoli specie di *uccelli* (dai grossi *pappagalli* straordinariamente variopinti ai piccoli *colibri*), varie famiglie di *primati* (*scimmie*), *serpenti*, *coccodrilli*, *ippopotami*. La foresta monsonica è popolata anche da numerosi esemplari di *elefanti* e *tigri*. Nelle foreste dei climi più freddi, invece, sono diffusi i *carnivori* di diverse specie (in particolare *lupi* e *volpi*), *ungulati* (*cervi*, *alci* etc.), *scoiattoli*, *orsi*, *renne*.

Nell'ambito della biocora della foresta, si distinguono *sette classi* di formazioni vegetali principali.

1. La **foresta pluviale equatoriale** è formata da alberi alti fino a 40 m, molto ravvicinati, stratificati, a corteccia liscia, privi di rami nella parte bassa, con foglie grandi e sempreverdi (di qui il nome di *foresta sempreverde a latifoglie*). Il suolo, interessato dal fenomeno di laterizzazione, si mostra poco fertile: il terreno risulta quasi perennemente ombreggiato, coperto soltanto da uno strato sottile di foglie e soggetto ad un'intensa azione batterica, che spiega la quasi totale assenza di *humus*. L'ambiente è popolato da un gran numero di specie vegetali, che convivono anche in spazi molto ristretti [7]: tipici sono il *noce del Brasile* e l'*albero del cotone serico*, diffusi nel bacino delle Amazzoni, le *magrovie* e le *palme*, comuni lungo le coste equatoriali. Abbondano le *liane*, piante rampicanti legnose, appese al tronco o ai rami, e le piante *epifite*, che utilizzano gli alberi per sostenersi (un esempio è costituito dal *fico strangolatore*). Laddove la foresta ha subito tagli o incendi, la vegetazione si riproduce mantenendosi più bassa e viene definita *jungla*: tipici di questa formazione sono il *bambù* e arbusti di vario tipo. La foresta pluviale equatoriale si sviluppa in zone con clima umido e caldo tutto l'anno: infatti è diffusa in Sud America, nel bacino del Congo e nelle Indie Orientali.
2. La **foresta pluviale tropicale** è per molti aspetti simile a quella equatoriale: la variazione del regime pluviometrico e termometrico (rispetto alle zone equatoriali, ai Tropici si alternano una stagione piovosa ed una a piovosità ridotta, e le temperature sono più basse), infatti, è tutto sommato poco influente nei riguardi della vegetazione. Le regioni interessate da questa classe di formazioni sono i Caraibi, la zona occidentale dell'India, le coste della Birmania, del Vietnam, delle Filippine, del Madagascar, del Brasile e dell'Australia nord-orientale. Il regime pedogenetico è quello della *laterizzazione*.
3. La **foresta monsonica** è predominante in Indonesia, in Thailandia, in Cambogia, nell'Africa occidentale, nell'Australia settentrionale, nel Centro e Sud America. Rispetto a quella equatoriale e tropicale, presenta un maggiore sviluppo della vegetazione più bassa, in quanto gli alberi più alti, sotto i 35 m, hanno le chiome più aperte e lasciano filtrare di più la luce. Prevalgono il *teak* e il *bambù*, ma sia le *liane* sia le *epifite* hanno buona diffusione. Durante la lunga stagione asciutta, tipica dei climi monsonici, gli alberi perdono le foglie (di qui il nome di *foresta tropicale decidua*), ma la forte lisciviazione nel periodo delle piogge non permette l'accumulo di *humus*: il suolo si forma, comunque, per *laterizzazione*. L'ambiente monsonico

[7] Gli individui di una singola specie risultano quindi isolati, e ciò implica un faticoso lavoro di selezione qualora si voglia tagliare a scopo commerciale.

è stato spesso trasformato dall'uomo, che vi ha impiantato piantagioni di *riso*, *tè*, *cotone* o *canna da zucchero*.

4. La **foresta pluviale temperata** (detta anche *foresta temperata sempreverde* o *foresta a laurifoglie*) è caratterizzata dalla presenza di poche specie di alberi e da vasti raggruppamenti di individui vegetali simili. Gli alberi, rispetto a quelli delle foreste già descritte, risultano più bassi, con le foglie più piccole, coriacee e meno addensate. Tipiche sono le *querce sempreverdi*, le famiglie del *lauro* e quelle della *magnolia*, nel Giappone meridionale e negli Stati Uniti; le *conifere*, i *faggi* e le grandi *felci*, in Nuova Zelanda. La vegetazione degli strati inferiori è costituita da *felci*, *bambù*, piccole *palme*, *arbusti* e *piante erbacee*; sono abbondanti *liane* ed *epifite* e, spesso, i tronchi risultano coperti anche da *muschi*. Il regime climatico in cui s'impone questo tipo di foresta è caratterizzato da piovosità abbondante, ben distribuita durante tutto l'anno, e da un'escursione termica molto moderata. Il regime pedogenetico è quello della *laterizzazione*, associato a processi di *podsolizzazione* nelle regioni più fredde, alle più alte latitudini.
5. La **foresta decidua a fogliazione estiva** (detta anche *foresta temperata decidua*) è tipica del Nord America, dell'Europa e dell'Asia orientale, ed è contraddistinta da alberi alti con grandi foglie a copertura continua che, tuttavia, cadono completamente d'inverno. Negli strati inferiori sono poco diffusi piccoli alberi e *arbusti*, mentre in primavera è ben sviluppato un manto erbaceo. Le specie più abbondanti sono la *quercia*, il *faggio*, la *betulla*, il *noce*, l'*acero*, il *tiglio*, l'*olmo*, il *frassino*, il *castagno* e il *carpino*. Nell'Europa centrale e occidentale predominano, invece, *quercia*, *frassino* e *faggio*. Quando queste foreste vengono diradate dall'uomo per produrre legname (si tratta di un'attività che ha avuto un'enorme diffusione in passato e che risulta oggi responsabile della quasi totale scomparsa di questa formazione), possono nascere, come vegetazione secondaria, alberi di *pino*. Le condizioni climatiche che determinano la formazione della foresta decidua sono le precipitazioni abbondanti in tutti i mesi (ma soprattutto in estate) e l'alternanza di una stagione fredda e una calda. I suoli generalmente sono di tipo *podsol*, ma verso latitudini più basse si può avere anche *laterizzazione*.
6. La **foresta ad aghifoglie** è costituita in massima parte da *conifere*, alberi a forma conica, generalmente molto ravvicinati a formare una copertura completa, dotati di foglie piccole e strette, a forma di ago. Questa formazione è diffusa in due grandi fasce, comprese tra i 45° e i 75° di latitudine, localizzate in Nord America e in Eurasia. Le specie più abbondanti sono il *peccio*, l'*abete* e il *pino*, tra le conifere sempreverdi, e il *larice*, tipico della Siberia, che perde i suoi aghi in inverno. Spesso associati alle conifere si trovano il *frassino*, il *pioppo*, il *salice* e la *betulla*. Le foreste ad aghifoglie della California sono invece caratterizzate dalle *sequoie*, che raggiungono anche i 100 m di altezza e i 20 m di circonferenza. Il clima propizio per la formazione di questo tipo di foresta è quello continentale, con abbondanti precipitazioni e temperature basse; i suoli che, di conseguenza, si formano sono di tipo *podsol*. Le foreste a conifere costituiscono per l'uomo un'importante e preziosa risorsa economica, in quanto forniscono il legname per le costruzioni e per l'industria della carta.
7. La **foresta a sclerofille** consiste in alberi bassi sempreverdi, con foglie piccole e dure, associati ad arbusti (quando questi prevalgono, si parla di **macchia**) e ad uno strato erbaceo in primavera. Prevalgono il *sughero*, il *leccio*, il *pino d'Aleppo*, il *pino da pinoli*, e l'*olivo*. Questa formazione, ampiamente alterata dall'intervento dell'uomo, risulta diffusa nei Paesi mediterranei, in Cile, in Sud Africa e in parte della California (dove predominano la *quercia bianca* e l'*agrifoglio*), in quanto è associata a un clima caratterizzato da un'estate calda e asciutta e da un inverno piovoso e mite. Il regime pedogenetico è quello della *calcificazione*, che produce, nelle regioni del Mediterraneo, la nota *terra rossa*.

b) Savana

La *savana* consiste in un'associazione di alberi, disposti a gruppi, in maniera discontinua, arbusti ed erbe, che formano gli strati più bassi; il suo sviluppo è legato a climi con scarse precipitazioni, distribuite irregolarmente durante l'anno. Numerose regioni «a savana» sono state completamente trasformate dall'uomo, che ha sostituito la vegetazione naturale con colture cerealicole o con piante industriali, come l'*arachide*, la *canna da zucchero*, il *cotone* e il *caffè*. Nelle savane, i grossi Felini (*tigri*, *pantere*, *leopardi*, *ghepardi*) predano piccoli e grandi erbivori (*giraffe*, *antilopi*, *bufali*, *gazzelle*, *zebre*, *elefanti*). Inoltre sono presenti grandi insetti (come le *cavallette* e le *termiti*), uccelli corridori (*struzzi*, *nandù*) e rettili.

Questa biocora comprende almeno sei classi di formazioni vegetali principali.

1. La **savana a parco** (o *bosco tropicale*) è formata da alberi di media altezza, con foglie piccole e spine (che possono cadere nella stagione asciutta), alquanto distanziati, tra i quali si sviluppano erbe e arbusti. Questa formazione comprende i boschi di *pino giallo* e di *ginepro* degli Stati Uniti occidentali, quelli di *eucalipto* dell'Australia meridionale, del Centro e Sud America, dell'Africa e dell'Asia sud-orientale. È associata ad un clima sufficientemente arido da impedire uno sviluppo continuo degli alberi. Tra i processi pedogenetici predomina la *laterizzazione*.
2. La **savana spinosa** (o *macchia tropicale*) è costituita da un'associazione di piccoli alberi e arbusti, che formano cespugli legnosi, spinosi e ravvicinati. Piante annuali o graminacee formano lo strato erbaceo inferiore. Il regime climatico in cui si sviluppa questo tipo di savana è quello tropicale, con alternanza di umidità ed aridità. Risulta dunque diffusa in Brasile e in Sud Africa. I suoli sono sottoposti a *calcificazione* e, spesso, a *salinizzazione*.
3. Se il periodo di siccità risulta alquanto fresco, mentre il caldo più forte si fa sentire nel periodo immediatamente precedente alle piogge, si forma la **savana**. È anch'essa costituita da alberi ed arbusti, e, nello strato inferiore, da graminacee con steli grandi, rigidi e duri, che raggiungono comunemente i 2 m di altezza. Nella savana africana, che prende il posto della foresta pluviale nelle zone di altopiano equatoriale, quali il Sudan, il Kenya e la Tanzania, predominano le *acacie*, il *baobab* e l'*erba elefante*. Esempi di savana si trovano anche in Sud America, nell'Australia settentrionale, in India e in Asia sudorientale. I processi pedogenetici attivi sono la *laterizzazione* e, alle latitudini maggiori, la *calcificazione*.
4. L'assenza di alberi caratterizza il **semideserto**, costituito solo da arbusti (come l'*artemisia*) e da uno strato erbaceo poco sviluppato. Diffuso nella regione delle Montagne Rocciose centrali e meridionali, è tipico di un clima arido con alte temperature, sia esso subtropicale sia delle medie latitudini. I processi pedogenetici associati sono la *calcificazione* e la *salinizzazione*.
5. La **landa** (o **brughiera**) è formata da uno strato basso di arbusti, in modo particolare della famiglia delle *Ericacee* e, spesso, da *muschi*. Ha origine in climi freschi, con precipitazioni ben distribuite e bassa escursione termica annua, come quelli che caratterizzano le coste occidentali delle medie e alte latitudini: è infatti tipica dell'Irlanda, delle Isole Britanniche, nonché delle coste dell'Europa occidentale e centro-settentrionale. Alla brughiera sono associati generalmente i *podsol*.
6. Nelle zone con clima subartico o della tundra, in cui gli inverni sono rigidi e l'acqua del suolo si mantiene ghiacciata per molti mesi dell'anno, la savana si presenta come **taiga**. Si tratta di una formazione prevalentemente arbustiva, con alberi bassi (*abete nero*, *larice americano*, *betulla stentata*) molto distanziati ed una fitta copertura di *muschi* e *licheni*. Le aree di diffusione sono il Canada, l'Alaska e la Scandinavia. I processi pedogenetici sono la *gleizzazione* e la *podsolizzazione*.

c) Prateria

La *prateria* è caratterizzata da vegetazione prevalentemente erbacea, continua o discontinua e, in alcuni casi, può presentare anche alberi sparsi; è tipica di un clima secco, da estremamente caldo a estremamente freddo. Le praterie ospitano grandi e piccoli erbivori (come *bisonti*, *bufali*, *antilopacpre*, *tassi*, *scoiattoli*) che sono preda di carnivori di varie specie, come i canidi e i rettili. Rappresentano ottimi pascoli per l'allevamento ma, spesso, vengono utilizzate dall'uomo anche per colture alimentari e industriali, come *mais*, *girasole*, *grano* e *barbabietola*.

Nell'ambito della biocora della prateria, si individuano *tre classi* principali.

1. La **prateria** propriamente detta è formata da erbe alte a foglia espansa, soprattutto da graminacee, mentre mancano quasi completamente alberi ed arbusti. È tipica dei climi continentali delle medie latitudini, in cui evaporazione e precipitazioni si bilanciano, e infatti risulta diffusa in Nord America, in Argentina (le famose *pampas*), in alcune regioni europee, come l'Ungheria, e nella Cina settentrionale. I suoli sono prevalentemente interessati da processi di *calcificazione*.
2. La prateria ad erbe basse è detta **steppa**. Costituita da molte specie di graminacee e di altre erbe, risulta povera di arbusti e di alberi. La sua diffusione è enorme: si spinge dall'Equatore fino ai 55° N e i 45° S di latitudine. È associata ad un clima continentale semiarido, con estati calde ed inverni freddi e asciutti. Per i suoli prevalgono *calcificazione* e *salinizzazione*.
3. Laddove il clima è molto freddo e umido, come nelle regioni subartiche, predomina la **tundra erbosa**, formata da piante erbacee, *muschi* e *licheni* e, in qualche caso, da alberi bassi come il *salice nano*. Il suolo ha origine per *gleizzazione* e si trova, di solito, in condizioni di *permafrost*. Ad altitudini elevate, tra il limite degli alberi e al di sotto della zona a roccia nuda con nevi perpetue, si sviluppa invece la **tundra alpina**, più simile a quella artica che a quella erbosa.

d) Deserto

Il *deserto* è per la massima parte costituito da terreno nudo, in quanto la vegetazione (piante piccole come erbe, *licheni* e *brioidi*) risulta assai rada. Questa biocora è associata a climi estremamente aridi, siano essi assai caldi o eccezionalmente freddi, e in entrambi i casi la fauna risulta molto povera. I deserti caldi sono popolati da rettili, roditori, mammiferi di piccola taglia e *struzzi*, che vivono nelle oasi o ai margini della steppa. Assai diffuso è il *cammello* che, però, non è un animale selvatico ma viene utilizzato dall'uomo per gli spostamenti da un'oasi all'altra. Molti deserti freddi risultano del tutto inospitali per gli animali; quelli più prossimi alla taiga o alla tundra sono abitati da grossi erbivori, come la *renna*, da *roditori*, da *foche* e *trichechi*, da alcuni uccelli rapaci e dai mammiferi con la pelliccia, come la *volpe artica*.

Il **deserto arido** è caratterizzato da una copertura quasi nulla del terreno: piccoli arbusti a foglie dure o con spine, piante succulente come i *cactus* o graminacee coriacee sono le forme vegetali più comuni. In realtà le specie variano a seconda della regione, dal momento che il deserto è diffusissimo su tutto il pianeta. Così, mentre negli Stati Uniti sud-occidentali si trovano spesso piante di grandi dimensioni, nel deserto del Sahara prevalgono le graminacee ruvide e le *tamerici*, e nel deserto costiero dell'Africa sudoccidentale la *tumboa*. I suoli associati si formano per *salinizzazione* o, in caso di terreni elevati, per *calcificazione*.

In corrispondenza degli estremi settentrionali delle terre emerse si trova il **deserto artico**, che consiste in superfici di terreno roccioso prodotte dalla *gelifrazione*, con scarsa vegetazione di *muschi*, *licheni* e piccoli cespugli. I suoli si formano essenzialmente per processi fisico-meccanici di disaggregazione della roccia.

1 L'equilibrio ecologico

L'insieme degli elementi naturali, inerti (abiotici) e/o viventi che, con rapporti di interazione, occupano una determinata area costituisce un **ecosistema**. L'**ecologia** è la scienza che si occupa di studiarne le caratteristiche e l'evoluzione. Un *ecosistema* rappresenta, dunque, il connubio tra l'ambiente, detto anche **biotopo** [1], e la totalità delle specie che in esso vivono, definita anche **biocenosi** [2]; esso comprende anche l'**insieme delle relazioni tra i vari elementi**, che si traducono generalmente in uno scambio di materia o di energia e si manifestano sotto forma di catene o di cicli.

Ne è un esempio la **catena alimentare**, simboleggiata da una piramide alla cui base si trovano le *piante*, organismi produttori che sfruttano l'energia solare per creare sostanza organica; quindi, muovendosi verso il vertice, gli *animali erbivori* che si nutrono delle piante, i *carnivori* che mangiano gli animali erbivori, gli *uomini* che si cibano degli uni e degli altri e, infine, i *microrganismi decompositori* che trasformano tutti i residui in sostanza organica: a questo punto il ciclo si chiude così come era iniziato. In tal senso, pertanto, nel concetto di «ecosistema», non ci sono soltanto «elementi», quali un fiume, una foresta o un'associazione di animali, ma sono compresi anche tutti i «fenomeni dinamici» che li mettono in relazione e che sono inevitabilmente legati fra loro.

Complessivamente, con il tempo, all'interno di ogni ecosistema, si raggiunge un **equilibrio ecologico**, per cui tutti gli oggetti fisici e gli organismi viventi in esso compresi convivono tranquillamente: si tratta chiaramente di un equilibrio non statico ma «dinamico», caratterizzato cioè da fenomeni continui che, tuttavia, mediamente non mutano in maniera significativa la situazione. Tale equilibrio, che costituisce la base per la vita di tutti gli organismi, può essere facilmente rotto dall'intervento di fattori esterni, o anche interni, all'ecosistema stesso, di origine sia naturale sia antropica. Tali **fattori**, definiti **limitanti**, rappresentano spesso, per gli organismi, il raggiungimento di un punto critico che può essere fatale per la sopravvivenza. Possono manifestarsi eventi normali o eccezionali: nel primo caso l'ecosistema evolve in modo graduale, così da raggiungere in tempi relativamente rapidi una nuova forma di equilibrio (è il caso di una variazione progressiva ed omogenea di temperatura, che porta lentamente al passaggio da una formazione vegetale ad un'altra simile, di modo che sia le piante sia gli animali abbiano il tempo di adattarsi al cambiamento); nel secondo caso l'ecosistema reagisce all'evento improvviso e traumatico in maniera violenta, trasformandosi completamente e in maniera irreversibile (è il caso di una foresta che subisce un incendio: anche in tempi lunghi, risulta praticamente impossibile il ritorno alla situazione originaria). Esempi di *fattori limitanti* sono le variazioni climatiche, la scomparsa di specie viventi, l'aumento eccessivo di individui, l'introduzione di una specie nuova, le calamità naturali, l'intervento dell'uomo.

Gli *ecosistemi* si distinguono in **naturali**, se non hanno subito alcun intervento antropico; **artificiali**, se sono nati per azione dell'uomo (come, ad esempio, le colture); e **artificiali non creati dall'uomo**, se risultano fortemente condizionati dalle opere antropiche (è il caso degli ambienti costieri, influenzati nella loro dinamica dalle costruzioni umane). Ormai gli ambienti naturali sono relegati quasi esclusivamente alle oasi e ai parchi protetti, mentre la biosfera si

[1] Dal greco: *bios* = vita e *topos* = luogo.

[2] Dal greco: *bios* = vita e *kóinosis* = comunanza.

sta trasformando sempre più in un insieme di ecosistemi artificiali. Se nelle civiltà primitive l'uomo si limitava a partecipare agli ecosistemi, con lo sviluppo della cultura agricola egli ha cominciato a trasformare il paesaggio e a incidere sull'evoluzione degli altri organismi viventi. Con il passaggio alla civiltà industriale la situazione è notevolmente peggiorata, tanto che oggi si teme fortemente per la scomparsa di molti ecosistemi [3]. Il pericolo risiede nel fatto che i mutamenti introdotti dall'uomo finiscono per essere irreversibili, causando delle involuzioni disastrose per la vita vegetale, animale e umana stessa.

2 Uomo e ambiente

Uno dei primi, significativi documenti internazionali sul rapporto uomo-ambiente, la *Conferenza sull'ambiente umano*, organizzata a Stoccolma nel 1972 dalle Nazioni Unite, produsse un'importante dichiarazione, basata su centonove raccomandazioni, incentrate su **cinque temi principali**:

- la pianificazione e la gestione degli insediamenti umani in relazione alla qualità dell'ambiente;
- l'effetto sull'ambiente della gestione delle risorse naturali;
- l'identificazione e l'eliminazione degli agenti inquinanti;
- gli aspetti educativi, informativi, sociali e culturali dei problemi dell'ambiente;
- l'ambiente e lo sviluppo.

Malgrado, nel frattempo, molti passi avanti siano stati fatti, soprattutto in materia di legislazione ambientale, le raccomandazioni delle Nazioni Unite restano drammaticamente attuali. La difesa dell'ambiente necessita ancora di innumerevoli sforzi e di notevole impegno. Purtroppo non basta più limitarsi ad una **difesa passiva**, che significa essenzialmente evitare d'intervenire negativamente sull'ambiente (cercando, ad esempio, di non sporcare e di non inquinare), ma occorre anche organizzare una **difesa attiva**, ovvero sensibilizzare le popolazioni sulle tematiche ambientali, sviluppare una coscienza ecologica, difendere le specie vegetali e animali a rischio di estinzione, ripristinare le condizioni naturali degli ecosistemi laddove l'uomo è intervenuto in maniera troppo violenta, utilizzare più razionalmente le risorse naturali, diminuire la quantità di rifiuti e il tasso di inquinamento, conciliare le esigenze umane con quelle dell'ambiente ispirandosi ai principi dello sviluppo sostenibile e della qualità della vita, provare a rimediare ai danni fatti in passato. I campi d'intervento ricadono, non a caso, nell'ambito dei cinque temi proposti già nel 1972 dalle Nazioni Unite. Di seguito saranno brevemente discussi i problemi principali relativi all'impatto che l'uomo, con le sue attività, ha sull'ambiente. La loro risoluzione costituirebbe un notevole passo avanti nella difesa della natura dal suo più pericoloso nemico: l'uomo stesso.

Senza dubbio l'**inquinamento** rappresenta il caso più emblematico di deterioramento dell'ambiente da parte dell'essere umano. Esso coinvolge praticamente l'intero pianeta, in tutte le sue componenti: **idrosfera, atmosfera e terra solida**. Nasce come risposta allo sviluppo tecnologico, che ha portato alla scoperta di sostanze, macchinari e strumenti i quali hanno apparentemente migliorato la nostra vita, diventando ormai indispensabili per tutti (si pensi all'automobile, ai pesticidi o alle centrali termoelettriche), ma che, nel contempo, si sono rivelati estremamente dannosi per l'habitat naturale. Gli insediamenti urbani e industriali costituiscono le principali fonti d'inquinamento.

[3] Risale ormai a circa trent'anni fa il celebre rapporto del MIT (Massachusetts Institute of Technology) sul futuro dell'uomo sulla Terra. Lo studio, basato sull'analisi di cinque fattori ritenuti fondamentali (popolazione, capitale, alimenti, risorse non rinnovabili e inquinamento), prevedeva tale peggioramento, imputandolo ad un'eccessiva crescita della popolazione, ad una progressiva diminuzione delle materie prime e delle risorse naturali in genere, oltre che all'incondizionato aumento dell'inquinamento.

Gli scarichi di liquami arricchiscono fiumi e mari di sostanze tossiche, detergenti sintetici, residui di cibo, rifiuti di vario genere — nella maggior parte dei casi non biodegradabili — che rappresentano elementi estremamente pericolosi per ogni forma di vita. Tutti gli organismi, vegetali ed animali, che abitano le **acque**, risultano contaminati e, indirettamente, lo è anche l'uomo: sia perché si nutre di tali organismi, sia perché beve l'acqua attinta dalle falde acquifere compromesse dalle infiltrazioni di acqua fluviale e marina. All'inquinamento delle falde e delle acque in genere contribuiscono anche i fertilizzanti chimici e i fitofarmaci (insetticidi, pesticidi, erbicidi etc.), sempre più utilizzati in agricoltura.

Molto frequenti sono gli episodi d'inquinamento del mare a causa del petrolio (fenomeni di perdite dai pozzi petroliferi, avaria delle petroliere, lavaggio dei serbatoi etc.): una tonnellata di greggio ha la capacità di contaminare un milione di tonnellate d'acqua, ricoprendole con una patina oleosa per un'estensione di diversi km². Ulteriori conseguenze sono la morte di numerosi organismi marini e un generale rallentamento dello sviluppo della fauna ittica, che risente del più lento scambio di ossigeno tra acqua e aria.

Un fenomeno assai particolare, legato all'inquinamento di mare e fiumi, è l'*eutrofizzazione* della vegetazione. Gli scarichi innalzano le temperature delle acque e, di conseguenza, i processi di fotosintesi degli *organismi termofili* (amanti del caldo) si fanno più intensi: queste specie crescono in modo innaturale, finendo per farne scomparire molte altre. Gli scarichi favoriscono, inoltre, lo sviluppo di alcune alghe maleodoranti che popolano in maniera sempre più massiccia i litorali.

Innumerevoli sono le regioni del mondo interessate all'inquinamento delle acque: dai Grandi Laghi americani al mar Mediterraneo, dal fiume Gange in India al Reno in Europa. In Italia, poi, la situazione si presenta davvero critica: quasi tutti i corsi d'acqua che attraversano zone industrializzate si rivelano inquinati, dall'Arno in Toscana al Tevere nel Lazio, al Sarno in Campania. Soltanto una politica basata sul controllo delle strutture inquinanti e sulle sanzioni da comminare in caso di mancato rispetto delle leggi potrebbe, almeno in parte, risolvere la questione.

Non meno grave è il problema dell'**inquinamento atmosferico**: le fonti dei fumi nocivi sono innumerevoli, dai veicoli di trasporto alle industrie, dalle centrali termoelettriche e nucleari agli inceneritori. Gli inquinanti sono polveri e particelle solide con ossido di carbonio, anidride solforosa, ammoniaca, idrocarburi e altri elementi tossici. Con il trasporto aereo, inoltre, non risultano più coinvolti soltanto gli strati più bassi dell'atmosfera ma anche quelli in quota. Le conseguenze per l'uomo sono micidiali: l'aumento di tutte le malattie di tipo respiratorio o canceroso è, in massima parte, imputabile proprio all'inquinamento dell'aria.

Ad esso è, inoltre, connesso il fenomeno noto come **effetto-serra**: lo strato di gas che permea l'atmosfera impedisce l'emissione dei raggi infrarossi dalla Terra, che, di conseguenza, si surriscalda. Secondo molti scienziati tale effetto risulta particolarmente rilevante e, col tempo, potrà causare danni enormi: il progressivo riscaldamento del pianeta provocherà lo scioglimento delle calotte glaciali, gli oceani si innalzeranno di alcune decine di metri e moltissime coste risulteranno quindi sommerse, con danni facilmente immaginabili.

Una conseguenza indiretta dell'inquinamento atmosferico è rappresentata dalle **piogge acide**: gli ossidi di azoto, l'anidride solforosa e l'anidride carbonica, a contatto con l'umidità e l'ossigeno dell'atmosfera, si trasformano in acido nitrico e solforoso, cadendo sulla superficie terrestre come piogge ad alto contenuto di acidi, che bruciano la vegetazione e danneggiano i monumenti. Le misure per salvaguardare l'aria dall'inquinamento compaiono nelle leggi più recenti in materia ambientale ma, purtroppo, si fa ancora troppo poco per controllare che vengano rispettate.

Una forma di inquinamento estremamente insidiosa è legata alla **radioattività** presente nell'aria. Gli esperimenti nucleari, consentiti fino ad alcuni decenni fa, hanno sprigionato nell'atmosfera particelle radioattive, che poi sono cadute al suolo attraverso le piogge. Gli effetti si rivelano spesso a distanza di anni, ma sono devastanti e portano direttamente alla morte.

L'incendio della *centrale nucleare di Chernobyl*, in Ucraina, nel 1986, ha causato infezioni terribili in gran parte degli abitanti della regione e malformazioni genetiche che si stanno manifestando nei figli di quanti furono esposti alle radiazioni. La nube si diffuse quasi in tutta Europa e, probabilmente, molte malattie cancerose degli europei sono il risultato di quella contaminazione. Altrettanto spaventoso è stato, più recentemente, l'incidente occorso — in seguito al terremoto e relativo *tsunami* del Tōhoku dell'11 marzo 2011 — alla *centrale nucleare giapponese di Fukushima*, una delle molte (troppe) attive in terra nipponica, che ha diffuso il panico nel Paese del Sol Levante (e non solo), i cui effetti nocivi per l'ambiente circostante e per le popolazioni non sono ancora chiaramente valutabili, ma che, purtroppo, non tarderanno a farsi sentire. L'unica soluzione per proteggersi da eventi di questo tipo, eliminando ogni possibilità di rischio, sarebbe la chiusura e lo smantellamento di tutte le centrali atomiche, come da più parti del mondo chiedono gli ecologisti.

Una forma d'inquinamento da cui l'uomo difficilmente riuscirà a liberarsi proviene dai **rifiuti**. Si tratta di un problema straordinariamente complesso e sempre attuale, che coinvolge inevitabilmente tutto il pianeta. Si è calcolato che una persona, mediamente, produce 1 kg di rifiuti al giorno, gran parte dei quali non biodegradabili. Le città si trovano a dover raccogliere e smaltire, attraverso discariche e inceneritori, quantità enormi di spazzatura. Il sistema più adottato, sebbene illegale, consiste purtroppo nel semplice accumulo in aree abbandonate di tali residui, senza alcun trattamento preventivo del materiale (**discarica selvaggia**). I rifiuti si decompongono, producendo liquami e gas altamente tossici che inquinano acque, aria e suolo. Le alternative approvate dalla legge sono le **discariche controllate**, in cui i liquidi e i gas vengono raccolti per essere smaltiti o addirittura utilizzati come fonte di energia; oppure gli **inceneritori**, nei quali i rifiuti vengono bruciati in maniera collaudata e i gas prodotti usati anch'essi per produrre energia.

Questi sistemi, spesso difficili da realizzare, comportano comunque dei rischi d'inquinamento, tuttavia comunemente considerati tollerabili. Un discreto aiuto per la soluzione del problema è fornito dalla **raccolta differenziata dei rifiuti e dal loro riciclaggio**: è un'operazione affidata, in gran parte, alla sensibilità di amministratori e cittadini, che può dare sempre ottimi risultati.

Lo **sfruttamento delle risorse naturali** da parte dell'uomo è un'attività antichissima, che, tuttavia, negli ultimi decenni sta producendo effetti devastanti sull'ambiente. In molti casi si ottiene un **deturpamento**, spesso irrecuperabile, **del paesaggio**: miniere a cielo aperto e cave si presentano come grosse ferite che solcano brutalmente colline e montagne, indebolendo spesso i versanti ed esponendoli maggiormente al rischio di frane. All'attività estrattiva, secondo quanto sancito dalle leggi, dovrebbe seguire la sistemazione dei pendii, con il riempimento e il rimboschimento delle cavità; ma si tratta di un evento raro: di solito cave e miniere vengono trasformate in discariche a cielo aperto, che non soltanto rovinano l'amenità dei luoghi, ma diventano anche fonte d'inquinamento. Allo stesso modo, gli impianti e le strutture per l'estrazione di petrolio e gas naturale alterano completamente l'ambiente e si rivelano estremamente pericolosi in caso di incendi o, in generale, di incidenti. Per controllare e mitigare le conseguenze ecologiche che la costruzione di una grande opera, come un gasdotto, una diga o una centrale, può avere sul paesaggio in termini di deturpamento e di pericolosità, recenti norme hanno introdotto l'obbligo di procedere ad uno studio sulla V.I.A. (*Valutazione di Impatto Ambientale*).

In altri casi si deve affrontare il problema dell'**esaurimento delle risorse**. L'esigenza continua e pressante di materiali da costruzione spinge sempre più frequentemente l'uomo a prelevare sabbie e ghiaie dai letti dei fiumi o dalle spiagge, con conseguenze disastrose sulle dinamiche costiere, ma anche la richiesta sempre maggiore di acque e la gestione spesso dissennata di quelle *potabili* potranno produrre in futuro l'esaurimento di tale risorsa o, quantomeno, aspri conflitti internazionali per il suo controllo. Anche da questo rischio è possibile difendersi: il riciclaggio delle acque per scopi agricoli e l'uso di impianti di riduzione dei consumi, soprattutto nelle industrie, rappresentano provvedimenti utili in tal senso.

Per sopperire alla **carenza d'acqua**, generalmente si ricorre a grandi opere, come le *dighe*, utilizzate per sbarrare i fiumi e creare invasi artificiali. Al di là degli effetti innaturali che esse provocano sul regime dei corsi d'acqua, senza voler considerare le variazioni indotte sul clima locale e l'impatto ambientale, questi invasi possono risultare anche estremamente pericolosi per le condizioni di equilibrio dei versanti. Le frane dei pendii di spalla di una diga o i fenomeni di tracimazione delle acque rappresentano eventi, purtroppo, abbastanza frequenti [4].

Lo sfruttamento degli alberi per produrre legname e il taglio dei boschi per la costruzione di strade o per l'apertura di miniere sono la causa del processo di **deforestazione**, che coinvolge, al ritmo di alcuni milioni di ettaro all'anno, tutti i polmoni verdi del pianeta, dall'Africa all'America. Un tentativo di difesa è stato sperimentato introducendo la *silvicoltura*, consistente nella gestione razionale dei boschi su base scientifica ed economica. Si procede ad una rotazione pianificata del taglio degli alberi, mentre le specie arboree a «crescita lenta» vengono sostituite da quelle a «crescita rapida».

Il taglio dei boschi sembra sia responsabile, in parte, anche del processo di **desertificazione** che sta interessando diverse regioni della Terra. La distruzione della vegetazione favorisce l'erosione del suolo da parte di vento e acque, contribuendo quindi allo sviluppo di questo fenomeno. Concausa altrettanto determinante è l'uso di moderne tecnologie per l'agricoltura, sperimentate su alcuni suoli e utilizzate su altri: il risultato consiste spesso, dopo un periodo di forte produttività, in un impoverimento del terreno fino alla sua completa sterilità.

3 Il dissesto idrogeologico: un problema anche, e soprattutto, italiano

Le bellezze paesaggistiche che hanno contribuito alla fama del nostro Paese sono allo stesso tempo frutto di un'**eterogeneità topografica tanto accentuata quanto potenzialmente pericolosa**: infatti, appena il 21% del territorio italiano è costituito da *pianure*, la restante parte è rappresentata da *colline* (il 40%) e *montagne* (il 39%). L'Italia ha il **primato europeo della franosità**, soprattutto a causa della sua *configurazione geomorfologica*, frutto delle *spinte geodinamiche* della crosta terrestre. Tali *forze endogene* sono ancora attive e i terremoti, nonché i fenomeni vulcanici che interessano ripetutamente la nostra Nazione ne sono la dimostrazione più concreta: in altri termini, lo Stivale è, nel suo complesso, uno dei territori geologicamente più giovani e dinamici del Vecchio Continente.

Secondo una recente ricerca condotta dal Servizio Geologico Nazionale (SGN), in un arco di tempo che va dal secondo dopoguerra agli anni Novanta del secolo scorso, il costo in termini di perdita di vite umane connesso ai **disastri idrogeologici** (*frane*) ha raggiunto la drammatica cifra di circa 3.500 unità. I Comuni italiani colpiti da questo dissesto, sempre nel medesimo periodo, sono non meno di 4.500 (più della metà del totale nazionale), cui corrispondono approssimativamente 194.500 km² (il 65% dell'intero territorio italiano).

Le conseguenze dei **fenomeni franosi** assumono una rilevanza tale da costituire un gravissimo **problema socio-economico**; tanto è vero che:

- con una media di quasi 60 vittime all'anno, nell'ultimo secolo, l'Italia è al quarto posto nel mondo tra i Paesi più colpiti, dopo quelli andini, la Cina e il Giappone;
- relativamente ai danni subiti, con un ammontare per frana stimato fra 1 e 2 miliardi di euro all'anno, il Belpaese è al secondo posto assoluto, a pari merito con Stati Uniti e India, e dopo il solo Giappone;
- in termini di rapporto danni/PIL, l'Italia (con l'1,5%) si colloca, fra le Nazioni tecnologicamente avanzate, al secondo posto, subito dopo l'arcipelago del Sol Levante.

[4] La frana staccatasi nel 1963 dai versanti sovrastanti la diga del Vajont, in Veneto, produsse una tracimazione dell'invaso: le acque precipitarono come un'enorme onda sui centri abitati a valle, causando la morte di circa 2.000 persone.

Altra temibile conseguenza dell'ormai tristemente noto *dissesto idrogeologico* è rappresentata dalle **alluvioni**, che spesso si manifestano *in concomitanza con gli eventi franosi*, riproducendo una sorta di binomio letale e inscindibile.

I **principali fattori** che provocano le alluvioni sono il **cambiamento del regime delle precipitazioni e dei parametri idrologici**, tra cui:

- la *diminuzione dei tempi di corrivazione* (vale a dire l'intervallo temporale che le acque di afflusso meteorico impiegano per raggiungere una data sezione fluviale che sottende un bacino, partendo dai punti più lontani dello stesso);
- l'*impermeabilizzazione del suolo*;
- l'*aumento della velocità delle acque*;
- la *minore capacità di ricezione e di contenimento di queste ultime*, dovuta alla riduzione degli spazi d'alveo;
- la *forte pendenza dei bacini dei fiumi e l'assenza di vegetazione capace di trattenere l'acqua che scorre*;
- la *diminuzione delle aree golenali*, naturale sede di espansione delle piene;
- l'*abbandono*, con conseguenze degrado, di *prati, pascoli e boschi*;
- la *realizzazione di opere senza rispetto dell'equilibrio preesistente* nei bacini idrologici (come la costruzione di strade a mezza costa sulla montagna con notevoli sbancamenti, oppure l'esecuzione di lavori di sistemazione dei torrenti mal progettati);
- la *realizzazione di sbarramenti e bacini* senza tutte le fondamentali cautele che simili azioni richiederebbero.

Il *mutamento del regime delle precipitazioni* è da collegarsi, a sua volta, ai **cambiamenti climatici** in atto, i quali determinano sempre più spesso un aumento della ricorrenza delle piogge «a carattere alluvionale», ossia particolarmente intense e concentrate in periodi di tempo sempre più brevi. La situazione peggiora nel caso in cui le piogge siano accompagnate da fenomeni di disgelo e/o da forti venti. A ciò va aggiunto l'**enorme sviluppo dell'urbanizzazione, delle infrastrutture e delle attività economiche**, che ha provocato un'estesa *impermeabilizzazione del territorio*, da cui ha avuto origine un *ruscellamento* superficiale incontrollato e di notevoli proporzioni. Infine, la drastica *riduzione delle aree golenali* (porzioni di territorio disposte nelle immediate circostanze degli alvei fluviali, in grado di accogliere le masse d'acqua in eccesso in seguito a una piena) è stata causata dalla creazione di vasti insediamenti produttivi e abitativi a ridosso dei medesimi letti dei fiumi.

Una serie di comportamenti umani così insensati, unitamente a una «natura sempre più ribelle», dà vita a una sorta di «retroazione» positiva: in sostanza, il verificarsi di un fenomeno produce un risultato che di per sé rinforza ulteriormente l'effetto del fenomeno stesso. Il problema della fragilità del territorio (nella fattispecie del nostro) e dell'esposizione al rischio di frane e alluvioni può essere arginato solamente con maggiori investimenti in termini di prevenzione, attraverso cui affermare una **nuova cultura dell'utilizzo del suolo**, che metta al primo posto la sicurezza della collettività, ponendo fine a usi speculativi e abusivi del territorio, al suo completo abbandono e alla sua gestione irrazionale.

V

Educazione alla salute

Sommario Sezione V

Introduzione	
1 Sessualità e riproduzione umana	4 La malattia come rottura dell'equilibrio biologico e psicologico
2 La gestione delle funzioni biologiche nell'uomo	5 La difesa della salute
3 Consumi voluttuari e prevenzione delle dipendenze	6 Conoscenza e prevenzione della patologia incidentale
	7 La salute mentale

Introduzione

La definizione formulata nel 1948 dall'Organizzazione Mondiale della Sanità (OMS) — sia pure un po' datata e, negli anni, criticata poiché tendenzialmente irrealista e utopistica — è a tutt'oggi alla base della concezione comune del termine *salute*: «uno stato di completo **benessere fisico, mentale e sociale**» che «**non consiste soltanto in un'assenza di malattia o di infermità**». Fino alla fine degli anni Quaranta del Novecento, infatti, il concetto di salute era legato a quello di malattia, definendo, appunto, la salute come semplice «assenza di malattia»; una condizione, cioè, in cui non esistono segni di processi patologici in atto. Il passaggio dalla centralità della malattia alla **centralità della salute** fu epocale, anche perché l'OMS riconosceva un ruolo centrale al benessere mentale e sociale.

Il concetto di salute prescinde, quindi, dalle semplici condizioni sanitarie, che considerano l'obiettività fisica, per collocarsi in un contesto più ampio, comprendente le **problematiche psicologiche e sociali** che scaturiscono dal **rapporto tra individuo e ambiente**. La difesa e la tutela della salute costituiscono una delle principali sfide della società attuale per il carattere fondante che questa assume rispetto a tutte le altre dimensioni dell'esistenza. Essere in salute presuppone, dunque, l'**essere in equilibrio** con sé stessi e col proprio ambiente di riferimento: si è sani se in armonia con la propria interiorità ed esteriorità; malati se, in tale rapporto, prevale la disarmonia.

La scuola ha tra i suoi compiti istituzionali l'Educazione alla salute, che assume particolare rilievo nell'ambito della formazione integrale della personalità degli alunni. Educare alla salute significa, infatti, **incoraggiare lo sviluppo delle singole personalità promuovendone le potenzialità** dell'intelligenza, dell'affettività, dell'integrazione sociale, in maniera tale da favorire la piena realizzazione della vita personale degli scolari e la massima espressione di tutta la loro umanità. Educare alla salute è, insomma, **educare alla vita**. Non stupisce, allora, che le problematiche inerenti l'Educazione alla salute siano assai complesse.

Come presidio di **prevenzione del disagio** e di **promozione del benessere psicofisico**, l'Educazione alla salute dovrebbe costituire una competenza centrale rispetto alla dimensione emotiva e affettiva delle giovani generazioni. Dal disagio, durante l'infanzia e l'adolescenza, possono scaturire vere e proprie emergenze sociali e sanitarie. Sono patologie del comportamento la dipendenza da sostanze stupefacenti, l'alcolismo, i disturbi alimentari e il bullismo. Di fatto, però, l'attenzione degli insegnanti finisce col focalizzarsi sulla prevenzione delle tossicodipendenze. Per poter educare alla salute la scuola dovrà, invece, farsi promotrice di una condizione di **benessere diffuso e condiviso** che riguardi gli alunni, i docenti, il dirigente scolastico, il personale tecnico-amministrativo, le strutture, le attrezzature, i laboratori e finanche la **didattica**, in un'ottica aperta e multidisciplinare.

Sessualità e riproduzione umana

1 La riproduzione negli animali superiori

Negli animali superiori, di solito, la riproduzione avviene per **fecondazione**. Tale fenomeno presuppone la presenza di sessi diversi che, con meccanismi biochimici e comportamentali si attraggono per assolvere alla funzione della riproduzione. I sessi diversi sono anche caratterizzati dalla presenza nell'organismo di ghiandole particolari, dette **gonadi**, che producono i *gameti*, nonché da una serie di altri organi che nel loro insieme vengono identificati come *caratteri sessuali primari e secondari*.

Nel periodo precedente la fecondazione si verificano, in questi organi e nel comportamento dell'animale, una serie di modifiche che facilitano l'approccio con un altro organismo di sesso diverso: quella successione di fenomeni che iniziano col corteggiamento e terminano con la fecondazione. Tale processo, in effetti, consta di due momenti: la **copulazione**, che porta all'incontro del gamete maschile con quello femminile, e la **cariogamia**, che comporta, una volta avvenuto l'incontro tra spermatozoo e ovulo, la fusione dei *nuclei aploidi* di queste due cellule e la formazione della prima cellula del nuovo organismo e cioè dello *zigote diploide*.

La fecondazione può avvenire all'esterno del corpo dell'individuo di sesso femminile, soprattutto per quegli animali che vivono in ambiente acquatico, oppure al suo interno. In entrambi i casi, essa è preceduta dal corteggiamento. I gameti maschili, gli *spermatozoi*, sono prodotti in gran numero a differenza dell'ovulo, perché deve essere assicurata la possibilità di incontro tra i due gameti: perché ciò avvenga, è necessario che si presentino condizioni bio-ambientali positive come la temperatura, la salinità, il grado di maturazione dei gameti stessi. Dopo che lo zigote si è formato, ha inizio il suo sviluppo che, secondo diverse modalità, porterà l'embrione a raggiungere la forma adulta.

2 Gonadi e sessualità

La formazione dei gameti (*gametogenesi*) avviene in speciali organi, le *gonadi*, che vengono chiamati *testicoli* quando producono *spermatozoi*, e *ovari* quando producono *uova*. In molti animali e nella maggioranza dei vegetali coesistono nello stesso individuo sia i testicoli (nei vegetali paragonabili agli *stami*) sia gli ovarii come gonadi singole, o talora anche un'unica gonade che ha la capacità di produrre spermatozoi e uova; a tale fenomeno si dà il nome di **ermafroditismo**.

Nella maggioranza degli animali, però, i testicoli sono presenti in un individuo, il **maschio**, o individuo di sesso maschile (indicato in biologia con il segno σ), e gli ovarii in un altro individuo, la **femmina** o individuo femminile (indicato in biologia con il segno δ). Si parla allora di **gonocorismo**. Tra ermafroditismo e gonocorismo esiste tutta una serie di condizioni di sessualità intermedie definite *intersessualità*.

Le gonadi, poiché identificano le caratteristiche fondamentali dei due sessi, sono considerate espressione dei *caratteri sessuali primari*, mentre tutte le altre caratteristiche, in rapporto o meno con la funzione riproduttiva, ma che contraddistinguono un sesso dall'altro, sono dette *caratteri sessuali secondari*. Tra questi ultimi bisogna poi distinguere quelli legati direttamente alla funzione riproduttiva, come il trasporto all'esterno degli spermatozoi o delle uova (vie genitali,

organi copulativi), che vengono chiamati *caratteri sessuali secondari*, da quelli presenti nelle forme gonocoriche, che non fanno parte del sistema riproduttore e si rinvencono in altre parti del corpo, detti *caratteri sessuali secondari extragenitali*; sono questi ultimi che permettono di distinguere il maschio dalla femmina e di affermare che la specie presenta un **dimorfismo sessuale**, cioè esiste nella specie una forma maschile ed una femminile ben distinte somaticamente.

Questo dimorfismo sessuale può essere accentuato fino a giungere a differenze tali tra maschio e femmina da rendere difficile la loro attribuzione ad una stessa specie. Casi di dimorfismo sessuale sono la forma e i colori degli uccelli (cresta e bargigli evidenti nel gallo, ridotti nella gallina), le corna nei maschi dei cervi etc. Anche l'uomo ha un vistoso dimorfismo sessuale: oltre ai genitali esterni, sono noti i baffi e la barba nei maschi, lo sviluppo del seno e la maggiore larghezza del bacino nella donna, la distribuzione dell'adipe in diversi distretti (distretto superiore nell'uomo, distretto inferiore nella donna).

3 Il processo di fecondazione dell'ovulo

La nascita dell'uomo è condizionata dalla realizzazione di una serie di eventi che ha origine nel processo di fecondazione dell'ovulo femminile da parte dello spermatozoo maschile ed evolve successivamente attraverso diversi stadi di accrescimento che si realizzano, nella fase iniziale, nell'ambiente uterino materno. L'inizio di una gravidanza è possibile per il verificarsi di alcune fasi biologiche che possono essere schematizzate come segue:

- la **migrazione** degli spermatozoi dal fondo vaginale, dove sono stati depositati durante il coito, al lume delle tube, dove si trova l'ovocita;
- la **capacitazione** degli spermatozoi a fecondare l'ovocita. Tale fenomeno si verifica grazie alla destabilizzazione della membrana plasmatica dello spermatozoo e alla reazione acrosomiale;
- la **captazione dell'ovocita** da parte del padiglione tubarico in seguito all'ovulazione;
- la **trasformazione** dell'ambiente biochimico tubarico al fine di renderlo adatto alla fecondazione;
- l'interazione e la **fusione dei due gameti**;
- il **trasporto dell'embrione** verso la cavità uterina mediante movimenti peristaltici della tuba e di flusso (3-5 giorni);
- l'**arrivo dell'embrione** in un opportuno stadio di differenziazione nella cavità uterina e il suo *annidamento nella mucosa uterina*.

Gli enzimi che si liberano al momento della *reazione acrosomiale* permettono allo spermatozoo di lisare i tessuti che circondano l'ovocita (*corona radiata*) e di aprirsi una strada che lo porterà verso la *zona pellucida*. Contemporaneamente, l'ovocita subisce una serie di trasformazioni che lo rendono idoneo ad essere fecondato. Lo spermatozoo, penetrato nella zona pellucida grazie all'*acrosina*, si ritrova nello *spazio perivitellino*, dove le membrane citoplasmatiche dei due gameti si fondono. Il processo meiotico riprende con l'emissione del secondo globulo polare (il primo era stato emesso prima dell'ovulazione) e, per coalescenza dei cromosomi, si formano il **pronucleo femminile** e il **pronucleo maschile**. L'avvicinamento dei due pronuclei è reso possibile da una serie di microtubuli convergenti a fuso. A questo punto ha luogo la replicazione del DNA; dopo circa ventiquattr'ore avviene la scomparsa della membrana dei due pronuclei e si forma lo *zigote*, prima cellula dell'embrione. Lo zigote per due o tre giorni resta nell'ampolla tubarica, andando incontro alle prime divisioni mitotiche (*morula*); successivamente, allo stadio di *blastocisti*, viene trasportato nell'utero, dove si impianterà.

Nel corso dell'ulteriore sviluppo si formerà il *disco embrionario*, dal quale si origineranno tre foglietti primitivi; da questi ultimi deriveranno poi i diversi organi e tessuti, più precisamente:

- dall'**ectoderma** deriveranno la cute e le sue appendici, il sistema nervoso, la midollare del surrene, l'ipofisi e le ghiandole salivari;

- dall'**endoderma** deriveranno il tratto gastrointestinale, il fegato, il pancreas, l'apparato respiratorio, la tiroide, le cellule germinali primitive;
- dal **mesoderma** deriveranno le ossa, i muscoli, le cartilagini, il tessuto connettivo, il peritoneo, la pleura, il sistema cardiovascolare, il sistema urogenitale, la corticale del surrene.

4 Lo sviluppo embrionale e fetale

Se il prodotto del concepimento presenta caratteristiche genetiche e biologiche compatibili con la vita e se l'impianto si è verificato in un ambiente che ne consente l'ulteriore maturazione, la crescita all'interno della *camera gestazionale* proseguirà fino al completamento dello sviluppo in un primo momento embrionale e poi fetale.

In base all'*età concezionale*, il prodotto del concepimento viene definito:

- **preembrione**, dallo stadio di zigote al 14° giorno dopo la fecondazione;
- **embrione**, dal 14° giorno sino alla fine dell'8ª settimana di sviluppo;
- **feto**, dall'8ª settimana di sviluppo sino al termine della gravidanza.

Nella specie umana la durata normale della gravidanza, calcolata a partire dal giorno di inizio dell'ultimo ciclo mestruale, è di 40 settimane, pari a 280 giorni.

Fino al 20° giorno di età concezionale l'embrione si trova nello *stadio presomitico*, durante il quale si sviluppano il disco embrionale, la cavità amniotica e la placca neurale.

Il periodo che va dal 20° al 30° giorno si definisce *stadio somitico*, in quanto avviene la formazione dei **somiti** dai quali deriveranno lo scheletro e i muscoli striati.

Durante la 4ª settimana di sviluppo compaiono gli abbozzi del cuore, degli occhi, del proencefalo, del fegato e del rene e alla fine della 5ª settimana sono evidenti gli abbozzi degli arti superiori e inferiori. Tale periodo corrisponde alla 6ª-7ª settimana di gestazione ed è la fase in cui è possibile rilevare ecograficamente il *battito cardiaco fetale*, perché il cuore presenta già pulsazioni ritmiche.

L'8ª settimana di sviluppo è caratterizzata dalla comparsa degli abbozzi delle ossa, dei muscoli, dei nervi e dei grossi vasi; l'estremità cefalica comincia a separarsi dal torace e vi si iniziano a distinguere il naso, le orecchie e la mandibola.

Alla 12ª settimana gli abbozzi oculari vengono ricoperti dalle palpebre, si verificano i primi movimenti attivi e si ha la completa differenziazione dei genitali interni.

Durante la 16ª settimana nel canale intestinale compare il *meconio*.

Alla fine della 20ª settimana i movimenti fetali attivi sono molto vivaci.

Alla 24ª settimana si formano i primi depositi adiposi.

Dalla 25ª alla 28ª settimana la maturità è tale da poter garantire l'eventuale sopravvivenza *pre-termine*. La cute è grinzosa, i capelli non sono ancora ben differenziabili dalla lanugine che ricopre il corpo, le palpebre sono separate, le orecchie addossate al cranio.

Alla 32ª settimana la cute è meno rugosa ma ancora ricoperta di lanugine.

Durante la 36ª settimana il corpo assume contorni tondeggianti per il deposito di tessuto adiposo sottocutaneo, mentre la cute perde l'aspetto rugoso.

Alla 40ª settimana *gestazionale* il feto è maturo: a questo punto, si verifica il parto e la gravidanza viene definita *a termine*.

Ma non tutte le gravidanze hanno la stessa durata e i feti, in base all'epoca gestazionale in cui si verifica la nascita, sono definiti:

- pre-termine (nati prima della 37ª settimana);
- a termine (nati tra la 38ª e la 42ª settimana);
- post-termine (nati oltre la 43ª settimana).

Il parametro più importante da valutare è quello del peso del feto alla nascita in rapporto all'età gestazionale. L'accrescimento ponderale può essere espresso in curve di *percentili*. In base a questo concetto, i feti vengono suddivisi in:

- proporzionati (fra il 10° e il 90° percentile);
- sproporzionati in eccesso (>90° percentile);
- sproporzionati in difetto (<10° percentile).

Caratteristiche del neonato proporzionato e a termine sono:

- lunghezza del corpo oscillante tra i 48 e i 52 cm;
- peso variabile tra i 2.800 e i 4.000 g;
- discesa dei testicoli nel maschio completata;
- grandi labbra in apposizione che ricoprono completamente le piccole labbra nella femmina;
- presenza del nucleo di ossificazione dell'epifisi distale del femore;
- presenza di alcune importanti attività funzionali, quali la possibilità di respirare e di servirsi del tubo gastroenterico.

La valutazione della *vitalità* del neonato al momento della nascita viene effettuata mediante la **scala di APGAR**.

5 Nascita e sviluppo extrauterino

L'**accrescimento** si può definire come l'insieme dei fenomeni che caratterizzano la trasformazione progressiva dell'organismo dal concepimento all'età adulta. Lo **sviluppo post-natale** si verifica nell'ambiente esterno extrauterino e può essere distinto in diverse fasi.

La *prima infanzia* è caratterizzata da un incremento staturale molto rapido, anche se progressivamente in riduzione. Alcuni organi importanti come il *cervello* completano durante tale periodo l'accrescimento quantitativo iniziato nello stadio precedente, mentre altri organi modificano di poco le proprie caratteristiche.

La *seconda e la terza infanzia* sono caratterizzate da un accrescimento lento e regolare (5-6 cm all'anno circa).

La terza fase o *adolescenza* è caratterizzata dalla maturazione degli organi dell'apparato genitale, che aumentano di volume; parallelamente, compaiono i caratteri sessuali secondari. Durante tutto il periodo è presente un incremento della velocità di crescita sia in altezza sia in peso (*picco adolescenziale*). Anche la struttura corporea si modifica in relazione al sesso e si verificano grossi cambiamenti a livello psichico. L'acquisizione della capacità procreativa e il raggiungimento della statura definitiva (scomparsa della cartilagine di coniugazione) rappresentano il punto di passaggio dall'adolescenza all'*età adulta*.

Nella prima infanzia le modificazioni morfologiche interessano in particolare il rapporto tra il cranio (caratterizzato da accrescimento notevole) e il resto del corpo, mentre è presente una sproporzione analoga tra la lunghezza del corpo e quella degli arti, che diminuisce con l'aumentare dell'età. L'accrescimento staturale presenta due periodi di incremento più intensi che corrispondono ai primi due anni di vita e all'adolescenza; lo sviluppo del *sistema nervoso*, invece, è legato ai primi anni di vita, mentre quello gonadico e dei caratteri sessuali secondari avviene essenzialmente in epoca puberale. Il *tessuto linfatico* si accresce in età infantile e si riduce in età adulta. Il contenuto percentuale di *acqua* decresce nel periodo successivo alla nascita, mentre parallelamente aumentano i grassi totali. Resta pressoché invariato il contenuto proteico corporeo.

L'intero **processo di accrescimento** avviene sotto il controllo di vari fattori:

- **genetici**, responsabili delle basi del processo di sviluppo e modulatori del programma biologico;

- **ormonali**, sostanze che permettono la realizzazione del programma biologico stimolando i vari organi (GH, T3, ormoni sessuali);
- **ambientali**, gli eventuali generatori di distorsioni più o meno gravi della realizzazione del programma di accrescimento (alimentazione, igiene, problemi psicologici, problemi familiari etc.).

6 Le malattie genetiche ed ereditarie

Il concepimento è reso possibile dalla fusione del pronucleo maschile (22 autosomi + X oppure Y) con il pronucleo femminile (22 autosomi + X). È evidente come da tale processo dipenda non solo la *determinazione del sesso* dell'individuo concepito, ma anche il suo *patrimonio genetico*. Lo sviluppo successivo, così come la trasmissione di eventuali malattie genetiche, dipenderanno dall'espressione fenotipica dei geni ereditati dai genitori. L'ereditarietà è condizionata da una serie di leggi riguardanti la *trasmissione dei caratteri* e dagli *eventi casuali di ricombinazione genetica* che possono verificarsi durante la meiosi.

L'esame dell'individuo con possibile difetto genetico (*caso probando*) prevede la raccolta dell'anamnesi personale e familiare e delle informazioni relative a ciascuno dei parenti di primo grado (genitori, fratelli, sorelle e figli del probando). Le **malattie genetiche** eventualmente riscontrate potranno essere **ereditarie** o **non ereditarie**. La presenza di malattie ereditarie consentirà di costruire un *albero genealogico* tipico per ciascun difetto. La modalità di trasmissione sarà di tipo *verticale* nei tratti *dominanti* (genitore e prole affetti), di tipo *orizzontale* nei tratti *recessivi* (fratelli e sorelle colpiti), di tipo *obliquo* nei tratti *recessivi legati al sesso* (zii e nipoti colpiti). La mancanza di ereditarietà mendeliana non esclude la **familiarità**, che origina dall'interazione complessa di geni multipli con fattori ambientali multipli.

Le malattie genetiche possono essere distinte in **tre categorie principali**:

- **cromosomiche**;
- **monogeniche** o **mendeliane**;
- **multifattoriali**.

a) Malattie cromosomiche

Le malattie cromosomiche sono dovute alla mancanza, all'eccesso o all'arrangiamento anomalo di uno o più cromosomi con conseguente eccesso o difetto di materiale genetico. Nell'*aneuploidia* si ha un'alterazione quantitativa del materiale genetico che può determinare gravi malformazioni o l'aborto spontaneo; si verifica per una mancata disgiunzione che può avvenire in fasi diverse. Se si manifesta durante la meiosi si genererà un uovo aneuploide, che, fecondato da uno spermatozoo normale, darà luogo ad un organismo con tutte le cellule aneuploidi; se avviene durante una divisione mitotica o durante lo sviluppo embrionale darà luogo ad un organismo che presenterà l'alterazione soltanto in una percentuale rispetto al totale delle proprie cellule (*mosaico*).

Le più frequenti sindromi correlate alla presenza di aberrazioni cromosomiche compatibili con la vita sono sintetizzate di seguito. Sono possibili, anche se meno frequenti, le monosomie del cromosoma 21 e del cromosoma 18, e le trisomie del cromosoma 18 e del cromosoma 13.

• Sindrome di Klinefelter

La maggior parte dei soggetti affetti da tale sindrome è normale, ad eccezione di alcune alterazioni fenotipiche secondarie. Si tratta, infatti, di individui con testicoli poco sviluppati, ginecomastia, alta statura, obesità, che presentano una *sterilità totale* per mancanza della spermatogenesi. Il corredo cromosomico è rappresentato da *44 autosomi + XXY*.

- **Sindrome di Turner**

È detta anche «disgenesia gonadica». Si manifesta in individui di aspetto femminile, di bassa statura, con collo tozzo e organi sessuali interni infantili; spesso sono presenti anche amenorrea e mancato sviluppo dei caratteri sessuali secondari. Il cariotipo è di 45 cromosomi: *44 autosomi + X0*.

- **Sindrome di Down (o trisomia del cromosoma 21)**

È tra le più frequenti ed importanti sindromi genetiche e ha origine, generalmente, da gameti anormali. Gli individui affetti da tale sindrome presentano caratteristiche fenotipiche patognomoniche, malformazioni multiple, ritardo mentale, alterazioni del sistema nervoso centrale. Il viso ha un tipico aspetto di luna piena, con fessure palpebrali oblique e occhi molto separati fra loro con una piega della pelle nella loro parte interna (*epicanto*), il naso è appiattito e la bocca costantemente aperta. È la più comune fra le malattie congenite, con una frequenza dello 0,1% che aumenta se la madre ha più di 35 anni di età. Si presenta come evento sporadico.

b) Malattie monogeniche (o mendeliane)

Le malattie monogeniche sono provocate dalla mutazione di un *gene singolo*. La frequenza di tali difetti è di circa 10:1000 nati vivi. Presentano una trasmissione ereditaria semplice o mendeliana e vengono classificate in:

- **autosomiche dominanti;**
- **autosomiche recessive;**
- **legate al cromosoma X.**

Il termine *autosomico* si riferisce alla localizzazione del difetto genetico su un autosoma. Il termine *dominante* o *recessivo* dipende dal livello di espressione fenotipica cui ci si riferisce. Da un punto di vista clinico, *dominante* è quella condizione in cui è sufficiente la mutazione di uno solo dei due alleli ad un certo *locus* per produrre la malattia; *recessiva* è, invece, quella condizione in cui entrambi gli alleli devono essere mutati affinché la malattia si verifichi.

Le caratteristiche principali dei soggetti che presentano **difetti autosomici dominanti** sono le seguenti:

- uno dei genitori è malato;
- il soggetto malato può produrre il 50% di figli malati e il 50% di figli sani;
- i due sessi possono essere interessati in proporzione uguale.

Nella maggior parte dei casi questi difetti sono legati ad una mutazione interessante proteine regolatrici di vie metaboliche o proteine strutturali non enzimatiche. Le malattie autosomiche dominanti, rispetto a quelle recessive, presentano una ritardata età di insorgenza e una variabilità nell'espressione clinica. Appartengono a questa categoria: l'*ipercolesterolemia familiare*, il *rene policistico*, la *sferocitosi ereditaria*.

Le caratteristiche principali dei soggetti che presentano **difetti autosomici recessivi** sono le seguenti:

- i genitori sono clinicamente normali;
- in una coppia portatrice vi è il rischio di avere il 25% dei figli malati, il 50% dei figli portatori e il 25% dei figli sani;
- frequentemente sono il prodotto di un incontro sessuale tra consanguinei.

I difetti autosomici recessivi si manifestano clinicamente nel solo stato omozigote; rispetto a quelli autosomici dominanti, hanno un quadro clinico più uniforme e un'insorgenza precoce, pertanto sono diagnosticate più frequentemente nei bambini. Il difetto riguarda in genere una proteina enzimatica di una via catabolica. Appartengono a questa categoria: la *fibrosi cistica*, la

beta talassemia (molto frequente per il vantaggio selettivo conferito dalla malaria), la *fenilchetonuria*, le *mucopolisaccaridosi*, le *glicogenosi*.

Le caratteristiche principali dei soggetti che presentano **difetti legati al cromosoma X** sono le seguenti:

- i soggetti di sesso femminile possono essere dominanti o recessivi per quanto riguarda l'espressione della malattia, in quanto la femmina presenta 2 cromosomi X;
- nei soggetti di sesso maschile, la presenza di un solo cromosoma X comporta sempre la completa espressione del carattere (dominante o recessivo che sia).

I geni responsabili delle malattie legate al sesso sono localizzati sul cromosoma X, perciò il rischio clinico e la gravità della malattia variano in funzione del sesso del soggetto.

Esempi di malattie legate al cromosoma X sono la *distrofia muscolare* tipo *Duchenne* e l'*emofilia*.

c) Malattie multifattoriali

Le malattie multifattoriali sono condizioni prodotte dall'interazione di *geni diversi* posti in loci molteplici (componenti poligenici) e *fattori ambientali*. Rispetto ai difetti mendeliani, tali disturbi colpiscono in percentuale molto inferiore (5-10%) i parenti di primo grado.

Dei disordini multifattoriali fanno parte due grandi gruppi di affezioni:

- le **malattie croniche**, ovvero diabete, ipertensione, gotta, schizofrenia, coronopatie;
- i **difetti congeniti**, ossia labio e palato-schisi, stenosi pilorica ipertrofica, difetti cardiaci congeniti.

Il calcolo del rischio si basa sull'analisi degli alberi familiari interessati. Il rischio è, comunque, di molto inferiore rispetto a quello che comportano i disordini monogenici.

d) Prevenzione e diagnosi delle malattie genetiche ed ereditarie

La prevenzione delle malattie genetiche richiede l'identificazione degli accoppiamenti che possono produrre dei genotipi difettosi. Tali accoppiamenti possono avvenire tra due individui portatori della mutazione, oppure tra un individuo sano e un individuo portatore del difetto genetico. La **consulenza genetica** può essere *retrospettiva* nel caso in cui si parta da un bambino affetto per giungere alla stima del *rischio di ricorrenza* della malattia nella futura prole della stessa coppia, oppure *prospettiva* nel caso in cui si parta da possibili portatori di geni recessivi prima che nasca un individuo affetto. In quest'ultimo caso si procede all'identificazione degli individui eterozigoti attraverso una procedura di *screening della popolazione*. La diagnosi di alcune malattie genetiche può essere effettuata anche prima della nascita, mediante l'*amniocentesi*. Tale tecnica consiste nell'aspirazione transaddominale di fluido amniotico e consente la valutazione del *cariotipo* (numero e struttura dei cromosomi) del feto per la visualizzazione di eventuali alterazioni genomiche non presenti nella popolazione normale. L'esame va condotto preferibilmente tra la 14^a e la 16^a settimana di gravidanza.

7 La contraccezione

Per *contraccezione* s'intende l'insieme dei **metodi contraccettivi**, ossia le diverse soluzioni (farmaci, dispositivi, metodi naturali...) utili a evitare il concepimento. I più diffusi sono:

- la **pillola anticoncezionale**, o **contraccettiva**, un farmaco da assumere per via orale che contiene l'associazione di due ormoni: l'etinilestradiolo ed un ormone progestinico. La differenza tra le numerose pillole in commercio consiste nel dosaggio dell'etinilestradiolo,

nel tipo di progestinico usato e nella combinazione tra i due ormoni. Esistono, infatti, *pillole monofasiche*, in cui il dosaggio dei due ormoni è fisso per tutti i giorni dell'assunzione, e *pillole bifasiche o trifasiche*, nelle quali il dosaggio di uno o di entrambi varia, rispettivamente, per due o tre volte, a seconda dei giorni del ciclo. A seconda dei quantitativi di etinilestradiolo che contengono, poi, le pillole possono essere *a basso dosaggio*. L'azione anticoncezionale della pillola si basa su tre effetti: l'inibizione dell'ovulazione (per evitare che l'ovocita venga liberato dall'ovaio e s'incontri con lo spermatozoo), l'alterazione del muco cervicale (che, diventando denso, inibisce il passaggio degli spermatozoi), l'alterazione dell'endometrio (il tessuto che riveste internamente l'utero, dove s'impianta un eventuale embrione). La cosiddetta **pillola del giorno dopo**, sostanzialmente un dispositivo micro-abortivo, impedisce invece all'embrione di impiantarsi nell'utero. L'effetto della pillola, che va prescritta dal ginecologo in base alle esigenze della singola paziente e va assunta sotto stretto controllo medico, dev'essere valutato caso per caso. In ambito ginecologico ed endocrinologico alcune pillole vengono prescritte a scopi terapeutici; altre, in ambito oncologico, concorrono alla prevenzione del tumore ovarico ed hanno effetti benefici sulla mammella nel caso di mastopatia fibrocistica. Tra le controindicazioni, il sospetto di favorire l'insorgenza del tumore alla mammella se si è iniziato ad assumere la pillola in giovane età e si è continuato per almeno 8-10 anni consecutivi prima della prima gravidanza; e il rischio di trombosi, in particolare per le fumatrici;

- il **cerotto transdermico**, dal funzionamento analogo a quello della pillola. Applicato settimanalmente rilascia in circolo, mediante assorbimento cutaneo, etinilestradiolo e il progestinico norelgestromina;
- l'**anello vaginale**, morbido e di piccole dimensioni, che, inserito in profondità nella vagina per 21 giorni, rilascia localmente etinilestradiolo ed etonogestrel, un progestinico di nuova generazione;
- la **spirale intrauterina**, un pezzetto di polietilene all'incirca di 2-3 centimetri, che, inserito ambulatorialmente nella cavità uterina dal medico, può restare in sede per anni. Alterando l'ambiente endouterino, ostacola il passaggio degli spermatozoi e inibisce i processi d'impianto embrionali. Poiché agisce quasi sempre sull'embrione (e non ne previene la formazione), la spirale viene considerata una metodica microabortiva. Non assicura, quindi, la funzione anticoncezionale, non protegge dalle malattie sessualmente trasmesse, può favorire infezioni (quelle delle tube determinano sterilità permanente) e, in caso di gravidanza, dare complicazioni;
- il **diaframma**, una piccola membrana in plastica inserita in profondità, prima del rapporto, all'interno della vagina. Impedisce meccanicamente il passaggio degli spermatozoi verso il canale cervicale. Ha una efficacia contraccettiva non eccellente e non protegge dalle malattie sessualmente trasmesse;
- il **profilattico**, «cappuccio» in plastica da applicare sull'organo genitale maschile durante il rapporto sessuale. Rappresenta un rimedio ottimale, se correttamente usato, sia per l'efficacia contraccettiva sia per la prevenzione delle malattie sessualmente trasmesse.

I **metodi naturali** sono perlopiù adatti alle coppie stabili e finalizzati al **controllo delle nascite**. Hanno, infatti, un'efficacia contraccettiva assai limitata (di contro, vengono utilizzati anche per favorire le gravidanze desiderate) e, soprattutto, non proteggono in alcun modo dalle malattie sessualmente trasmissibili. Ve ne sono di vari tipi; tra i più comuni:

- il **coito interrotto**, consistente nell'interrompere, appunto, il rapporto sessuale prima dell'eiaculazione per evitare — o, comunque, limitare — il contatto tra il liquido seminale maschile ed i genitali femminili;
- il **metodo Ogino-Knaus**, che prevede di calcolare le date dei giorni fertili in base alla durata dei cicli mestruali precedenti. La fase fertile viene identificata individuando il più breve e il più lungo ciclo mestruale dell'anno precedente, sottraendo 19 giorni alla lunghezza del ciclo

più breve e 11 a quella del ciclo più lungo, per stabilire rispettivamente il primo e l'ultimo giorno della fase fertile. Non sempre, però, i cicli sono regolari;

- il **metodo della temperatura basale**, che prescrive la misurazione quotidiana della temperatura corporea della donna. In prossimità dell'ovulazione, infatti, la temperatura subisce un incremento per alcuni giorni. Il metodo prevede la possibilità di rapporti sicuri dal terzo giorno successivo al massimo aumento della temperatura, fino alla comparsa del successivo ciclo mestruale;
- il **metodo Billings**, ossia l'autovalutazione, da parte della donna, della quantità e della qualità delle caratteristiche secrezioni vaginali in prossimità dell'ovulazione, approssimativamente a metà di ogni ciclo mestruale.

La gestione delle funzioni biologiche nell'uomo

1 Le funzioni biologiche

Il corpo umano è costituito da un armonico e complesso insieme di organi e di apparati, ciascuno finalizzato ad una specifica funzione biologica; schematicamente, queste ultime possono essere così suddivise:

- funzioni della **vita vegetativa**;
- funzioni della **vita di relazione**.

Le funzioni della *vita vegetativa* sono finalizzate alla conservazione dell'individuo e, pertanto, comprendono **la nutrizione** e le sue varie fasi (digestione, assimilazione, respirazione, escrezione), le funzioni relative alla conservazione della specie e, quindi, **la riproduzione**.

Le funzioni della *vita di relazione* comprendono le funzioni necessarie alla conoscenza e all'interazione con il mondo esterno e, dunque, riguardano la **sensibilità**, la **psiche** e il **movimento** (riflesso, istintivo, automatico, volontario).

2 L'energia

L'energia è la capacità potenziale di compiere un lavoro. Essa viene fornita all'uomo dalla **catena alimentare**, grazie alla quale l'energia fisica si trasforma in energia chimica mediante la realizzazione di un flusso energetico. La trasformazione inizia con la conversione dell'energia luminosa solare (*irraggiamento*) in energia chimica dei vegetali (*fotosintesi*); i composti vegetali sono ingeriti e modificati dagli **erbivori**, i quali, a loro volta, sono la fonte energetica dei **carnivori**. I vegetali, gli erbivori e i carnivori costituiscono il substrato plastico ed energetico degli **onnivori** (tra cui l'uomo). Nell'uomo il flusso di energia continua attraverso le trasformazioni degli alimenti nei singoli nutrienti e prosegue, successivamente, nella metabolizzazione di questi ultimi sino alla sintesi dell'adenosintrifosfato (**ATP**). Alla fine si giungerà ad una nuova trasformazione in energia fisica per il compimento di lavoro meccanico.

L'energia *decade da un passaggio all'altro*. Lo svolgimento del lavoro muscolare — compiuto a spese di una trasformazione di energia chimica in energia meccanica (da alimenti a movimento) — presenta un rendimento di circa il 25%, accompagnato da una dispersione di calore nell'ambiente interno che si manifesta come aumento della temperatura corporea e da una dispersione di calore nell'ambiente esterno che determina un aumento dell'entropia. L'energia utilizzata come lavoro rappresenta la quota necessaria a mantenere un equilibrio fra l'ordine e il disordine del nostro sistema corporeo. Agli estremi di tale equilibrio possono essere configurati l'inattività da una parte e il lavoro eccessivo dall'altra.

La possibilità che il corpo umano possa produrre **lavoro esterno** dipende da due presupposti fondamentali:

- l'efficienza del sistema neuro-muscolo-osteo-teno-cartilagineo;
- la disponibilità di nutrienti energetici mediante i quali giungere alla sintesi dell'ATP e della CP (*creatinafosfato*). Infatti, la rottura dei legami fosfato ad alta energia di tali composti libera

l'energia necessaria alla produzione del lavoro meccanico e permette la realizzazione della sequenza:

contrazione muscolare → spostamento leve ossee → movimento

Il cibo e i meccanismi preposti alla sua trasformazione hanno, dunque, una grande importanza nelle diverse fasi dell'accrescimento corporeo, nelle attività di movimento lavorative e relazionali. La branca delle Scienze biologiche che si occupa della scelta degli alimenti, della loro combinazione e preparazione, della degustazione dei cibi, della digestione e riduzione a nutrienti, dell'assorbimento e accumulo di questi ultimi nei tessuti si chiama **Scienza dell'alimentazione**.

Lo **stile di vita**, cioè l'insieme dei comportamenti tenuti, è determinante per lo stato di salute. In particolare, l'eccessiva assunzione di *alimenti*, insieme al consumo improprio di *tabacco* e di *alcol*, ha un rapporto di causa-effetto nell'insorgenza di diverse malattie, tanto da essere frequentemente oggetto dei moderni piani educativi sanitari.

3 Nutrizione ed equilibrio idro-elettrolitico

Per *nutrizione* s'intende l'introduzione dei nutrienti nell'organismo attraverso l'alimentazione.

Un'**alimentazione corretta** prevede:

- la scelta degli alimenti adatti al singolo individuo;
- la pulizia e la cernita delle parti commestibili di tali alimenti;
- la preparazione e la cottura secondo opportuni principi nutrizionali.

L'*alimentazione personalizzata* è resa necessaria dalla diversa distribuzione tra gli individui dei costituenti fondamentali del corpo umano (*acqua, massa magra, massa grassa, massa minerale*), legata all'età, al sesso, ai fattori genetici, ai fattori ambientali, all'attività lavorativa svolta, all'attività sportiva praticata. Dai rapporti esistenti tra età, peso e altezza, su appositi *nomogrammi* è possibile calcolare il grado di variazione patologica del peso corporeo per ciascun individuo in base al sesso.

La quantità di calorie che un individuo deve assumere quotidianamente per far fronte ai propri bisogni di energia viene definita **fabbisogno calorico** ed è rappresentato dall'insieme del *metabolismo basale*, necessario al mantenimento dei processi vitali, e del fabbisogno indispensabile per la termoregolazione, le attività quotidiane, il lavoro muscolare. Nei soggetti adulti (18-35 anni di età) il *fabbisogno energetico* è di circa 3.000 calorie per gli uomini e di circa 2.000 calorie per le donne. L'esigenza calorica è maggiore nell'adolescenza e diminuisce con l'avanzare dell'età. Per elaborare correttamente la dieta bisogna aver raccolto la storia alimentare dell'individuo, calcolato il fabbisogno calorico medio quotidiano, valutato la distribuzione percentuale media dei carboidrati, dei lipidi e delle proteine, considerato eventuali condizioni di eccesso o difetto di peso. Le calorie necessarie sono fornite dalla combustione dei nutrienti contenuti nei diversi alimenti, cioè zuccheri, grassi e proteine.

Ogni *nutriente* ha una propria funzione specifica, che è:

- **energetica;**
- **plastica;**
- **regolatrice.**

Gli zuccheri e i grassi sono fattori prevalentemente energetici; le proteine, i sali minerali e l'acqua sono elementi prevalentemente plastici, in quanto coinvolti nei processi di costruzione e rimaneggiamento dei tessuti; le vitamine, alcuni sali e cofattori possono essere considerati prevalentemente elementi regolatori.

Nella ripartizione quotidiana dei nutrienti bisogna tener conto del fatto che:

- le **proteine** devono rappresentare il 10-15% della quota energetica giornaliera e devono comprendere proteine vegetali e animali, in quanto queste ultime sono ricche degli otto *amminoacidi essenziali* che l'organismo non è in grado di sintetizzare;
- i **grassi** devono costituire il 25-30% della dieta giornaliera, in quanto sono la fonte energetica più conveniente (9Cal/g). Si dividono in *saturi*, *monoinsaturi* (olio d'oliva) e *polinsaturi* (oli di cereali); questi ultimi sono considerati essenziali, in quanto in parte non sintetizzabili dall'organismo;
- gli **zuccheri** devono costituire almeno il 50% della quota calorica giornaliera. Sono di provenienza prevalentemente vegetale e rappresentano la fonte di combustione più rapida. I carboidrati si dividono in *zuccheri semplici* (fruttosio, glucosio) e *zuccheri complessi* (pasta, pane); questi ultimi sono preferibili, in quanto stimolano in misura minore la secrezione insulinica rispetto agli zuccheri semplici.

Le linee guida da seguire per un'**alimentazione corretta** sono le seguenti:

- pesarsi settimanalmente e, in caso di sovrappeso, ridurre la quantità degli alimenti e il consumo di grassi e zuccheri;
- preferire pane integrale e pasta condita con salse semplici a base di pomodoro, ortaggi o verdure;
- limitare le carni rosse e i formaggi, preferendo il pesce e le carni bianche;
- limitare il consumo di dolci e zucchero;
- abolire il consumo di alcol;
- consumare abbondantemente frutta e verdure;
- condire con olio d'oliva;
- limitare l'uso del sale da cucina, preferendo aromatizzare i cibi con spezie naturali e succo di limone.

L'**acqua** è il costituente principale delle cellule e quindi dei tessuti, nonché dei compartimenti extra e pericellulari. L'abbondanza di acqua è funzionale alle necessità dell'organismo, che sono:

- la diluizione delle sostanze fondamentali (ossigeno, proteine strutturali, lipidi, zuccheri, enzimi, ormoni, vitamine e sali);
- il trasporto e la distribuzione ai tessuti di tali sostanze;
- gli scambi diretti tra cellula e ambiente extracellulare;
- l'attivazione dei processi di scissione dell'ATP;
- il mantenimento della termoregolazione;
- la lubrificazione del pericardio, della pleura, del peritoneo e delle articolazioni.

Il contenuto di acqua nell'organismo presenta modeste variazioni nei due sessi, mentre subisce una progressiva e costante diminuzione con il progredire dell'età; infatti, l'invecchiamento può essere considerato un processo di lenta disidratazione e di decadimento delle capacità funzionali. Normalmente un'adeguata idratazione consente di mantenere costante l'equilibrio idro-elettrolitico dell'individuo, tuttavia in situazioni estreme si possono manifestare sindromi da disidratazione o da eccesso o difetto di sali. La causa più frequente di perdita acuta di liquidi è rappresentata dal lavoro muscolare eccessivo; meno frequentemente le disidratazioni sono collegate a situazioni di patologia, come la febbre (e dunque la sudorazione profusa), il vomito ripetuto, la diarrea, l'abuso di diuretici. In tali situazioni è possibile riscontrare un'alterazione dell'equilibrio distributivo dell'acqua e dei sali minerali fra i compartimenti intra e pericellulari. Infatti, all'interno della cellula, nel plasma sanguigno e nei fluidi extracellulari sono presenti **sali**, detti *elettroliti* per la loro proprietà di possedere una carica elettrica e di rendere elettroconduttive le soluzioni. I sali più importanti sono: il *sodio*, il *cloro*, il *potassio*, il *magnesio*, il *calcio*. L'eccita-

bilità delle membrane cellulari è legata alla distribuzione delle cariche elettriche all'interno e all'esterno della cellula determinate dagli ioni di tali sali. Pertanto i tessuti che maggiormente risentono delle variazioni saline sono il tessuto nervoso e il tessuto muscolare, ossia i tessuti più *eccitabili* presenti nel corpo umano.

4 I disordini nutrizionali

Per *malnutrizione* si intende l'assunzione inadeguata di alimenti, sia in difetto sia in eccesso, rispetto alle esigenze fisiologiche.

I disordini nutrizionali, in passato, sono dipesi prevalentemente da malattie dovute a carenze alimentari. Da alcuni decenni, il miglioramento dei redditi familiari e l'arricchimento dei cibi con vitamine e sali minerali hanno permesso di limitare la patologia carenziale alimentare nella popolazione generale con, all'opposto, un incremento relativo delle **patologie da eccesso alimentare**.

I *disordini alimentari* possono essere suddivisi in:

- **difetti alimentari;**
- **eccessi alimentari.**

I difetti alimentari riguardano situazioni in cui, per l'inadeguata introduzione dei nutrienti globalmente considerati, oppure per la mancata assunzione di elementi dietetici *essenziali*, si manifestano delle **sindromi carenziali nutrizionali**. In passato, la maggior parte dei difetti carenziali riguardava le *avitaminosi* determinate dalle carestie e dall'impossibilità di assumere gli alimenti in maniera variata per lunghi periodi di tempo. Le patologie più frequentemente riscontrate erano: la *xerofthalmia*, dovuta alla mancanza di vitamina A; il *beriberi*, dovuto a deficit della vitamina B1; la *pellagra*, dovuta a carenza di vitamina PP; lo *scorbut*, che si manifestava nei marinai per una dieta carente di vitamina C; il *rachitismo*, determinato da una insufficiente azione della vitamina D. Tali situazioni carenziali selettive si affiancavano a manifestazioni di malnutrizione abbastanza frequenti nella popolazione generale. Attualmente la malnutrizione calorica e proteica si riscontra soltanto nei bambini dei Paesi del Terzo e del Quarto Mondo, dove può evidenziarsi attraverso due sindromi: il *marasma*, causato dalla mancanza di calorie, che si manifesta con l'arresto della crescita; il *kwashiorkor*, causato da carenza di proteine, che si manifesta con edemi per l'ipoalbuminemia. Nella popolazione generale è possibile riscontrare sindromi carenziali negli alcolisti, nei tossicodipendenti, negli anziani non controllati, nelle giovani donne colpite da **anoressia nervosa** che si sottopongono a digiuno volontario.

La società occidentale, invece, è caratterizzata perlopiù da un'alimentazione costante e abbondante; pertanto, la capacità di immagazzinare grasso è diventata spesso un fattore che influisce negativamente sulla sopravvivenza. Il consumo globale di cibo è eccessivo e, di conseguenza, la situazione patologica alimentare più diffusa è l'**obesità**, che rappresenta un fattore di rischio per diverse malattie, tra le quali quelle cardiovascolari, neoplastiche e del sistema osteoarticolare. Tale condizione, inoltre, presenta un'associazione molto stretta con le malattie del ricambio o metaboliche. Per *ricambio* si intende il complesso metabolico attraverso il quale una determinata sostanza viene elaborata e la rapidità con la quale viene consumata e reintegrata. Le principali **malattie del ricambio**, provocate da situazioni di accumulo, sono il diabete, la gotta e le dislipidemie. La maggior parte di queste patologie può essere prevenuta attraverso l'eliminazione o la riduzione di alcuni principi alimentari che, introdotti in eccesso per lunghi periodi di tempo, possono determinare alterazioni a carico dei capillari degli organi con microcircolo terminale, come il cuore (circolo coronarico), il cervello (circolo cerebrale), il rene, la retina. Le sostanze principalmente responsabili di tali alterazioni vascolari sono il colesterolo, i trigliceridi e il glucosio. Quando vi si accompagnano altri fattori, quali l'ipertensione, il

fumo, l'obesità, il rischio aumenta in maniera esponenziale. L'educazione alimentare, affiancata dall'educazione al movimento, può creare le basi per avviare programmi di prevenzione delle patologie di origine dismetabolica. Il raggiungimento di tale obiettivo sulla popolazione generale non può prescindere da un'adeguata educazione alla gestione del corpo, attuata anche nella scuola mediante progetti interdisciplinari integrati.

5 I disturbi alimentari

Con l'espressione *disturbi del comportamento alimentare* (DCA) s'identificano **una serie di patologie complesse** che possono compromettere in maniera irrimediabile tutti gli organi e gli apparati (cardiovascolare, gastrointestinale, endocrino,ematologico, scheletrico, sistema nervoso centrale, dermatologico etc.), avere evoluzioni nefaste ed effetti letali. Ciascuna variante si presenta con modalità specifiche; tuttavia, le caratteristiche comuni consistono nell'**ossessione per**:

- il **cibo**;
- il **peso**, che, però, non è un marcatore clinico imprescindibile, giacché anche persone con «peso forma» possono esserne colpite;
- l'**immagine corporea**.

Colpiscono perlopiù **donne giovani** ma il numero dei maschi è in crescita, soprattutto in età adolescenziale e pre-adolescenziale. L'età di esordio, inoltre, nel corso degli ultimi anni, è sensibilmente calata, per cui non è raro individuare i DCA tra **bambini e pre-adolescenti**. Le cause sono multifattoriali, comprendendo **fattori sia psicologici sia biologici**. Sono state individuate, comunque, alcune **condizioni predisponenti**, ovvero:

- familiarità per disturbi del comportamento alimentare, depressione, abuso di sostanze;
- possibili eventi avversi/traumatici, malattie croniche dell'infanzia e difficoltà alimentari precoci;
- appartenenza a gruppi sociali in cui la magrezza rappresenta un'indispensabile prerogativa (sport, danza, moda etc.);
- percezione e interiorizzazione dell'ideale di magrezza;
- insoddisfazione della propria immagine corporea;
- scarsa autostima e tendenza al perfezionismo;
- stati emotivi negativi.

Accanto ad alcune forme secondarie — i cosiddetti «disturbi sottosoglia», le forme ibride e gli Ednos, ossia i disturbi alimentari non altrimenti specificati — se ne distinguono **tre forme principali**:

- l'**anoressia nervosa**;
- la **bulimia**;
- il **binge eating** (disturbo da *alimentazione incontrollata*).

Una citazione a parte merita l'**ortoressia**, ossia l'ossessione di alimentarsi in maniera sana. Un comportamento, questo, definito e codificato come patologia nel 1997 dal dietologo statunitense **Steve Bratman**, attraverso l'osservazione di una serie di atteggiamenti patologici assunti da lui stesso e poi riscontrati in alcuni dei suoi pazienti, vale a dire:

- non riuscire a mangiare **verdure** colte da più di un quarto d'ora;
- masticare per almeno cinquanta volte ogni singolo **boccone**;
- **temere di ammalarsi mangiando** qualcosa di non perfettamente sano.

L'ortoressia è un disturbo alimentare particolarmente subdolo, poiché il confine tra alimentarsi in maniera sana e assumere comportamenti maniacali al riguardo appare labile e sfumato. Bratman ha messo a punto il **test** proposto qui di seguito per facilitarne la diagnosi: chi risponde

affermativamente ad almeno quattro domande viene classificato «potenzialmente» ortoressico, in un crescendo di gravità della patologia.

1. Spendi più di tre ore al giorno riflettendo sulla tua alimentazione?
2. Pianifichi i tuoi pasti diversi giorni prima?
3. La possibilità che i cibi assunti ti facciano ingrassare è sempre più importante del piacere di mangiarli?
4. Lo stato d'ansia nella tua vita è aumentato da quando hai riflettuto sulla tua alimentazione?
5. Sei diventato più severo con te stesso nei confronti del tuo comportamento quotidiano e alimentare?
6. La tua autostima aumenta quando ti alimenti correttamente?
7. Hai eliminato radicalmente diversi cibi che ti piacevano in favore di cibi più salutari?
8. Ti riesce più difficile mangiare fuori casa?
9. Ti senti in colpa quando non mangi in modo corretto?
10. Ti senti in pace con te stesso e in pieno controllo quando mangi in modo corretto?

Per la riuscita del trattamento dei disturbi del comportamento alimentare la **diagnosi precoce** è determinante. Gli interventi terapeutici dipendono dal tipo di disordine e dall'età. Per l'anoressia nervosa e per la bulimia si può ricorrere ai farmaci (gli antidepressivi, ad esempio, sono efficaci per ridurre i sintomi nel breve termine), ma i pazienti necessitano d'interventi mirati sia agli aspetti psichiatrici e psicologici, sia a quelli nutrizionali, fisici e socio-ambientali. Fondamentale, quindi, la collaborazione di un'**équipe di specialisti** che comprenda:

- il medico internista, il medico o il pediatra di famiglia;
- lo psichiatra o il neuropsichiatra infantili;
- lo psicologo e lo psicoterapeuta;
- il nutrizionista;
- il dietista.

È altrettanto importante che nella riabilitazione nutrizionale dei bambini e degli adolescenti vengano coinvolti tutti i membri della **famiglia**. Se il malato è drammaticamente sottopeso o se è in pericolo di vita è necessario il **ricovero ospedaliero**, per il quale si propende anche per i soggetti a elevato rischio di suicidio o di comportamento autolesionistico.

a) Anoressia nervosa

L'anoressia nervosa è la patologia psichiatrica col rischio di **mortalità** più elevato (entro i primi 10 anni dall'esordio è 10 volte maggiore che nella popolazione generale di pari età). I livelli di **autostima** degli anoressici sono esageratamente dipendenti dalla loro **immagine corporea**, per cui tendono a mantenere il loro peso inferiore di almeno il 15% rispetto al peso forma. Terrorizzati dalla sola idea d'ingrassare, nonostante il grave sottopeso percepiscono in maniera distorta la loro immagine corporea e negano di trovarsi in pericolo di vita.

Per perdere peso, oltre che a diete estremamente restrittive, possono ricorrere ai seguenti sistemi:

- praticare **esercizio fisico** in modo esagerato;
- ricorrere a **purghe, farmaci anoressizzanti, diuretici**;
- indursi il **vomito**, soprattutto dopo i pasti.

Se ne distinguono **due tipi**:

- con **restrizioni** (dieta, digiuno e attività fisica intensa);
- con **abbuffate (binge)** e **condotte di eliminazione** (uso inappropriato di lassativi, diuretici, clisteri; vomito autoindotto anche dopo l'assunzione di piccole quantità di cibo).

Le persone affette da **anoressia nervosa** presentano i seguenti sintomi:

- magrezza estrema con scomparsa dei depositi adiposi e atrofia muscolare;
- cute secca, rugosa, comparsa di lanugine sul volto e sugli arti; riduzione della produzione sebacea e del sudore; colorazione giallastra della cute;
- estremità delle mani e dei piedi bluastre in seguito all'esposizione a freddo (*acrocianosi*);
- cicatrici o callosità sul dorso delle dita (*segno di Russell*), dovute al continuo mettersi le dita in gola per provocarsi il vomito;
- capelli opachi, spenti e sottili;
- carie dentali ed erosioni dello smalto, infiammazioni delle gengive, ingrossamento delle ghiandole parotidi (dovuti al vomito autoindotto di frequente e al conseguente aumento di acidità nella bocca);
- bradicardia (rallentamento del battito cardiaco), aritmie, palpitazioni e ipotensione;
- crampi allo stomaco, ritardato svuotamento gastrico;
- stipsi, emorroidi, prolasso rettale;
- alterazioni del sonno;
- amenorrea (scomparsa delle mestruazioni per almeno tre cicli consecutivi);
- perdita d'interesse sessuale;
- osteoporosi e aumentato rischio di fratture;
- perdita di memoria, difficoltà di concentrazione;
- depressione (possibile ideazione suicidaria e tentati suicidi), atti autolesionistici, ansia, comportamento ossessivo-compulsivo;
- possibili rapide fluttuazioni del livello degli elettroliti, con importanti ripercussioni sul cuore (fino all'arresto cardiaco).

b) Bulimia

La bulimia si distingue dall'anoressia per la presenza di **ricorrenti abbuffate**. I bulimici tendono, infatti, a consumare grandi quantità di cibo in poco tempo e riferiscono la sensazione di non riuscire a smettere di mangiare. Per non ingrassare, successivamente all'episodio bulimico, i pazienti mettono in atto **condotte compensatorie** consistenti nel procurarsi il vomito o ricorrendo a lassativi, purghe, farmaci anoressizzanti, diuretici e ormoni tiroidei. Perché si possa avere una diagnosi di bulimia, abbuffate e condotte compensatorie devono verificarsi, in media, almeno due volte a settimana nel corso di tre mesi. I pazienti sono terrorizzati dall'idea di ingrassare ma, allo stesso tempo, sono attratti in maniera irresistibile dal cibo. Anche i livelli di autostima dei bulimici, come quelli degli anoressici (talvolta nella loro storia è presente una forma di anoressia nervosa), sono esageratamente dipendenti dalla loro **immagine corporea**.

Se ne distinguono **due tipi**:

- con condotte di eliminazione (uso inappropriato di lassativi, diuretici, clisteri, vomito autoindotto);
- senza condotte di eliminazione (con altri comportamenti compensatori ma senza il ricorso abituale all'uso inappropriato di lassativi, diuretici, clisteri o al vomito autoindotto).

Le persone affette da **bulimia** presentano i seguenti sintomi:

- carie dentali ed erosioni dello smalto, infiammazioni delle gengive, ingrossamento delle ghiandole parotidi (dovuti al vomito autoindotto di frequente e al relativo aumento di acidità nella bocca);
- cicatrici o callosità sul dorso delle dita (*segno di Russell*), dovute al continuo mettersi le dita in gola per provocarsi il vomito;
- stipsi e rallentamento del transito intestinale, con conseguente dipendenza dai lassativi;

- possibili rapide fluttuazioni del livello degli elettroliti, con importanti ripercussioni sul cuore (fino all'arresto cardiaco).

c) Disturbi del comportamento alimentare non altrimenti specificati

I disturbi del comportamento alimentare che non rientrano nei criteri diagnostici dell'anoressia nervosa e della bulimia si definiscono «non altrimenti specificati» e hanno le seguenti caratteristiche:

- sono presenti tutti i criteri dell'anoressia nervosa nelle donne, a eccezione dell'amenorrea (la paziente, cioè, ha normali cicli mestruali);
- sono presenti tutti i criteri dell'anoressia nervosa, a eccezione del peso, ridotto ma nella norma;
- sono presenti tutti i criteri della bulimia, a eccezione delle abbuffate incontrollate e dei meccanismi compensatori inappropriati, che si verificano con frequenza inferiore a due volte a settimana o per meno di tre mesi;
- il peso è normale, tuttavia il paziente ricorre abitualmente a comportamenti compensatori inappropriati dopo aver ingerito piccole quantità di cibo;
- il paziente mastica abitualmente grandi quantità di cibo per poi sputarlo, senza inghiottirlo;
- il paziente fa ricorrenti abbuffate, non accompagnate, però, dai regolari comportamenti compensatori inappropriati, caratteristici dei pazienti con bulimia nervosa. In tal caso si parla di **disturbo da alimentazione incontrollata (*binge-eating*)**, patologia che è spesso, ma non necessariamente, associata all'obesità;
- il paziente presenta anoressia al mattino, aumento dell'appetito nel corso della sera e insonnia. Può anche non ricordarsi di aver mangiato durante la notte. In tal caso si tratta di **sindrome da alimentazione notturna (*night eating syndrome*)**.

6 Igiene dell'apparato locomotore e importanza del movimento

L'igiene dell'apparato locomotore va praticata sotto due aspetti:

- **correzione delle abitudini;**
- **educazione al movimento.**

La correzione delle abitudini è finalizzata a determinare nei soggetti l'assunzione di atteggiamenti corretti nella postura e nel movimento. Nella scuola secondaria di I e II grado il compito del docente è quello di prevenire la comparsa di *paramorfismi* (scoliosi, lordosi etc.) e/o ipotonie muscolari, frequenti nei soggetti in fase di crescita, illustrando anche le conseguenze spesso irreversibili che tali alterazioni possono comportare.

L'educazione al movimento si realizza in maniera complementare all'insegnamento dell'**Educazione fisica** e mira ad un'equilibrata maturazione psico-fisica dei soggetti in età evolutiva, concorrendo parallelamente a facilitare l'inserimento sociale e a sollecitare un armonico sviluppo corporeo. Nella prassi dell'azione educativa, le esercitazioni connesse andranno combinate logicamente, in modo che la successione di sforzi e di carichi risponda a rigorose leggi fisiologiche. Gli **obiettivi** dell'Educazione fisica e delle attività sportive di tempo libero (calcio, tennis, nuoto etc.) sono rappresentati fondamentalmente dal *potenziamento fisiologico*, ottenuto attraverso il miglioramento della funzione cardiorespiratoria, il rafforzamento della potenza muscolare, la mobilità articolare e la velocità dei movimenti.

Da un punto di vista fisico va attentamente commisurata la **fatica**, per cui l'attività di movimento e/o di sforzo va intervallata con periodi di **riposo**. Particolare attenzione va posta anche alle condizioni ambientali in cui l'attività fisica viene svolta, verificando che l'umidità relativa,

la temperatura, la ventilazione e i livelli di ossigeno siano adeguati. L'attività fisica va evitata nel periodo post-prandiale. L'organismo recupera le energie perse per le funzioni vitali e per le attività della vita di relazione attraverso il *riposo* e l'*alimentazione*.

7 Fatica e riposo

Il **riposo** va «organizzato» e realizzato quotidianamente riservando un periodo di tempo congruo al **sonno**, che deve essere sufficiente ed adeguato e deve avvenire in condizioni idonee. Il fabbisogno di sonno varia in funzione dell'**attività svolta** e dell'**età**, diminuendo con l'aumentare degli anni. Il sonno deve essere soddisfatto su letti comodi e in microclimi adeguati; sarebbe auspicabile, inoltre, distribuire una quota del riposo quotidiano nel periodo post-prandiale.

Il *lavoro mentale* corrisponde, in termini fisiologici, al lavoro muscolare: ogni richiesta determina un consumo di energie e la produzione di metaboliti tossici. La **fatica mentale** può, quindi, condizionare le funzioni circolatorie, le funzioni respiratorie, la funzione digestiva e la tensione muscolare. Esistono, pertanto, dei limiti superati i quali si possono realizzare due tipi di fatica mentale:

- la **fatica fisiologica (o piccola fatica)** si manifesta con disattenzione, senso di svogliatezza, irrequietezza, ed è un fenomeno fisiologico naturale conseguente a un dato lavoro;
- la **fatica patologica (o grande fatica)** si manifesta con la comparsa di disturbi organici, che incidono sull'equilibrio psico-fisico, sull'appetito, sulla capacità di riposo e che, se persistenti, possono condurre a perturbazioni dello stato di salute quali insonnia, irritabilità, cefalea, depressione, fino a giungere, nei casi estremi, all'anemia e al dimagrimento.

L'**attività fisica**, la vita all'aria aperta in condizioni di salubrità e l'attività sportiva garantiscono ottima salute. Praticare uno sport rappresenta senz'altro una valida compensazione ad una vita stressante e sedentaria. Essere sportivi significa *praticare attività libera o agonistica*, cioè realizzare un rapporto intimo e dinamico con la natura, traendone vantaggi spirituali e fisici. Per mancanza di una educazione specifica, in Italia si tende a considerare lo sport solo da spettatori e troppo spesso si finisce con lo scaricare, in occasione di manifestazioni sportive, le tensioni e gli stress accumulati nella vita quotidiana.

La sedentarietà è il presupposto a vari disturbi organici. Infatti si definisce *ipocinesi* una tipica malattia del nostro tempo che è causata, come suggerisce il termine scientifico, da *manca di movimento*. L'uso degli ascensori, dell'automobile, occupazioni sempre più sedentarie, la meccanizzazione del lavoro industriale, l'automazione delle faccende domestiche hanno favorito l'insorgenza di diverse patologie. L'obesità o la tendenza all'obesità, la scoliosi, la ridotta capacità respiratoria e la scarsa efficienza cardiocircolatoria sono malformazioni dipendenti dalla poca attività fisica. Il vecchio detto «*mens sana in corpore sano*» conferma quanto l'*attività motoria* sia benefica per lo *sviluppo fisico* ma anche, e soprattutto, per l'equilibrio mentale. Tutti i sistemi e gli apparati del corpo umano traggono vantaggio dall'attività sportiva, sempre che sia condotta con cautela e nel rispetto di alcune norme igienico-sanitarie.

Ogni corpo, muovendosi, allaccia intimi rapporti con lo spazio e con il tempo e realizza, in tal senso, la propria autonomia fisica e psichica. Lo spirito agonistico esalta tale condizione e favorisce lo sviluppo del carattere, permettendo inoltre la socializzazione. Per poter praticare una disciplina sportiva, soprattutto a livello agonistico, è necessario imparare le nozioni tecniche di ogni particolare sport e praticare l'*allenamento*. Questo abitua l'organismo a compiere determinati movimenti e, riducendo la spesa energetica, garantisce un maggior rendimento. È opportuno eseguire l'attività in luoghi adatti e sani, palestre, campi, piscine, perché soltanto così l'organismo resta protetto da spiacevoli inconvenienti. L'apparato respiratorio ne sarà avvantaggiato perché ad un aumento della capacità respiratoria corrisponderà una maggiore ventilazione e, dunque, un maggior consumo di ossigeno.

Le articolazioni e le strutture scheletriche saranno mantenute in esercizio, il sistema muscolare e il sistema cardiocircolatorio acquisteranno maggior potenza e resistenza; la più intensa attività muscolare stimolerà una migliore circolazione del sangue nei vasi.

Anche il cuore, allenandosi allo sforzo, raggiunge una potenza maggiore. Infatti, il cosiddetto «cuore d'atleta» è caratterizzato da un minore numero di frequenza di battiti e quindi delle compressioni, ma da una maggiore potenza nella spinta.

Consumi voluttuari e prevenzione delle dipendenze

1 La dipendenza

Le strategie di prevenzione della *dipendenza da sostanze tossiche* non possono prescindere dalla complessità dei fattori che concorrono a determinare il fenomeno e devono necessariamente coinvolgere tutti gli organi di governo, nonché le istituzioni che cooperano alla costruzione dell'individuo e del tessuto sociale. Tali strategie hanno come presupposto fondamentale la costituzione di programmi integrati e interdisciplinari. Particolarmente importante in tal senso risulta il ruolo dei Comuni, della scuola, della famiglia e degli operatori dei servizi socio-sanitari.

La scuola ha un'importanza cruciale nello sviluppo degli individui; e il disadattamento scolastico, la cui prevenzione, per la parte di «competenza» della scuola, si traduce nella **lotta contro le tossicodipendenze e la devianza in generale**, è presente nella storia personale della gran parte degli individui problematici. La scuola può e deve fornire agli alunni strumenti di lettura e di interpretazione delle relazioni umane, in una fase decisiva dello sviluppo qual è quella dell'adolescenza. Le attività di educazione sanitaria a livello scolastico devono, comunque, coinvolgere insegnanti e studenti attraverso interventi interdisciplinari.

2 Il tabagismo

Per confezionare le sigarette vengono utilizzate le foglie del tabacco essiccate e trattate con additivi che ne rendono più gradevole l'aroma; in genere si utilizzano miscele composte da diversi tipi di tabacco e altri ingredienti. Quando il fumatore «tira», inspira direttamente il fumo prodotto dal tabacco che brucia, il quale agisce su labbra, cavità orale, gola, trachea; il fumo prodotto dalla combustione della sigaretta, attraverso il naso, raggiunge anch'esso la trachea, dove si concentrano i due tipi di fumo, che contengono, fra le altre, sostanze considerate cancerogene.

Le sigarette contengono, inoltre, **sostanze irritanti** la cui azione si esplica specialmente sulla mucosa di rivestimento dei bronchi e, in particolare, su quelle cellule provviste di ciglia che difendono il polmone da polveri, germi e sostanze inquinanti. Venendo meno i meccanismi di difesa dei polmoni, aumenta il rischio di infezioni, perché gli organi sono più esposti e indeboliti; se l'azione irritante continua, viene alterato anche il normale processo di rinnovamento cellulare dei polmoni. Insorgono la tosse e il catarro e, con il passare del tempo, possono comparire la bronchite cronica e l'enfisema polmonare. Altro ingrediente della sigaretta è il **catrame**, composto a sua volta da centinaia di sostanze di sicuro effetto cancerogeno sull'apparato respiratorio, sul cavo orale, sulla gola e sulle corde vocali; tra i suoi effetti immediati: l'ingiallimento dei denti, l'amaro in bocca e l'irritazione delle vie respiratorie.

Il **monossido di carbonio** si sprigiona dalla sigaretta accesa e viene immesso nei polmoni con la respirazione, «bloccando» l'emoglobina del sangue, proteina che serve al trasporto dell'ossigeno dai polmoni a tutti gli organi o tessuti del nostro corpo, in una percentuale persino del 15% nel fumatore incallito. Il sangue trasporta così meno ossigeno ai tessuti, con le seguenti conseguenze: i tessuti sono parzialmente privati di ossigeno e il cuore deve lavorare con un impe-

gno maggiore, si ha un indurimento dei vasi sanguigni e la loro calcificazione, la pelle invecchia precocemente, i capelli si indeboliscono e il rendimento fisico cala.

Il filtro della sigaretta può parzialmente ridurre la quantità delle sostanze che arrivano nelle vie respiratorie, ma certo non è in grado di eliminarle: gli esperti raccomandano, di conseguenza, di **fumare la sigaretta fino a un massimo di 2/3**, essendo le prime boccate meno dannose grazie al fatto che il filtro e il tabacco assorbono la nicotina e il catrame. Nella parte finale della sigaretta il fumo di tabacco che viene inalato è così saturo di nicotina e di altri composti chimici da inondare le pareti polmonari con catrame liquido in una quantità tale da causare secrezioni maggiori di quelle avute durante il fumo degli altri 2/3 di sigaretta e il filtro non ha più capacità assorbenti, perché saturo e riscaldato.

L'elemento che determina assuefazione e che fa del fumo una delle tossicodipendenze più pericolose è la **nicotina**, un eccitante che accelera i battiti cardiaci e sottopone il cuore ad un maggior lavoro. Ma i suoi effetti sul sistema nervoso sono anche più duraturi: l'inalazione del fumo è un metodo straordinariamente efficace per apportare nicotina al cervello, dove si verificano alterazioni a livello psichico e si crea la dipendenza.

Dai polmoni la nicotina entra nel sistema circolatorio e raggiunge in pochi secondi il cervello, dove agisce sulla trasmissione degli impulsi nervosi attivando funzioni cerebrali legate all'attenzione, alla concentrazione e all'umore, favorendo il rilassamento e riducendo lo stress.

Dopo questo effetto iniziale, che dà alla mente e al corpo l'impressione di un potenziamento, subentra una fase di depressione che spinge ad accendere nuovamente una sigaretta e ad aumentare la dose di fumo per mantenere costante il livello di nicotina nell'organismo: questo è il segnale della **dipendenza da nicotina**. Il fumatore accanito, infatti, arriva ad accendere una sigaretta ogni 15-30 minuti, perché proprio questo è il tempo necessario per eliminare la dose di nicotina assunta. I meccanismi con i quali si instaura la dipendenza da nicotina non sono molto diversi da quelli di altre droghe: anzi, sembra che questo ingrediente della sigaretta abbia persino una capacità maggiore di indurre alla dipendenza rispetto ad altre sostanze.

In Italia i fumatori sono circa 12 milioni e sono attribuibili al tabacco 85.000 morti l'anno. In Europa un quindicenne su cinque fuma. I fumatori nel mondo sono circa un miliardo e centomila. Trecento milioni in Occidente e ottocento milioni nei Paesi non industrializzati. Nei Paesi industrializzati fuma il 41% degli uomini e il 21% delle donne. Nei Paesi del Terzo Mondo fumano il 50% degli uomini e l'8% delle donne.

Il fumo è responsabile del 30% di tutti i tumori, mentre l'alimentazione incide per il 35% e l'alcol per il 3%. Un fumatore, rispetto a un non fumatore, si sottopone al rischio di contrarre il cancro di 11 volte maggiore per il tumore al polmone; 7 volte maggiore per i tumori del cavo orale (bocca, faringe, esofago); 2 volte maggiore per quello alla vescica. E più si inizia a fumare presto, più i rischi aumentano.

a) Il fumo come fenomeno sociale

Se chiediamo a un fumatore quali sono i «benefici» del fumo, ci dirà che favorisce il rilassamento e la concentrazione, che attenua la tensione nervosa, che il gesto di accendere una sigaretta è di per sé soddisfacente. Nonostante le sempre più diffuse campagne antifumo, non esiste intorno al **tabagismo** la disapprovazione sociale che, invece, circonda l'uso di droghe e l'alcolismo. Anzi, il fumo viene considerato spesso un'abitudine che facilita le relazioni sociali. Alla sigaretta, per anni, sono stati attribuiti vari significati; spesso fumare è considerato un «rito», poiché segue una sequenza di movimenti obbligati, al punto di diventare, per molte persone, un calmante.

Nell'adolescenza si inizia a fumare per curiosità, per imitare i coetanei, per ostentare un atteggiamento che è tipico degli adulti, o semplicemente perché i genitori lo vietano. Se, infatti, l'immagine del fumatore viene percepita come attraente, l'adolescente accetta di conformarsi

ai coetanei fumatori, prova per curiosità, usa questo atteggiamento per trasgredire alle norme imposte dagli adulti e così via. Se, invece, l'adolescente (e poi l'adulto) modifica la percezione di questo modello, arrivando a valutare come positiva l'immagine del non fumatore, potrà smettere di fumare, facendo leva sul senso di indipendenza dal conformismo, sull'aumento dell'autostima, sul senso di benessere che ne deriva.

b) Normativa antifumo

È universalmente riconosciuto che il **fumo passivo** provoca gravissimi danni alla salute. In molti Paesi sono in corso campagne di sensibilizzazione che hanno lo scopo non solo di convincere i fumatori a smettere, ma soprattutto di tutelare i non fumatori. Un esempio delle iniziative antifumo è l'obbligo, imposto ai produttori di sigarette, di stampare sui pacchetti diciture del tipo «*Il fumo uccide*», intendendo con questa espressione che il rischio riguarda non solo i consumatori di tabacco, ma anche chi il fumo lo respira senza fumare.

Per tutelare i non fumatori la legge prevede misure sempre più restrittive. Nel nostro Paese è **vietato fumare** in uffici pubblici, ospedali e locali aperti al pubblico (pubblici e privati), compresi *bar, ristoranti, luoghi di svago, palestre e centri sportivi*, con la sola eccezione dei **locali riservati ai fumatori** e in casa propria. È quanto stabilito dall'art. 51 della **L. 16 gennaio 2003, n. 3**, che tutela la salute dei non fumatori e prevede, tra l'altro, multe per i trasgressori. L'importo della sanzione cresce se si fuma in presenza di **donne in gravidanza** o di **bambini** fino a dodici anni. Sono anche previste sanzioni per chi deve far rispettare il divieto. Sono chiamati a vigilare i titolari degli esercizi pubblici e il datore di lavoro. Quest'ultimo può delegare a terzi tale compito; in questo caso, i nomi delle persone preposte vanno segnalati sull'apposito cartello di divieto di fumo. La legge stabilisce che i ristoratori devono adibire ai non fumatori uno o più locali di superficie prevalente, rispetto alla superficie complessiva dell'esercizio, mentre non sono obbligati a destinare spazi attrezzati ai fumatori. Le **sale fumatori** sono soggette a pesanti restrizioni: devono essere ventilate separatamente, con specifici valori circa il tasso di ricambio dell'aria. La pressione atmosferica, al loro interno, deve essere costantemente più bassa di quella delle sale adiacenti. Devono, inoltre, essere munite di porte scorrevoli a chiusura automatica.

Ulteriori disposizioni sono contenute nel **Decreto Legislativo n. 6 del 12 gennaio 2016**, entrato in vigore il 2 febbraio 2016, che, recependo la **Direttiva europea 2014/40/UE**, riguarda le confezioni e i pacchetti di tabacco e sigarette, il divieto di vendita ai minori, le sigarette elettroniche con nicotina. Più dettagliatamente, esso concerne:

- l'introduzione sulle confezioni di sigarette, di tabacco da arrotolare e di tabacco per pipa ad acqua delle «**avvertenze combinate**» **relative alla salute**: testo, fotografia a colori e *numero verde* contro il fumo (800.554.088). Le avvertenze arrivano, così, ad occupare il 65% del fronte e del retro delle confezioni e dell'eventuale imballaggio esterno;
- il **divieto di additivi** (come caffeina, vitamine, coloranti delle emissioni, nonché additivi che facilitino l'inalazione o l'assorbimento di nicotina e che abbiano proprietà cancerogene, mutageniche o tossiche) che rendono più «attrattivo» e «nocivo» il prodotto del tabacco;
- l'abolizione dei pacchetti da 10 sigarette e delle confezioni di tabacco da arrotolare contenenti meno di 30 grammi di tabacco;
- il **divieto di utilizzare nell'etichettatura elementi promozionali e fuorvianti**, come riferimenti a benefici per la salute o per lo stile di vita, a un gusto o un odore etc.;
- il **divieto di apporre sulle etichette informazioni relative al contenuto di catrame, nicotina o monossido di carbonio**, ritenute ingannevoli per il consumatore che, nel confronto tra più prodotti, tende a preferire quello con minori quantità di tali sostanze, ritenendolo meno nocivo;

- il **divieto di «aromi caratterizzanti»** (odori o gusti chiaramente distinguibili, dovuti a un additivo o a una combinazione di additivi, ad esempio frutta, spezie, erbe) nelle sigarette e nel tabacco da arrotolare;
- il **divieto di vendita a distanza transfrontaliera** (cioè *on line*) ai consumatori di prodotti del tabacco, sigarette elettroniche e contenitori di liquido di ricarica con presenza di nicotina;
- il **divieto di vendita ai minori** dei prodotti del tabacco di nuova generazione, di sigarette elettroniche e di liquido di ricarica con presenza di nicotina, già precedentemente disposto da un'ordinanza del Ministro della Salute;
- il **divieto di fumo in autoveicoli in presenza di minori e donne in gravidanza**;
- il **divieto di fumo nelle pertinenze esterne degli ospedali e degli istituti di ricovero e cura**, nonché nelle pertinenze esterne dei singoli reparti pediatrici, ginecologici, di ostetricia e neonatologia;
- l'**inasprimento delle sanzioni per la vendita e la somministrazione ai minori** di prodotti del tabacco, sigarette elettroniche e prodotti di nuova generazione;
- la **verifica periodica dei distributori automatici**, al fine di controllare il corretto funzionamento dei sistemi di rilevamento dell'età dell'acquirente;
- l'introduzione di **requisiti di sicurezza per le sigarette elettroniche** e i contenitori di liquido di ricarica con nicotina, a prova di bambino e di manomissione, corredati da un «foglietto illustrativo» recante istruzioni d'uso, controindicazioni, informazioni su eventuali effetti nocivi etc.

3 I rischi dell'alcolismo (o etilismo)

Il termine *alcolismo* indica la dipendenza dall'uso di bevande alcoliche e gli effetti da questa provocati. L'**OMS** consiglia di sostituire il termine alcolismo con l'espressione **Sindrome da Dipendenza Alcolica** (SDA), specificando che *«la dipendenza alcolica esiste quando il consumo di alcol in una persona supera i limiti che sono accettati dalla sua cultura, se egli consuma alcol in orari che sono considerati inappropriati nella sua cultura o il suo consumo di alcol diviene così grande da danneggiare la sua salute o da disturbare le sue relazioni sociali»*.

Il **vino** è costituito da moltissime sostanze, fra le quali alcune in grado di inibire lo sviluppo di cellule cancerogene, la formazione di radicali liberi e di attivare un enzima, la MAP-chinasi (*mitogen activated protein*), che agisce positivamente nel cervello sul meccanismo della memoria e dell'apprendimento. Naturalmente, gli effetti benefici si ottengono consumandone modiche quantità.

La *quantità di alcol che si può assumere senza danni* corrisponde a circa 3 bicchieri di vino al giorno (400 ml), considerando che un bicchiere medio ha una capienza di 120 ml e contiene all'incirca 80 calorie. La stessa quantità di alcol viene introdotta con una coppa di spumante (100 ml, circa 60-70 calorie), un boccale medio di birra (330 ml, circa 50-100 calorie), un bicchierino di grappa o altro liquore a 40° (40 ml, circa 90 calorie). Quando si superano queste quantità, l'alcol diventa una sostanza dannosa per l'organismo.

L'alcolismo non è solo un problema che riguarda la salute: è soprattutto una **piaga sociale** se ne consideriamo la ricaduta sulla vita affettiva, familiare e lavorativa. Le cosiddette «*stragi del sabato sera*» sono, nella quasi totalità dei casi, dovute allo stato di ebbrezza dei giovani automobilisti all'uscita dalle discoteche. Il limite legale di alcol, compatibile con la guida, è di 0,5 grammi per litro di sangue. Ma trovarsi alla guida di un'auto o di una moto dopo aver bevuto alcolici è pericoloso anche se le quantità non sono elevate, perché una piccola dose di alcol, pur non compromettendo la capacità di condurre un veicolo, diminuisce la percezione del rischio e rallenta i riflessi.

a) Cause dell'alcolismo

L'alcolismo era considerato fino a non molto tempo fa un *vizio*, cioè un comportamento volontario negativo e quindi moralmente riprovevole. Con il progresso delle conoscenze scientifiche sul problema si è compreso, invece, che la dipendenza dalle bevande alcoliche è dovuta non tanto alla mancanza di volontà del soggetto, ma ad una serie di fattori molto più complessi e collegati tra loro, che si possono così classificare:

- **fisici**, ovvero problemi genetici, metabolici, neurologici;
- **psichici**, ossia disturbi psichici di varia natura che provocano sofferenza e depressione, spingendo a cercare conforto nell'alcol;
- **sociali**, vale a dire cultura del bere, pressione sociale, abitudini e stili di vita dettati dalle mode e dal consumismo.

Ciascuno di questi fattori, da solo, non è in grado di determinare il disturbo. Perché il problema si manifesti in modo patologico è necessario che siano contemporaneamente presenti più fattori predisponenti e che sopravvenga una causa occasionale scatenante (una disgrazia, una delusione etc.). Una volta che il disturbo è apparso, il suo mantenimento sarà reso possibile da altri fattori, detti *perpetuanti*.

b) Effetti dell'alcolismo

Le bevande alcoliche possono essere a bassa o ad alta gradazione. Nel primo caso, l'alcol che contengono deriva dalla *fermentazione* di sostanze organiche come il mosto di uva per il vino o l'orzo, aromatizzato con il luppolo, per la birra; nel secondo caso, invece, l'alcol è il risultato di un processo di *distillazione*. Con questo sistema si ricavano la grappa dalle vinacce, il whisky dal malto, il rum dalla canna da zucchero, il gin dal ginepro, la vodka dalla segale.

Gli **effetti acuti** dell'alcol dipendono dalla gradazione della bevanda, dalla quantità introdotta, dal peso, dall'età e dal sesso del bevitore; consistono, inizialmente, in un'azione depressiva sul sistema nervoso centrale, con alterazioni del comportamento (scarso autocontrollo, aggressività), nausea, mal di testa. Successivamente si verifica una riduzione della vigilanza, della vivacità e dei riflessi, con peggioramento della coordinazione neuro-muscolare, percezione sbagliata delle distanze, aumento dell'abbagliamento notturno, riduzione del campo visivo, allungamento dei tempi di reazione.

Gli **effetti cronici** dell'alcol provocano danni gravissimi, quali:

- diminuzione delle difese immunitarie, con maggiore incidenza di malattie respiratorie;
- alterazioni gravi e irreversibili al fegato e alle mucose gastro-intestinali;
- disturbi neurologici;
- malattie cardiovascolari;
- rischio di tumori (orofaringe, laringe, esofago, fegato);
- in caso di gravidanza, mancato sviluppo del feto.

L'*alcolista*, o *etilista*, tende a negare la dipendenza ed è molto raro che chieda aiuto spontaneamente. Di solito è il medico curante, constatando i danni organici provocati dall'alcol, il primo a intervenire. Bisognerebbe, pertanto, attuare un'azione preventiva ed educativa sui soggetti a rischio (soprattutto in età evolutiva), mediante informazioni corrette e dettagliate sui danni diretti e indiretti causati dall'etilismo, sulla facilità con cui dall'uso dell'alcol si può passare al suo abuso, nonché all'assunzione di altre sostanze in grado di causare dipendenza.

4 La tossicodipendenza

L'Organizzazione Mondiale della Sanità classifica come *droga* qualsiasi sostanza che abbia **potere psicoattivo**, capace cioè di provocare effetti sul sistema nervoso dando assuefazione e dipendenza sia psichica sia fisica. L'uso delle sostanze stupefacenti tra i giovani è in crescita costante, con le conseguenze devastanti che tutti conosciamo e che non riguardano solo la salute. Per procurarsi una dose, infatti (e questo vale soprattutto per le sostanze stupefacenti costose, come l'eroina e la cocaina), molti giovani sono costretti a rubare, a prostituirsi o a diventare spacciatori, incrementando ulteriormente un mercato clandestino gestito dalla criminalità organizzata.

La droga è una sostanza capace di alterare gli equilibri dei processi biochimici e psicologici dell'organismo, interferire sulla trasmissione e sull'elaborazione degli impulsi nervosi, annullare la capacità di adattamento all'ambiente esterno compromettendo seriamente le dinamiche relazionali e l'inserimento sociale.

a) Processo di assuefazione

La droga produce, in genere, una marcata sensazione di benessere e di euforia. Ma interrompendone l'assunzione o mantenendone costante la dose, via via, gli effetti piacevoli svaniscono e compaiono i sintomi della cosiddetta **sindrome da astinenza**, con lacrimazione, gocciolamento di muco dal naso, sudorazione, diarrea, battito cardiaco accelerato e spesso irregolare, febbre e insonnia.

A questo punto, il soggetto è diventato assuefatto e dipendente dalla sostanza, per cui, per riprovare benessere ed euforia, dovrà aumentare ulteriormente la dose e così via, intrappolato in un circolo vizioso. Ma i sintomi d'astinenza saranno sempre più gravi, il bisogno della droga sempre più intenso e, continuandone l'abuso, oltre alla necessità di aumentare sempre più la dose, si giungerà ad un punto in cui l'euforia non sarà più raggiungibile, mentre sarà sempre più grave il malessere.

b) Droghe

A seconda degli effetti che producono sul sistema nervoso centrale, le droghe vengono classificate in diversi gruppi.

c) Stupefacenti

Gli stupefacenti comprendono l'**oppio** e i suoi derivati. L'oppio è estratto dal papavero sonnifero e contiene 20 tipi di alcaloidi fra cui la **morfina**, dalla quale viene ricavata l'**eroina**. Questa polvere bianca, iniettata nelle vene, provoca dipendenza fisica, crisi di astinenza e dimagrimento a vista d'occhio. L'*overdose*, spesso favorita dai prodotti usati per il «taglio» o assunti per amplificare l'effetto dell'eroina, provoca coma da oppioidi, con depressione respiratoria e cardiovascolare.

d) Stimolanti

Gli stimolanti comprendono principalmente la **cocaina**, alcaloide estratto dalle foglie di coca, una pianta originaria dell'America Meridionale, ma anche le amfetamine, il tabacco, il caffè, il tè.

La cocaina si assume iniettandola in vena, inalandola o fumandone le foglie essiccate. In una prima fase, stimolando il sistema nervoso centrale, la cocaina produce euforia, aumento della percezione, dell'attenzione e della forza fisica, resistenza al sonno e alla fame. Queste sensazioni però durano poco e sono seguite da una fase (stato di *down*) di ansia e depressione, che induce alla dipendenza. Il consumo abituale di cocaina provoca convulsioni, psicosi, malattie cerebrali (ictus) e cardiovascolari (tachicardia, ipertensione, aritmie); l'intossicazione acuta può essere letale.

Le **amfetamine** (o *anfetamine*), nate come farmaci contro l'asma (broncodilatatori), producono un miglioramento delle prestazioni fisiche e un effetto euforizzante, con eliminazione della fatica, del sonno e dell'appetito. I rischi vanno dal semplice tremore al colpo di calore, fino a eventi cardiovascolari simili a quelli indotti dalla cocaina.

e) Sedativi o ipnotici

Rientrano nella categoria dei sedativi (o ipnotici) gli **psicofarmaci**, che riducono l'ansia e/o inducono il sonno, e l'**alcol**.

f) Inebrianti

Gli inebrianti comprendono sostanze naturali e altre di uso industriale come eteri, solventi, colle. Fra le sostanze naturali, la più conosciuta è la *canapa indiana*, pianta tropicale da cui si estraggono la **marijuana** (miscela di foglie, fiori e steli) e l'**hashish** (resina estratta dal polline). L'hashish e la marijuana producono un senso di distacco dal mondo, sensazioni di benessere, euforia, sonnolenza, disturbi dell'orientamento e della vista, talvolta illusioni e allucinazioni. Provocano **dipendenza non fisica ma psicologica**.

g) Allucinogeni

Fra gli allucinogeni naturali troviamo la **mescalina**, estratta dal cactus peyote, originario del Centro America; la *psilocibina* e la *psilocina*, ricavate da un fungo sudamericano. Fra quelli sintetici troviamo l'**LSD**, l'acido lisergico, scoperto in laboratorio nel 1943 dallo scienziato Albert Hoffmann, e l'**ecstasy**. Quest'ultima sostanza è commercializzata sotto forma di pasticche colorate che, negli ultimi anni, sono diventate la tipica «droga da discoteca» ma, in realtà, erano già usate durante la prima guerra mondiale, quando venivano somministrate ai soldati per superare la fame e il sonno e per la loro potente azione socializzante. I consumatori abituali di allucinogeni hanno disturbi di apprendimento e di memoria, confondono passato, presente e futuro, mostrano demotivazione, apatia, scarso interesse per la vita e anche tachicardia, astenia, occhi iniettati di sangue, disturbi respiratori.

h) Nuove droghe

Il panorama delle sostanze stupefacenti si arricchisce costantemente di nuovi prodotti, alcuni estratti da piante, altri di origine animale, altri ancora da sostanze chimiche altamente tossiche. Di alcune forniamo una breve descrizione.

BROWN - È un liquido da iniettare, di colore giallo o marrone scuro. Produce un effetto di piacere e inibisce il senso di fame. Viene prodotto soprattutto nella Repubblica Ceca.

BUFO - Diffuso soprattutto negli Stati Uniti, è un alcaloide estratto dal liquido delle ghiandole del rospo Bufo e di alcune varietà di pesci. Contiene numerosi principi attivi, tra cui l'*adrenalina* e la *bufotenina*, un allucinogeno che provoca effetti eccitanti simili a quelli delle amfetamine.

CHAT - È una pianta coltivata nel Corno d'Africa e nello Yemen, regioni in cui è usanza comune masticarne le foglie fresche. Il suo principio attivo, il *catinone*, resiste nella pianta estirpata da meno di 48 ore. Per questa ragione, lo spaccio è concentrato attorno agli aeroporti internazionali, soprattutto nel Nord Italia. A consumarla sono perlopiù gli immigrati.

CRACK - Molto diffuso negli Stati Uniti, è un *derivato della cocaina* che viene fumato e ha un effetto molto più immediato e forte della polvere bianca.

D.M.T. - Sigla della *dimetiltriptamina*, sostanza cristallina inodore che si estrae dai semi di alcune piante delle Antille e dell'America del Sud. Si ottiene anche per sintesi in laboratorio e si assume per via inalatoria. Nell'isola caraibica di Haiti è nota con il nome di *cohoba*.

DROGHE DEI POVERI - In Brasile, in alcuni Paesi asiatici e da qualche tempo anche nell'Est europeo sono diventate «droghe dei poveri», per il loro basso costo, alcuni *collanti* e prodotti chimici generalmente utilizzati nella lavorazione artigianale dei pellami. Si assumono per inalazione.

EFEDRINA - Alcaloide estratto, fin dall'inizio del Novecento, dall'*ephedra*, una pianta coltivata per uso farmaceutico contro alcune malattie delle vie respiratorie e patologie cardiovascolari. Oggi si produce in laboratorio in alcuni Paesi dell'Est Europa con un procedimento meno costoso dell'estrazione dalla pianta. Poiché ha effetti simili a quelli dell'adrenalina, è utilizzata anche per il *doping*. Provoca gravi danni al sistema nervoso centrale.

PEYOTIL - Si ricava da una specie di cactus che cresce spontaneamente nelle zone desertiche del Messico. Conosciuto fin dall'antichità, era utilizzato dalle popolazioni indigene durante le cerimonie rituali per i suoi effetti allucinogeni, visivi e uditivi: contiene, infatti, mescalina, un'alcaloide generalmente assunto per via orale.

PERVITIN - È una polvere bianca, solubile in acqua, da iniettare. Produce sensazione di benessere, euforia, carica energetica, perdita del senso di fame e di sete, resistenza al sonno.

PSILOCIBINA - È un allucinogeno che si estrae da alcune varietà di funghi che si trovano soprattutto in Messico, dove era usato dagli sciamani della regione di Oaxaca.

i) Terapia e reinserimento sociale

Le modalità di recupero dei tossicodipendenti (in particolare gli eroinomani) possono essere distinte in quattro grandi categorie di trattamento:

- **sostitutivo** (metadone);
- **con antagonisti** (naltrexone);
- **psicoterapeutico**;
- **residenziale** (psicoterapia in ambiente protetto, ricovero in comunità terapeutiche).

Gli **psicologi** sono coinvolti da anni nel lavoro di ricerca per la comprensione, la prevenzione e il recupero della tossicodipendenza. Sono stati analizzati, tra l'altro, gli aspetti psicologici e la dinamica delle varie fasi del rapporto dell'individuo con la droga.

j) Doping

Il termine *doping*, di origine inglese, indica l'assunzione volontaria da parte dell'atleta — o la somministrazione da parte di altri — di **sostanze proibite** dai regolamenti sportivi e dalla legge, allo scopo di **incrementare le prestazioni agonistiche**. Quello del doping è un problema etico che parte da un'interpretazione errata del concetto di gara sportiva, perché il rispetto delle regole è il principio su cui si fonda ogni disciplina agonistica. Il ricorso a mezzi illeciti per incrementare la prestazione è inaccettabile dal punto di vista etico perché contrasta con le regole sportive, creando delle disuguaglianze fra i concorrenti e vanificando il significato della competizione.

Sotto il profilo legislativo con l'entrata in vigore della Legge 14/12/2000, n. 376, ovvero *Disciplina della tutela sanitaria delle attività sportive e della lotta contro il doping*, quest'ultimo non è più solo un reato sportivo ma anche penale. L'assunzione e la somministrazione di sostanze dopanti è punibile con la reclusione da tre mesi a tre anni e con multe fino a 50mila euro. La ricettazione e il traffico di sostanze dopanti è, invece, un reato che prevede la reclusione da due a sei anni e multe fino a 75mila euro. I *controlli antidoping* vengono eseguiti da una Commissione appositamente istituita presso il Ministero della Salute. Le sanzioni sportive nei confronti degli atleti trovati positivi, o che si rifiutano di sottoporsi ai controlli, spettano al CONI e alle Federazioni Sportive.

La malattia come rottura dell'equilibrio biologico e psicologico

1 Il complesso equilibrio dell'organismo umano

Il nostro organismo è sempre in un **equilibrio dinamico**. Ciò significa che il corpo umano, come quello di qualsiasi altro essere vivente, è in una situazione di instabilità e che solo grazie a continui aggiustamenti e modifiche riesce a sopravvivere alle modificazioni e alle sollecitazioni dell'ambiente interno ed esterno.

Ad esempio, se ad una persona sopravviene uno svenimento che è determinato da una interruzione momentanea di attività di aree specifiche del cervello, con istantanea carenza di sangue all'encefalo, l'organismo reagisce rientrando rapidamente nelle normali condizioni di funzionamento. Questo perché ogni organismo ha continui **scambi di sostanze chimiche** con l'esterno: *cioè è un sistema aperto in equilibrio con l'ambiente che lo circonda* e ricostruisce il proprio equilibrio interno laddove questo sia stato alterato da eventuali cambiamenti esterni. Gli esseri viventi hanno, quindi, dei **meccanismi di autocontrollo** tali da impedire che un disturbo esterno determini l'alterazione dell'equilibrio organismo-ambiente.

a) Concetto di salute

L'uomo, nel corso della storia, ha sempre cercato di migliorare le proprie condizioni di vita ma, in realtà, è riuscito perlopiù a *prolungare l'aspettativa di vita, senza migliorarne la qualità*. Nella società attuale è andata, tuttavia, affermandosi la convinzione che nella gestione della salute debba prevalere il momento di *prevenzione* e di *riabilitazione* su quello di cura o, per meglio dire, di *medicalizzazione*. Oggi, la *salute* va considerata alla stregua di un *bene da proteggere*.

Mettere in atto *misure di prevenzione e di salvaguardia ambientale*, evitando che i componenti di una comunità si ammalino, è compito altamente etico ed è, oltretutto, un modo corretto di impedire aggravi economici alla collettività; inoltre, comporta certamente minor danno, anche psicologico, in quanto evita le ripercussioni — sull'individuo e sulla società — del verificarsi della malattia.

D'altra parte, la stima dello stato di salute di un individuo — così come di una popolazione — non può prescindere da una valutazione delle condizioni del territorio. La salvaguardia della salute, o il suo recupero, non può essere quindi limitata all'intervento del medico nel caso di infermità o all'inserimento del malato in una struttura sanitaria, ma deve coinvolgere l'uomo nella sua globalità di interessi e di partecipazione.

Educare alla salute, in quest'ottica, significa anche aiutare a **scoprire il valore di un bene** che richiede impegno e controllo di sé e degli altri, nonché conoscenza delle regole, dei comportamenti e degli stili di vita che influiscono sulla salute di tutti i cittadini. La convinzione che i problemi della salute, la programmazione e la gestione della realtà sanitaria di un Paese siano aspetti prioritari delle attività degli organi di governo ha determinato, in Italia, la realizzazione del **Servizio Sanitario Nazionale (SSN)**, che ha compiti di **prevenzione, diagnosi, cura e recupero** nei confronti delle infermità che possono colpire il cittadino e di gestione del territorio,

soprattutto in riferimento alla prevenzione della degradazione dell'ambiente e al risanamento delle zone inquinate.

I problemi inerenti alla salute, che afferiscono all'ambito disciplinare della *Sociologia della salute*, sono diventati negli ultimi anni di stringente attualità e stanno più o meno confusamente emergendo nel nostro Paese in concomitanza con l'attuazione del Servizio Sanitario Nazionale, che ha in parte rivoluzionato i termini del rapporto *cittadino-salute*. Se possedere adeguati strumenti e capacità di difesa della salute, nonché di prevenzione e cura delle malattie è un diritto-dovere di ogni cittadino, è pur vero che per raggiungere tale fine è necessario l'impegno politico, culturale, tecnico e finanziario del Paese.

Sempre più spesso, infatti, vengono considerate concause di molte patologie i condizionamenti e le strutture sociali; pertanto, la realizzazione di uno stato di *non malattia* significa prevenire le cause di malessere, interpretando i bisogni sanitari degli individui o intervenendo nell'anticipare, con ipotesi e proposte, la richiesta delle popolazioni stesse, tesa ad ottenere una migliore *qualità della vita*.

Si può senz'altro dire che le condizioni ambientali e sociali possono influenzare il sopraggiungere di talune patologie: basti pensare alle influenze socio-ambientali evidenti nell'insorgere delle malattie psichiatriche, al rapporto tra condizioni ambientali e sviluppo psico-fisico del bambino, alle malattie professionali legate a particolari ambienti di lavoro, alla carenza di strutture protettive indispensabili per taluni procedimenti industriali, al degrado ambientale, alle carenze igieniche etc.

b) Concetto di malattia

I concetti di *salute* e di *malattia*, così come la valutazione del rapporto tra il medico e il paziente, si sono andati evolvendo nel corso degli anni in seguito alla pressione culturale esercitata dalle Scienze umane e sociali. Come abbiamo visto, oggi è da intendersi «malato» chi *non si sente sano*: la malattia corrisponde, cioè, alla *perdita della salute*; quest'ultima è finalmente considerata *in positivo*, e cioè come *un diritto da salvaguardare*.

La medicina, intesa come *scienza della salute*, deve approfondire le questioni relative alla prevenzione e alle complesse interazioni tra ambiente, condizioni di lavoro e struttura della società, intesa come comunità di individui. La prevenzione non deve essere considerata un'azione esclusivamente sanitaria (*screening* o *depistage* di massa, vaccinazioni, diagnosi precoci etc.), ma deve abbracciare interventi multidisciplinari che vanno dall'individuazione dei bisogni reali dell'individuo alle scelte cui orientare la popolazione tutta; il focus va quindi spostato: *dall'anatomia dell'uomo all'anatomia della società*.

La **malnutrizione** spesso è concausa di alcune malattie (obesità, diabete, ipertensione etc.); l'eccesso di comfort determina, ad esempio, il *sedentarismo*; lo smog e — paradossalmente — l'eccesso di igiene favoriscono l'insorgere delle allergie; la corsa al successo e al guadagno è causa di stress angosciosi e di tensioni aggressive. Vengono trascurati, invece, *bisogni essenziali*, come condurre una vita sana, disporre di strutture sociali di assistenza e di ricreazione, godere del tempo libero con attività distensive e piacevoli, intessere rapporti interpersonali tesi alla valorizzazione delle reciproche personalità. Rivolgersi al medico soltanto per eliminare i sintomi attraverso le prescrizioni comporta l'*abuso di farmaci* e la conseguente *farmacodipendenza*.

2 Le malattie metaboliche

Il *metabolismo* è l'insieme di tutti i processi fisici e chimici per cui una sostanza viene organizzata, prodotta e mantenuta (*anabolismo*) oppure trasformata per produrre energia e renderla disponibile per le necessità dell'organismo (*catabolismo*). I processi metabolici sono controllati

da due fattori integrati: i **geni**, che definiscono le funzioni di ogni data cellula, e l'**ambiente**, che determina le modalità con cui i geni si esprimono. Tutti i disordini metabolici originano da qualche disturbo nell'interazione tra fattori genetici e ambientali. In base alle modalità di trasmissione genetica, le malattie metaboliche possono essere divise in **malattie metaboliche ereditarie** e **malattie metaboliche ad eziologia multifattoriale**.

Attualmente si conoscono centinaia di malattie metaboliche ereditarie o «**errori congeniti del metabolismo**», come definite originariamente dal medico inglese Archibald Garrod.

Per *errore congenito del metabolismo* si intende un'alterazione biochimica geneticamente determinata in cui un difetto enzimatico specifico determina un blocco metabolico che può comportare conseguenze patologiche alla nascita o in una fase successiva della vita. Tale gruppo di affezioni può interessare tutte le fasi del metabolismo e ha contribuito alla conoscenza delle vie metaboliche normali. Gli elementi che le diverse patologie metaboliche hanno in comune tra loro sono l'**ereditarietà** e l'**anormalità funzionale di una proteina specifica**, mentre differiscono enormemente per incidenza, prevalenza razziale o etnica, manifestazioni cliniche e letalità. Nonostante negli ultimi anni la ricerca scientifica abbia fatto considerevoli progressi, molto resta ancora da scoprire su tali malattie. In genere le mutazioni responsabili dei disordini metabolici ereditari colpiscono i geni strutturali che codificano la struttura primaria delle proteine, oppure le mutazioni possono modificare la velocità con cui una proteina viene costruita. Sono stati descritti errori congeniti per tutti i tipi di proteine, ma principalmente: **alterazioni degli enzimi intracellulari, disordini ereditari dei costituenti della membrana cellulare, alterazioni delle proteine circolanti**. Infatti, errori congeniti delle proteine di trasporto, riguardanti l'intestino e i reni, possono selettivamente ostacolare gli spostamenti attraverso le membrane degli zuccheri, degli aminoacidi, dei fosfati, delle vitamine o dell'acqua (cistinuria, glicosuria); altri difetti di trasporto portano ad un legame anormale di macromolecole con i recettori di membrana (ipercolesterolemia familiare) oppure a legami anomali dei complessi proteino-fissatori; altre mutazioni ancora alterano le proteine circolanti (analbuminemia, abetalipoproteinemia). La maggior parte degli errori congeniti è associata con una diminuita attività (o contenuto) di una proteina. La carenza può essere completa o parziale. La maggior parte degli **errori enzimatici congeniti** è caratterizzata dalla carenza parziale dell'attività.

Tale carenza può derivare da differenti meccanismi, vale a dire da:

- una ridotta sintesi delle molecole enzimatiche;
- un'accelerata distruzione di un enzima strutturalmente alterato;
- una ridotta affinità degli enzimi attivi per i substrati o i cofattori.

Gli **errori congeniti del metabolismo** possono essere suddivisi come segue:

- disordini ereditari del metabolismo degli aminoacidi (fenilchetonuria, alcaptonuria);
- difetti ereditari del trasporto di membrana (cistinuria, M. di Hartnup, sferocitosi);
- disordini del metabolismo degli acidi nucleici (iperuricemia, gotta);
- disordini dei metalli e delle metallo-proteine (emocromatosi, M. di Wilson, porfirie);
- disordini del metabolismo lipidico (iperlipoproteinemie);
- malattie da accumulo lisosomiale;
- disordini del metabolismo dei carboidrati (glicogenosi, galattosemia, diabete mellito).

Le malattie metaboliche più frequenti nella popolazione generale e, quindi, nei riguardi delle quali vanno attuati programmi di educazione alla salute e di prevenzione, sono: il *diabete mellito*, le *iperlipoproteinemie* e la *gota*. Nella scuola, l'educazione alla prevenzione delle malattie metaboliche deve essere attuata mediante **programmi di educazione alimentare e di educazione all'esercizio fisico**, condotti contemporaneamente all'illustrazione scientifica delle cause e delle conseguenze di tali patologie.

a) Diabete mellito

Il **diabete mellito** rappresenta il più comune disturbo metabolico: la sua frequenza nella popolazione generale si aggira, infatti, intorno al 2-6%. Da un punto di vista clinico, si tratta di una sindrome complessa determinata da una lenta, ma inesorabile, **distruzione delle cellule b delle isole pancreatiche**, conseguenza di interazioni fra molteplici fattori: predisposizione genetica, fattori ambientali e reazioni immunitarie cellulo-mediate. Il risultato è una deficienza insulinica, prima relativa e poi assoluta, che si instaura in un periodo di tempo variabile. L'iperglicemia che consegue al deficit insulinico determina, se non ben controllata, gravi danni vascolari: *microangiopatia diabetica* (retinopatia, nefropatia, neuropatia) e *macroangiopatia diabetica* (cardiomiopatia). La malattia si manifesta con poliuria, sete, aumento dell'appetito, dimagrimento oppure obesità, iperglicemia e glicosuria. Il diabete giovanile può anche esordire improvvisamente, con il quadro del coma cheto-acidosico.

Clinicamente il diabete viene classificato in:

- diabete mellito tipo I o insulino-dipendente (10%);
- diabete mellito tipo II o non-insulino-dipendente o dell'età adulta (85%);
- diabete secondario (endocrinopatie, malattie del pancreas, alterazioni genetiche);
- diabete MODY (*Maturity Onset Diabetes of the Young*), che si manifesta in età giovanile;
- diabete chimico o intolleranza ai carboidrati;
- diabete gestazionale (compare in gravidanza).

L'esistenza del diabete mellito può essere sospettata in base alla presenza di alcuni sintomi (astenia, dimagrimento, polifagia, polidipsia, impotenza sessuale) e alla comparsa di manifestazioni riferibili alla microangiopatia (retinopatia, aree cutanee di necrosi, proteinuria), soprattutto se vi è presenza di familiarità diabetica. Questa caratteristica sintomatologia corrisponde però alla fase conclamata della malattia, che segue ad una fase di latenza, nella quale la scoperta della malattia è spesso casuale.

Nella fase di latenza sono gli esami di laboratorio che svelano o confermano l'esistenza e la gravità di un'intolleranza al glucosio. Le indagini di laboratorio fondamentali per accertare e controllare un diabete nel tempo sono:

- ricerca della glicosuria;
- determinazione della glicemia basale e del profilo glicemico;
- determinazione della chetonuria;
- curva da carico di glucosio;
- test dell'emoglobina glicosilata.

Il trattamento consiste nel combinare una dieta priva di zuccheri semplici a ipoglicemizzanti orali nel diabete di tipo II e alla terapia insulinica nel diabete di tipo I.

b) Iperlipoproteinemie

Le **iperlipoproteinemie** sono **disturbi del trasporto lipidico** dovuti ad anomalie nella sintesi o degradazione delle lipoproteine plasmatiche. Alcune iperlipoproteinemie vengono definite *primarie* perché sono il diretto risultato dei difetti del metabolismo delle particelle lipoproteiche; altre, invece, vengono dette *secondarie*, in quanto l'elevato livello plasmatico delle lipoproteine compare come conseguenza di anomalie causate da un concomitante disordine in un sistema metabolico collegato (ad esempio, deficit tiroideo o di insulina). Le iperlipoproteinemie rappresentano un fattore di rischio «maggiore» per l'arteriosclerosi in generale e per la malattia coronarica in particolare.

Le iperlipoproteinemie **primarie** possono essere divise in due vaste categorie:

- disordini di un gene unico, che vengono trasmessi con un semplice meccanismo dominante o recessivo;
- disordini multifattoriali con modalità complesse di eredità, in cui molteplici geni varianti interagiscono con fattori ambientali per produrre vari gradi di iperlipoproteinemia nei membri di una famiglia.

Le forme **secondarie** possono essere trovate associate ad un enorme numero di malattie (Sindrome di Cushing, ipotiroidismo); tuttavia, le forme che più frequentemente si incontrano ed hanno importanza da un punto di vista sociale ed epidemiologico sono le iperlipoproteinemie collegate al diabete mellito, all'obesità, all'uso di alcol e di contraccettivi orali.

Le *iperlipoproteinemie* sono state **classificate** da Frederickson in **5 tipi**, sulla base della migrazione elettroforetica, ovvero:

- **Tipo I:** aumento dei chilomicroni e dei trigliceridi;
- **Tipo II:** aumento del colesterolo, delle lipoproteine beta e dei trigliceridi;
- **Tipo III:** iperlipemia essenziale (larga banda beta);
- **Tipo IV:** aumento dei trigliceridi e delle lipoproteine beta;
- **Tipo V:** iperlipidemia mista.

Da un punto di vista clinico, le iperlipidemie, in linea di massima, si manifestano con xantomi cutanei e tendinei e con lesioni precoci aterosclerotiche, sia periferiche sia coronariche. Tali disordini metabolici sono aggravati dalla concomitanza di ipertensione arteriosa e dal fumo di tabacco.

c) Gotta e altri disordini del metabolismo purinico

«**Gotta**» è un termine che rappresenta un eterogeneo gruppo di malattie che si manifestano esclusivamente negli esseri umani e che, nel loro completo sviluppo, si presentano con:

- aumento sierico della concentrazione uratica ($>7\text{mg}/100\text{ml}$);
- attacchi ricorrenti di un tipo caratteristico di artrite acuta, in cui nei leucociti del liquido sinoviale sono dimostrabili cristalli di urato di sodio monoidrato;
- depositi di aggregati di urato monosodico monoidrato (*tofi*) nelle articolazioni delle estremità, che talvolta producono gravi limitazioni e deformità;
- una malattia renale che interessa il tessuto interstiziale e i vasi sanguigni;
- nefrolitiasi (calcolosi renale) da acido urico.

Queste manifestazioni possono essere presenti isolatamente o contemporaneamente. La prevalenza della gotta nella popolazione generale è inferiore a quella della iperuricemia ed è grandemente variabile nel mondo; tale patologia si manifesta principalmente negli uomini e la sua punta di incidenza è nella quinta decade di vita. Gli studi genetici suggeriscono per tale malattia un'ereditarietà poligenica. L'iperuricemia della gotta può essere dovuta ad un eccessivo ritmo di produzione di acido urico, ad una diminuzione nell'escrezione renale di acido urico o ad un'associazione di entrambe le modalità.

Nel corso della sua **storia naturale**, la gotta passa attraverso quattro stadi:

- iperuricemia asintomatica;
- artrite gottosa acuta;
- gotta intercritica;
- gotta tofacea cronica.

La nefrolitiasi può manifestarsi in ogni stadio al di fuori del primo. Condizioni spesso associate sono l'*obesità*, l'*ipertrigliceridemia*, l'*ipertensione*. Le finalità terapeutiche nella gotta consistono

nel risolvere l'attacco acuto al più presto con farmaci antinfiammatori; nel prevenire le recidive dell'artrite e le complicazioni della malattia mediante farmaci antiuricemici, come l'*allopurinolo*.

3 Le malattie degenerative

Le malattie degenerative sono rappresentate da un gruppo di patologie che hanno come caratteristiche principali:

- la **cronicità**;
- l'**inguaribilità**.

Il declino dell'incidenza della patologia infettiva, dovuto al miglioramento delle condizioni igienico-sanitarie e all'introduzione dell'*antibiotico-terapia* e della *vaccinoprofilassi*, ha fatto assumere alle malattie cronico-degenerative le caratteristiche di un problema di rilevanza socio-economica. Tale problema incide in: clinica, epidemiologia, medicina preventiva e medicina delle comunità.

Attualmente le patologie cronico-degenerative rappresentano, infatti, la più frequente causa di:

- disabilità;
- assenza dal lavoro per malattia;
- diminuzione e compromissione della qualità di vita;
- mortalità.

Rientrano in questo gruppo, le malattie:

- non sostenute da microrganismi;
- a decorso prolungato con fasi di remissione e recidive;
- ad eziologia multifattoriale;
- con forte impronta sociale (tossicodipendenza e alcolismo);
- con conseguenze a lungo termine di fenomeni acuti (incidenti).

Di particolare importanza, per queste patologie, è la **prevenzione**, da attuare come:

- prevenzione **primaria**, mediante l'individuazione e la prevenzione dei fattori di rischio;
- prevenzione **secondaria**, attraverso l'individuazione in fase precoce della malattia;
- prevenzione **terziaria**, tramite la cura, la riabilitazione e la prevenzione delle possibili recidive.

L'**agente causale** di tali malattie può essere: chimico, fisico e sociale.

I **fattori di rischio** possono essere:

- *genetici* (ereditarietà e familiarità);
- *ambientali* (occupazionali e abitudini personali).

L'**osteoartrosi** è la malattia degenerativa più frequente nella popolazione adulta e che maggiormente incide in termini di disabilità e di spesa sanitaria. È la malattia più comune in reumatologia, costituendo, da sola, circa il 50% di tutte le patologie osteoarticolari. È espressione di un processo degenerativo della cartilagine articolare, determinato dall'azione di enzimi litici prodotti dai condrociti, dai sinoviociti e dai granulociti, e seguito dall'esposizione dell'osso sub-condrale con la formazione conseguente di *osteofiti*. La denominazione *artropatia degenerativa* sottolinea la natura essenziale di quest'affezione.

L'artrosi, secondo la Società Italiana di Reumatologia, può essere classificata, in base alle cause che l'hanno determinata, in:

1. **artrosi primaria:**

- generalizzata,
- localizzata (ad esempio: discartrosi, gonartrosi, coxartrosi primaria);

2. artrosi secondaria:

- a fattori generali (ad esempio: l'obesità),
- a fattori locali (ad esempio: cifosi, osteocondriti, artriti),
- a traumi (ad esempio: idroartrosi e meniscoli, secondarie a traumi professionali e no);

3. artrosi dismetabolica (ad esempio: alcaptonurica, uratica, condrocalcinosa).

L'*osteoartrosi primaria* colpisce di preferenza il sesso femminile (10:1), con età media di comparsa fra i 50 e i 60 anni. Interessa quasi sempre le piccole articolazioni delle mani: interfalangee distali (*noduli di Heberden*), interfalangee prossimali (*noduli di Bouchard*), metacarpo-falangea del 1° dito (*rizartrosi di Forestier*).

La **sintomatologia** clinica è costituita:

- in **fase iniziale**, dal dolore diurno, dal dolore da carico meccanico e dalla riduzione della mobilità articolare;
- in **fase evolutiva**, dal dolore (come in fase iniziale) e dai segni di flogosi con dolore vivo, spontaneo o provocato, e dall'ulteriore riduzione della mobilità articolare;
- in **fase avanzata**, dalla prevalenza della limitazione dei movimenti sulla sintomatologia dolorosa e dalla presenza di scrosci articolari.

Il dolore insorge al mattino e si attenua con il movimento; è sensibile alle variazioni meteorologiche; la rigidità muscolare è tipica dopo il riposo (rigidità mattutina). Nel caso dell'*artrosi cervicale* si sviluppa una sindrome caratterizzata da vertigini, cefalea e dolore agli arti superiori. Il trattamento dell'osteoartrosi si basa sull'associazione di farmaci antinfiammatori e antidolorifici con cicli di fisiochinesiterapia.

4 Le malattie tumorali

I **tumori** rappresentano uno dei problemi sociali più gravi del nostro tempo per l'**incidenza** e la **mortalità**, per le profonde alterazioni che arrecano, per le ripercussioni e i danni economici che ne derivano alla collettività. Contro tutto ciò sono impegnati ricercatori, clinici, mezzi di comunicazione e associazioni di volontariato: tuttavia, ancora manca, nel nostro, come in molti altri Paesi, un'aggiornata *cultura del tumore*.

La parola «tumore» — e gli equivalenti «cancro» e «neoplasia» — evocano nella maggior parte delle persone un senso di rifiuto totale, di condanna inevitabile o, nella migliore delle ipotesi, un alone di mistero collegato all'impressione che di fronte a ciò la medicina si arresti impotente. È importante, invece, che si sappia che di cancro oggi si guarisce tanto più facilmente quanto più presto lo si scopre. Si è in grado di guarire dal 35 al 40% di tutti i tumori, che, però, potrebbero essere curati con successo fino al 60% se la diagnosi fosse posta precocemente. Non pochi possono essere prevenuti: basti ricordare che, astenendosi dal **fumo**, si prevengono la maggior parte dei tumori del polmone e della laringe, e si riduce di gran lunga la frequenza del cancro della bocca, dell'esofago, del pancreas e della vescica.

I tumori sono, dunque (è bene ribadirlo con forza), una malattia, non un mistero e tanto meno una condanna: una malattia, o meglio un complesso gruppo di malattie, vecchie quanto l'uomo (ne sono state trovate tracce nelle mummie egizie), di cui oggi si ammala un individuo su quattro. Questo dato sembra scoraggiante, ma va posto in relazione con una serie di variabili che in qualche modo incidono sulle statistiche, come l'allungamento della vita media e l'incremento della popolazione.

Ciò non toglie che i tumori siano in effettivo aumento, soprattutto in rapporto all'incontrollata **industrializzazione della società** (pare, infatti, che l'origine ambientale abbia un ruolo predominante nella genesi di circa l'80% dei tumori): è un dato noto, ad esempio, che nel Nord Italia la

mortalità per cancro sia maggiore rispetto al Sud e risulti altresì maggiore nei grandi agglomerati urbani rispetto ai piccoli centri. Tuttavia, se nel 1960 sopravviveva alle neoplasie un malato su quattro, oggi ne sopravvivono tre su quattro. Infatti, negli ultimi decenni l'enorme sforzo compiuto dai ricercatori di tutto il mondo ha portato alla luce una serie di nuove conoscenze sulla biologia di questa malattia e sulle possibilità di guarirne con successo. Accanto agli ammirabili progressi della diagnostica, che hanno permesso di individuare alcune lesioni molto piccole, e quindi aggredibili radicalmente dai chirurghi, le ricerche farmacologiche hanno introdotto nella pratica clinica nuove efficaci sostanze antineoplastiche. Se si tiene poi conto di come siano stati meglio definiti i ruoli ed i tempi delle singole terapie (chirurgia, radioterapia, chemioterapia) nel trattamento della maggior parte dei tumori, ben si comprende perché neoplasie che un tempo non lasciavano alcuna speranza di sopravvivenza possono oggi essere curate definitivamente.

Ci si è però resi conto di quanto sia importante il momento in cui si inizia la terapia, in rapporto allo sviluppo del tumore: scoprire una lesione tempestivamente è, ai fini della possibilità di cura, più importante della scelta di una terapia.

Gli ultimi anni hanno rappresentato un periodo singolare per la **prevenzione** del cancro: si è compreso, infatti, in via definitiva, che senza prevenzione servirebbero a ben poco le costose attrezzature e le ricerche avanzate che si vanno approntando in tutto il mondo. Perché la prevenzione sia possibile è necessario un buon livello di informazione: fondamentale, in tal senso, l'opera di enti come l'*Associazione italiana per la lotta contro i tumori* e le campagne di sensibilizzazione periodicamente programmate a livello istituzionale.

Alcuni **virus** sono fortemente sospettati di provocare, o comunque di favorire, lo sviluppo del carcinoma del collo uterino, nonché di altri tumori delle vie respiratorie e/o del sistema linfatico. Anche numerose **sostanze chimiche**, con le quali entriamo quotidianamente in contatto, sono considerate cancerogene. Per comprendere l'entità del danno da **radiazioni** basta ricordare la tragedia di Chernobyl e anche la comune abbronzatura comporta dei rischi, poiché l'eccesso di **radiazioni ultraviolette** può provocare tumori della pelle. Se aggiungiamo a ciò le insidie delle sostanze potenzialmente cancerogene assunte attraverso i cibi, comprenderemo bene come siano molteplici le cause da cui possono avere origine i tumori. Un ulteriore, classico esempio è la prevenzione primaria del carcinoma del polmone, che ha nel fumo di sigaretta la causa principale. Circa l'85-90% delle neoplasie polmonari è causata dal **fumo di sigaretta**. L'abitudine al fumo è, come già detto, la causa principale anche del cancro dell'esofago, della laringe, della cavità orale, e contribuisce a determinare il cancro della vescica, del pancreas e del rene; complessivamente, il fumo è coinvolto almeno nel 30% delle morti per cancro.

Fra le donne, a causa della crescente percentuale di fumatrici, il cancro del polmone raggiunge tassi di mortalità sovrapponibili a quelli del tumore mammario.

Fortunatamente, il rischio di cancro da fumo può essere ridotto. Studi clinici ed epidemiologici indicano che una popolazione che smette di fumare ha un rischio di cancro al polmone più basso di una che persiste nell'abitudine del fumo: gli ex fumatori da almeno 15 anni mostrano un rischio di poco superiore ai non fumatori. Perciò evitare di fumare è il metodo più efficace per prevenire il cancro; in questo senso, sia il medico di famiglia sia l'insegnante rivestono un ruolo di primaria importanza.

5 Le malattie infettive

Nel corso della Storia si sono susseguite tante epidemie terribili. Peste, vaiolo, colera, tifo erano i più gravi flagelli che potessero affliggere l'umanità insieme alle carestie, alle guerre e ai disastri naturali, come inondazioni e terremoti. Per secoli gli uomini sono rimasti quasi impotenti di fronte al diffondersi delle epidemie, ma il progresso, con il **miglioramento delle condizioni igieniche** e le **scoperte della medicina**, ha contribuito a farne scomparire alcune o a limitare al

massimo gli effetti devastanti delle malattie epidemiche negli Stati industrializzati. Questo non vale per i Paesi del Terzo e del Quarto Mondo, dove le epidemie continuano a mietere vittime a causa della denutrizione, delle scarse condizioni igieniche e della difficoltà di curare i malati.

La possibilità di contrarre malattie infettive dipende da due fattori:

1. la carica microbica, cioè la tipologia e la quantità di virus e batteri che si introducono all'interno dell'organismo;
2. le difese immunitarie, vale a dire il sistema di difesa che l'organismo oppone a questi agenti patogeni.

Se la **carica dei virus** (la cosiddetta *virulenza*) è forte, mentre le difese immunitarie, cioè la resistenza che il corpo oppone all'attacco del virus, sono deboli, si creano le condizioni per infezioni e malattie.

Le vie attraverso le quali i microrganismi entrano nel corpo umano sono essenzialmente tre:

- **via aerea:** cavità nasali, bronchi, polmoni;
- **via oro-fecale:** bocca, esofago, stomaco, intestino;
- **via parentale e sessuale:** sangue e liquidi biologici.

La trasmissione della malattia può avvenire in modo diretto, da uomo a uomo, o attraverso un altro organismo che la ospita.

Tra le malattie che si trasmettono per **via aerea**, citiamo:

- la **tubercolosi**, che colpisce i polmoni, le ossa, i reni o le meningi. Molto diffusa nei secoli passati, è oggi in ripresa in seguito all'immigrazione di extracomunitari con problemi di malnutrizione;
- il **morillo**, comune malattia dell'infanzia nei Paesi industrializzati, nel Terzo e Quarto Mondo è ancora causa di morte per i bambini denutriti. Si manifesta con febbre e macchioline rosse sul tutto il corpo;
- la **varicella**, che colpisce in genere i bambini dai 5 ai 9 anni e si manifesta con febbre unita alla comparsa di vescicole su tutto il corpo. Il virus responsabile si chiama Varicella Zoster perché, quando lo si incontra per la prima volta, provoca la varicella, ma rimane annidato nell'organismo per tutta la vita e, in assenza di difese, può causare l'Herpes Zoster (il cosiddetto Fuoco di Sant'Antonio);
- la **meningite**, un'infezione delle meningi, ossia le membrane che ricoprono il cervello e il midollo spinale, che si manifesta con febbre alta e forte mal di testa. Se non curata può essere mortale.

Tra le malattie che si trasmettono attraverso gli **alimenti**, citiamo:

- la **salmonellosi**, che si contrae in seguito al consumo di alimenti contaminati e può colpire a qualsiasi età, ma è più rischiosa nei lattanti e negli anziani. Provoca diarrea, febbre e dolori addominali. I maggiori distributori di salmonella nell'ambiente sono gli insetti. In generale, la presenza di tale agente batterico su un alimento è dovuta ad una contaminazione crociata (contatto cibi cotti e crudi, presenza di individui infetti, superfici non pulite);
- la **toxoplasmosi**, della quale esistono due forme: una che viene contratta al momento della nascita e che può provocare danni irreversibili, e un'altra, meno pericolosa, che può essere causata dal contatto con le feci di animali infetti o dal consumo di carni crude;
- l'**epatite A**, di cui ci si può ammalare consumando cibi o acqua contaminati con le feci umane di soggetti che stanno per ammalarsi o che ne sono già affetti. Si presenta con nausea, vomito, stanchezza, colorazione gialla degli occhi e/o della pelle (*ittero*), urine scure e feci chiare. Gli individui a rischio sono coloro i quali viaggiano in aree dove l'incidenza della malattia è medio-alta, specialmente se il cibo e le condizioni igienico-sanitarie sono scadenti; le persone

(specialmente i tossicodipendenti) con malattie epatiche croniche; gli emofiliaci; il personale di laboratorio medico che lavora a diretto contatto con il virus; nonché gli omosessuali;

- le **gastroenteriti**, che consistono in infiammazioni acute o croniche dello stomaco e dell'intestino, causate da virus, batteri o parassiti, e che provocano nausea, diarrea e vomito. Si pensa, di solito, che questi disturbi siano tipici dei Paesi caldi o delle zone dove le condizioni igieniche sono precarie, invece a scatenarle possono essere microrganismi differenti, alcuni dei quali presenti anche nelle nostre regioni. Le più frequenti nei Paesi sviluppati sono quelle di tipo non alimentare, trasmesse dal Rotavirus soprattutto durante la stagione invernale. Le gastroenteriti di origine alimentare sono una conseguenza del consumo di cibi contenenti microrganismi vivi, i quali, una volta raggiunto l'intestino, si moltiplicano. Chi è colpito da queste infezioni accusa per qualche giorno disidratazione, stanchezza e inappetenza. In alcuni casi, se non si interviene con una cura antibiotica, il disturbo può diventare cronico.

Tra le malattie che si trasmettono per contatto con **liquidi organici**, ricordiamo:

- le **epatiti B e C**, le quali possono provocare una malattia del fegato acuta o cronica, anche se in alcuni portatori sani sono del tutto asintomatiche. I sintomi sono simili a quelli dell'epatite A, ma le conseguenze sulla salute possono essere talvolta molto gravi. Si previene evitando di scambiare oggetti personali che possono venire a contatto con il sangue (lame da barba, siringhe, forbici) e usando i profilattici nei rapporti a rischio. Attualmente, in Italia, tutti i neonati sono sottoposti a vaccinazione obbligatoria;
- l'**AIDS**, di cui si dirà qui di seguito.

6 Le malattie a trasmissione sessuale

La gravità dell'AIDS e le campagne pubblicitarie effettuate per evitare il contagio e la diffusione di tale malattia hanno comportato una riduzione dei flussi informativi ed educativi nei confronti delle altre affezioni a trasmissione sessuale. L'acquisizione di conoscenze sulle manifestazioni e sui rischi che esse comportano è un aspetto importante dell'Educazione alla salute. Infatti, anche se tali patologie non sono necessariamente letali, possono determinare **danni d'organo** e situazioni di **cronicizzazione** con la compromissione della qualità della vita di chi ne viene contagiato.

Tra le malattie sessuali storicamente più diffuse bisogna ricordare: la **gonorrea** o **blenorragia**, che si manifesta con bruciori, fastidi e perdite purulente, e si può complicare nella sterilità; la **sifilide**, causata dal *Treponema pallidum*, che si presenta, nella fase iniziale, molto contagiosa, con un'ulcerazione genitale, e successivamente, quando diventa sistemica, con un esantema rubeolico generalizzato, mentre nella fase finale avviene l'interessamento di vari organi (sistema nervoso centrale e periferico, articolazioni, cuore e vasi).

Le malattie veneree oggi più frequenti sono rappresentate dalle **condilomatosi virali genitali**, provocate da tipi specifici di *papillomavirus* e dalle **infezioni virali erpetiche**: vanno diagnosticate e curate in fase precoce, in quanto, dopo periodi di latenza più o meno lunghi, possono contribuire all'insorgenza di alcuni tipi di carcinoma dell'apparato genitale. Esistono poi **numerosi microrganismi** come la *Chlamydia*, la *Gardnerella*, il *Trichomonas*, la *Candida albicans*, lo *Stafilococco*, che sono stati ritrovati frequentemente, da soli o combinati, associati a uretriti, vaginiti, cervicitì e rettocoliti recidivanti e il cui ruolo, nei casi di sterilità apparentemente inspiegata, non è ancora chiaro.

Mentre per queste patologie esistono rimedi efficaci se la diagnosi viene fatta in fasi non avanzate, molto gravi possono essere le conseguenze di altre patologie come l'**epatite B** e l'**epatite C**, che possono essere ugualmente contratte durante rapporti sessuali non protetti. Il rischio sembra essere maggiore se il rapporto sessuale è occasionale.

Per ridurre o annullare i rischi di contagio delle malattie veneree si consiglia di limitare il numero dei partner; di usare il preservativo soprattutto nei rapporti non stabili; in caso di contagio, di estendere il trattamento al partner; di prevenire l'epatite B con la vaccinazione. Nonostante le ormai numerose campagne informative e pubblicitarie, l'uso del preservativo non è sufficientemente diffuso tra gli adolescenti e le malattie veneree rappresentano tuttora un problema. Il ruolo del docente di Scienze nella Scuola secondaria di I e II grado deve anche essere quello di educare la popolazione giovanile alla conoscenza delle cause e dei rischi connessi alla trasmissione delle malattie sessuali, analizzando, in collaborazione con gli operatori sanitari (medici, psicologi etc.), i problemi alla base di eventuali resistenze e tenendo conto del fatto che esporsi ai rischi in genere, e a quelli sessuali in particolare, può essere espressione dei **comportamenti autodistruttivi** tipici delle personalità più fragili durante l'adolescenza.

L'**AIDS** è una malattia infettiva, causata dal **virus HIV** (*Human Immunodeficiency Virus*), che provoca la distruzione e la perdita di funzionalità dei linfociti T del sistema immunitario, per cui i soggetti affetti diventano progressivamente sensibili a *infezioni opportunistiche* e ad altre affezioni, le quali, nel loro complesso, costituiscono la sindrome clinica dell'AIDS (*Acquired Immune Deficiency Syndrome*). Attualmente si conoscono almeno due ceppi virali responsabili dell'AIDS: l'HIV-1 e l'HIV-2. In entrambi i casi, si assiste ad una perdita progressiva dei linfociti T-CD4 (*helper*) e ad un aumento della quantità di HIV circolante nel sangue.

I primi casi di tale sindrome si manifestarono agli inizi degli anni Ottanta del Novecento in maschi omosessuali; nel 1983 il virologo francese Luc Montagnier isolò per primo il virus responsabile. Il retrovirus umano dell'HIV è costituito da una capsula lipoproteica all'interno della quale si trova il materiale genetico virale (RNA). Sulla capsula sono esposte numerose glicoproteine che consentono al virus di aderire alle cellule del sistema immunitario; la più importante è la glicoproteina **gp120**, che si lega alla molecola CD4 e, in tal modo, consente l'infezione e la moltiplicazione del virus nei linfociti T-CD4.

L'HIV si può trasmettere mediante rapporto sessuale con una persona infetta, tramite la condivisione di aghi e siringhe tra i soggetti che fanno uso di droghe iniettabili, dalle madri malate ai figli durante il parto. La trasmissione attraverso trasfusioni o emoderivati è oggi rara, infatti sono disponibili test in grado di rilevare anche piccole quantità di anticorpi contro il virus nel siero dei soggetti **sieropositivi**.

Sono state, inoltre, definite 25 infezioni opportunistiche come «**malattie tipiche dell'AIDS**», che definiscono la diagnosi clinica di AIDS negli individui sieropositivi. L'infezione più comune è rappresentata dalla polmonite da *Pneumocystis carinii*; tra le altre: le polmoniti da Streptococco e da Haemophilus; la tubercolosi; le infezioni dell'apparato gastrointestinale da Salmonella, *Campylobacter*, *Shigella*; le micosi da *Candida albicans* e da *Histoplasma capsulatum*; le meningiti da *Cryptococcus*; le infezioni dovute a virus erpetici.

La progressione della malattia segue una **serie di fasi** ben definite ed elencate a seguire.

1. **Sindrome retrovirale acuta.** Dopo 1-3 settimane dal contagio, il soggetto manifesta sintomi simil-influenzali (febbre, cefalea, linfadenopatie e malessere generale) per circa 2 settimane. In questa fase il virus si moltiplica nell'organismo, soprattutto a carico delle stazioni linfonodali. Il numero dei linfociti T-CD4 diminuisce, per poi risalire quando il sistema immunitario riesce a limitare la moltiplicazione virale.
2. **Fase asintomatica.** È il periodo in cui gli individui affetti godono di salute normale e questa condizione può durare anche più di un decennio. Le cellule del sistema immunitario, però, sono in continua lotta con il virus, che si moltiplica; nel momento in cui l'efficienza del sistema immunitario si esaurisce, il paziente evolve verso lo stadio successivo.
3. **Fase sintomatica precoce.** Può avere durata variabile ed è caratterizzata da una rapida caduta dei linfociti T-CD4 circolanti e dalla comparsa di infezioni opportunistiche non letali.

4. **Fase sintomatica tardiva.** Anche questa fase può durare pochi mesi o diversi anni ed è contraddistinta dalla compromissione totale della funzionalità del sistema immunitario con la comparsa di malattie gravi, deperimento fisico, perdita progressiva di peso, mancanza di energie.
5. **AIDS avanzato.** Il numero dei linfociti T-CD4 è minore di 50. I pazienti muoiono nel giro di un anno, in genere per infezioni gravi o tumori (Sarcoma di Kaposi, linfomi etc.).

Nonostante l'impegno economico, scientifico e sociale, l'AIDS resta a tutt'oggi una malattia incurabile. Risultati incoraggianti si sono avuti utilizzando farmaci inibitori della trascrittasi inversa, come l'AZT, la ddI, la ddC, combinati ad altri inibitori delle proteasi; l'OMS, inoltre, sta sostenendo la sperimentazione di nuovi vaccini nei Paesi dove il tasso d'incidenza della malattia è più elevato (Thailandia, Brasile, Uganda).

Diversamente dalle altre principali cause di morte, come il cancro e le malattie cardiache, l'AIDS colpisce perlopiù i **bambini**, i **giovani adulti** e le **donne**. La maggior parte delle infezioni da HIV e delle morti dovute all'AIDS avvengono nelle persone al di sotto dei 25 anni. Oltre due milioni di bambini al di sotto dei 15 anni sono infettati dal virus HIV-1, mentre le donne rappresentano la metà dei casi di HIV nel mondo. Negli Stati Uniti, l'AIDS è la seconda causa di morte tra i giovani (gli incidenti stradali e altri incidenti vengono prima). Nell'Africa subsahariana, l'AIDS è la prima causa di morte tra le persone di tutte le età. Si stima che le persone infette ammontino globalmente a quasi 40 milioni.

Le difficoltà nel mettere a punto un vaccino efficace rendono necessari interventi preventivi volti a limitare la diffusione del virus nella popolazione. La scuola rappresenta il luogo ideale per la realizzazione di programmi educativi finalizzati all'acquisizione di norme igieniche da adottare a livello personale per ridurre il rischio di infezione. La popolazione giovanile, tuttavia, non si percepisce a rischio di contagio; pertanto, sarà compito dei docenti organizzare lezioni e seminari utili ad approfondire le relative tematiche.

7 Le malattie da lavoro

Lo studio delle malattie da lavoro ha origini antiche (Paracelso 1493-1541; Ramazzini, 1633-1714). Tali patologie dipendono quasi sempre dalle influenze nocive dell'**ambiente** in cui il lavoro si svolge o dall'azione patogena diretta delle **sostanze** maneggiate. Oggigiorno, sono legalmente riconosciute e assicurate.

Particolarmente importanti sono le **intossicazioni professionali**, dato l'accresciuto uso di svariate sostanze chimiche nelle lavorazioni. Le più frequenti sono le intossicazioni croniche causate da piombo (*saturnismo*), mercurio, fosforo, cromo, acido nitrico, composti dello zolfo, idrocarburi alifatici e aromatici, ossido di carbonio, amine alifatiche e aromatiche, silice e amianto. Le manifestazioni cliniche variano dall'anemia alle alterazioni mucose e cutanee, fino a gravi manifestazioni dell'apparato gastrointestinale e del sistema nervoso centrale e periferico. Per alcuni composti, inoltre, è stato riconosciuto il ruolo causale nelle patologie neoplastiche; l'esempio più noto è dato dall'elevata incidenza di morti per carcinoma polmonare e mesotelioma pleurico tra i lavoratori esposti alle polveri di amianto.

Tra le **patologie professionali non dovute a intossicazione**, ricordiamo: le malattie da radiazioni ionizzanti e da onde elettromagnetiche, le dermatosi professionali, le malattie osteo-articolari e angioneurotiche causate da vibrazioni meccaniche, l'ipoacusia e la sordità da rumori, le broncopneumopatie da inalazione.

8 L'educazione alla solidarietà nella sofferenza, nella vecchiaia e nell'inabilità temporanea o permanente

La **vecchiaia** è caratterizzata da un progressivo decadimento fisico e psicologico dell'organismo. Ma il più grave problema della cosiddetta *terza età* non è di ordine fisiologico, in quanto questa decadenza naturale avviene di solito gradualmente, bensì di tipo psicologico. L'anziano non dev'essere abbandonato né escluso dalla vita sociale attiva, onde evitare un troppo brusco decadimento delle facoltà psico-fisiche. Per mantenersi in buona salute le persone anziane dovrebbero condurre una vita serena, coltivando con continuità i propri interessi.

L'emarginazione degli anziani, così come quella dei disabili, è purtroppo frutto del mito dell'efficienza fisica, troppo spesso propagandato a sproposito dall'attuale società dei consumi. Con i progressi della medicina e della farmacologia, tra l'altro, l'età media si è allungata considerevolmente, acuendo ancora di più il problema.

I disturbi che affliggono le persone in là con gli anni sono determinati perlopiù da carenze di affetto e dalla solitudine, causate spesso dall'indifferenza della famiglia e della società. Storicamente a capo delle società patriarcali, oggi i *vecchi*, di frequente, vengono considerati inutili e improduttivi; ciò induce crisi esistenziali che fungono da fattore scatenante di situazioni patologiche spesso irrisolvibili.

Concentrare in un solo luogo (ospizi, case di riposo etc.) coetanei in età avanzata non è, peraltro, una soluzione soddisfacente né dal punto di vista sanitario né da quello sociale, poiché ciascuno porta con sé il suo vissuto, le sue ansie, le sue malattie e la convivenza forzata non fa altro che moltiplicare tali incertezze; pertanto, l'anziano si isola, tendendo a diventare sempre più scorbutico e nevrotico.

La soluzione da adottare sarebbe quella già messa in atto con successo in alcuni Paesi: lasciare loro un minimo di responsabilità e autonomia, reinserendoli (soprattutto se in buona salute e autosufficienti), con un ritmo di lavoro ridotto e con mansioni adatte, nel mondo del lavoro. Gli ospizi, di conseguenza, dovrebbero essere sostituiti da *residence* con miniappartamenti collegati ai servizi centralizzati, con assistenza medica e sociale (ristorante, cinema, sale da gioco, di lettura etc.). Anche l'attività culturale e la cura dei propri hobbies possono rappresentare una possibilità di recupero per l'anziano, non solo sul piano psicologico ma anche su quello morale.

1 L'educazione sanitaria

La predominanza delle malattie cronico-degenerative nella popolazione è la risultante delle condizioni di rischio biologico e ambientale in cui si svolge la vita moderna. Esistono situazioni nei confronti delle quali l'individuo da solo non può esercitare alcuna influenza (ad esempio, l'inquinamento ambientale), mentre in altri casi la conoscenza delle cause di alcune patologie è in grado di prevenirne l'insorgenza assumendo stili di vita più sani o partecipando a programmi di prevenzione specifici (abolizione del fumo, screening periodici etc.).

Il modello preventivo di Tannahill ha introdotto il concetto di **promozione della salute** accanto a quelli tradizionali di *diagnosi, cura e riabilitazione*. In particolare, tale concetto riunisce quelli di *prevenzione delle malattie, protezione della salute ed educazione sanitaria*.

La **prevenzione delle malattie** si attua secondo i principi classici della **medicina preventiva**.

La **protezione della salute** riguarda le regolamentazioni e gli interventi legali e amministrativi atti a influenzare i cambiamenti economici, sociali e ambientali in favore della salute del singolo e della comunità (leggi antifumo, sicurezza stradale, sicurezza sul lavoro).

L'**educazione sanitaria**, secondo la definizione di Smith, è «*un'attività di comunicazione intesa ad incrementare la salute, ad eliminare i fattori di rischio e a prevenire le malattie, rivolta a soggetti singoli o ad intere comunità e realizzata influenzando positivamente le conoscenze, gli atteggiamenti e i comportamenti del singolo, delle comunità e dei detentori del potere*». È evidente come l'attuazione di tale tipo di educazione preveda una grossa preparazione e maturità della classe politica, dell'intera popolazione di un Paese e degli specialisti del settore. L'**educazione sanitaria nella scuola** viene attuata mediante l'insegnamento dell'*Educazione alla salute* e gli specialisti del settore diventano, nel caso specifico, i docenti delle diverse discipline e di Scienze in particolare. Un linguaggio semplice e il lavoro di gruppo sono i mezzi a disposizione della comunità scolastica per raggiungere obiettivi che si tradurranno, in futuro, nel miglioramento della salute della popolazione adulta.

La **salute**, secondo la definizione dell'**OMS**, è uno stato di completo benessere fisico, sociale e mentale, e non consiste semplicemente nell'assenza di malattie o menomazioni.

L'educazione sanitaria è un **processo attivo bi-direzionale** che si svolge attraverso tre momenti fondamentali, ovvero:

- l'acquisizione delle *conoscenze*;
- la modifica degli *atteggiamenti*;
- l'adozione dei *comportamenti*.

La fase dell'*acquisizione delle conoscenze* riguarda l'apprendimento delle nozioni scientifiche di base sul corpo umano, sulle patologie e le cure disponibili, sulle possibilità di utilizzare i servizi sanitari, sui fattori di rischio personali e ambientali. I messaggi trasmessi per avere successo dovranno essere credibili e gradevoli, accurati e pertinenti, semplici, nuovi e, soprattutto, non «terroristici» ma positivi. La modifica delle conoscenze attuata attraverso l'informazione è piuttosto semplice da ottenere, se paragonata alle difficoltà che si incontrano nell'indurre le modifiche degli atteggiamenti e, ancor più frequentemente, nel determinare l'adozione di nuovi

comportamenti (ad esempio, smettere di fumare). L'educazione sanitaria mira, quindi, a suscitare nella popolazione atteggiamenti e comportamenti atti a **migliorare la salute** e a **combattere le malattie**. Per *comportamento* s'intendono anche le azioni volontarie e involontarie, le abitudini, i costumi, le risposte predisposte, i riflessi condizionati, ciascuno dei quali, oltre che l'interesse del medico, deve richiamare quello dell'educatore sanitario. Più recentemente, infatti, si è accettato che le *condizioni morbose* nelle quali il comportamento dell'uomo assume un ruolo eziologico preminente diventano sempre più numerose e sempre meglio definite (cancro, cardiopatie, malnutrizioni, tossicomanie).

A seconda del destinatario, l'educazione sanitaria considera:

- il **paziente** (*approccio tradizionale*);
- il **soggetto a rischio** (*approccio intermedio*);
- la **popolazione sana** (*approccio innovativo*).

L'educazione del *paziente* è quella che ha il miglior indice costi-benefici. Il paziente (e la sua famiglia, nel caso del bambino) è ottimamente disposto ad ascoltare suggerimenti personalizzati riguardanti la propria salute (la gestione del diabete, dell'asma e così via).

L'educazione del *soggetto a rischio* ha probabilità minori di successo, strettamente connesse alla professionalità e alle capacità didattiche dell'operatore (pressione arteriosa su valori limite, figli di asmatici).

L'educazione dell'intera popolazione si può anche definire come rivolta al *soggetto sano*. Richiede lo sforzo maggiore ed è destinata ad avere il minore ascolto; deve utilizzare messaggi gradevoli e produrre cambiamenti nello stile di vita finalizzati al mantenimento e al potenziamento della propria e della altrui salute.

Dal punto di vista storico, l'educazione sanitaria si connota come un'attitudine culturale abbastanza recente. Nel passato, nel classico triangolo della malattia costituito dall'**ospite**, dall'**agente causale** e dall'**ambiente**, la maggiore attenzione era rivolta agli ultimi due fattori. Oggi il centro della malattia è riconosciuto nell'ospite, quindi l'uomo è l'oggetto principale dei programmi di Educazione alla salute.

Un **intervento di educazione sanitaria** si attua in **quattro fasi**, che possono essere schematizzate nella maniera che segue.

1. **Analisi dei problemi e dei bisogni.** Tale fase riguarda la raccolta dei dati relativi alla comunità verso la quale è diretto l'intervento educativo e la rilevazione dei bisogni oggettivi e soggettivi, espressi e non espressi, individuali e di gruppo. I bisogni oggettivi emergeranno dalle indagini epidemiologiche (ad esempio, il fumo nella scuola), quelli soggettivi saranno espressi direttamente dalla collettività (ad esempio, i pidocchi nella scuola). I bisogni oggettivi e quelli soggettivi non sempre coincidono. Sulla loro base vanno definiti gli obiettivi comportamentali.
2. **Programmazione.** In questa fase vanno analizzati vantaggi e svantaggi delle possibili soluzioni e devono essere definiti gli obiettivi educativi. Tali obiettivi possono essere a breve, medio e lungo termine; devono essere definiti, misurabili e verificabili alle scadenze previste, al fine di consentire la scelta degli interventi successivi e la valutazione dei risultati. I gruppi sociali ai quali si rivolge l'intervento educativo devono essere il più possibile omogenei rispetto alle esigenze sanitarie che si vogliono trattare. Possono essere costituiti sulla base del rischio (ad esempio, le donne gravide), su base territoriale (ad esempio, gli anziani di un distretto), su gruppi già formati (ad esempio, le classi di una scuola). Nell'organizzazione dell'attività educativa bisognerà tenere conto del livello culturale, della situazione sociale, delle resistenze del gruppo *target*.
3. **Attuazione dell'intervento educativo.** Consiste nella messa a punto della soluzione prescelta, tenendo conto delle priorità e della fattibilità. L'intervento educativo nasce dalla combinazione di tre elementi: l'informazione, il rapporto educatore-utente, la durata. L'*informazione* può

essere epidemiologica e medica *di base*, oppure *specifica* in merito al comportamento di cui ci si sta occupando. L'informazione *di base* è costituita dal complesso di nozioni di anatomia e fisiologia del corpo umano, nonché di notizie riguardanti il tipo e la diffusione delle malattie correlate, inerenti all'argomento oggetto di educazione sanitaria (ad esempio, la contraccezione). Tale tipo di informazione può essere fornito anche da mezzi di comunicazione uni-direzionali (i tradizionali *mass-media*). L'informazione *specifica* è quella riguardante il comportamento in esame e deve essere fornita sempre attraverso mezzi di comunicazione bi-direzionali, in maniera da consentire un flusso di informazioni utente-educatore ed educatore-utente. I mezzi di comunicazione bi-direzionali sono costituiti da incontri individuali e/o di gruppo, seminari interattivi, dibattiti, questionari e altro. L'operatore, in questa fase, deve avvalersi di sussidi didattici quali grafici, tabelle, fotografie, diapositive, film, sistemi multimediali. L'informazione specifica deve essere reale e completa: deve, cioè, contenere tutti i motivi che giustificano una certa scelta di comportamento. La natura del *rapporto educatore-utente* è essenziale ai fini educativi. Il requisito fondamentale è che l'educatore sia il naturale punto di riferimento per l'utente su quel determinato argomento. Al fine di realizzare la «simbiosi culturale» che costituisce la premessa indispensabile per la credibilità e l'accettazione del messaggio, il rapporto tra utente ed educatore deve essere: costante, naturale e di estrema fiducia. Se l'argomento oggetto di educazione è l'uso dei farmaci, ad esempio, l'interlocutore più adeguato dell'utente è il suo medico di base; se, invece, si opera all'interno di una scuola, l'interlocutore naturale dello scolaro è il suo insegnante. La *durata* dell'intervento educativo è tanto più lunga quanto più ampi sono gli obiettivi prefissati e dipende dalle risorse, dal personale impegnato, dall'atteggiamento dell'utenza e dalle resistenze opposte.

4. **Valutazione dei risultati raggiunti.** Rappresenta il momento di verifica del raggiungimento degli obiettivi, nonché dell'*efficienza* e dell'*efficacia* di un intervento di educazione sanitaria. L'*efficienza* sarà stata ottenuta se il programma educativo è stato condotto con la minore spesa possibile. L'*efficacia* andrà, invece, valutata a livello individuale, per verificare se vi è stato un aumento delle conoscenze e una modifica dei comportamenti, e a livello comunitario, rilevando se sono avvenute modifiche nei regolamenti, nella legislazione e nei comportamenti collettivi (riduzione dei tassi di incidenza).

2 La medicina preventiva

La **medicina preventiva** si pone gli **obiettivi** di:

- proteggere il singolo individuo dalle malattie;
- controllare e limitare la diffusione delle malattie nella popolazione generale;
- eliminare le malattie.

La *storia naturale di una malattia* si articola in diversi stadi:

- lo **stadio della suscettibilità**, in cui l'individuo non ha ancora sviluppato la malattia ma è esposto a fattori che ne favoriscono l'insorgenza. Questi fattori, la cui presenza è associata ad una maggiore probabilità di sviluppare in futuro la malattia, senza però determinarne obbligatoriamente lo sviluppo, si definiscono «**fattori di rischio**».

I *fattori di rischio* possono essere distinti, a loro volta, in diverse categorie:

- *componenti genetiche o ereditarie* (errori congeniti del metabolismo, alterazioni genomiche), che fanno parte dell'individuo e pertanto non sono modificabili;
- *componenti ambientali di vita e di lavoro* (inquinamento atmosferico, inquinamento del microambiente), modificabili con il cambiamento di ambiente e mediante l'eliminazione di sorgenti inquinanti;

- *componenti comportamentali individuali* (dieta iperlipidica, fumo, farmaci), la cui modificazione è legata alla volontà del soggetto di voler cambiare le proprie abitudini;
- lo **stadio subclinico**, in cui la malattia non si manifesta ancora clinicamente, anche se sono presenti già alterazioni patogenetiche (lesioni precancerose, placche aterosclerotiche);
- lo **stadio clinico**, in cui i danni anatomici e funzionali determinano il manifestarsi della malattia con segni e sintomi;
- lo **stadio della disabilità**, in cui il paziente è disabile in virtù dei danni residuati a patologie non completamente guarite o divenute croniche.

In riferimento ai diversi stadi, la medicina preventiva individua tre livelli di intervento, rispettivamente di *prevenzione primaria, secondaria e terziaria*.

a) Prevenzione primaria

La *prevenzione primaria* è rivolta a soggetti sani e ha lo scopo d'impedire l'insorgenza di nuovi casi di malattia; deve, quindi, agire quando l'affezione si trova nello stadio della suscettibilità e può essere attuata mediante:

- l'inattivazione o la rimozione dell'agente causale;
- il risanamento e la protezione ambientale;
- l'aumento delle difese individuali;
- l'adozione di uno stile di vita che favorisca il mantenimento dello stato di salute.

Non sempre un singolo intervento è sufficiente per ottenere rapidi e consistenti vantaggi; spesso, infatti, occorre attuare contemporaneamente i diversi interventi preventivi sopra menzionati. Per alcuni di essi è sufficiente l'accettazione passiva da parte della popolazione, come nel caso delle campagne di vaccinazioni; altri, invece, richiedono la partecipazione attiva della collettività e dell'individuo, come, ad esempio, si verifica nella campagna contro il fumo.

b) Prevenzione secondaria

La *prevenzione secondaria* consiste nell'individuazione della malattia prima che si manifesti clinicamente e nella correzione di condizioni di rischio o di stati patologici in persone apparentemente sane. La selezione (*screening*) delle persone a rischio deve essere effettuata secondo procedimenti e metodologie corrette. Ad esempio, le neoplasie e le malattie cardiovascolari aterosclerotiche hanno un lungo periodo di fase preclinica, durante il quale un intervento mirato può prolungare la sopravvivenza in uno stato di salute più accettabile.

c) Prevenzione terziaria

Alcune malattie mostrano la tendenza alla progressione e all'invalidità. La *prevenzione terziaria*, definita in alcuni casi *riabilitazione*, che ha l'obiettivo di riportare l'individuo disabile a un ruolo socialmente utile, soddisfacente e, se possibile, autosufficiente, attraverso l'utilizzo ottimale delle sue capacità residue, consiste nell'attuazione di tutte quelle misure di ordine medico, sociale e psicologico atte a ridurre danni e invalidità. La realizzazione dei programmi di prevenzione richiede il consenso e la partecipazione dei singoli e della comunità. Il razionale impiego dei presidi, nonché dei servizi di cura e riabilitazione necessita di una sufficiente conoscenza dei problemi sanitari da parte della popolazione. Tale obiettivo può essere raggiunto attraverso l'*educazione sanitaria*.

d) Educazione ambientale ed ecologica

Sui principi sopra esposti si basano gli obiettivi dell'**igiene**, che può essere orientata in tre **direzioni**: *preventiva, curativa, educativa*; e secondo tre **oggetti**: *l'individuo, la collettività, l'am-*

biente. L'igiene dell'individuo (o *personale*) e quella della collettività sono oggetto dei programmi informativi di Educazione sanitaria e di Medicina preventiva. L'igiene ambientale fa ormai parte del settore più ampio dell'**Educazione ambientale ed ecologica**. Tale disciplina si basa su alcuni principi di tutela ambientale dettati dall'UNESCO nel 1980 e rafforzati successivamente dalla *Carta di Francoforte* delle Nazioni Unite, che possono essere sintetizzati come segue:

- considerare l'**ambiente nella sua globalità** (naturale e artificiale);
- instaurare un **processo educativo** che inizi a livello prescolastico e duri per tutta la vita, all'interno e all'esterno della scuola;
- seguire un **approccio interdisciplinare**;
- esaminare i principali problemi ambientali dai punti di vista **locale, regionale, nazionale e internazionale**;
- mettere in **relazione all'età degli allievi** e alla loro crescente sensibilità ambientale i contenuti dell'educazione ecologica.

3 Visite periodiche per la diagnosi precoce dei difetti sensoriali, delle malattie del ricambio e dei tumori

La difesa della salute personale va attuata seguendo le comuni regole di igiene personale e le indicazioni ormai obbligatorie in termini di vaccinoprofilassi, nonché adottando stili di vita corretti. L'Educazione alla salute deve però riguardare anche la possibilità di sensibilizzare l'individuo nei confronti della prevenzione di malattie specifiche. È noto, infatti, che numerose patologie possono aggravarsi e determinare la comparsa di danni permanenti se non diagnosticate precocemente. In particolare, i difetti degli organi sensoriali andrebbero individuati nell'età evolutiva, mentre altre patologie, come le malattie metaboliche e i tumori, insorgono prevalentemente in età adulta.

La vista e l'udito sono i sensi più evoluti e raffinati, più delicati e importanti. La loro cura richiede, quindi, particolari attenzioni igieniche, che vanno attuate fin dalla primissima età. Tra i bambini in età scolare e gli adolescenti è frequente la presenza di **difetti visivi** non diagnosticati: hanno un'incidenza del 30-33% e, nella maggior parte dei casi, sono rappresentati da *difetti di rifrazione* (astigmatismo, miopia, ipermetropia) che possono progressivamente aggravarsi e arrecare disturbi secondari (cefalea, astenia). Meno spesso, ma con conseguenze più gravi, si verifica l'*ambliopia*, talvolta provocata da strabismo non curato o non diagnosticato in tempo. È compito del docente di Scienze educare gli allievi ai possibili danni alla vista e alla necessità di sottoporsi periodicamente alla visita oculistica per individuare precocemente eventuali difetti visivi.

Tali obiettivi potranno essere raggiunti trasmettendo le seguenti **regole igieniche**:

- effettuare una corretta pulizia degli occhi, evitando di toccarli se le mani sono sporche;
- assicurare sempre l'idonea illuminazione all'ambiente e al piano di lavoro, intervallando le ore di attività che prevedono l'uso del sistema visivo con adeguati periodi di riposo;
- adottare le norme consigliate sull'uso del computer;
- sottoporsi periodicamente a visite specialistiche di controllo, soprattutto se c'è familiarità per disturbi visivi e se, per motivi di studio (o di lavoro), il sistema visivo è continuamente sollecitato.

Per quanto riguarda l'udito, l'igiene deve interessare sia il condotto uditivo esterno (da pulire con acqua e sapone, evitando l'introduzione di corpi estranei) sia il controllo della funzione uditiva. Infatti, anche se non è possibile attuare alcun tipo di prevenzione per i **difetti uditivi** congeniti, molto si può fare per i difetti acquisiti in conseguenza di malattia, stati tossici o traumi; in questi casi la diagnosi precoce può rallentare l'evoluzione verso la sordità, se non portare alla guarigione. Le conseguenze sull'apprendimento e sul rendimento saranno tanto meno reversibili

quanto più è avanzata l'età in cui si sono manifestati i segni della malattia. Anche qui l'unica difesa è rappresentata dal controllo audiometrico immediato e/o periodico in quei soggetti che, per motivi patologici (portatori di otiti croniche, otosclerosi) o professionali, sono a rischio.

Le visite periodiche per la prevenzione delle **malattie del ricambio** e per le **malattie tumorali** sono, invece, effettuate nella popolazione adulta su sottogruppi a rischio.

È buona regola anche nella popolazione giovane sottoporsi periodicamente a visite mediche di controllo e a indagini ematochimiche se esiste familiarità. Rientrano tra queste patologie il *diabete*, le *iperlipidemie*, l'*ipertensione*, le *epatopatie*, le *nefropatie*.

Socialmente assai rilevante è l'incidenza dei tumori, laddove la diagnosi precoce rappresenta uno dei pochi mezzi a disposizione per garantire la sopravvivenza. Per le neoplasie più diffuse si effettuano *screening* sui gruppi a rischio (ad esempio: pap-test nella cervice uterina, mammografia nel carcinoma mammario). I soggetti con situazioni di rischio aspecifico dovrebbero sottoporsi a visita oncologica, dopo aver consultato il medico di base, notando un cambiamento significativo nel proprio organismo (ad esempio: comparsa di noduli, modificazioni di nei, stipsi improvvisa).

4 Epidemiologia e profilassi delle malattie infettive

Le malattie infettive rappresentano un vasto gruppo di affezioni, contagiose o meno, causate da numerosi microrganismi parassiti dell'uomo: *batteri*, *virus*, *protozoi*, *miceti*, *rickettsie*, *micoplasmi*, *clamidie*. I casi di malattie infettive sono soggetti a **denuncia obbligatoria** (D.M. 5-7-1975 e aggiornamenti successivi) secondo una disposizione decretata dal Testo Unico delle Leggi Sanitarie e, più precisamente, dal D.M. del 15-12-1990; le malattie che devono essere segnalate sono 47, suddivise in 5 classi e con tempi di segnalazione differenziati.

Per **infezione** si intende l'invasione e la moltiplicazione, nei tessuti e/o nel sangue di organismi superiori, di microrganismi, indipendentemente dalla via di penetrazione. La **malattia infettiva** è costituita dalle lesioni anatomiche e funzionali correlate alla patologia. Le infezioni e le malattie infettive possono avere un'origine:

- **esogena**, se si manifestano come fenomeno spontaneo e non controllato nella comunità o come manifestazione nosocomiale nei servizi ospedalieri;
- **endogena**, se si manifestano come il risultato del viraggio del commensalismo e/o simbiosi tra agente e ospite.

Le **infezioni opportunistiche** sono causate dai patogeni *opportunisti*, cioè da microrganismi che determinano l'insorgenza dello stato di malattia in condizioni particolari in cui essi riescono a colonizzare sedi normalmente sterili o a superare i sistemi di difesa dell'organismo.

Esistono *malattie infettive contagiose* causate da agenti patogeni che si trasmettono da un ospite all'altro e malattie infettive *non contagiose* che si trasmettono solo attraverso vettori o modalità particolari (malaria, tetano).

Le **fasi di una malattia contagiosa** sono rappresentate da: invasione, incubazione, fase prodromica, fase clinica, convalescenza, ritorno allo stato di salute.

La **catena contagionistica** è una sequenza di fattori tra loro correlati che devono esistere o sono necessari affinché un individuo possa divenire infetto. Gli elementi della catena contagionistica sono: l'**ospite**, l'**agente**, l'**ambiente**.

Rispetto all'**ospite** possiamo distinguere i microrganismi in:

- **saprofiti**, che vivono nell'ambiente esterno;
- **commensali**, che vivono sulle superfici di rivestimento del corpo, ossia pelle e mucose;
- **parassiti**, i quali possono aggredire e vivere a spese dell'organismo ospite causandogli danno;
- **patogeni**, che inducono all'ospite uno stato di malattia.

Le caratteristiche dell'**agente** sono:

- **carica microbica**, cioè la quantità di agenti microbici capaci di causare l'infezione;
- **patogenicità**, cioè l'attitudine dell'agente patogeno a provocare la malattia (è una caratteristica geneticamente determinata);
- **virulenza**, cioè l'attitudine dell'agente a provocare malattie gravi, anche mortali;
- **contagiosità**, cioè l'attitudine dell'agente a passare da un soggetto ad un altro nel corso del processo infettivo.

Le caratteristiche dell'**ambiente** da considerare sono:

- temperatura;
- umidità;
- presenza di nutrienti;
- assenza di ossigeno;
- radiazioni UV;
- pH.

Le caratteristiche di **difesa dell'ospite** riguardano:

- i fattori di resistenza aspecifici (cute e mucosa, risposta infiammatoria);
- i fattori di resistenza specifici (immunità).

Le **sorgenti di infezione** sono rappresentate dai mezzi animati o inanimati dai quali l'agente passa direttamente all'ospite. Classiche sorgenti di infezione sono l'uomo malato, il portatore sano, l'animale malato, il terreno; in alcuni casi l'infezione può avere un'origine endogena.

La **trasmissione** descrive il movimento dell'agente patogeno dalla sorgente all'ospite e può avvenire per *contatto diretto* (*person to person*), per *contatto indiretto* attraverso veicoli (acqua) o *vettori* (zanzare), per *trasmissione verticale* dalla madre al figlio.

Viene definita **epidemia** la situazione in cui più casi della stessa malattia si presentino in una determinata popolazione in un breve periodo di tempo. Se la diffusione avviene tra diversi continenti si parla di **pandemia** (ad esempio, l'influenza spagnola nel 1918). La presenza stabile di una malattia (ad esempio, il morbillo) in una popolazione si definisce **endemia**.

La **profilassi delle malattie infettive** è l'insieme delle norme tendenti a prevenire nell'uomo la comparsa delle affezioni, proteggendolo in vario modo dall'attacco di microrganismi patogeni. Può essere attuata mediante la distruzione del germe, oppure mediante la protezione e il conferimento di resistenza ai soggetti recettivi.

Le tipologie di profilassi si distinguono in:

- **profilassi diretta**, la quale comprende norme che si adottano direttamente in presenza di un caso attuale di malattia infettiva in atto;
- **profilassi indiretta**, che ha lo scopo di modificare l'ambiente umano, fisico, sociale, in maniera tale da renderlo inadatto all'insediamento e alla propagazione dei germi;
- **profilassi immunitaria o specifica**, la quale ha l'obiettivo di rendere gli organismi refrattari o meno sensibili all'azione degli agenti patogeni. Può essere conferita mediante: vaccinoprofilassi (artificiale attiva) e/o sieroprofilassi (artificiale passiva).

Per quanto riguarda la profilassi delle malattie infettive, in generale, vigono i seguenti principi e disposizioni:

- **denuncia** dei casi accertati o sospetti;
- **accertamento diagnostico**;
- **isolamento**, per separare il soggetto ammalato o sospetto dal contatto con terzi. Può essere domiciliare o in reparto ospedaliero;

- **disinfezione**, per distruggere i germi patogeni, causa della malattia in atto. Va operata nei locali in cui ha soggiornato l'ammalato, sugli oggetti e gli indumenti da lui usati, sui suoi prodotti di escrezione;
- **vaccinazione**, allo scopo di immunizzare le persone esposte al contagio in maniera «attiva» (cioè mediante la produzione di anticorpi propri).

La **prevenzione primaria** delle infezioni deve essere, quindi, effettuata attraverso le seguenti fasi:

- identificazione e bonifica delle sorgenti e dei serbatoi di infezione;
- isolamento e contumacia;
- disinfezione e sterilizzazione, disinfestazione;
- scoperta e inattivazione dei portatori;
- eradicazione dei serbatoi naturali;
- bonifica dell'ambiente;
- modificazioni dei comportamenti;
- immunoprofilassi e chemiopprofilassi.

a) Vaccinoprofilassi

Un **vaccino** è un preparato in grado di conferire al soggetto recettivo ad una data malattia una protezione immunitaria specifica. I vaccini sono costituiti da agenti infettivi (o loro componenti) che hanno perso il potere patogeno e conservato le proprietà immunogene.

Quelli attualmente utilizzabili possono essere costituiti da:

- microrganismi vivi attenuati;
- microrganismi uccisi;
- frazioni di microrganismi;
- antigeni microbici purificati;
- anatossine;
- vaccini antidiotipo;
- vaccini da manipolazioni genetiche.

b) Vaccinazioni

In Italia sono obbligatorie, **per tutti i nuovi nati**, le vaccinazioni contro **difterite, tetano, poliomelite, epatite virale B**. Sebbene il ricorso all'obbligo di legge possa apparire anacronistico, va ricordato che lo strumento legale ha garantito il diritto alla salute e alla prevenzione di ogni bambino sul territorio nazionale e ha fornito la copertura finanziaria delle spese di vaccinazione, permettendo un ottimo controllo delle malattie così prevenibili.

Per quanto riguarda la **popolazione adulta**, alcune vaccinazioni sono obbligatorie per **determinate categorie** di persone e di lavoratori:

- la *vaccinazione antitetanica* è obbligatoria, oltre che per tutti gli sportivi affiliati al CONI, per i lavoratori agricoli, i metalmeccanici, gli operatori ecologici, gli stradini, i minatori, gli sterratori etc., secondo l'elenco riportato nella L. 5-3-1963, n. 292;
- le vaccinazioni *antimeningococcica*, *antitifica*, *antidiftotetanica*, *antimorbillo-parotite-rosolia* sono obbligatorie per tutte le reclute all'atto dell'arruolamento (D.M. 19-2-1997).

Sono raccomandate, nel nostro Paese, **per i nuovi nati**, le vaccinazioni contro **morbillo, parotite, rosolia** (MPR), **pertosse, infezioni da *Haemophilus influenzae b* (Hib), antipneumococcica ed antimeningococcica C, anti-HPV** (per le ragazze nel corso del 12° anno di età), **antivaricella** (a partire dal 2015).

Per quanto riguarda la **popolazione adulta**, esistono vaccinazioni **raccomandate** per:

- alcune categorie professionali considerate maggiormente a rischio per talune infezioni;
- alcune categorie di persone suscettibili di andare incontro a serie complicazioni in caso di infezione.

La vaccinazione contro l'*epatite virale B* è raccomandata, e offerta gratuitamente, agli operatori sanitari e al personale di assistenza degli ospedali e delle case di cura private, alle persone conviventi con portatori cronici del virus dell'epatite B, agli operatori di pubblica sicurezza, ai politrasfusi, agli emodializzati e a tutte le altre categorie indicate nel D.M. 4-10-1991.

La vaccinazione contro l'*influenza* è raccomandata a tutte le persone di età superiore ai 65 anni e a coloro che sono sofferenti di malattie croniche e debilitanti a carico dell'apparato cardiovascolare, broncopolmonare, renale etc., nonché agli addetti a servizi di pubblica utilità.

La vaccinazione contro le infezioni da *pneumococco* è consigliabile alle persone di età superiore ai 65 anni, o sofferenti di malattie croniche e debilitanti a carico dell'apparato cardiovascolare, broncopolmonare, renale, o con asplenia (mancanza della milza) funzionale o a seguito di intervento chirurgico.

La vaccinazione *antitifica*, così come quella antiepatite virale A e B, l'*antipoliomielitica*, l'*antitetanica*, l'*antimeningococcica*, l'*antirabbica* e quella contro la febbre gialla possono essere indicate per i viaggiatori che si rechino all'estero in zone endemiche o comunque considerate a rischio, dopo attenta valutazione.

c) Calendario delle vaccinazioni

Per *calendario delle vaccinazioni* si intende la successione cronologica con cui vanno effettuate le vaccinazioni dei bambini, sia obbligatorie sia raccomandate, fino all'adolescenza. Il calendario va costantemente aggiornato in considerazione delle conoscenze scientifiche, della situazione epidemiologica relativa alle diverse malattie, dell'evoluzione delle esigenze organizzative e delle nuove preparazioni vaccinali messe a disposizione dall'industria (vedi Tab. 1).

d) Classificazione dei vaccini

I vaccini possono essere costituiti da:

- agenti infettivi interi;
- componenti del microrganismo.

I vaccini costituiti dall'agente infettivo intero si possono suddividere in due grandi gruppi, a seconda che siano composti da:

- batteri o virus attenuati;
- batteri o virus inattivati (cioè uccisi).

I vaccini contenenti componenti del microrganismo sono suddivisi, a loro volta, in due tipologie, ovvero quelli:

- costituiti da componenti naturali del microrganismo o da sostanze da lui stesso sintetizzate (tossine);
- preparati con antigeni ottenuti sinteticamente, spesso con la tecnica del DNA ricombinante.

In base al tipo di microrganismo che determina la malattia (batterio o virus), i vaccini possono essere validi per:

- malattie batteriche;
- malattie virali.

Vaccino	Nascita	3° mese	5° mese	6° mese	11° mese	13° mese	15° mese	5-6 anni	11-18 anni	> 65 anni	Ogni 10 anni
Difterite - Tetano - Pertosse		DTPa	DTPa	DTPa	DTPa	DTPa		DTPa ¹			
Poliomelite		IPV	IPV	IPV	IPV	IPV		IPV			
Epatite B	HBV ²	HBV	HBV	HBV	HBV	HBV					
Haemophilus Influenzae b		Hib	Hib	Hib	Hib	Hib					
Morbillo - Parotite - Rosolia							MPR	MPR	MPR ³		
Pneumococco		PCV	PCV	PCV	PCV	PCV					
Meningococco C						Men C ⁴			Men C ⁴		
Infezione papillomavirus umano									HPV ⁵ (2dosi)		
Influenza										Inf	
Varicella									Var ⁶ (2dosi)		

LEGENDA

DTPa: vaccinazione antidifterico-tetano-pertossica; **dTPa:** vaccino difto-tetano-pertossico acellulare per adulti; **IPV:** vaccino antipoliomielitico iniettabile-inattivato; **HBV:** vaccino anti epatite B; **Hib:** vaccino contro le infezioni invasive da *Haemophilus influenzae b*; **MPR:** vaccino antimorbillo-parotite-rosolia; **PCV:** vaccino pneumococcico coniugato eptavalente; **Men C:** vaccino meningococcico C coniugato; **HPV:** vaccino antipapilloma virus umano; **VAR:** vaccino anti varicella.

3° mese: si intende dal 61° giorno di vita; **5-6 anni:** si intende dal 5° compleanno (5 anni e 1 giorno) ai 6 anni e 364 giorni (7° compleanno); **12° anno:** si intende da 11 anni e 1 giorno (11° compleanno) fino a 11 anni e 364 giorni (12° compleanno); **11-18 anni:** si intende da 11 anni e 1 giorno (11° compleanno) fino a 17 anni e 364 giorni (18° compleanno).

NOTE

1. Dopo il compimento dei 7 anni è necessario utilizzare la formulazione con vaccino antidiftetico-pertossico acellulare di tipo adolescenziale-adulto (dTPa). Gli adulti con anamnesi incerta per il ciclo primario di vaccinazione con dT devono iniziare o completare la vaccinazione primaria. Un ciclo primario per adulti è composto da 2 dosi di vaccino contenente tetano e difterite (dT) e una terza dose con vaccino dTPa. Le prime 2 dosi devono essere somministrate a distanza di almeno 4 settimane l'una dall'altra e la terza dose 6-12 mesi dopo la seconda. I successivi richiami devono essere effettuati ogni 10 anni (a partire dal completamento della serie primaria) e almeno una delle dosi *booster* di vaccino dT dovrebbe essere rimpiazzata da 1 dose di vaccino dTPa.
2. Per i bambini nati da madri positive HBsAg: somministrare entro le prime 12-24 ore di vita, contemporaneamente alle immunoglobuline specifiche anti epatite B, la prima dose di vaccino anti-HBV; il ciclo andrà completato con una seconda dose a distanza di 4 settimane dalla prima, con una terza dose dopo il compimento dell'ottava settimana e con la quarta fase in un periodo compreso tra l'undicesimo ed il dodicesimo mese di vita, anche in concomitanza con le altre vaccinazioni.
3. In riferimento ai focolai epidemici in corso, si ritiene opportuno, oltre al recupero dei soggetti suscettibili in questa fascia d'età (*catch up*), anche una ricerca attiva ed un'immunizzazione dei soggetti conviventi o venuti a contatto, non vaccinati (*mop up*).
4. Dose singola. La somministrazione a 11-18 anni va considerata nei soggetti non vaccinati nell'infanzia.
5. Per il sesso femminile, nel corso del 12° anno di vita, seguendo una scheda a 2 dosi. Vaccino bivalente (contro i genotipi 16 e 18 di HPV): 0, 1 e 6 mesi; vaccino quadrivalente (contro i genotipi 6, 11, 16 e 18 di HPV): 0, 2 e 6 mesi.
6. Nei soggetti anamnesticamente negativi e non precedentemente vaccinati, è prevista la somministrazione di due dosi a distanza l'una dall'altra.

Tab. 1 - Piano nazionale di prevenzione vaccinale 2012-2014

e) Vaccini per malattie batteriche

• Tetano e vaccino antitetano

Il **tetano** è una malattia infettiva non contagiosa. L'agente eziologico è il *Clostridium tetani*, che vive solitamente nell'intestino degli animali e viene eliminato con le feci. La spora penetra nell'organismo attraverso ferite lacero-contuse e può provocare la malattia a qualsiasi età. La spora, per germinazione, origina le forme vegetative responsabili della produzione di una *esotossina neurotropica* che è la vera causa della malattia tetanica. La sintomatologia esordisce in genere con il **trisma** (ipertonìa dei muscoli masseteri e pterigoidei che causa impedimento all'apertura della bocca). Lo stato di ipertono tende poi a generalizzarsi, fino ad interessare tutti i muscoli scheletrici, compresi quelli respiratori. La prognosi è sempre riservata, con un elevato indice di letalità (30-60%). In Italia si registrano oltre 100 casi di tetano l'anno con una letalità di circa il 60%. Dal 1998 ad oggi sono stati introdotti in commercio numerosi prodotti combinati ad altri componenti, in cui l'anatossina tetanica è associata, oltre che ai vaccini antidifterico e antipertosse acellulare, anche all'*Haemophilus influenzae b* (Hib), all'antipolio (IPV) e all'antipatite B (vaccino esavalente). La vaccinazione antitetanica, a ciclo fasale ultimato, conferisce una protezione pressoché totale ed è efficace al 100%. L'immunità dopo il richiamo è di circa 10 anni, le reazioni locali sono rare; tale vaccino determina un'immunizzazione attiva contro il tetano e può essere somministrato a qualsiasi età a partire dal 2°-3° mese di vita. Rientra tra le vaccinazioni obbligatorie nel bambino, in associazione a quella antidifterica; negli adulti è obbligatoria per le categorie più esposte al rischio di infezione.

• Difterite e vaccinazione antidifterica

La **difterite** è una tossinfezione provocata dal *Corynebacterium diphtheriae*, che esplica la propria azione patogena attraverso una potente esotossina. La trasmissione è interumana e avviene mediante le secrezioni faringee e laringee, o per contagio indiretto. Il periodo di incubazione varia dai 2 ai 6 giorni e inizia con malessere, astenia, bruciore alla gola; dopodiché la febbre sale rapidamente a 38-39°C e si formano le caratteristiche pseudomembrane, che ricoprono dapprima le tonsille e, successivamente, possono invadere laringe e trachea, provocando ostruzione respiratoria talmente grave da richiedere talvolta la tracheotomia. Le forme cliniche attraverso cui tale patologia si può manifestare sono: rinite, angina, angina maligna e laringite. La prognosi varia in funzione della manifestazione e della localizzazione della malattia. La vaccinazione per tale patologia è obbligatoria e viene praticata insieme a quella antitetanica. A ciclo di base ultimato, conferisce una protezione pressoché totale (100%), la cui durata è di almeno 10 anni ed è ulteriormente garantita dall'esecuzione dei richiami.

• Pertosse e vaccinazione antipertosse

La **pertosse** è una malattia infettiva altamente contagiosa, provocata dalla *Bordetella pertussis*; si trasmette per via aerogena. Il periodo di incubazione varia tra i 5 e i 20 giorni; il quadro clinico che segue è suddivisibile in tre fasi: catarrale, convulsiva e di declino. Le prime due fasi sono caratterizzate dalla presenza di tosse secca che, aumentando progressivamente, culmina nei tipici accessi (spesso associati a vomito). Durante il periodo di declino, gli accessi si riducono come intensità e numero, fino a giungere alla guarigione clinica. Tale patologia è grave nel lattante e nella prima infanzia, nei bambini già asmatici e nei casi in cui insorgono complicanze come, ad esempio, broncopolmoniti. Negli ultimi anni è disponibile il nuovo vaccino «acellulare», contenente cioè soltanto alcuni componenti del batterio. Questo nuovo vaccino, rispetto al vecchio preparato ottenuto con cellule batteriche intere, provoca un minor numero di reazione nei vaccinati, pur conservando un'elevata efficacia protettiva. Il vaccino antipertossico è solita-

mente combinato con il vaccino antitetanico e antidifterico, al quale si accomuna per modo e calendario di somministrazione. Per l'immunizzazione dei nuovi nati, di solito, oggi, viene utilizzato il vaccino esavalente, che, oltre a proteggere contro la pertosse, previene anche il tetano, la difterite, la poliomielite, l'epatite virale B e le infezioni invasive da Hib. La protezione conferita dalla vaccinazione antipertossica è di circa l'85%. La sua efficacia si riduce con il passare del tempo ed è, quindi, importante eseguire i richiami previsti. In genere, nei vaccinati la pertosse si presenta in forma meno grave.

- **Meningite meningococcica e vaccino antimeningite meningococcica**

La **meningite meningococcica** è una malattia batterica caratterizzata da un processo infiammatorio a carico delle meningi e provocato dalla *Neisseria meningitidis*. La trasmissione avviene per via respiratoria. I microorganismi di gruppo A e C sono responsabili delle maggiori epidemie. Al breve periodo di incubazione seguono: febbre elevata, cefalea intensa, nausea, vomito, rigidità nucale e, nei casi più gravi, delirio e coma. L'incidenza più alta si ha in inverno e in primavera. Attualmente la meningite meningococcica tende a colpire ragazzi e giovani adulti che vivono in condizioni di sovraffollamento, come le caserme. La somministrazione del vaccino liofilizzato provoca un'immunizzazione attiva contro la meningite cerebrospinale da meningococco di gruppo A e C. Ha un'efficacia superiore al 90%, conferisce un'immunità della durata di almeno 2 anni, può essere praticata a partire dal secondo anno di vita ed è particolarmente suggerita per:

- soggetti esposti al rischio di infezioni in zone epidemiche e altamente endemiche;
- componenti di gruppi familiari, collegi, caserme, comunità in cui si sia manifestato uno o più casi di malattia meningococcica.

Recentemente la presenza di meningiti non meningococciche nelle comunità scolastiche infantili ha condotto alla somministrazione facoltativa di vaccini antimeningite nella fascia di età neonatale.

f) Vaccini per malattie virali

- **Poliomielite e vaccino antipoliomielite orale secondo Sabin**

La **poliomielite** è una malattia infettiva causata dai *Poliovirus hominis 1, 2, 3* appartenenti al gruppo dei *Picornavirus*. È frequentemente asintomatica, ma in alcuni casi si può manifestare con paralisi muscolare e grave insufficienza respiratoria, spesso letale. Il contagio avviene per trasmissione interumana, sia diretta (via oro-fecale) sia mediata da oggetti contaminati. I poliovirus sono dotati di potere antigene specifico, pertanto l'immunità che si determina dopo la malattia è riferita solo al tipo di virus responsabile. Esistono due tipi di vaccino, entrambi efficaci nel prevenire la malattia: il vaccino IPV (vaccino antipolio inattivato o ucciso, di Salk) e il vaccino OPV (vaccino antipolio orale vivo attenuato, di Sabin). In Italia, attualmente, si utilizza il vaccino di Salk; quello di Sabin è stato impiegato fino a quando era importante ridurre la circolazione dei virus selvaggi. Per l'immunizzazione dei nuovi nati, di solito, viene utilizzato il vaccino esavalente, di cui abbiamo già parlato.

I vaccini di Salk e di Sabin sono molto efficaci e la loro protezione, che verosimilmente dura tutta la vita, è praticamente assoluta.

- **Morbillo e vaccino antimorbilloso**

Il virus del **morbillo** appartiene al gruppo dei *Paramyxovirus* e colpisce principalmente la prima e la seconda infanzia. Il morbillo presenta sia un andamento endemico, con riaccensioni epidemiche ogni 2-3 anni, sia un andamento stagionale, in quanto raggiunge il suo apice tra la fine dell'inverno e l'inizio della primavera. Questa malattia conferisce un'immunità permanente. L'incubazione dura circa 10 giorni. Il decorso clinico comprende dapprima un periodo catarrale o

di invasione di 3-5 giorni, in cui prevalgono febbre e mucositi, e successivamente: interessamento della mucosa congiuntivale (con arrossamento, fotofobia e lacrimazione); della mucosa nasale (con starnuti e rinorrea); della mucosa orale (enantema); della mucosa laringea e tracheale (con tosse stizzosa). Due giorni prima dell'esantema si formano, nella parte interna della bocca, all'altezza delle guance, piccole e caratteristiche papule biancastre con alone rosso vivo (*macchie di Koplik*). L'**esantema** compare circa 14 giorni dopo il contagio. La febbre raggiunge i 38°-39°C; il malato presenta fotofobia, lacrimazione, cefalea, sonnolenza, sete intensa. L'esantema, costituito da macule e papule rosee o rosse, compare prima sul viso (inizia in regione retroauricolare), per estendersi poi al torace, all'addome, agli arti e svanire al 6°-7° giorno. Il morbillo ha di per sé carattere benigno, ma possono essere gravi le complicanze, specie quelle a carico dell'apparato respiratorio (laringite, bronchiolite, broncopneumite); altre complicanze possono interessare l'apparato digerente, sotto forma di vomito e diarrea. Raramente (1 su 1.000 malati circa) il paziente può essere colpito da meningo-encefalite a patogenesi autoimmune, con il 10-15 % di decessi e il 40% di turbe mentali permanenti o epilessia. Esiste, inoltre, una correlazione tra virus del morbillo, panencefalite sclerosante subacuta (PESS) e sclerosi multipla, in quanto, negli individui affetti da queste malattie, si è rilevata la presenza, sia in circolo sia nel liquor, di anticorpi neutralizzanti il virus del morbillo. La vaccinazione contro il morbillo (vaccino MPR), effettuata con due dosi, ha un'efficacia del 98-99%. L'immunità dura tutta la vita. Poiché tale vaccino è disponibile solo nella forma trivalente, cioè associata con i vaccini antirosolia e antiparotite, con la medesima inoculazione si ottiene anche la protezione permanente contro la rosolia (efficacia del 100%) e la parotite (efficacia del 95%). Esso è costituito da virus attenuati, cioè modificati in modo da renderli innocui, ma capaci di stimolare le difese naturali dell'organismo.

- **Vaccinazione antiepatite B da DNA ricombinante**

L'attuale vaccino antiepatite B è costituito da una proteina della superficie del virus (HBsAg) sintetizzata utilizzando la tecnica del DNA ricombinante. Per l'immunizzazione dei nuovi nati, di solito, viene utilizzato il vaccino esavalente. Nei bambini si somministrano tre dosi di vaccino al 3°, 5° e 11° mese di vita. Nei neonati da madre infetta (HBsAg positiva), invece, sono necessarie quattro dosi: alla nascita, al 1°, 2° e 11-12° mese di vita; assieme alla prima dose di vaccino si somministrano le immunoglobuline. Negli adolescenti e negli adulti bastano tre dosi al tempo 0, e dopo 1 e 6 mesi. I vaccini antiepatite B sono efficaci al 95%, non necessitano di richiami e la loro protezione dura, in linea di massima, tutta la vita. L'importanza di dover praticare la **vaccinazione antiepatite B** risiede nel fatto che il virus in questione, contrariamente a quello dell'epatite A (a trasmissione oro-fecale), una volta contratto può determinare cronicizzazione della patologia, con prolungata possibilità di contagio della popolazione sana, nonché di evoluzione dell'epatite cronica in cirrosi epatica e/o in epatocarcinoma. Un andamento clinico simile viene assunto dall'epatite C, contro la quale però non esiste ancora una vaccinazione efficace.

- **Rosolia e vaccino antirubeolico**

La **rosolia** è una malattia infettiva acuta molto contagiosa causata dal *Rubella virus*, del gruppo dei *Togavirus*. Colpisce in genere i bambini, ma può interessare tutte le persone non immunizzate di qualsiasi età. Il contagio avviene con l'inalazione di goccioline di secrezione respiratoria disseminate dai malati. Le caratteristiche di questa malattia sono: esantema maculo-papuloso, con elementi di colorito roseo a diffusione cranio-caudale; tumefazione linfoghiandolare diffusa, prevalentemente retronucleare; plasmocitosi ematica; febbre incostante.

La malattia ha prognosi fondamentalmente benigna e le possibili complicazioni, artrite e nevrositi, sono estremamente rare. Un serio problema è invece determinato dalle donne che contraggono la malattia durante la gravidanza; quest'evenienza può infatti condurre a morte,

a malformazioni varie o a malattie congenite del feto. La patologia conferisce immunizzazione permanente. Il vaccino contro la rosolia è combinato con quello contro il morbillo e la parotite (vaccino MPR), ed è costituito da virus vivi attenuati, cioè modificati in modo da renderli innocui ma capaci di stimolare le difese naturali dell'organismo. La vaccinazione, effettuata con due dosi, ha un'efficacia del 100% nei confronti della rosolia, per la quale garantisce l'immunità per tutta la vita e consente anche la protezione permanente contro il morbillo (efficacia del 98-99%) e contro la parotite (efficacia superiore al 95%).

- **Influenza e vaccinazione antinfluenzale**

L'**influenza** è una malattia infettiva contagiosa a diffusione endemica ed epidemica, dovuta ad un virus del gruppo dei *Mixovirus*. Si tratta di una patologia dell'apparato respiratorio a contagio aerogeno (il virus si propaga attraverso goccioline di secrezioni respiratorie). In genere ha andamento benigno, ma può presentare gravi complicanze in soggetti particolarmente esposti, come anziani e bambini. Il periodo di incubazione è variabile, ma di solito breve (24-48 ore); la sintomatologia insorge improvvisamente con malessere, mal di testa, brividi e febbre fino a 39°C. Si ha poi accentuazione del malessere e comparsa di sintomi a carico dell'apparato respiratorio. Il malato si presenta sofferente, astenico, sonnolento. L'affezione ha decorso alquanto breve, con convalescenza caratterizzata da astenia e leggeri rialzi termici. L'influenza può presentarsi anche in forme più gravi, con decorso protratto; sono da temere soprattutto le complicanze a livello polmonare, che possono causare gravi ripercussioni sull'apparato cardiocircolatorio. La vaccinazione antinfluenzale è rappresentata dal vaccino polivalente tipi A e B adsorbito, il quale ha un'efficacia di circa il 75% e conferisce una durata dell'immunità in genere non superiore ad un anno. Per tale motivo e anche perché possono cambiare i ceppi in circolazione, è necessario ripetere la vaccinazione all'inizio di ogni stagione influenzale. Raramente essa può determinare la comparsa di effetti collaterali, quali reazioni locali, generali, allergiche, e può essere praticata a qualsiasi età, a partire dal 3° mese di vita. È indicata per tutte le persone ad alto rischio: broncopneumopatici cronici, cardiopatici, diabetici, soggetti con insufficienza renale, anziani al di sopra dei 65 anni, componenti di collettività industriali o di servizi pubblici e privati o di primario interesse (servizi sanitari, trasporti pubblici, scuole, collegi, case di riposo).

- **Vaccinazione antipapilloma**

Il **Papilloma virus (HPV)** è l'agente che può essere causa di infezioni genitali femminili e, a lunga distanza, anche del tumore della cervice uterina, il quale provoca ogni anno circa mille decessi. Oggi sono disponibili due vaccini, entrambi destinati ai due ceppi a maggior prevalenza, il 16 e il 18, ritenuti responsabili di circa il 70% dei tumori della cervice. Il vaccino ha un'elevata efficacia (compresa tra il 98 e il 100%) nei tre anni successivi alla somministrazione. Questo livello di protezione si riduce in modo rilevante quando consideriamo anche le lesioni prodotte da ceppi virali non coperti da vaccino (efficacia del 27%), e ancor più quando nell'analisi si includono ragazze già sessualmente attive (efficacia del 18%). L'utilizzo del vaccino, comunque, affianca ma non sostituisce lo **screening periodico attraverso il PAP test**, attualmente raccomandato per le donne di età compresa tra i 25 e i 64 anni.

g) Esantemi del bambino

L'**esantema**, letteralmente dal greco «fiorire fuori», è espressione sulla pelle del corpo di svariate condizioni patologiche: da stati infettivi a situazioni di allergia, a connettiviti, a vasculiti. Le malattie esantematiche sono quelle che più frequentemente si presentano nella prima e nella seconda infanzia, dopo l'influenza e le patologie delle prime vie respiratorie. Alcune di esse conservano ancora l'etichetta numerica (come la *quinta* o la *sesta malattia*).

Le **malattie esantematiche** dell'infanzia sono:

- I malattia o morbillo;
- II malattia o scarlattina;
- III malattia o rosolia;
- IV malattia o scarlattinetta;
- V malattia o megaloeritema infettivo;
- VI malattia o esantema critico.

• Scarlattina

La **scarlattina** è una malattia endemica con saltuarie annate epidemiche, causata da ceppi di *Streptococco beta emolitico di gruppo A*, produttori di una tossina eritrogenica responsabile dell'esantema. Quest'ultimo è diffuso, a piccole papule scarlatte confluenti, e si estende rapidamente al volto (risparmiando naso e mento: *triangolo di Filatow*), al tronco e agli arti (risparmiando palmo della mano e pianta del piede). L'esantema è preceduto da una febbre elevata, con angina eritematosa, arrossamento della punta e dei bordi della lingua che, successivamente, desquama, assumendo un tipico aspetto «a lampone». Segni patognomonicamente sono: la cute giallastra alla pressione, le strie rubre alla piega degli arti (*segno di Pastia*), il dermatografismo bianco allo strisciamento (*segno di Borsieri*), le petecchie dopo l'applicazione di laccio emostatico (*segno di Rumpell-Leed*). La diagnosi di certezza può essere effettuata con la positività del tampone faringeo per lo *Streptococco beta emolitico di gruppo A*. Le eventuali complicanze settiche e tossiche, dovute all'azione diretta della tossina, sono piuttosto rare. La malattia può recidivare.

• Quarta malattia o scarlattinetta

La **quarta malattia** (o **scarlattinetta**) è più unanimemente ritenuta una scarlattina ad espressione clinica attenuata. Recentemente è stata considerata come una conseguenza di un'infezione da *Stafilococco aureo*, le cui tossine (*tossina epidermolitica*) possono dare manifestazioni cutanee, che vanno da un esantema scarlattiniforme a formazioni bollose sulla pelle, con marcata desquamazione successiva.

• Quinta malattia o megaloeritema infettivo

La **quinta malattia**, o **megaloeritema infettivo**, è una malattia provocata dal *Parvovirus B19*. L'esantema, nelle forme tipiche, è preceduto da un periodo di incubazione perlopiù asintomatico, variabile da 4 a 20 giorni. La patologia colpisce, di solito, i bambini tra i 2 e i 12 anni e l'andamento si sviluppa in **tre fasi**:

- nella prima viene colpito il volto, con arrossamento delle guance, cute calda e leggermente rilevata (come se schiacciato); può essere presente la febbre;
- dopo 1-4 giorni inizia la seconda fase, durante la quale compaiono caratteristiche maculopapule, talora larghe alcuni centimetri, localizzate al tronco e alle superfici estensorie degli arti. L'esantema, al principio, si accompagna a modesto prurito, spesso confuso con un'allergia a farmaci, il quale, alla fine della seconda fase, assume un aspetto reticolare;
- la terza fase varia da 1 a 3 settimane ed è caratterizzata da attenuazione e recrudescenza dell'esantema, che resta il sintomo predominante, anche se ne possono essere presenti altri aspecifici.

Poiché l'esantema si manifesta nella metà dei soggetti colpiti, è consigliabile integrare le indagini di laboratorio con le *IgM anti parvovirus B19* in caso di sospetto clinico. Le complicanze più frequenti sono le artriti, che rappresentano la norma negli adulti, mentre interessano solo il 10% dei bambini. Le articolazioni più colpite sono: il ginocchio, l'anca e le interfalangee prossi-

mali. L'agente di questa malattia è in grado di attraversare la placenta; pertanto, anche se non si hanno informazioni complete sulla teratogenicità (sviluppo anomalo di organi nell'embrione) di tale virus, è consigliabile evitare contatti tra i soggetti affetti e le donne in gravidanza.

- **Sesta malattia o esantema critico**

L'agente eziologico della **sesta malattia**, o **esantema critico**, è l'*Herpes virus umano tipo 6* (HHV-6), identificato nel 1988. La malattia ricorre in piccole epidemie e colpisce i soggetti dai 6 mesi ai 3 anni di età. Il decorso clinico è caratteristico: dopo un periodo febbrile, di solito della durata di 3 giorni, con il ritorno della temperatura a valori normali compare un esantema, rappresentato da macule o maculo-papule roseoliformi non confluenti, non pruriginose e non desquamanti, localizzato prevalentemente al tronco e al collo. Poiché il quadro clinico è aspecifico, la diagnosi è effettuata in genere «a posteriori». In casi relativamente frequenti, la malattia si complica con convulsioni febbrili. Grazie all'uso delle *IgM anti HHV-6* in laboratorio, recentemente sono state identificate altre forme di manifestazione clinica della malattia: esantema senza febbre, febbre senza esantema, epatite fulminante, epato-splenomegalia marcata, sindrome emofagocitaria a decorso fatale, nevrassite e sue sequele.

- **Malattia di Kawasaki o sindrome linfo-muco-cutanea**

Una patologia importante, non inclusa nella classificazione delle malattie esantematiche, è la **malattia di Kawasaki**.

Identificata e descritta nel 1967 dal pediatra giapponese Tomisaku Kawasaki, è una *vasculite sistemica* caratterizzata da:

- febbre elevata, di durata superiore a 5 giorni, resistente al trattamento antibiotico;
- esantema a morfologia variabile, particolarmente diffuso al tronco;
- congiuntivite bilaterale non purulenta;
- alterazioni delle labbra, della mucosa orale e della lingua;
- alterazioni delle estremità, con eritema e/o edema delle mani e dei piedi, e successiva desquamazione lamellare;
- tumefazione dei linfonodi laterocervicali, di solito monolaterale.

A tutto ciò si può associare un corteo di *sintomi aspecifici* di tipo gastrointestinale (diarrea, vomito, dolori addominali, idrope della colecisti), artralgie e sintomi neuromeningei. In un simposio internazionale tenutosi a Tokio nel 1988, sono stati stabiliti precisi *criteri clinici per la diagnosi* di tale patologia, in quanto non è riconoscibile l'agente causale: VES molto elevata, PCR positiva, leucocitosi neutrofila, alfa-2 globuline aumentate, ipertransaminasemia, piastrinosi. Tale malattia pediatrica deve essere precocemente riconosciuta per consentire il trattamento specifico con acido acetilsalicilico e gamma-globuline, in quanto è causa di patologia coronarica (*aneurisma*) e cardiaca in una percentuale elevata di casi (20%). Nelle forme atipiche (*similinfluenzali*), il mancato riconoscimento può far progredire le alterazioni coronariche fino a conseguenze gravi.

Conoscenza e prevenzione della patologia incidentale

1 Patologia incidentale e malattie traumatiche

L'evoluzione tecnologica ha consentito il miglioramento dell'attività manuale sia in ambiente domestico sia in ambiente lavorativo; tuttavia, ciò ha determinato un aumento del numero degli *infortuni*, soprattutto per inadempienza alle misure precauzionali che andrebbero, invece, adottate in situazioni di rischio. Parallelamente, gli *incidenti stradali* rappresentano la più frequente causa di morte e di inabilità temporanea o permanente nella popolazione giovanile.

La **patologia incidentale** può essere ricondotta, quindi, a tre ampie categorie:

- quella degli *incidenti domestici*;
- quella degli *infortuni sul lavoro*;
- quella degli *incidenti stradali*.

L'educazione contro l'infortunistica, in generale, e contro l'infortunistica stradale, in particolare, rappresenta uno dei settori dell'*Educazione alla salute* attualmente in espansione. La scuola è la migliore sede per lo svolgimento di programmi educativi finalizzati alla conoscenza delle cause degli incidenti e alla loro prevenzione. La realizzazione di tali obiettivi contribuirebbe, infatti, sia alla riduzione delle malattie traumatiche, spesso conseguenti agli incidenti, sia alla diminuzione della mortalità globale.

Per **malattia traumatica** s'intende una patologia organica per la quale si riconosce l'esistenza di un rapporto causale tra la sua comparsa e un trauma. Tra le affezioni riconosciute come sicuramente collegabili ad un evento traumatico ricordiamo: le angiopatie, le artropatie, le fratture, le epilessie conseguenti a traumi cranici, le polmoniti e gli pneumotoraci, le spondilopatie, le ulcere da stress, le nevrosi post-traumatiche. Malattie che, invece, possono subire un aggravamento in seguito ad un evento traumatico sono: le cardiopatie, il diabete mellito, le nefropatie, le epatopatie, le psicosi e i tumori.

2 Educazione alla conoscenza delle cause e alla prevenzione della patologia incidentale

Per **incidenti domestici** si intendono eventi traumatici accidentali che si verificano in casa, in ambienti pubblici (scuole, palestre, discoteche etc.) e per la strada, escludendo gli incidenti stradali e quelli sul lavoro. In base alle richieste di Pronto Soccorso, ogni anno si contano oltre 1.500.000 di incidenti domestici, che, nella maggior parte dei casi, sono rappresentati da *traumi da caduta*. Questi ultimi possono essere molto gravi, in alcuni casi letali, e si verificano più frequentemente nei bambini di età inferiore ad un anno e nelle persone anziane. Molto diffuse sono anche le lesioni traumatiche causate da oggetti contundenti e le ferite dovute all'uso di strumenti da taglio. Alla prevenzione di tali infortuni si può provvedere soltanto con un'educazione volta alla responsabilizzazione personale nell'ambito delle azioni pratiche.

La prevenzione degli incidenti in ambiente domestico e di lavoro richiede, prima di tutto, l'attuazione delle norme di sicurezza sulle strutture edilizie, previste ormai a livello internazio-

nale. Gli impianti elettrici e quelli del gas, gli ascensori, le scale, gli elettrodomestici, gli utensili di lavoro, i prodotti chimici utilizzati dovranno essere tutti realizzati e resi funzionali secondo le normative vigenti nei Paesi dell'UE. L'ambiente domestico, inoltre, dovrà essere particolarmente controllato e reso non pericoloso soprattutto laddove se ne preveda la frequentazione da parte di bambini e anziani. La patologia incidentale più spesso riscontrata, in **casa** e sul **lavoro**, è rappresentata da *traumi, folgorazioni, ustioni e intossicazioni*.

a) Traumi

Il **trauma** è un'alterazione prodotta da una causa lesiva meccanica. Si parla di *lesione meccanica* anche quando si verifica una situazione in cui viene impiegata energia muscolare per vincere una resistenza (si compie cioè uno *sforzo*), che provoca danni organici a causa del cedimento di strutture interne.

Classificazione delle lesioni traumatiche		
Lesioni da corpi contundenti	+	escoriazioni ecchimosi ferite lacero-contuse rottture di visceri e fratture ossee
Grandi traumatismi	+	lesioni da precipitazione lesioni da investimento lesioni del conducente e del passeggero lesioni da esplosione lesioni da frane
Lesioni da strumenti speciali	+	da punta da taglio da fendente da punta e taglio da fuoco

Tab. 1 - Tipologie di lesioni traumatiche

Questa classificazione non va, tuttavia, intesa in senso stretto, in quanto spesso i vari eventi lesivi possono combinarsi tra loro.

Gli *effetti* di una lesione traumatica sull'organismo si distinguono in:

- *locali o diretti*;
- *a distanza o indiretti*;
- *general*i.

Le lesioni locali tendono, in genere, alla guarigione; quando permangono alterazioni anatomiche o funzionali si hanno i cosiddetti *reliquati*, che possono essere transitori o permanenti. In casi estremi, ad una lesione traumatica può conseguire la morte immediata (emorragia, embolia etc.). La morte non immediata può essere determinata, invece, da emorragie secondarie, shock traumatico, infezioni.

b) Folgorazioni e lesioni da elettricità

Le **folgorazioni** e le **lesioni da elettricità** sono provocate da *fonti di scariche elettriche*. Tra queste ultime, quelle che possono provocare danni all'organismo sono due:

- *l'elettricità atmosferica* (con le *folgori*);
- *le correnti elettriche industriali* (CEI).

Le CEI, coinvolte nell'infortunistica in misura nettamente superiore rispetto agli eventi atmosferici, possono essere distinte sia in *correnti ad alta e bassa tensione*, sia in *correnti continue e alternate*. Gli effetti sull'organismo dipendono principalmente dall'intensità della corrente in questione, tuttavia rappresenta un fattore molto importante anche la *resistenza* offerta dal corpo al passaggio della corrente stessa. Tale resistenza dipende dalla pelle del soggetto, dal grado di umidità, dalla superficie di contatto, dalla «messa a terra» e dalla durata dell'applicazione.

Le lesioni provocate in un organismo dal passaggio di una corrente elettrica possono essere *interne* o *esterne* e, nei casi più gravi, risultare letali. Le **lesioni interne** riguardano principalmente il muscolo cardiaco, gli altri muscoli in genere, i sistemi vascolare e nervoso. Le **lesioni esterne** sono quelle che si localizzano in corrispondenza dei punti di ingresso e, talvolta, di uscita delle correnti elettriche; sono rappresentate fondamentalmente da ferite somiglianti alle ustioni (*flittène* ed *èscare*) provocate dal calore prodotto nel punto di contatto (*effetto Joule*). Le ustioni possono giungere fino alla carbonizzazione; le tensioni molto alte possono comportare la fusione dell'osso.

c) Ustioni

Le **ustioni** si formano, in genere, per il contatto con la superficie corporea di gas, liquidi e solidi a temperatura elevata, o per azione diretta della fiamma o del calore irradiante. Vanno distinte le *scottature* dalle *bruciature*, in quanto solo le seconde causano la carbonizzazione dei tessuti.

Le ustioni rappresentano un'emergenza molto frequente, che, spesso, riguarda i bambini. I danni tissutali sono causati dal contatto con agenti termici, elettrici o chimici che — a seconda della gravità — coinvolgono l'epidermide, il derma, il tessuto sottocutaneo e i tessuti profondi. Possono essere interessate anche le mucose delle vie aeree e quelle digestive. Il danno fisiopatologico locale è costituito dalla denaturazione delle proteine con lesione e necrosi cellulare. Gli effetti sistemici possono provocare un grave stato di shock, che può essere *primario*, di tipo neurogeno, o *secondario*, di carattere ipovolemico (la diminuzione volumetrica del sangue nell'organismo). Frequente è, inoltre, l'insorgenza di infezioni gravi per la mancanza dell'azione di difesa esercitata normalmente dalla barriera cutanea integra, nei confronti dei microrganismi esterni.

A seconda del livello di profondità della distruzione cutanea, si distinguono *4 gradi di ustioni*, anche se spesso si trovano le diverse forme combinate tra loro:

- **ustioni di 1° grado**, che riguardano solo l'epidermide e provocano arrossamento e dolore (*eritema*);
- **ustioni di 2° grado**, le quali coinvolgono anche il derma e si manifestano con vescicole piene di liquido (*flittene* o bolle) che, alla rottura, cicatrizzano, guarendo spontaneamente. Frequentemente possono andare incontro ad infezione;
- **ustioni di 3° grado**, che sono molto profonde, si verificano quando la temperatura del corpo urente supera i 180°C e sono caratterizzate dalle *èscare* (croste). Spesso non sono dolorose per la distruzione delle terminazioni nervose e lasciano cicatrici permanenti (*cheloidi*);
- **ustioni di 4° grado**, le quali determinano una vera e propria carbonizzazione tissutale, che può riguardare anche muscoli e ossa.

La gravità dipende, oltre che dalla profondità dell'ustione, anche dalla sua estensione. In genere lo shock si verifica per ustioni che occupano un'area estesa all'incirca il 15-20% della superficie corporea, ma nei bambini anche per superfici minori.

d) Intossicazioni chimiche

Le **intossicazioni chimiche** (o *avvelenamenti*) accidentali possono verificarsi in ambiente domestico o nei luoghi di lavoro. L'*intossicazione domestica*, evenienza non sporadica, nella maggior parte dei casi consegue all'ingestione di sostanze tossiche e si manifesta spesso nei

bambini. Le sostanze più frequentemente causa di avvelenamento acuto sono i farmaci, i detersivi, i solventi, gli insetticidi, i derivati del petrolio. Le *intossicazioni alimentari* (ad esempio, il botulismo) e gli *avvelenamenti da ossido di carbonio* per il malfunzionamento di stufe e simili possono avvenire, ugualmente, in casa o negli ambienti comuni (caserme, ospizi). L'*intossicazione in ambiente lavorativo*, nella maggior parte dei casi, è dovuta all'esposizione continua a sostanze chimiche tossiche, le quali producono manifestazioni cliniche subdole, che assumono l'andamento di patologia cronica e sono appannaggio della Medicina del lavoro.

I principali sintomi delle **intossicazioni acute** sono rappresentati da: lievi disturbi psichiatrici o neurologici, dolori addominali, emorragie, febbre, ipotensione, congestione polmonare, eruzioni cutanee. La diagnosi è difficile: in genere, i sintomi si instaurano di colpo e possono essere collegati ad un evento specifico. Per il trattamento dell'intossicazione bisogna dapprima individuare l'agente tossico responsabile, quindi adottare una serie di norme:

- prevenire l'ulteriore assorbimento del veleno (ad esempio, mediante lo svuotamento dello stomaco);
- rimuovere dal corpo il veleno già assorbito (ad esempio, con la dialisi);
- effettuare una terapia sintomatica della fase acuta;
- usare antidoti sistemici specifici (ad esempio, antagonisti).

In Italia esistono centri specializzati per gli avvelenamenti e sarebbe buona norma tenerne a portata di mano i recapiti telefonici e la localizzazione. La prevenzione delle intossicazioni acute prevede l'educazione ad un corretto utilizzo dei prodotti chimici e dei farmaci nelle proprie abitazioni; è necessario tenere ciascun prodotto nel contenitore di origine che, secondo le norme attuali, deve essere dotato di un'apertura difficilmente accessibile ai bambini e mantenere ben visibile l'etichetta con le indicazioni su composizione, modalità d'uso e data di scadenza.

e) Incidenti stradali

Si definiscono **incidenti stradali** i sinistri in cui sono coinvolti veicoli a motore (automobili, motocicli, autocarri, autobus). L'incidente può coinvolgere il conducente, i passeggeri, i pedoni. La maggior parte della mortalità dovuta a traumi è provocata dagli infortuni stradali. L'energia meccanica dei veicoli in movimento è la causa principale dei danni provocati alle vittime di tali sinistri.

I **fattori di rischio** sono da ricondurre sostanzialmente:

- alle **condizioni dell'autoveicolo**: instabilità, scarsa capacità di ammortizzazione degli urti, insufficienza nei sistemi di protezione dei passeggeri (*airbag*, cinture di sicurezza), abitacolo a rischio;
- alle **condizioni della strada**: presenza di curve e dislivelli nel tracciato, manto stradale dissestato, traffico intenso, segnaletica inadeguata, condizioni atmosferiche avverse;
- al **comportamento** e allo **stato del conducente**: velocità eccessiva, guida imprudente, mancata osservanza delle norme del *Codice della Strada*, guida in stato di ebbrezza e/o sotto l'effetto di sostanze stupefacenti, difetti visivi.

Le tre **cause principali** di incidenti stradali sono:

- l'eccesso di velocità;
- il mancato uso dei dispositivi di sicurezza (cinture e casco);
- il mettersi alla guida dopo aver assunto alcol o droghe.

La **prevenzione** ha tre obiettivi fondamentali:

- limitare il numero di incidenti;
- renderne meno gravi le conseguenze;
- riabilitare i soggetti già colpiti da lesioni invalidanti.

Essa viene attuata attraverso due *canali preferenziali*:

- l'applicazione delle sanzioni amministrative e penali a coloro che non rispettano le normative dettate dal *Codice della Strada*;
- gli interventi educativi.

Da un punto di vista legislativo, dal 1° gennaio 1993 è in vigore il **nuovo Codice della Strada** (Decreto legislativo 30 aprile 1992, n. 285) che ha consentito all'Italia di adeguarsi agli standard europei in termini di educazione e di sicurezza stradale. La novità principale è stata rappresentata dalla trasformazione in illeciti di rilievo penale di alcuni comportamenti scorretti, in passato puniti con la sola sanzione amministrativa. Sono, infatti, riconosciuti quali *reati di condotta*: la marcia contromano, l'inversione su strade principali e autostrade, la circolazione su veicoli non idonei, l'omissione di soccorso. Particolarmente attuali appaiono poi l'**art. 186**, l'**art. 186bis** e l'**art. 187** che riguardano, rispettivamente, la guida in stato di ebbrezza; la guida sotto l'influenza dell'alcol per conducenti di età inferiore ai 21 anni, per i neo-patentati e per chi esercita professionalmente l'attività di trasporto di persone e cose; la guida in stato di alterazione psicofisica per uso di sostanze stupefacenti: le norme prevedono il divieto di guida in tali condizioni, sanzionandone l'inosservanza sul piano penale.

Più recentemente, nella seduta del 2 marzo 2016, il Senato ha introdotto nel *Codice penale* il delitto di **omicidio stradale** (art. 589bis), attraverso il quale, a titolo di colpa, è punito con la reclusione (di diversa entità, in ragione del grado della colpa) il conducente di veicoli a motore la cui condotta imprudente costituisca causa dell'evento mortale. In particolare:

- è confermata la fattispecie generica di omicidio colposo commesso con violazione delle norme sulla circolazione stradale (la pena rimane la reclusione da 2 a 7 anni);
- è punito con la reclusione da 8 a 12 anni l'omicidio stradale colposo commesso da conducenti di un veicolo a motore in stato di ebbrezza alcolica grave (tasso alcolemico superiore a 1,5 grammi per litro) o di alterazione psicofisica conseguente all'assunzione di sostanze stupefacenti o psicotrope. Se si tratta di conducenti professionali, per l'applicazione della stessa pena è sufficiente essere in stato di ebbrezza alcolica media (tasso alcolemico compreso tra 0,8 e 1,5 grammi per litro);
- è, invece, punito con la pena della reclusione da 5 a 10 anni l'omicidio stradale colposo commesso da conducenti di un veicolo a motore in stato di ebbrezza alcolica media, autori di specifici comportamenti connotati da imprudenza, ossia superamento di limiti di velocità; attraversamento di incroci con semaforo rosso; circolazione contromano; inversione di marcia in prossimità o in corrispondenza di intersezioni, curve o dossi; sorpassi azzardati.

La pena è diminuita fino alla metà quando l'omicidio stradale, pur cagionato dalle suddette condotte imprudenti, non sia esclusiva conseguenza dell'azione (o dell'omissione) del colpevole. La pena è, al contrario, aumentata se l'autore del reato non ha conseguito la patente (o ha la stessa sospesa o revocata) o non ha assicurato il proprio veicolo a motore. È previsto un aumento anche nel caso in cui il conducente provochi la morte di più persone, ovvero la morte di una o più persone e le lesioni di una o più persone. In queste circostanze si applica la pena che dovrebbe infliggersi per la più grave delle violazioni commesse, aumentata fino al triplo; il limite massimo di pena viene stabilito, però, in 18 anni. È stabilita, poi, una specifica **circostanza aggravante** nel caso in cui il conducente, responsabile di un omicidio stradale colposo, si sia dato alla *fuga*.

L'*idoneità fisica* alla guida deve essere verificata soprattutto per quanto riguarda le capacità visive del conducente e la presenza di eventuali malattie che potrebbero causare incidenti (epilessia, diabete etc.). Tale problema sta assumendo proporzioni progressivamente maggiori in considerazione dell'aumento dell'età media della popolazione generale e del numero di anziani alla guida di autovetture.

Gli **interventi formativi** di prevenzione dell'infortunistica stradale vanno attuati precocemente: l'educazione stradale del bambino e dell'adolescente deve essere praticata nella scuola dell'obbligo in maniera interattiva e interdisciplinare, soprattutto come integrazione all'educazione civica. L'utilizzo di sussidi multimediali consentirà all'insegnante di trasmettere in tempo reale dati aggiornati relativi all'incidenza e alla prevalenza degli incidenti stradali mondiali, oltre che alle principali cause, fornendo la possibilità di discutere sulle differenze col nostro Paese, dove le abitudini di guida sono diverse in confronto a molti altri. In particolare, i giovani dovranno essere sensibilizzati al rispetto dei limiti di velocità, nonché all'uso delle cinture di sicurezza nelle automobili e del casco sui motocicli. Sarà utile, inoltre, se possibile, far conoscere ai ragazzi le conseguenze a breve e a lungo termine dei danni provocati dai traumi stradali, attraverso interviste a soggetti coinvolti in esperienze reali.

Le campagne educative dovranno riguardare, infine, le principali regole da adottare quando ci si trova coinvolti direttamente o indirettamente in un incidente stradale. Sarà importante considerare cosa «fare» e cosa «non fare» nel prestare i primi soccorsi per ridurre i rischi di morte e di invalidità; valutare i comportamenti da adottare per favorire lo sgombero della carreggiata e lo scorrimento del traffico; aver chiare le informazioni da fornire per consentire un rapido e corretto aiuto da parte dei servizi di pronto soccorso e di primo intervento.

1 Lo sviluppo neuromotorio e psichico

Le tappe principali della *neurogenesì*, quali la proliferazione e la migrazione delle cellule, la differenziazione cellulare in senso neuronale, la formazione di dendriti, assoni e sinapsi, la mielinizzazione, si realizzano durante il periodo di sviluppo embrionale e fetale.

Nel neonato, dopo la nascita, continuano:

- la moltiplicazione dei neuroni cerebellari (fino alla fine del 1° anno di vita);
- la moltiplicazione delle cellule gliali (fino al 2° anno di vita);
- la mielinizzazione (fino al 4° anno di vita);
- la differenziazione della placca motrice.

In conseguenza della progressiva maturazione anatomica e funzionale del sistema nervoso, si verifica un cambiamento graduale dei parametri clinici man mano che dall'età neonatale si passa agli stadi successivi. Le modificazioni comportamentali relative alle diverse fasi di maturazione del sistema nervoso si realizzano in funzione dell'età concezionale del bambino, indipendentemente dall'età cronologica.

Nel feto, l'attività motoria inizia con la comparsa dei riflessi esterocettivi del viso che, poco dopo la 12ª settimana, lo rendono capace di succhiare. Successivamente tali riflessi si sviluppano in senso cranio-caudale e a 36 settimane compaiono i riflessi tonico-miotonici; dopodiché si attivano i muscoli flessori in senso caudo-craniale e si manifestano le prime reazioni motorie complesse. Dopo la nascita, il neonato giace abitualmente in posizione supina, con i quattro arti in flessione; si ha un'ipotonìa relativa dei muscoli estensori del collo e del tronco, e i movimenti sono afinalistici.

Nel neonato è possibile evocare numerose **reazioni motorie complesse caratteristiche**, che, in condizioni normali, scompaiono dopo il 1° anno di vita. Tra queste ricordiamo:

- il **riflesso di Moro**, che consiste in un sussulto del bambino, come un sobbalzo, accompagnato dall'apertura delle braccia, con le mani e le dita che si allargano e poi si richiudono a pugno. Osservando il piccolo si ha proprio l'idea dello spavento. Proprio per le modalità con cui si manifesta, è chiamato anche *riflesso di abbraccio o di caduta*;
- il **riflesso di prensione palmare e plantare**, che consiste nell'evocare una flessione delle dita applicando una leggera pressione sul palmo della mano o sulla pianta del piede;
- il **riflesso dei punti cardinali**, che si evoca stimolando la cute vicina alla labbra e consiste nella rotazione del capo e nell'apertura della bocca verso il lato stimolato;
- il **riflesso di raddrizzamento**, che si evoca sostenendo il neonato verticalmente, poggiandolo con i piedi su di un piano, e consiste nell'estensione e nell'irrigidimento degli arti e del tronco.

La valutazione della presenza delle reazioni motorie complesse è importantissima per verificare l'integrità e il livello di maturità funzionale del sistema nervoso.

Successivamente all'età neonatale, lo sviluppo neuromotorio continua, nel bambino sano, secondo una determinata sequenza, che dipende in parte dalla maturazione del sistema nervoso, in parte dall'apprendimento e dalle esperienze. La motilità diventa progressivamente sempre più finalistica, scompaiono gli automatismi neonatali e si manifestano, invece, fenomeni più complessi (ad esempio: *riflesso del paracadute*, *riflesso di Landau*). Parallelamente alla maturazione neuromotoria si assiste a quella delle funzioni psichiche, per cui lo *sviluppo* nel suo insieme può essere definito **sviluppo psicomotorio** e deve essere considerato nei quattro settori principali della *postura*, della *motilità fine*, del *linguaggio* e della *socialità*.

Tappe principali dello sviluppo psicomotorio nel bambino da 6 mesi a 3 anni	
6 - 9 mesi	— lallazione (ma-ma, pa-pa) — volge la testa verso un suono o se chiamato per nome — sta seduto anche senza sostegno — passa un oggetto da una mano all'altra — manipola gli oggetti con opposizione del pollice — inizia a mangiare da solo — manifesta ansia quando la madre si allontana — gattona
9 - 12 mesi	— inizia la comunicazione intenzionale (già dai 6 mesi comprende quando lo si rimprovera e gli si dice no) — si gira se qualcuno lo chiama — lallazione consonantica variata (sembra che conversi) — adopera parecchie vocali e consonanti — tentativi di imitazione di vocaboli prodotti dall'interlocutore — prime parole di senso compiuto — cammina lungo le pareti — sta in piedi da solo per un breve periodo — imita i suoni — indica un oggetto — afferra un oggetto opponendo pollice e indice
12 - 18 mesi	— prime parole riferite ad un oggetto — inizia la parola-frase — sta in piedi da solo — cammina senza sostegno con base allargata — dice qualche parola con significato (oltre a mamma e papà) — restituisce un oggetto su richiesta — assembla una torre di due pezzi con le costruzioni — cammina senza sostegno — corre e si arrampica — sale i gradini tenuto per mano — si toglie le scarpe — lancia una pallina — si alimenta da solo con le mani — riconosce alcune parole — costruisce una torre con tre pezzi

(Segue)

18 - 24 mesi	— cammina, corre, si arrampica — lancia la palla — calcia la palla — sale e scende le scale se aiutato — si alimenta da solo col cucchiaino — solleva una tazza — scarabocchia — cammina all'indietro — riconosce diverse parole — costruisce una torre con quattro pezzi — riconosce o indica alcuni animali domestici — riconosce o indica alcune parti del corpo — si toglie scarpe e calzini, apre la cerniera lampo
2-3 anni	— si spoglia — si arrampica su una sedia per raggiungere un oggetto — sfoglia un libro girando una pagina per volta — apre e chiude la porta — nomina gli oggetti più familiari — pronuncia alcune frasi — saltella sul posto — comincia a controllare gli sfinteri — usa un certo numero di parole — nomina gli oggetti semplici — associa alcune parole in frasi — nomina e indica un'immagine — inizia a porre le prime domande (dove?, cosa?) — ritrova un oggetto anche se non lo vede (la parola diventa simbolo) — inizia a pronunciare le frasi nucleari con competenza — utilizza i pronomi personali — il vocabolario si arricchisce
3 anni	— usa la struttura grammaticale dell'adulto (singolare, plurale, maschile, femminile, articolo, avverbi etc.) — articola frasi composte da più parole — sa farsi capire — va in triciclo — conosce il suo nome, l'età e il sesso — disegna un cerchio — si sbottona i vestiti — ripete i numeri — comincia a svestirsi — controlla gli sfinteri
dopo i 3 anni	— si arricchiscono le strutture lessicali e morfosintattiche

Tab. 1 - *Momenti fondamentali dello sviluppo psicomotorio infantile*

a) Screening dello sviluppo psicomotorio

Nel lattante e nel bambino non è sufficiente ricercare, mediante un esame neurologico, eventuali deficit, ma bisogna attuare un esame dello sviluppo psicomotorio confrontando, in

aree funzionali differenti, le capacità con i valori normali per l'età. Da un punto di vista pratico, tra le numerose prove di sviluppo a disposizione viene raccomandato il **test di Denver**, uno strumento di *screening* che dovrebbe essere utilizzato dai pediatri o dagli operatori sociali per individuare probabili casi di ritardo, per poi sottoporli agli specialisti.

In tale test, valido da 0 a 6 anni, il comportamento del bambino viene suddiviso in 4 settori: *postura, motricità fine e adattamento, linguaggio e socialità*. Il vantaggio del test di Denver è quello di offrire i *percentili* per l'età di acquisizione di certe abilità.

Nelle successive fasi di crescita e durante l'età adulta possono essere utilizzati numerosi test per la valutazione dei diversi aspetti psichici, mentre l'intelligenza del soggetto può essere determinata mediante l'**esame del Q.I.** (*quoziente di intelligenza*), che indica il rapporto tra l'età mentale e l'età cronologica. Tale quoziente, introdotto dallo psicologo e filosofo tedesco William Stern nel 1912, è un completamento della scala di misura dell'età mentale precedentemente definita dal pedagogo francese Alfred Binet.

b) Età scolare ed età mentale

La **valutazione clinica** dello sviluppo neuromotorio e delle capacità psichiche maturate è competenza del medico, in generale, e del pediatra o del neuropsichiatra infantile, in particolare. Un ruolo determinante viene svolto tuttavia dall'*insegnante*, il quale si trova quotidianamente a verificare che sia avvenuta una corretta armonizzazione tra apprendimento scolastico e sviluppo psichico. Ogni età scolastica coincide, infatti, con una sostanziale fase di sviluppo della capacità di apprendimento. Le componenti principali da considerare costantemente nella valutazione delle tappe dello sviluppo nel bambino e nell'adolescente sono l'*intelligenza*, le *motivazioni*, la *percezione*, il *linguaggio*, il *pensiero* e la *personalità*. Tali elementi vanno inquadrati secondo i diversi stadi di sviluppo sociale corrispondenti ai tre cicli principali dell'età scolare, che possono essere schematizzati come segue:

Cicli dell'età scolare
3-6 anni — scuola dell'infanzia (comunemente denominata <i>scuola materna</i>) — fase di autonomia funzionale — Importanza dell' <i>insegnante come modello</i> da imitare — Importanza del <i>gioco</i> per la promozione dello sviluppo creativo del bambino
6-11 anni — scuola primaria (comunemente denominata <i>scuola elementare</i>) — fase del pensiero reversibile — Acquisizione delle regole morali e sociali, nonché formazione del <i>Super Io</i>
11-14 anni — scuola secondaria di primo grado (comunemente denominata <i>scuola media</i>) — fase del pensiero operativo formale — Acquisizione delle capacità di ragionamento astratto — Importanza della pubertà e formazione dell' <i>Io ideale</i> — Importanza dell'atteggiamento e della comprensione degli insegnanti

Tab. 2 - Fasi di sviluppo sociale, corrispondenti ai tre cicli principali dell'età scolare

c) Principali cause di turbamento dello sviluppo neuromotorio e psichico

Numerosi fattori possono interferire con il normale sviluppo neuromotorio e psichico: componenti tossiche, infettive, genetiche e ambientali. In base al fattore di rischio determinante e al periodo in cui esso va ad agire, si possono distinguere diverse manifestazioni cliniche e funzionali, caratterizzate da situazioni di maggiore o minore gravità. I principali fattori da prendere

in considerazione e le tappe preferenziali della crescita sulle quali essi incidono possono essere riepilogati come segue:

Fattori legati al concepimento o al periodo preconcezionale
— malattie familiari, ereditarie, congenite (fibrosi cistica, distrofia muscolare) — sordità o cecità geneticamente determinate — anomalie cromosomiche — isoimmunizzazione materno-fetale (anticorpi anti RH)
Fattori legati al periodo della gravidanza
— scarsa o nulla tutela medica e sociale della gravidanza — lavori nocivi — radiazioni — fattori tossici (alcol, farmaci, droghe) — malattie infettive della madre (rosolia, toxoplasmosi) — altre malattie della madre (diabete, nefropatia, insufficienza cardiovascolare)
Fattori legati al momento della nascita o alla fase perinatale
— parto prematuro oppure oltre il termine — insufficiente o nulla assistenza ostetrica — presentazioni anomale del feto o anomalie del canale del parto — parti strumentali — anomalie della placenta o del cordone ombelicale — asfissia o anossia neonatale
Fattori legati al periodo neonatale
— ospedalizzazioni — assenza o carenza di cure materne — qualsiasi malattia o infezione grave nel neonato (meningiti, encefaliti) — turbe del metabolismo (ipoglicemia, ittero grave)
Fattori che insorgono in epoche successive
— meningiti ed encefaliti virali — intossicazioni — traumatismi — sindromi psichiche su base organica — sindromi psichiche su base funzionale (carenze dell'ambiente familiare e socio-educativo, che non permettono oppure ostacolano uno sviluppo armonico della personalità)

Tab. 3 - Fattori che possono alterare lo sviluppo neuromotorio e psichico di un individuo

2 La prevenzione delle nevrosi favorite dallo stile di vita

a) Stress

Lo **stress** è una situazione funzionale che si verifica nell'organismo quale reazione ad un pericolo o ad un evento che non rientrano nella vita ordinaria del soggetto. La **reazione di allarme** che si attua determina la **liberazione di adrenalina** dalle ghiandole surrenali, la quale provoca contrazione muscolare e riduzione dei tempi di reazione su tutti i distretti. Lo stress può essere considerato positivo (*eustress*) quando l'organismo, trovando un nuovo equilibrio, si adatta alla situazione; negativo (*distress*) quando l'adattamento non si verifica, per cui il disagio diviene cronico fino a sfociare nella patologia. I primi studi, effettuati sulle cavie di laboratorio, risalgono agli anni Sessanta del Novecento, quando venne evidenziato che lunghi periodi di

stress provocavano danni all'organismo e che le cellule maggiormente colpite erano quelle del sistema immunitario (timo, milza, linfonodi) e del sistema nervoso (ippocampo). Infatti, in condizioni di stress, l'ippocampo e l'ipotalamo inducono l'ipofisi a liberare l'ACTH (ormone adrenocorticotropo), che, a sua volta, stimola il surrene a produrre il cortisolo. Quest'ultimo, se sviluppato in eccesso, va a danneggiare i normali processi metabolici energetici di alcuni sistemi cellulari di tipo neuronale o immunitario, provocando disfunzioni che, a lungo termine, possono condurre alla depressione e ad un abbassamento delle difese immunitarie, aumentando il rischio di infezione. Quindi, se le occasioni di stress si ripetono o le situazioni stressanti si prolungano, l'individuo cade in uno stato di malessere che ne condiziona la salute. I sintomi e le conseguenze più frequenti sono rappresentati da: mal di testa, predisposizione alle malattie da raffreddamento, problemi cardiaci e gastrici, prurito, asma, insonnia, ansia e depressione, sudorazione anormale, impotenza o sterilità.

Prima di arrivare a tali manifestazioni cliniche dello stress, il soggetto incomincia ad avvertire segni aspecifici come:

- sensazione di stanchezza, anche appena sveglio;
- problemi relazionali, deterioramento dei rapporti con i familiari e con i colleghi;
- scoramento, mancanza di iniziativa, abulia, tendenza all'isolamento.

Lo stress può avere cause sia fisiche sia psichiche. Quelle più frequenti nella vita quotidiana sono rappresentate da:

- situazioni in cui il soggetto si trova a dover fronteggiare più difficoltà allo stesso tempo (ad esempio: i problemi di coppia insieme al carico di lavoro);
- situazioni in cui un problema persiste per un tempo troppo prolungato (ad esempio: il conflitto permanente con i colleghi di lavoro, la disoccupazione);
- situazioni in cui bisogna tollerare eventi estremamente gravi (ad esempio: un tracollo finanziario, una diagnosi di malattia).

L'individuo sottoposto a situazioni di questo genere può prevenire l'insorgenza dello stress patologico attraverso tre passaggi fondamentali:

- la presa di coscienza dei segnali d'allarme inviati dal proprio corpo;
- l'analisi delle soluzioni possibili ai problemi che hanno generato la situazione di stress;
- la compensazione attraverso ambiti definiti di vita (nuovi hobby, sport, impegni sociali).

La comparsa dello stress patologico può essere controbilanciata mediante l'utilizzo di **tecniche di rilassamento e metodiche specifiche** (training autogeno, visualizzazione, yoga, musica, farmaci). La prevenzione va attuata nelle scuole e negli ambienti di lavoro attraverso l'educazione a comportamenti adeguati, in termini qualitativi e quantitativi, nei confronti dei **ritmi incalzanti e innaturali** cui la società attuale costringe il nostro organismo.

b) Nervosismo, ansia, depressione, nevrosi e psicosi

Un termine molto diffuso nel linguaggio comune è **nervosismo**, con il quale s'intende una situazione di disagio, irrequietezza, tensione, irritabilità, facile emozionabilità, che spesso si associa ad alterazioni del sonno e dell'appetito, oltre che a cefalea. Quando lo stato di nervosismo diventa intenso e prolungato, si parla di **ansia**. Secondo Freud, l'ansia «è il fenomeno fondamentale e il problema centrale della nevrosi...» e per la Psicologia occidentale rappresenta la principale condizione mentale della società attuale. Da un punto di vista clinico, l'ansia si manifesta con sensazioni di paura e apprensione associate a sintomi neurovegetativi, come affanno, palpitazione, tensione muscolare, vertigini, tremori, sudorazioni, determinati da reazioni indotte nel sistema nervoso simpatico, nella tiroide e nelle ghiandole surrenali. L'ansia può

essere *episodica*, cioè verificarsi in una situazione particolare, o *prolungata*; frequentemente si associa o sfocia nella depressione.

La **depressione** è una normale reazione di scoraggiamento e disperazione della persona sana nei confronti di situazioni dolorose, frustranti o limitanti la propria libertà o i propri impulsi (lutti, separazioni, licenziamenti). In questi casi essa viene definita *esogena*, ha una durata limitata nel tempo, spesso non necessita di terapia aggressiva e difficilmente pone a rischio di suicidio. Quando lo stato depressivo è grave e immotivato rispetto alle cause che lo determinano, quando il paziente è infelice, lamenta una riduzione dell'energia fisica e psichica, della memoria e della capacità di concentrazione, manifesta tendenze paranoiche e/o ipocondriache, tenta il suicidio, allora ci si trova di fronte alla depressione *endogena*.

Sia l'ansia sia la depressione possono presentarsi come semplici reazioni a situazioni ambientali, oppure come vere e proprie nevrosi. Non bisogna, inoltre, sottovalutare che entrambe possano essere i primi sintomi di *patologie demenziali e involutive* nella popolazione anziana.

La **nevrosi** è un'affezione mentale caratterizzata da turbe funzionali senza compromissione della personalità; rappresenta l'interiorizzazione di un conflitto non risolto, che si manifesta sotto forma di *sintomo*. Essa è una condizione in cui il nervosismo, l'ansia e la depressione sono legate a diverse combinazioni di paure irrazionali, pensieri ossessivi, pulsioni, stanchezza e fatica, insonnia, preoccupazioni per sintomi insignificanti, malattie somatiche per le quali non si riesce a identificare una causa precisa. Le nevrosi sono malattie mentali funzionali, che possono comparire nell'adolescenza o in altre fasi della vita adulta; i pazienti affetti da tali disturbi sono vigili, correttamente orientati e vanno distinti dai pazienti con psicosi, i quali, invece, presentano sensazioni irreali e allucinazioni.

Esistono diversi tipi di nevrosi, in funzione della maggiore incidenza di determinati sintomi rispetto ad altri; in particolare, quelle più frequentemente riscontrate sono:

- la *nevrosi d'ansia*;
- la *nevrosi isterica* (tipica dell'adolescenza);
- la *nevrosi fobica*;
- la *nevrosi ossessiva*;
- la *nevrosi depressiva*;
- la *nevrosi ipocondriaca*.

Le **psicosi** sono malattie mentali che si manifestano con esordio improvviso, gravi disturbi del comportamento, dissociazione del pensiero. Lo *psicotico* è incapace di valutare adeguatamente la realtà e di rapportarsi correttamente alle altre persone. Il linguaggio utilizzato è incomprensibile o quasi nullo.

Le **cause delle malattie mentali**, in generale, e delle nevrosi, in particolare, non sono ancora note. Sono state elaborate numerosissime ipotesi su probabili ragioni organiche, genetiche, ambientali o psicologiche, ma nessuna sino ad oggi è stata confermata. Le nevrosi d'ansia e la depressione presentano, in alcuni casi, un andamento di tipo familiare. Secondo Freud, la nevrosi sarebbe scatenata dall'incapacità dell'individuo o dall'impossibilità a risolvere un proprio *conflitto interno*, che si origina quando due o più forze interiori si contrappongono tra loro; tali forze sono generate dai vari desideri e dalle differenti componenti psichiche della personalità (ad esempio, *Es* e *Super Io*). Le difese messe in atto per arginare la pulsione indesiderata o gli stati di angoscia dovuti all'emergere costante della pulsione rifiutata possono provocare la nevrosi.

È evidente come la vita quotidiana, oggi, sia continuamente causa di frustrazioni e conflitti tra quelle che sono le aspirazioni personali di tipo lavorativo, economico, estetico e quelle che, invece, sono le realizzazioni pratiche di tali obiettivi. I modelli proposti dai mass media, l'elevata scolarizzazione, il consumismo hanno determinato un aumento delle *aspettative* della popolazione giovanile nei confronti di sé stessi, dei genitori, degli insegnanti e della società in generale, con conseguenze spesso drammatiche qualora — e cioè molto spesso — esse vadano deluse.

Gli insuccessi farmacologici e psicoterapeutici conseguiti nel campo della terapia delle malattie mentali fanno assumere una consistente importanza agli **aspetti preventivi** dei disturbi psichici. La prevenzione delle nevrosi e dei disturbi della personalità, tuttavia, non potrà essere attuata in maniera corretta fino a quando non saranno stabilite le cause di tali patologie. Sicuramente appare importante il ruolo svolto dall'ambiente familiare e dall'istituzione scolastica nell'identificazione precoce dei disturbi mentali e di disadattamento. I genitori e gli insegnanti dovranno mostrarsi quali modelli di sicurezza da imitare, evitando comportamenti che potrebbero influire negativamente sullo sviluppo neuropsichico dei ragazzi e ricorrendo in prima persona ad operatori specialisti del settore, laddove se ne avverta la necessità. La scuola potrà contribuire alla crescita di una società psicologicamente meno fragile attraverso l'inserimento dei soggetti disadattati e portatori di disturbi mentali, attenuando la competizione tra i soggetti giovani normali ed educandoli a una maggiore tolleranza nei confronti delle frustrazioni.

Le malattie mentali sono oggetto di studio della **neuropsichiatria**, che si interessa di due categorie principali di disturbi:

- le **malattie del sistema nervoso**;
- le affezioni basate su **reazioni psicologiche anormali**.

Con il termine *malattia* si definisce una qualunque condizione che produca una lesione visibile, o presumibile, al sistema nervoso. Per *reazione psicologica anormale* si intende un disordine della vita psichica o della condotta, causato da esperienze di vita atipiche, con relativa incapacità a relazionarsi a livello sociale. Molte affezioni nervose, tuttavia, non sono così facilmente classificabili. Infatti, la personalità psicopatica, la nevrosi d'ansia o la depressione possono essere considerate sia come manifestazioni di malattia cerebrale sia come reazioni psicologiche anomale. Allo stesso modo, la schizofrenia e la psicosi maniaco-depressiva possono essere classificate tanto come malattie geneticamente determinate quanto come stati reattivi.

3 I problemi relativi alla malattia mentale e alla disabilità psicofisica

Nel mondo, dai dati dell'OMS, risulta che le persone non completamente autosufficienti rappresentano circa il 15% della popolazione. Nel nostro Paese i disabili sono circa 3 milioni, pari più o meno al 5% della popolazione. Ogni tentativo di catalogazione dell'handicap rappresenta una convenzione settoriale e riduttiva per facilitare l'approccio alla tematica. La **classificazione** che segue è stata proposta, tenendo conto di elementi medici, socio-biologici e psicologici, da Anna Chiodini e Maria Grazia Pedretti.

DISABILI FISICI	affetti da minorazione fisica
— distrofici	a causa di malattie muscolari degenerative
— motulési	a causa di malattie del sistema nervoso o di traumatismi
— epilettici	a causa di alterazioni cerebrali dovute a traumatismi, alterazioni metaboliche o senza causa accertata
— cardiopatici	affetti da gravi malformazioni cardiache
— idrocefali, microcefali e macrocefali	affetti da malformazioni craniche o cranico-facciali
DISABILI PSICHICI	affetti da minorazione psichica
— insufficienti mentali	per cause di carattere biopatico o encefalopatico, con manifestazione di incompleto o insufficiente sviluppo delle capacità intellettive
— soggetti Down	anomalie congenite (cromosomiche) caratterizzate da deficienza nello sviluppo psichico e da alterazioni somatiche

(Segue)

DISABILI SENSORIALI	affetti da minorazioni sensoriali
— ciechi, subvedenti, sordi, sordastri, sordomuti	
PLURIDISABILI	affetti da molteplici minorazioni di carattere fisico e/o psichico e/o sensoriale (spastico, insufficiente mentale, cieco etc.)

Tab. 4 - Classificazione delle disabilità

Gli epilettici, più specificamente, possono essere classificati come disabili psichici solo nei casi in cui la loro condizione comporti un'alterazione delle funzionalità mentali o del comportamento.

Negli ultimi anni si è verificato un progressivo incremento di normative e azioni pubbliche finalizzate all'inserimento dei disabili psicofisici nella società e nell'istituzione scolastica. L'edilizia, i mezzi di trasporto, i servizi e i locali pubblici, le strutture sanitarie sono stati progressivamente modificati o costruiti in funzione anche di tale parte della popolazione, sino a non molti anni fa del tutto ignorata. Le questioni istituzionali legate a tale problematica sono appannaggio della classe politica e amministrativa; le caratteristiche cliniche e patologiche, nonché le possibilità terapeutiche e riabilitative sono di competenza della classe medica; l'integrazione dei diversamente abili negli istituti formativi normali è compito della scuola, in generale, e dei docenti in particolare.

a) Disadattamenti scolastici e disabilità

Il problema degli alunni disabili e disadattati e della loro integrazione scolastica deve coinvolgere tutta la collettività. Nella società attuale è più che mai indispensabile non ritenere il disabile un diverso, ma accettarlo e considerarne le infinite possibilità di recupero. Il coinvolgimento degli insegnanti, delle famiglie, degli enti assistenziali, di sociologi, psicologi, medici e pedagogisti è indispensabile per affrontare il tema e per intervenire in modo continuativo e coordinato.

Affinché l'**inserimento dei soggetti portatori di handicap** nella scuola risulti corretto ed efficace, deve avvenire attraverso i seguenti momenti fondamentali:

- definizione dei ruoli e delle competenze del personale presente nella scuola e nel territorio coinvolto nell'intervento educativo (insegnante di sostegno, operatori sanitari);
- qualificazione professionale specifica del personale;
- coordinamento delle diverse professionalità per la realizzazione dell'unitarietà dell'intervento educativo;
- coinvolgimento dei genitori per garantire la continuità scuola-famiglia;
- adeguamento delle strutture scolastiche e del territorio sulla base delle necessità dei disabili;
- utilizzazione di specifici materiali didattici;
- educazione degli alunni sani all'integrazione dei soggetti diversamente abili.

I **programmi scolastici** offrono una serie di indicazioni utili al docente. Essi, infatti, consentono: una pianificazione che preveda la costruzione e la realizzazione di processi individuali di apprendimento; la preparazione di una diagnosi funzionale dei diversamente abili predisposta da servizi specialistici; interventi qualificati di didattica differenziata, con il coinvolgimento di sistema sanitario, scuola e strutture specializzate. L'istituzione scolastica deve poi darsi una maggiore articolazione e flessibilità, valorizzando esperienze formative come le classi aperte, i gruppi di studio e di attività, i gruppi di interclasse, la messa a punto di piani individuali di lavoro.

È importante che il docente sia consapevole che tra il soggetto normale e il soggetto portatore di handicap psicofisico esiste una certa differenza nella capacità di adattamento, che talvolta può anche manifestarsi in forma sfumata. Tali situazioni devono essere identificate precocemente, per consentire l'attuazione di programmi di recupero e di integrazione.

b) Instabilità psicomotoria

L'**instabilità psicomotoria** rappresenta una forma abbastanza frequente di *disadattamento scolastico*. Può essere dovuta a **immaturità neurologica** oppure a **disturbi psichici**. Il bambino — che spesso proviene da un ambiente familiare problematico — appare sempre in movimento, è di umore instabile, presenta scarsa capacità di concentrazione. Tale difficoltà, tuttavia, può essere indotta anche da un ambiente scolastico eccessivamente costrittivo.

c) Carenze nella coordinazione motoria, disgrafia e dislessia

Altre carenze molto frequenti, che il docente deve riconoscere precocemente sono rappresentate da:

- **difficoltà di coordinazione motoria**, per cui i bambini effettuano movimenti goffi e disordinati e non riescono ad inserirsi nei giochi con i compagni. Possono essere imputabili ad una disfunzione nella maturazione neuropsichica;
- **disgrafia** (fonologica o superficiale), ovvero un disturbo della scrittura che può essere recuperato attraverso la rieducazione psicomotoria;
- **dislessia**, ossia un disturbo della lettura attribuibile a cause congenite o a lesioni cerebrali. Esistono speciali addestramenti alla lettura per i soggetti dislessici.

d) Disabilità sensoriali

Le **disabilità sensoriali** più frequenti sono quelle *uditiv*e e quelle *visiv*e. L'**handicap uditivo** può essere *conduttivo*, se dipende da anomalie dell'orecchio esterno, o *perettivo*, se dipende da alterazioni delle strutture interne. In rari casi si può avere sordità totale. L'ipoacusia infantile deve essere precocemente riconosciuta per i gravi deficit di apprendimento e di inserimento sociale che ne possono derivare. L'**handicap visivo** è dovuto ad anomalie fisiche o meccaniche dell'occhio, spesso congenite. Anche in questo caso il deficit va riconosciuto al più presto per le gravi conseguenze che può comportare sia sull'apprendimento, sia sullo sviluppo emotivo, sia per l'aggravamento che la vista può subire a causa delle sollecitazioni esercitate dalla scrittura e dalla lettura.

e) Disturbi emotivi e cognitivi

I più frequenti **disturbi emotivi e cognitivi** sono la *fobia della scuola* e il *disadattamento scolastico*. Il bambino, come l'adulto, può esprimere attraverso il corpo una sofferenza psicologica. La sintomatologia riscontrata è molto ampia e presenta perlopiù *disturbi psicosomatici*, quali astenia, febbricola, pallore, sudorazione, anoressia, vomito, cefalea, dolori addominali e agli arti. Se tali affezioni si presentano ripetutamente al mattino prima di andare a scuola, si parla di **fobia della scuola**: insorge in soggetti normali tra i 5 e i 13 anni di età, è espressione della paura di distaccarsi dalla famiglia, si verifica più frequentemente nei figli unici e nei soggetti con rapporti atipici con i genitori (eccessiva protezione o severità). Questa ossessione è legata a sensazioni di panico e di angoscia scatenati dall'ambiente scolastico e, quando si manifesta in età più avanzata (ad esempio, nella scuola media), può essere espressione di *disadattamento*. La comprensione e l'accettazione da parte dei docenti e dei compagni di classe possono favorire il superamento delle crisi fobiche e l'inserimento sociale dell'alunno.

Le cause più frequenti di **disadattamento scolastico** possono essere:

- **personali**, se legate allo sviluppo fisico, psichico ed emotivo del soggetto e caratterizzate dalla presenza di situazioni come malattia, obesità, debolezza mentale, disturbi del linguaggio;
- **familiari**, se rappresentate dalle condizioni familiari del soggetto (iperprotezione, autoritarismo, separazione dei genitori);

- **socio-culturali**, se dipendenti dall'estrazione sociale e dalle condizioni economiche della famiglia;
- **scolastiche**, se riguardanti l'ambiente formativo, i metodi didattici, i programmi adottati. È determinante, in tal senso, il livello di integrazione raggiunto con i compagni di classe; tuttavia, il ruolo centrale in queste dinamiche deve essere assunto dal corpo docente, che dovrà mostrarsi adeguatamente preparato e sufficientemente maturo per raggiungere gli obiettivi prefissati dalla programmazione educativa e didattica.

f) Disabilità di tipo nevrotico e psicotico

Le **disabilità di genere nevrotico e psicotico** sono rappresentate da *disturbi della sfera affettivo-emotiva* e si presentano con caratteristiche diverse rispetto a quelle normalmente riscontrate nella popolazione adulta. Possono essere ricondotte a tre situazioni di disordine mentale: i *disturbi di reazione*, le *nevrosi* e le *psicosi*.

I **disturbi di reazione** sono comportamenti scatenati da situazioni specifiche e scompaiono una volta modificate le condizioni ambientali che li hanno evocati (ad esempio, fobia e disadattamento scolastici).

Le **nevrosi** dei soggetti in età evolutiva sono collegabili a conflitti inconsci irrisolti, provocati da una relazione inadeguata con uno o entrambi i genitori. Sintomi nevrotici dei soggetti giovani possono essere considerati la fobia della scuola, la claustrofobia, il rifiuto del cibo, l'agitazione motoria, i tic nervosi e così via.

Nei bambini le **psicosi** difficilmente possono essere distinte dalle nevrosi; si manifestano più frequentemente sotto forma di inespressività e di mancanza di risposta alle sollecitazioni verbali e comportamentali, in assenza di patologia organica evidente (ad esempio, la sordità). La forma di psicosi più frequente nella popolazione infantile è l'**autismo**: il bambino autistico appare completamente isolato dal mondo esterno, presenta incapacità di comunicare anche con i genitori, manifesta forme di attaccamento patologico nei confronti di oggetti, è apatico, distaccato, autoaggressivo, alternativamente immobile o ipercinetico, effettua spesso movimenti meccanici continui e ripetitivi (ad esempio, il dondolio). Le cause e le possibilità terapeutiche delle psicosi infantili sono tuttora oggetto di studio.

Indice

PARTE I Scienze matematiche

Introduzione Apprendimento matematico: l'importanza della didattica

1 Insegnare la Matematica.....	Pag.	7
2 La didattica della Matematica.....	»	7
3 Il contratto didattico.....	»	8
4 Le situazioni «didattiche» e quelle «a-didattiche».....	»	8
5 L'apprendimento della Matematica.....	»	9
6 Le indicazioni nazionali.....	»	9
Traguardi per lo sviluppo delle competenze al termine della scuola secondaria di primo grado.....	»	10
Obiettivi di apprendimento al termine della classe terza della scuola secondaria di primo grado.....	»	11

Capitolo 1 I momenti principali dello sviluppo del pensiero matematico

1 Introduzione.....	»	13
2 Dalle origini.....	»	13
3 La trigonometria: da ombra retta a cateto.....	»	16
4 Achille e la tartaruga.....	»	17
5 Le cifre arabe, ovvero le cifre indiane.....	»	18
6 Lo zero e i numeri negativi.....	»	18
7 I numeri razionali e il teorema di Pitagora.....	»	19
8 I numeri reali: la continuità.....	»	19
9 3,141592653589793..., ovvero l'irrealizzabile quadratura del cerchio.....	»	20
10 2,71828182845... oppure 10?.....	»	20
11 Il quadrato negativo.....	»	21
12 Fermat: da congettura a teorema.....	»	22
13 Un'infinità di infiniti.....	»	22
14 Il binomio di Newton, ovvero di Pascal.....	»	23
15 Il Teorema di L'Hôpital, ovvero di Bernoulli.....	»	23
16 Archimede e il segmento parabolico.....	»	23

Capitolo 2 Il linguaggio della teoria degli insiemi ed elementi di combinatoria

1 Gli insiemi.....	»	25
2 Le operazioni sugli insiemi.....	»	26
a) Unione.....	»	26
b) Intersezione.....	»	27
c) Differenza.....	»	28
3 Il prodotto cartesiano.....	»	28
4 Le relazioni.....	»	29
a) Relazione di equivalenza.....	»	30
b) Relazione d'ordine.....	»	30
5 Le strutture d'ordine.....	»	30

6 Le funzioni o applicazioni	Pag.	31
7 Cardinalità di un insieme, insiemi finiti e insiemi infiniti	»	32
8 Confronto tra insiemi infiniti e potenza di insiemi	»	33
9 Elementi di calcolo combinatorio	»	34
a) Disposizioni	»	34
b) Permutazioni	»	34
c) Combinazioni semplici.....	»	35
d) Formula del binomio di Newton.....	»	35
e) Regole per lo sviluppo della potenza di un binomio	»	36

Capitolo 3 Elementi di logica matematica

1 La logica.....	»	37
2 La dimostrazione.....	»	37
3 Nozioni di logica matematica	»	38
4 Le proposizioni	»	38
5 I connettivi	»	39
a) Congiunzione.....	»	39
b) Alternazione.....	»	40
c) Implicazione	»	40
d) Coimplicazione.....	»	41
e) Negazione	»	42
6 Dimostrazione, teorema, lemma e corollario	»	43
7 La proprietà transitiva della deduzione	»	44
8 La teoria assiomatica o ipotetico-deduttiva (i postulati).....	»	44
9 I postulati fondamentali della logica	»	44
10 Il metodo di riduzione all'assurdo.....	»	45
11 Le implicazioni derivate.....	»	46
12 Il teorema inverso o reciproco.....	»	46
13 Le proposizioni equivalenti	»	46
14 La prima legge delle inverse	»	47
15 La seconda legge delle inverse	»	47
16 Concetti primitivi e definizioni	»	49
17 Il concetto di astrazione.....	»	49
18 Il principio di induzione	»	56

Capitolo 4 La geometria euclidea del piano e dello spazio

I - La geometria euclidea del piano

1 Gli enti fondamentali della geometria del piano.....	»	59
a) Rette e loro porzioni	»	59
b) Angoli	»	61
2 La circonferenza e il cerchio	»	67
3 I poligoni	»	71
4 I triangoli.....	»	72
5 Alcuni quadrilateri.....	»	76
a) Parallelogramma	»	77
b) Trapezio	»	78
c) Quadrilateri inscrittibili e circoscrivibili, poligoni regolari.....	»	78
6 Il concetto di area.....	»	79
7 La similitudine.....	»	83

II - *La geometria euclidea dello spazio*

1 Punti, rette e piani nello spazio	Pag.	86
2 I poliedri.....	»	92
a) Alcuni poliedri.....	»	92
b) Poliedri regolari.....	»	97
3 I solidi di rotazione.....	»	99
4 Il concetto di volume.....	»	103

Capitolo 4 I sistemi numerici N, Z, Q, R, C e le strutture algebriche fondamentali

1 Estensione del concetto di numero: dai naturali ai complessi	»	105
a) Numeri naturali	»	105
b) Numeri relativi.....	»	105
c) Numeri razionali.....	»	106
d) Numeri reali	»	106
e) Numeri complessi.....	»	107
2 Numeri algebrici e numeri trascendenti	»	108
3 Le strutture algebriche.....	»	108
4 Le proprietà delle strutture algebriche	»	109
5 Struttura abeliana e struttura regolare.....	»	109
6 Semigrupp e gruppi	»	110
7 L'anello.....	»	110
8 Il campo.....	»	110
<i>Le equazioni diofantee.....</i>		(on line)

Capitolo 6 Il linguaggio dell'algebra lineare e il calcolo vettoriale

1 Le matrici.....	»	111
2 Matrici particolari.....	»	111
3 Le operazioni sulle matrici.....	»	112
a) Somma di due matrici.....	»	112
b) Differenza di due matrici.....	»	113
c) Prodotto di due matrici.....	»	113
d) Prodotto di una matrice per uno scalare	»	113
4 I determinanti.....	»	114
5 Le proprietà dei determinanti	»	115
6 L'inversa di una matrice.....	»	115
7 I sistemi di equazioni	»	116
8 La regola di Cramer	»	116
9 Il metodo di eliminazione di Gauss	»	117
10 Il teorema di Rouché-Capelli	»	118
11 I sistemi omogenei	»	119
12 Gli spazi vettoriali.....	»	119
13 La combinazione lineare nello spazio vettoriale	»	120
14 Le basi.....	»	120
15 I vettori.....	»	121
16 Le coordinate cartesiane di vettori	»	122
17 Le operazioni sui vettori.....	»	123
a) Somma di due vettori.....	»	123
b) Differenza di due vettori.....	»	123

c) Prodotto scalare	Pag. 124
d) Prodotto vettoriale di due vettori	» 125
e) Prodotto di un vettore per uno scalare	» 126

Capitolo 7 Il metodo delle coordinate per la descrizione di luoghi geometrici

1 La geometria analitica	» 127
2 Le coordinate sulla retta	» 127
3 Le coordinate cartesiane nel piano	» 127
4 La distanza di due punti	» 128
5 Le coordinate del punto medio di un segmento	» 129
6 La traslazione d'assi	» 130
7 La rappresentazione grafica di funzioni	» 130
8 L'equazione generale o implicita della retta	» 132
9 Le rette rispetto all'origine degli assi cartesiani	» 133
10 L'equazione della retta passante per un punto assegnato o per due punti assegnati	» 133
11 Rette parallele e rette perpendicolari	» 134
12 La distanza di un punto da una retta	» 136
13 Le coniche	» 137
14 La circonferenza	» 137
a) Mutua posizione di una circonferenza e di una retta	» 138
b) Mutua posizione di due circonferenze	» 139
c) Tangenti ad una circonferenza	» 139
15 L'ellisse	» 140
a) Eccentricità dell'ellisse	» 141
b) Tangenti ad un'ellisse	» 141
16 L'iperbole	» 141
a) Iperbole equilatera	» 143
b) Tangenti ad un'iperbole	» 144
17 La parabola	» 145
a) Equazione della parabola simmetrica rispetto all'asse delle y	» 145
b) Equazione della parabola simmetrica rispetto all'asse delle x	» 146
c) Parabola e funzione di secondo grado	» 147
d) Mutua posizione di una retta e di una parabola	» 147
e) Tangenti ad una parabola	» 148
18 L'equazione parametrica e cartesiana di un piano	» 149
19 L'equazione cartesiana di un piano passante per tre punti (non allineati)	» 149
20 Tre vettori complanari	» 150
21 Quattro punti complanari	» 151
22 Piani e vettori paralleli	» 151
23 Le rette nello spazio euclideo: equazioni parametriche di una retta	» 151
24 La direzione di una retta espressa in forma cartesiana	» 151
25 I fasci di rette nel piano euclideo	» 152
26 I fasci di piani nello spazio euclideo	» 153
27 La distanza fra due punti nello spazio	» 153
28 Il punto medio di un segmento nello spazio	» 154
29 Le superfici nello spazio	» 154
a) Equazione della sfera	» 154
b) Equazione di una superficie cilindrica che ha per asse l'asse z	» 155
c) Equazione segmentaria del piano	» 155
d) Equazione canonica di un ellissoide	» 156
e) Equazione canonica di un iperboloide a una falda e a due falde	» 156

f) Paraboloide ellittico	Pag. 157
g) Paraboloide iperbolico (o a sella)	» 158

Capitolo 8 Gli algoritmi

1 Nozioni di base	» 159
2 La traduzione di un algoritmo in linguaggio di programmazione	» 160
3 Le strutture elementari per la descrizione degli algoritmi	» 161
4 La rappresentazione grafica degli algoritmi	» 161
5 Il controllo della correttezza	» 162
6 La complessità degli algoritmi	» 162
7 Cenni sulla computabilità e sulla tesi di Church	» 162

Capitolo 9 Elementi di trigonometria e funzioni trigonometriche

1 La circonferenza trigonometrica	» 164
2 La relazione fondamentale	» 165
3 Le relazioni fra le funzioni trigonometriche	» 165
4 Le formule di addizione, sottrazione e duplicazione	» 165
5 Le relazioni trigonometriche applicate ai triangoli rettangoli	» 166
6 Le funzioni goniometriche	» 166
7 La funzione $y = \sin(x)$ e la sua inversa $y = \arcsin(x)$	» 166
8 La funzione $y = \cos(x)$ e la sua inversa $y = \arccos(x)$	» 167
9 La funzione $y = \tan(x)$ e la sua inversa $y = \arctan(x)$	» 168

Capitolo 10 Le funzioni reali di una variabile reale

1 Le funzioni	» 170
2 Intervallo e intorno	» 170
3 Il campo di esistenza di una funzione	» 170
4 Le funzioni limitate	» 171
5 Le funzioni crescenti e quelle decrescenti	» 172
6 Le funzioni composte e quelle inverse	» 172
7 Le funzioni elementari	» 172
a) Funzione potenza	» 172
b) Funzione radice	» 173
c) Funzione esponenziale di base a	» 174
d) Funzione logaritmo in base a	» 175
e) Funzione valore assoluto	» 175
8 I limiti di funzioni	» 175
9 I limiti destro e sinistro	» 176
10 Funzioni, limiti e infinito	» 176
11 I teoremi sui limiti di funzioni	» 177
12 Le operazioni sui limiti di funzioni	» 178
13 Il confronto di infinitesimi e di infiniti	» 179
14 Le funzioni continue	» 180
15 Le funzioni discontinue	» 180
16 I teoremi sulle funzioni continue	» 181
<i>Le funzioni polinomiali</i>	(on line)

Capitolo 11 Il calcolo differenziale per funzioni di una variabile reale

1 La derivata	Pag.	183
2 Le derivate destra e sinistra	»	183
3 Il significato geometrico della derivata	»	184
4 Il differenziale	»	184
5 Le regole di derivazione	»	185
6 Le derivate di funzioni composte e di funzioni inverse	»	186
7 Le derivate di ordine superiore	»	187
8 I teoremi sulle derivate	»	187
9 Il Teorema di L'Hôpital	»	188
10 Le relazioni tra derivate e funzioni crescenti o decrescenti	»	188
11 Massimi e minimi	»	189
12 Le concavità di una curva	»	190
13 Gli asintoti	»	190
14 Lo studio del grafico di una funzione	»	191

Capitolo 12 Calcolo integrale per funzioni di una variabile reale ed elementi di teoria della misura

1 L'integrale indefinito	»	192
2 L'integrazione per sostituzione	»	193
3 L'integrazione per decomposizione	»	194
4 L'integrazione per parti	»	195
5 L'integrale definito	»	196
6 Le proprietà dell'integrale definito	»	197
7 La relazione tra integrale indefinito e integrale definito	»	198
8 Le aree di superfici	»	199
9 I volumi dei solidi di rotazione	»	201

Capitolo 13 Successioni, serie numeriche ed equazioni differenziali

1 Insiemi e limiti	»	203
2 Le successioni	»	203
3 I limiti di successioni	»	203
4 I teoremi sui limiti di successioni	»	204
5 Le serie numeriche	»	205
6 Le serie geometriche	»	206
7 Le serie di funzioni	»	207
8 Le serie di potenze	»	208
9 La serie di Fourier	»	208
10 Le equazioni differenziali	»	209
11 Tipi di equazioni differenziali	»	210
a) Equazioni differenziali a variabili separate	»	210
b) Equazioni differenziali lineari del primo ordine	»	210
c) Equazioni differenziali lineari omogenee del secondo ordine a coefficienti costanti	»	211

Capitolo 14 I processi di approssimazione e di stima degli errori

1 Introduzione	»	212
2 Gli errori	»	212
a) Tipi di errori	»	213
b) Propagazione degli errori	»	214

3	L'interpolazione	Pag. 216
a)	Formula di interpolazione di Lagrange	» 216
b)	Formula di interpolazione di Newton	» 217
4	La risoluzione approssimata di equazioni	» 217
a)	Metodo di bisezione	» 217
b)	Metodo delle iterate successive	» 218
c)	Metodo di Newton	» 219
5	L'integrazione numerica	» 220
a)	Metodo dei rettangoli	» 220
b)	Metodo dei trapezi	» 221
c)	Metodo di Simpson	» 222

Capitolo 15 Elementi di statistica descrittiva

1	Indagine statistica e tabelle	» 224
2	Le distribuzioni statistiche semplici	» 225
a)	Variabili statistiche	» 225
b)	Mutabili statistiche	» 227
3	Le rappresentazioni grafiche	» 228
a)	Riferimento cartesiano ortogonale	» 228
b)	Ortogrammi	» 229
c)	Areogrammi per cerchi e per settori circolari	» 230
d)	Istogrammi	» 231
4	Gli indici statistici per variabili quantitative	» 232
5	Gli indici di posizione	» 232
a)	Media aritmetica	» 232
b)	Media quadratica	» 234
c)	Media armonica	» 235
d)	Media geometrica	» 236
e)	Relazioni tra le medie	» 238
f)	Cenni sulla media di somme di potenze	» 238
g)	Moda	» 238
h)	Mediana	» 240
i)	Percentili	» 242
6	Indici di variabilità	» 242
a)	Campo di variazione	» 242
b)	Differenza interquartilica	» 242
c)	Scostamento semplice medio dalla media aritmetica	» 242
d)	Scostamento semplice medio dalla mediana	» 243
e)	Scarto quadratico medio	» 243
f)	Devianza e varianza	» 244
g)	Differenze medie	» 246
h)	Indici rapportati al massimo della variabilità	» 246
i)	Indici di concentrazione	» 247
j)	Momenti	» 250
7	Indici di forma	» 251
a)	Indici di asimmetria	» 251
b)	Indice di curtosi	» 252
8	I rapporti statistici	» 253
9	Distribuzioni statistiche doppie	» 255
a)	Variabili statistiche doppie o distribuzioni bivariate	» 255
b)	Mutabili doppie	» 255

10 Connessione e concordanza/discordanza	Pag.	256
11 Regressione	»	256
a) Regressione lineare semplice	»	256
b) Regressione lineare multipla	»	258

Capitolo 16 Elementi di statistica inferenziale e di calcolo delle probabilità

1 Gli schemi di campionamento.....	»	260
a) Campioni casuali	»	260
b) Campioni ragionati.....	»	261
2 Gli eventi aleatori	»	261
a) Operazioni tra eventi.....	»	262
3 La probabilità	»	262
a) Impostazione oggettivista	»	263
b) Impostazione soggettivista.....	»	263
4 Probabilità composte, condizionate e totali	»	263
a) Teorema delle probabilità composte	»	263
b) Probabilità condizionate.....	»	264
c) Teorema delle probabilità totali.....	»	264
5 Il teorema di Bayes.....	»	265
6 Le variabili aleatorie o casuali	»	265
7 La distribuzione casuale binomiale	»	266
8 La distribuzione ipergeometrica	»	267
9 La distribuzione di Poisson.....	»	267
10 La distribuzione casuale normale o di Gauss	»	268
11 La distribuzione di Student	»	269
12 La distribuzione χ^2	»	269
13 La distribuzione F di Fisher-Snedecor.....	»	270
14 Teorema del limite centrale e legge dei grandi numeri.....	»	270
15 La stima dei parametri.....	»	271
a) Stima puntuale e stima per intervallo.....	»	271
b) Proprietà di uno stimatore puntuale.....	»	271
c) Come si ottiene una stima per intervallo	»	272
d) Diversità nei metodi di stima	»	272
e) Metodo della massima verosimiglianza	»	273
f) Metodo dei minimi quadrati	»	273
g) Metodo dei momenti.....	»	273
h) Metodo di Bayes	»	274
16 L'adeguatezza di un modello di regressione.....	»	274

Capitolo 17 Esami, problemi e concetti di interesse interdisciplinare

1 I numeri primi e la crittografia.....	»	276
2 Sistema decimale e sistema binario	»	277
3 Le operazioni nel sistema binario	»	278
4 Gli operatori booleani	»	278
5 I contributi di Hilbert, Turing e Godel.....	»	279
6 Il calcolatore elettronico	»	279
7 Un modello teorico della geometria del biliardo	»	281
8 Le trasformazioni	»	286
9 Funzioni di due variabili e loro applicazioni	»	288

10	La retta di bilancio	Pag.	289
11	Il tasso di cambio nominale.....	»	291
12	L'ottimizzazione vincolata	»	292
	a) Nozioni generali.....	»	292
	b) Programmazione lineare	»	292
	c) Il duale	»	292
13	L'interesse composto annuo	»	294
14	Lo spazio percorso da un punto	»	295
15	Il lavoro di una forza.....	»	295
16	L'equazione oraria del moto rettilineo uniforme.....	»	296
17	Equazioni differenziali e funzione armonica	»	296

Capitolo 18 I principali software per imparare e sperimentare la matematica

1	Derive	»	299
2	Lo schermo di Derive.....	»	299
3	Alcune funzioni di Derive	»	300
4	Cabri.....	»	302
5	Avvio di Cabri.....	»	303
6	Elenco di siti web da cui trarre materiale inerente ai principali software utilizzati in matematica	»	306

PARTE II Scienze

Introduzione Metodologie didattiche per lo studio delle Scienze

1	Il metodo tradizionale: lezione frontale.....	»	313
2	Il metodo operativo: laboratorio	»	314
3	Linguaggio comune e linguaggio scientifico	»	315
4	Traguardi per lo sviluppo delle competenze e obiettivi di apprendimento.....	»	315
	Traguardi per lo sviluppo delle competenze al termine della scuola secondaria di primo grado.....	»	316
	Obiettivi di apprendimento al termine della classe terza della scuola secondaria di primo grado.....	»	316

SEZIONE I Scienze chimiche

Capitolo 1 Il mondo della chimica

1	Il campo di studio della chimica.....	»	321
2	La materia.....	»	321
3	L'energia	»	322
4	L'equivalenza tra materia ed energia	»	323
5	Grandezze fisiche e sistemi di unità di misura	»	323
6	Il Sistema Internazionale delle unità di misura	»	324

Capitolo 2 La costituzione della materia

1 La natura elettrica della materia	Pag.	326
2 L'atomo	»	326
3 Gli elementi e i loro isotopi	»	327
4 Gli ioni	»	327
5 Le molecole	»	327
6 I pesi atomici e molecolari	»	328
7 Concetto di mole e numero di Avogadro	»	328

Capitolo 3 La struttura dell'atomo

1 Il modello di Rutherford	»	330
2 Il modello di Bohr	»	330
3 Le modifiche al modello di Bohr	»	331
4 La meccanica quantistica	»	332
5 Il modello ondulatorio dell'atomo	»	332
6 La costruzione delle configurazioni elettroniche degli elementi	»	336

Capitolo 4 Il sistema periodico degli elementi

1 Caratteristiche e struttura	»	338
2 Proprietà periodiche e comportamento chimico	»	340
a) Raggio atomico	»	340
b) Energia di ionizzazione	»	340
c) Affinità elettronica	»	341
d) Carattere metallico	»	341
3 La regola dell'ottetto	»	341
4 La notazione di Lewis	»	342

Capitolo 5 Il legame chimico

1 La specificità dei legami	»	343
2 Il legame atomico	»	343
a) Legame covalente omopolare	»	344
b) Legame covalente eteropolare	»	346
3 L'elettronegatività	»	346
4 Il legame dativo o di coordinazione	»	347
5 L'ibridizzazione degli orbitali	»	348
a) Ibrido sp o lineare	»	348
b) Ibrido sp^2 o planare	»	349
c) Ibrido sp^3 o tetraedrico	»	350
d) Ibridi sp^3d e sp^3d^2	»	350
6 Il legame a elettroni delocalizzati	»	350
7 Il legame ionico	»	352
8 I legami dipolari	»	352
9 Il legame idrogeno	»	353
10 Il legame metallico	»	354

Capitolo 6 Composti chimici e loro nomenclatura

1 Numero di ossidazione e valenza	Pag. 355
2 Le categorie di composti inorganici	» 356
3 Formule e nomenclatura	» 357
a) Nomenclatura di composti binari.....	» 357
b) Nomenclatura di ioni.....	» 357
c) Nomenclatura di sali	» 358
d) Nomenclatura di acidi.....	» 358

Capitolo 7 Stati di aggregazione e passaggi di stato

1 Lo stato solido	» 360
2 Lo stato liquido.....	» 361
a) Tensione superficiale	» 361
b) Viscosità	» 361
c) Pressione di vapore	» 361
3 Lo stato gassoso	» 362
4 Il modello del gas ideale	» 362
a) Legge di Maxwell-Boltzmann.....	» 363
b) Leggi di Boyle, di Charles e di Gay-Lussac.....	» 363
c) Equazione di stato del gas perfetto.....	» 364
d) Legge di Avogadro	» 365
e) Legge di Dalton	» 365
f) Legge di diffusione di Graham	» 366
g) Legge di Proust	» 367
5 L'equazione di Van der Waals	» 367
6 I passaggi di stato.....	» 368
7 I diagrammi di stato	» 369

Capitolo 8 La materia e le sue trasformazioni

1 Reazioni complete e di equilibrio	» 371
2 Equazioni chimiche e loro bilanciamento	» 371
3 Le reazioni di ossidoriduzione.....	» 373
4 Il bilanciamento delle reazioni di ossidoriduzione	» 373
a) Metodo della variazione del numero di ossidazione	» 374
b) Metodo ionico-elettronico	» 375
5 Ruggine e ossidazione dello zinco	» 377

Capitolo 9 I calcoli stechiometrici

1 Le reazioni complete.....	» 380
2 Le reazioni non complete: resa di reazione e resa percentuale.....	» 382

Capitolo 10 Le soluzioni

1 Definizione e composizione.....	» 384
2 Passaggio in soluzione e solubilità.....	» 384

3 Le proprietà di soluzioni di non elettroliti	Pag. 385
a) Variazione della pressione di vapore: legge di Raoult	» 385
b) Innalzamento ebullioscopico	» 386
c) Abbassamento crioscopico	» 386
d) Pressione osmotica	» 386
4 Le proprietà di soluzioni di elettroliti	» 387
5 L'elettrolisi dell'acqua	» 388

Capitolo 11 La termodinamica chimica

1 I concetti fondamentali	» 390
2 Entalpia e legge di Hess	» 390
3 Stato standard ed entalpie di reazione	» 391
4 I cicli termochimici	» 392
5 L'entropia: una misura del «grado di disordine»	» 393
6 Energia libera e spontaneità delle trasformazioni	» 394

Capitolo 12 La cinetica chimica

1 I meccanismi di reazione	» 395
2 La velocità di reazione	» 396
3 La catalisi	» 397

Capitolo 13 L'equilibrio chimico

1 Gli stati fisici degli equilibri chimici	» 399
2 La legge dell'azione di massa	» 399
3 L'influenza della pressione sull'equilibrio gassoso	» 400
4 La dipendenza della costante di equilibrio dalla temperatura	» 401
5 Lo spostamento dell'equilibrio	» 402
6 Gli equilibri eterogenei	» 403

Capitolo 14 Acidi e basi

1 Definizioni scientifiche degli acidi e delle basi	» 404
a) Definizione di Arrhenius	» 404
b) Definizione di Brønsted-Lowry	» 404
c) Definizione di Lewis	» 406
2 La forza degli acidi e delle basi	» 406
3 Elettroliti anfoteri e autoprotolisi dell'acqua	» 408
4 Il pH	» 408
5 Il calcolo del pH per soluzioni di acidi e basi	» 409
a) Acidi e basi forti diluitissime ($C_a < 10^{-6}M$)	» 410
b) Acidi e basi forti diluite ($10^{-5} < C_a < 10^{-2}M$)	» 410
c) Acidi e basi forti concentrate ($10^{-2} < C_a < 2M$)	» 410
d) Acidi e basi deboli diluite ($10^{-4} < C_a < 10^{-1}M$)	» 411
e) Acidi e basi deboli concentrate ($10^{-1} < C_a < 2M$)	» 411

Capitolo 15 Gli equilibri ionici in soluzione acquosa

1 L'idrolisi salina.....	Pag. 412
a) Idrolisi basica.....	» 412
b) Idrolisi acida	» 413
c) Idrolisi doppia.....	» 414
2 I sistemi tampone.....	» 414
a) Tipi di soluzioni tampone	» 414
b) Meccanismo di funzionamento delle soluzioni tampone	» 415
3 Le titolazioni acido-base.....	» 416
a) Titolazioni acido forte-base forte	» 417
b) Titolazioni acido debole-base forte	» 417
c) Indicatori di pH	» 419
4 Il prodotto di solubilità	» 420

Capitolo 16 I fondamenti della chimica organica

1 I legami tra atomi di carbonio	» 422
2 I gruppi funzionali	» 422
3 Gli idrocarburi	» 424
a) Alcani.....	» 424
b) Alcheni e alchini	» 425
c) Idrocarburi aromatici	» 426
4 Alcoli, fenoli ed eteri	» 428
5 Aldeidi, chetoni e acidi carbossilici	» 428
6 Gli esteri.....	» 429
7 I composti azotati.....	» 430

Capitolo 17 Scienza e tecnologia dei materiali

1 I materiali e le loro proprietà.....	» 432
2 Dalle risorse ai prodotti	» 434
3 Le materie seconde.....	» 435
4 Il legno	» 436
a) Proprietà	» 436
b) Legname	» 437
c) Il legno nell'arredamento: massello, pannelli impiallacciati e nobilitati	» 437
d) Attrezzature da falegnameria.....	» 438
5 La carta	» 438
a) Produzione artigianale.....	» 439
b) Produzione industriale.....	» 439
c) Tipologie	» 440
d) Supporti scrittori alternativi alla carta.....	» 442
6 Le fibre tessili.....	» 443
a) Cotone.....	» 443
b) Lino	» 444
c) Lana.....	» 444
d) Seta	» 444
e) Rayon	» 444
f) Fibre sintetiche	» 444

7 Il vetro.....	Pag. 445
a) Proprietà	» 445
b) Lavorazione	» 445
c) Produzione degli oggetti in vetro con incavi.....	» 446
d) Produzione delle lastre di vetro.....	» 446
8 I metalli	» 447
a) Proprietà	» 447
b) Come si producono e ricavano	» 448
c) Ferro, ghisa e acciaio.....	» 449
d) Alluminio.....	» 450
e) Il rame e le sue leghe.....	» 450
f) Altri metalli	» 451
9 I materiali fittili (ceramiche)	» 451
10 Le plastiche.....	» 452
a) Proprietà	» 453
b) Lavorazione	» 453
11 I materiali compositi da produzione	» 455

SEZIONE II Scienze fisiche

Capitolo 1 I sistemi di unità di misura

1 Grandezze fondamentali e grandezze derivate	» 459
2 Il Sistema Internazionale (S.I.).....	» 459
a) Grandezze fondamentali.....	» 459
b) Grandezze derivate.....	» 460
c) Prefissi moltiplicativi.....	» 463
d) Regole di scrittura	» 463
3 Il sistema C.G.S.	» 464

Capitolo 2 I vettori

1 Vettori e scalari.....	» 465
2 I versori	» 465
3 Il vettore opposto.....	» 465
4 La somma di vettori	» 465
a) Metodo geometrico	» 465
b) Metodo analitico.....	» 466
c) Proprietà della somma.....	» 467
5 La differenza di vettori	» 467
a) Metodo geometrico	» 468
b) Metodo analitico.....	» 468
c) Proprietà della differenza	» 469
6 La scomposizione di un vettore.....	» 469
7 La somma di più vettori	» 470
8 Il prodotto scalare o prodotto interno	» 470
a) Prodotto scalare tra i versori degli assi cartesiani.....	» 471
b) Proprietà del prodotto scalare	» 471
c) Prodotto scalare di due vettori in funzione delle loro componenti.....	» 471

9 Il prodotto vettoriale o prodotto esterno	Pag. 472
a) Regola della mano destra	» 472
b) Regola della vite destrorsa	» 472
c) Prodotto tra i versori degli assi cartesiani.....	» 473
d) Proprietà del prodotto vettoriale	» 473
e) Prodotto vettoriale di due vettori in funzione delle loro componenti.....	» 473

Capitolo 3 Le forze

1 La composizione di forze concorrenti	» 474
2 Il momento di una forza.....	» 474
a) Braccio di una forza	» 475
b) Calcolo del modulo di $\vec{M}$	» 475
c) Proprietà del momento di una forza	» 476
d) Momento di una forza in funzione delle componenti di $\vec{r}$ e $\vec{F}$	» 476
e) Momento di più forze concorrenti	» 476
3 Le forze applicate ad un corpo rigido.....	» 477
a) Coppia di forze.....	» 477
b) Momento della forza che agisce lungo la retta r_1 : $\vec{M}(r_1) = \vec{r} \times \vec{F}$	» 478
c) Momento della forza che agisce lungo la retta r_2 : $\vec{M}(r_2) = \vec{r}_2 \times \vec{F}$	» 478
4 Composizione di forze parallele e centro delle forze parallele	» 478
a) Baricentro di un corpo	» 479
b) Equilibrio di un corpo.....	» 480
c) Equilibrio stabile, instabile e indifferente	» 481
5 Le leve	» 482
a) Leva di primo genere: Forza resistente - Fulcro - Forza motrice	» 482
b) Leva di secondo genere: Fulcro - Forza resistente - Forza motrice	» 482
c) Leva di terzo genere: Fulcro - Forza motrice - Forza resistente	» 483

Capitolo 4 La cinematica

1 I termini della cinematica	» 484
2 Il moto rettilineo uniforme	» 485
a) Accelerazione media e accelerazione istantanea	» 485
b) Velocità media e velocità istantanea.....	» 485
c) Spazio percorso	» 485
3 Il moto rettilineo uniformemente accelerato	» 486
a) Accelerazione media e accelerazione istantanea	» 486
b) Velocità media e velocità istantanea.....	» 487
c) Spazio percorso	» 487
4 La caduta dei gravi	» 489
5 Il moto curvilineo	» 490
a) Velocità	» 490
b) Vettore spostamento e velocità in funzione delle componenti	» 491
c) Moto curvilineo e accelerazione	» 491
d) Vettore accelerazione in funzione delle componenti	» 492
e) Moto piano con accelerazione costante (moto uniformemente accelerato)	» 492
f) Moto di un proiettile.....	» 492
g) Studio delle componenti di $\vec{v}_0$	» 494
h) Componente tangenziale e normale dell'accelerazione.....	» 495

6 Il moto circolare uniforme	Pag. 497
a) Periodo T	» 497
b) Frequenza ν	» 497
c) Velocità tangenziale e velocità angolare.....	» 497
d) Accelerazione centripeta.....	» 498
7 Il moto circolare uniformemente accelerato	» 499
8 La composizione degli spostamenti.....	» 499
9 La composizione delle velocità	» 500
10 La composizione delle accelerazioni.....	» 501

Capitolo 5 Il moto relativo

1 La composizione degli spostamenti.....	» 502
2 La composizione delle velocità	» 502
3 La composizione delle accelerazioni.....	» 503
4 Le trasformazioni di Galileo.....	» 503
a) Il problema della velocità della luce	» 504
5 Le trasformazioni di Lorentz.....	» 505
6 Le conseguenze delle trasformazioni di Lorentz	» 507
a) Contrazione delle lunghezze	» 507
b) Dilatazione dei tempi	» 507

Capitolo 6 La dinamica

1 Il primo principio della dinamica (principio d'inerzia)	» 509
2 Il secondo principio della dinamica (legge di Newton).....	» 509
a) Massa inerziale	» 509
b) Equazione di Newton: $\vec{F} = m\vec{a}$	» 509
c) Peso di un corpo.....	» 510
d) Chilogrammo massa e chilogrammo peso	» 510
3 Il terzo principio della dinamica.....	» 510
4 Il moto di un corpo su un piano inclinato.....	» 511
a) Asse x	» 511
b) Asse y	» 511
5 L'attrito.....	» 512
6 Forza centripeta e forza centrifuga.....	» 512

Capitolo 7 Moto oscillatorio e molle

1 Proprietà generali.....	» 514
2 L'equazione oraria del moto armonico semplice	» 514
a) Equazione della velocità	» 516
b) Equazione dell'accelerazione	» 516
3 Il moto di un corpo soggetto a una forza elastica: la legge di Hooke	» 518
4 Il lavoro della forza elastica.....	» 518
5 L'energia potenziale elastica	» 520

Capitolo 8 Lavoro ed energia

1 Il lavoro	Pag. 521
a) Dimensioni	» 521
b) Unità di misura	» 521
c) Lavoro positivo e lavoro negativo	» 521
d) Lavoro di una forza variabile	» 522
2 Energia cinetica e teorema dell'energia cinetica	» 522
3 La potenza	» 523
a) Kilowattora	» 523
4 Forze conservative e forze non conservative	» 523
5 L'energia potenziale gravitazionale	» 524

Capitolo 9 Impulso e quantità di moto

1 Quantità di moto e impulso di una forza	» 526
a) Quantità di moto e secondo principio della dinamica	» 526
b) Impulso di una forza	» 526
c) Conservazione della quantità di moto	» 527
d) Forze esterne e forze interne in un sistema	» 527
2 La quantità di moto di un sistema di particelle	» 527
a) Centro di massa di un sistema	» 528
3 Il momento angolare	» 529
4 Gli urti	» 530
a) Urti elastici	» 530
b) Urti anelastici	» 531
c) Coefficiente di restituzione	» 532

Capitolo 10 La dinamica di un corpo rigido

1 Il momento angolare di un corpo rigido	» 533
a) Definizione di momento angolare	» 534
b) Momento d'inerzia	» 535
c) Differenza tra momento d'inerzia e massa inerziale	» 535
d) Calcolo del momento d'inerzia	» 536
e) Teorema di Steiner	» 539
2 La dinamica rotazionale di un corpo rigido	» 540
a) Energia cinetica rotazionale	» 540
b) Lavoro e potenza rotazionale	» 540
3 Il moto rototraslatorio di un corpo rigido	» 542
4 La conservazione del momento angolare	» 543

Capitolo 11 La gravitazione universale

1 Un po' di storia	» 545
2 Le leggi di Keplero	» 545
a) Prima legge	» 545
b) Seconda legge	» 546
c) Terza legge	» 546
3 Da Keplero a Newton	» 547
a) Generalizzazione della legge di gravitazione universale	» 548
4 Massa inerziale e massa gravitazionale	» 549

5 Il campo gravitazionale	Pag. 550
a) Variazioni di accelerazione dovute alla distanza dal centro della Terra	» 550
6 Lavoro ed energia potenziale gravitazionale	» 551
a) Lavoro della forza gravitazionale	» 551
b) Energia potenziale gravitazionale	» 552
c) Conservazione dell'energia meccanica	» 553
7 Il lancio di un satellite terrestre	» 553
<i>Cavendish e la costante G</i>	(on line)

Capitolo 12 La statica dei fluidi

1 Densità e pressione	» 555
a) Densità	» 555
b) Pressione	» 555
2 La trasmissione delle forze nei fluidi	» 556
a) Principio di Pascal	» 556
3 Variazione di pressione in un fluido a riposo e legge di Stevino	» 556
a) Pressione idrostatica	» 558
b) Conseguenza della legge di Stevino: il paradosso idrostatico	» 558
4 Il principio di Archimede	» 559
a) Enunciato	» 559
b) Corpi che affondano e corpi che galleggiano	» 560

Capitolo 13 La dinamica dei fluidi

1 Fluidi stazionari e non stazionari	» 561
2 L'equazione di continuità	» 562
3 L'equazione di Bernoulli	» 562
4 Il teorema di Torricelli	» 564

Capitolo 14 Onde e loro propagazione

1 Onde e moto ondulatorio	» 565
2 Onde trasversali e longitudinali	» 566
a) Onde trasversali	» 566
b) Onde longitudinali	» 566
3 Le grandezze che caratterizzano un'onda	» 566
a) Creste, gole, ampiezza e lunghezza	» 566
b) Frequenza, periodo e velocità	» 567
c) Energia trasportata e ampiezza	» 568
d) Equazione d'onda	» 568
4 Riflessione, rifrazione e diffrazione delle onde	» 568
a) Fronte o superficie d'onda	» 568
b) Raggi d'onda	» 568
c) Riflessione	» 569
d) Rifrazione	» 569
e) Diffrazione	» 570
5 Il principio di Huygens	» 570
6 Il principio di sovrapposizione	» 571
7 L'interferenza	» 571
8 Le onde stazionarie	» 571

Capitolo 15 Il suono

1 Le onde sonore	Pag. 574
a) Suono come onda elastica.....	» 574
b) Suono come onda longitudinale	» 574
c) Suono come onda di pressione	» 575
2 La velocità di propagazione delle onde sonore.....	» 575
a) Velocità del suono nei gas.....	» 578
b) Velocità del suono nei solidi	» 578
c) Velocità del suono nei liquidi.....	» 578
3 Le proprietà del suono	» 578
a) Frequenza	» 578
b) Tono	» 579
c) Intensità.....	» 579
d) Rimbombo ed eco	» 581
4 L'interferenza.....	» 581
5 I battimenti.....	» 582
a) Frequenza di battimento	» 582
6 L'effetto Doppler.....	» 583
7 Gli effetti supersonici	» 584

Capitolo 16 Luce e colori

1 La luce: un'onda elettromagnetica.....	» 585
2 La percezione dei colori.....	» 586
a) Riflessione, assorbimento e trasmissione della luce.....	» 586
b) Addizione dei colori.....	» 586
c) Sottrazione dei colori	» 587
d) Pigmenti.....	» 587

Capitolo 17 Luce e riflessione

1 Corpi luminosi e corpi illuminati	» 588
a) Linea visiva	» 588
b) Legge della riflessione.....	» 588
2 Gli specchi piani	» 589
a) Caratteristiche delle immagini formate dagli specchi piani	» 590
3 Gli specchi concavi.....	» 590
a) Formazione di un'immagine.....	» 590
b) Regole di riflessione	» 591
c) Come ricavare l'immagine riflessa	» 591
d) Posizione e dimensioni dell'immagine di un oggetto	» 592
4 Gli specchi convessi	» 594
a) Formazione dell'immagine.....	» 594
b) Regole di riflessione	» 594
c) Come ricavare l'immagine riflessa	» 595
d) Posizione e dimensioni dell'immagine di un oggetto	» 596

Capitolo 18 Luce e rifrazione

1 Il fenomeno della rifrazione	Pag. 597
a) Cause della rifrazione.....	» 597
b) Effetti ottici dovuti alla rifrazione.....	» 598
c) Materiali e rifrazione: densità ottica e indice di rifrazione.....	» 598
d) Direzione della rifrazione	» 599
2 La legge di Snell	» 600
3 La riflessione totale.....	» 601
a) Angolo limite.....	» 602
b) Legge di Snell e lunghezze d'onda.....	» 602
4 Le lenti	» 603
a) Lenti convergenti e lenti divergenti.....	» 603
b) Rifrazione delle lenti	» 603
c) Rifrazione da una lente convergente	» 604
d) Rifrazione da una lente divergente	» 604

Capitolo 19 La luce: interferenza e diffrazione

1 L'interferenza	» 606
a) Onde coerenti	» 606
2 I metodi di osservazione dell'interferenza della luce	» 608
a) L'esperimento della doppia fenditura di Young.....	» 608
3 La diffrazione.....	» 611
a) Diffrazione da una singola fenditura (diffrazione di Fraunhofer).....	» 611
b) Reticolo di diffrazione	» 613

Capitolo 20 Temperatura e dilatazione termica

1 Calore e temperatura	» 614
2 Il termometro	» 614
3 Le sostanze termometriche	» 615
a) Dipendenza della pressione dalla temperatura	» 615
b) Zero assoluto	» 616
4 Gli effetti della temperatura sul volume dei gas.....	» 616
5 Gli effetti della temperatura sul volume e sulla pressione di un liquido.....	» 617
6 Gli effetti della temperatura sul volume e sulla pressione di un solido	» 617
a) Dilatazione lineare	» 617
b) Dilatazione superficiale	» 618
c) Dilatazione cubica	» 619

Capitolo 21 Il calore

1 Calore, temperatura ed equilibrio termico	» 620
a) Principio zero della termodinamica.....	» 620
2 Capacità termica e calore specifico di un corpo.....	» 621
a) Capacità termica	» 621
b) Calore specifico	» 621
c) Caloria.....	» 621
3 Energia, calore e lavoro	» 622

4	La propagazione del calore	Pag. 623
a)	Conduzione.....	» 624
b)	Convezione	» 624
c)	Irraggiamento	» 625
5	Calore e cambiamenti di stato	» 625
a)	Fusione e solidificazione	» 626
b)	Condensazione e vaporizzazione	» 626
6	Il calore latente.....	» 627

Capitolo 22 I gas perfetti

1	Atomo, molecola e mole	» 629
2	Le proprietà dei gas perfetti	» 629
3	Le variabili termodinamiche.....	» 629
4	Trasformazioni termodinamiche e leggi dei gas perfetti	» 630
a)	Trasformazioni isobare e prima legge di Gay-Lussac	» 631
b)	Trasformazioni isocore e seconda legge di Gay-Lussac	» 632
c)	Trasformazioni isoterme e legge di Boyle	» 632
d)	Principio di Avogadro	» 633
e)	Legge di Dalton	» 633
5	L'equazione di stato dei gas perfetti	» 633

Capitolo 23 Il primo principio della termodinamica

1	Calore, lavoro ed energia.....	» 636
2	L'energia interna di un sistema termodinamico.....	» 638
3	Trasformazione termodinamica, trasformazione inversa e ciclo termico di un sistema	» 638
a)	Trasformazione termodinamica.....	» 639
b)	Trasformazione inversa.....	» 639
c)	Ciclo termico.....	» 639
4	Il lavoro in una trasformazione termodinamica.....	» 640
a)	Lavoro di una trasformazione isobara	» 640
b)	Lavoro di una trasformazione isocora.....	» 641
5	Il primo principio della termodinamica	» 642
6	L'energia interna di un gas perfetto	» 642
a)	Trasformazione a volume costante	» 643
b)	Trasformazione a pressione costante	» 643
c)	Trasformazione ciclica	» 644
d)	Trasformazione isocora.....	» 644
e)	Trasformazione adiabatica	» 644
f)	Processo di ebollizione	» 645

Capitolo 24 Il secondo principio della termodinamica

1	Gli enunciati di Kelvin-Planck e di Clausius.....	» 646
a)	Enunciato di Kelvin-Planck.....	» 646
b)	Enunciato di Clausius.....	» 646
c)	Equivalenza fra l'enunciato di Kelvin-Planck e l'enunciato di Clausius	» 647
d)	Rendimento di una macchina termica.....	» 648

2 Il teorema di Carnot	Pag. 649
a) Ciclo di Carnot	» 649
b) Efficienza di una macchina termica.....	» 651
3 L'entropia.....	» 651
a) Entropia dell'Universo	» 653

Capitolo 25 Carica elettrica e campi elettrici

1 La struttura atomica	» 655
2 La quantità di carica.....	» 655
3 Conduttori, isolanti e dielettrici	» 656
a) Conduttori	» 656
b) Isolanti	» 656
c) Dielettrici	» 656
4 La legge di Coulomb.....	» 656
5 Confronto tra forza elettrica e forza gravitazionale.....	» 657
6 Le forze elettriche in un sistema di cariche.....	» 658
7 Il campo elettrico	» 658
a) Campo elettrico di una carica puntiforme	» 660
b) Linee di forza	» 661
c) Campi elettrici generati da più cariche	» 662
8 Il flusso del campo elettrico.....	» 663
9 Il teorema di Gauss.....	» 664
a) Flusso positivo e flusso negativo	» 664
b) Come scegliere la superficie gaussiana.....	» 665
10 La relazione tra il teorema di Gauss e la legge di Coulomb	» 666
11 Le applicazioni del teorema di Gauss	» 667
a) Campo generato da una distribuzione sferica di cariche.....	» 667
b) Gabbia di Faraday.....	» 668
c) Campo generato da una superficie piana su cui sono uniformemente distribuite cariche ..	» 668
d) Campo generato da una doppia lastra carica.....	» 669

Capitolo 26 Il potenziale elettrico

1 Cos'è il potenziale elettrico?	» 670
2 La differenza di potenziale.....	» 671
a) Lavoro motore e lavoro resistente	» 671
b) Elettrovolta.....	» 672
3 Il calcolo del potenziale elettrico	» 673
a) Campo elettrico radiale	» 673
b) Campo elettrico uniforme.....	» 674
c) Campo elettrico di N cariche puntiformi.....	» 675
4 La capacità di un conduttore.....	» 676
a) Capacità di un conduttore sferico	» 677
b) Capacità di un conduttore carico in presenza di un conduttore neutro.....	» 678
5 I condensatori.....	» 680
a) Capacità di un condensatore sferico	» 680
b) Capacità di un condensatore piano	» 681
6 Condensatori in parallelo e condensatori in serie.....	» 681
a) Condensatori in parallelo	» 681
b) Condensatori in serie	» 682

7 L'energia immagazzinata in un condensatore	Pag. 683
a) Condensatore piano: energia del campo elettrico	» 684

Capitolo 27 La corrente elettrica

1 Cos'è la corrente elettrica?	» 685
a) Intensità di corrente elettrica	» 685
b) Ampere	» 686
2 I generatori di forza elettromotrice	» 686
3 La caduta di tensione	» 687
4 Resistenza elettrica e leggi di Ohm	» 687
a) Prima legge di Ohm	» 688
b) Seconda legge di Ohm	» 688
5 Resistività e campo elettrico	» 689
6 I circuiti elettrici	» 689
a) I circuiti elettrici in corrente continua	» 690
7 Un'applicazione della legge di Ohm	» 690
8 Generatori ideali e generatori reali	» 691
9 I metodi di risoluzione dei circuiti elettrici	» 692
a) Risoluzione mediante i principi di Kirchhoff	» 692
b) Risoluzione mediante il metodo di Maxwell	» 695
10 Resistenze in serie e in parallelo	» 695
a) Collegamento in serie	» 695
b) Collegamento in parallelo	» 696
11 I partitori di tensione e di corrente	» 697
a) Partitore di tensione	» 697
b) Partitore di corrente	» 698
12 Lavoro e potenza elettrica	» 698
a) Effetto Joule	» 699
b) Resistenza equivalente	» 699

Capitolo 28 Il campo magnetico

1 Cos'è un campo magnetico?	» 700
a) Campo magnetico terrestre	» 700
2 La forza di Lorentz	» 701
a) Regola della mano destra	» 701
b) Campo magnetico perpendicolare alla velocità della particella	» 702
c) Campo magnetico parallelo alla velocità della particella	» 702
3 Il moto di una particella con velocità perpendicolare alla direzione del campo	» 702
a) Moto di una particella con velocità obliqua rispetto al campo magnetico	» 704
b) Particella in moto in un campo elettromagnetico	» 705
4 L'effetto Hall	» 705
5 La legge di Laplace	» 706
a) Spira percorsa da corrente immersa in un campo magnetico	» 707
b) Momento di dipolo magnetico	» 709
c) Momento di un magnete permanente	» 709
d) Motore elettrico	» 710

Capitolo 29 Correnti elettriche e campi magnetici: l'elettromagnetismo

1	Correnti e campi magnetici	Pag.	713
2	Il campo magnetico generato da un filo percorso da corrente.....	»	713
	a) Seconda regola della mano destra.....	»	713
	b) Intensità del campo magnetico prodotto	»	714
	c) Forze tra due fili percorsi da corrente.....	»	714
3	Il campo magnetico generato da due fili paralleli percorsi da corrente.....	»	715
4	La legge di Biot-Savart	»	716
	a) Filo rettilineo percorso da corrente.....	»	717
	b) Campo magnetico nel punto centrale di un filo a forma di arco	»	719
	c) Campo magnetico nel punto centrale di un cerchio percorso da corrente.....	»	719
5	La legge di Ampere	»	720
6	Le applicazioni della legge di Ampere.....	»	721
	a) Campo magnetico in un conduttore cilindrico percorso da corrente	»	721
	b) Campo magnetico su un piano infinito	»	722
	c) Campo magnetico in un solenoide.....	»	723
	d) Campo magnetico in un toroide.....	»	725
7	Proprietà magnetiche della materia: il campo magnetico $\vec{H}$	»	725
	a) Permeabilità magnetica.....	»	726
8	La teoria microscopica del magnetismo.....	»	727
	a) Ciclo d'isteresi e punto di Curie	»	728

Capitolo 30 Induzione magnetica ed elettromagnetismo

1	L'induzione elettromagnetica.....	»	730
2	La legge di Faraday-Neumann.....	»	731
3	La legge di Lenz	»	733
4	I generatori di corrente alternata	»	735
	a) Forza elettromotrice indotta da un generatore	»	736
5	La mutua induzione	»	737
	a) Trasformatori.....	»	738
6	L'autoinduzione.....	»	739
	a) Induttanza.....	»	741
	b) Calcolo dell'induttanza in un solenoide.....	»	741
7	I circuiti RL	»	742

Capitolo 31 Quanti, materia, radiazione

1	Radiazione del corpo nero e ipotesi di Planck.....	»	746
2	L'effetto fotoelettrico	»	747
3	I raggi X	»	748
4	L'effetto Compton.....	»	749
5	La lunghezza d'onda di De Broglie	»	749
6	Il principio di indeterminazione di Heisenberg	»	750
7	L'equazione di Schrödinger	»	750
8	I modelli atomici	»	751
	a) Teoria di Dalton	»	751
	b) Modello di Thomson.....	»	752
	c) Modello di Rutherford	»	752
	d) Modello di Bohr	»	753
9	La teoria moderna.....	»	754

Capitolo 32 La fisica del nucleo e delle particelle

1 La composizione dei nuclei atomici	Pag. 756
a) Numero atomico.....	» 756
b) Numero di massa	» 756
2 Gli isotopi.....	» 757
a) Isotopi dell'idrogeno.....	» 757
b) Fissione e fusione nucleare.....	» 757
c) Classificazione delle particelle	» 758
3 Il Modello Standard.....	» 758
a) Interazioni fondamentali	» 759
b) Antimateria.....	» 760
c) Decadimento radioattivo	» 761

SEZIONE III Scienze biologiche

Capitolo 1 La biologia come scienza

1 La biologia: scienza della vita.....	» 765
2 La componente biotica e abiotica della Terra	» 765
3 Il metodo sperimentale.....	» 766
4 La dottrina della biogenesi.....	» 766

Capitolo 2 La chimica dei viventi

1 Gli atomi	» 767
2 La struttura elettronica degli elementi	» 767
3 Il legame chimico	» 769
4 Le trasformazioni chimiche	» 770
5 Le trasformazioni dell'energia: primo e secondo principio della termodinamica	» 771
6 La velocità delle reazioni chimiche	» 772

Capitolo 3 I costituenti fondamentali della materia vivente

1 La molecola d'acqua	» 774
a) Struttura e proprietà	» 774
b) Autoprotolisi e pH.....	» 775
2 Le molecole organiche presenti negli organismi viventi.....	» 775
a) Carboidrati.....	» 776
b) Lipidi	» 777
c) Proteine	» 779
d) Acidi nucleici.....	» 781

Capitolo 4 La cellula come base della vita

1 La teoria cellulare.....	» 784
2 Le dimensioni cellulari: procarioti ed eucarioti.....	» 784

3 La membrana cellulare	Pag.	785
4 Il nucleo	»	786
5 Il citoplasma	»	786
6 Il reticolo endoplasmatico	»	786
7 I mitocondri	»	787
8 L'apparato di Golgi	»	787
9 I lisosomi	»	787
10 Il citoscheletro	»	787
11 Ciglia e flagelli	»	788
12 La parete cellulare	»	788
13 Vacuoli e vescicole	»	789
14 I cloroplasti	»	789

Capitolo 5 Scambi di materia ed energia tra le cellule e al loro interno

1 Gli scambi di materia	»	790
a) Diffusione e osmosi	»	790
b) Proteine di trasporto	»	791
c) Endocitosi ed esocitosi	»	792
d) Giunzioni cellulari	»	793
2 La produzione di energia	»	793
a) Reazioni di ossidoriduzione nei viventi	»	793
b) ATP: la «valuta energetica» della cellula	»	794

Capitolo 6 La bioenergetica

1 L'ossidazione del glucosio	»	795
a) Glicolisi	»	795
b) Respirazione cellulare	»	797
c) Fermentazione	»	799
2 Il metabolismo cellulare	»	799
3 Il processo di fotosintesi	»	800
a) Reazioni luce-dipendenti	»	800
b) Reazioni luce-indipendenti	»	801

Capitolo 7 La riproduzione cellulare

1 Divisione cellulare e riproduzione asessuata	»	804
a) Divisione nei procarioti	»	804
b) Divisione negli eucarioti	»	805
2 La riproduzione sessuata	»	806
a) Aploidia e diploidia	»	806
b) Meiosi	»	807

Capitolo 8 Genetica mendeliana e genetica classica

1 Le leggi di Mendel	»	809
2 La riscoperta del lavoro di Mendel	»	812
3 Interazioni alleliche e geniche	»	812

4 Interazioni con l'ambiente e mutazioni	Pag. 814
5 La teoria cromosomica dell'ereditarietà.....	» 814
a) Cromosomi sessuali	» 814
b) Caratteri legati al sesso	» 815
c) Gruppi di associazione e mappe cromosomiche.....	» 816

Capitolo 9 La genetica molecolare

1 La natura chimica dell'informazione ereditaria: DNA o proteine?	» 818
a) Esperimenti di Griffith	» 818
b) Esperimenti con batteriofagi	» 819
c) Ulteriori conferme a favore del DNA	» 819
2 La struttura del DNA: modello di Watson e Crick.....	» 819
3 La duplicazione del DNA	» 820
4 Dal DNA alle proteine: la sintesi proteica.....	» 821
a) Un gene, una proteina.....	» 821
b) RNA messaggero e processo di trascrizione	» 822
c) Traduzione del codice genetico	» 822
5 Le mutazioni geniche	» 824

Capitolo 10 La regolazione dell'espressione genica

1 Cosa avviene nei procarioti	» 826
2 Cosa avviene negli eucarioti.....	» 827
3 Il DNA eucariote.....	» 828
4 I processi di trascrizione e di traduzione in procarioti ed eucarioti	» 829
5 Lo spostamento dei geni.....	» 829
a) Plasmidi e coniugazione	» 829
b) Virus come vettori d'informazione	» 830
c) Trasposoni	» 831

Capitolo 11 Le nuove frontiere della genetica

1 La trasmissione dei caratteri mono e polifattoriali	» 832
2 Le malattie genetiche ereditarie	» 833
a) Talassemia (o «anemia mediterranea»)	» 833
b) Emofilia	» 834
c) Daltonismo	» 834
d) Fibrosi cistica	» 835
e) Albinismo	» 835
f) Distrofia muscolare di Duchenne	» 836
g) Neurofibromatosi.....	» 836
3 DNA ricombinante e sue possibili applicazioni	» 836
a) Biotecnologia	» 836
b) Ingegneria genetica.....	» 837
c) Geni, piante e animali	» 838
d) Applicazioni e conseguenze sull'ambiente	» 839

Capitolo 12 Ereditarietà e ambiente

1 Le teorie di Lamarck e di Darwin	Pag. 840
2 La teoria dell'evoluzione.....	» 840
3 Variabilità e selezione naturale	» 841
4 L'equilibrio di Hardy-Weinberg.....	» 841
5 Il mantenimento della variabilità.....	» 842
6 La selezione naturale	» 843
7 L'adattamento	» 844
8 La speciazione.....	» 844
9 L'isolamento genetico delle specie	» 845
10 Le prove a sostegno dell'evoluzione	» 845

Capitolo 13 La classificazione dei viventi

1 La sistematica scientifica.....	» 847
2 La classificazione gerarchica	» 848

Capitolo 14 I cinque regni

1 Il regno delle monere	» 850
2 Il regno dei protisti	» 851
3 Il regno dei funghi	» 852
4 Il regno delle piante	» 853
a) Briofite	» 854
b) Piante vascolari	» 854
5 Il regno animale.....	» 855
a) Invertebrati	» 856
b) Vertebrati	» 859

Capitolo 15 I tessuti animali

1 Le funzioni degli organismi viventi.....	» 862
a) Omeostasi	» 862
b) Metabolismo	» 862
c) Difesa	» 862
d) Regolazione e controllo	» 863
e) Riproduzione	» 863
2 Dalle cellule ai tessuti	» 863
a) Tessuto epiteliale	» 864
b) Tessuto connettivo	» 864
c) Tessuto muscolare	» 865
d) Tessuto nervoso	» 865

Capitolo 16 Il movimento

1 Lo scheletro umano	» 867
a) Ossa.....	» 868
b) Articolazioni.....	» 868

c) Scheletro assiale	Pag. 869
d) Scheletro appendicolare.....	» 869
2 Il sistema muscolare	» 870
a) Classificazione dei muscoli	» 870
b) Fisiologia della contrazione muscolare	» 871

Capitolo 17 Digestione e respirazione

1 Alimentazione e digestione	» 873
2 L'apparato digerente.....	» 873
a) Bocca.....	» 874
b) Esofago.....	» 874
c) Stomaco.....	» 875
d) Intestino.....	» 875
e) Fegato e pancreas	» 876
3 Respirazione e scambi gassosi	» 876
4 L'apparato respiratorio	» 877
a) Struttura e organi costitutivi.....	» 877
b) Trasporto dei gas.....	» 878
c) Regolazione della respirazione.....	» 878

Capitolo 18 Circolazione sanguigna ed escrezione

1 Il sistema circolatorio negli animali	» 879
2 L'apparato circolatorio umano	» 879
a) Sangue	» 879
b) Vasi sanguigni	» 880
c) Cuore.....	» 880
d) Regolazione del battito cardiaco.....	» 881
e) Pressione sanguigna e centro di controllo cardiovascolare	» 881
3 Regolazione dell'equilibrio chimico: l'apparato escretore	» 882
a) Funzionalità dei reni.....	» 882
b) Anatomia e fisiologia del rene	» 883

Capitolo 19 I sistemi nervoso ed endocrino

1 I messaggeri chimici dell'organismo.....	» 885
2 Sistema endocrino e regolazione ormonale.....	» 885
a) Regolazione ormonale.....	» 885
b) Tiroide e paratiroidi.....	» 886
c) Pancreas.....	» 886
d) Ghiandole surrenali.....	» 887
3 Il sistema nervoso	» 887
a) Sistema nervoso centrale	» 888
b) Sistema nervoso periferico	» 888
c) Trasmissione dell'impulso nervoso.....	» 889
4 Sistema nervoso centrale e organi di senso	» 891
a) Encefalo	» 891
b) Orecchio.....	» 892
c) Occhio	» 893

d) Pelle.....	Pag. 894
e) Naso e bocca.....	» 894

Capitolo 20 Riproduzione e sviluppo

1 L'apparato riproduttore maschile	» 896
2 L'apparato riproduttore femminile	» 897
3 Fecondazione e sviluppo embrionale	» 898
4 Gravidanza e parto	» 898

Capitolo 21 Difesa dell'organismo, igiene e Microbiologia

1 L'epidemiologia	» 900
2 Il sistema immunitario	» 900
a) Rete linfatica	» 901
b) Milza e timo	» 901
3 Le difese aspecifiche	» 901
4 Le difese specifiche	» 902
a) Linfociti B	» 902
b) Linfociti T	» 903
5 La profilassi	» 903
6 Sieri e vaccini	» 903
7 Gli accertamenti delle malattie infettive	» 904
8 Allergie e disfunzioni del sistema immunitario	» 905
9 Cos'è la «Microbiologia» e di cosa si occupa?	» 905

Capitolo 22 La biosfera

1 La vita nelle acque.....	» 907
a) Fiumi	» 907
b) Acque ferme	» 907
c) Oceani	» 907
d) Litorali	» 907
2 La vita sulle terre emerse.....	» 908
a) Foresta decidua	» 908
b) Foresta di conifere	» 908
c) Tundra artica	» 908
d) Prateria	» 909
e) Savana	» 909
f) Macchia mediterranea	» 909
g) Deserto	» 909
h) Foreste tropicali.....	» 909

Rispettare e preservare la «biodiversità» nei sistemi ambientali, realizzando esperienze quali, ad esempio, la «costruzione di nidi per uccelli selvatici» oppure l'«adozione di un orto» ... (on line)

SEZIONE **IV** Astronomia e Scienze della Terra

Capitolo 1 L'Universo

1 I metodi d'indagine in astrofisica	Pag. 913
2 La sfera celeste	» 913
3 Stelle e galassie	» 914
4 L'origine dell'Universo	» 916

Capitolo 2 Il Sistema Solare

1 Caratteri generali.....	» 917
2 Il Sole	» 917
3 Pianeti, leggi di Keplero e legge di gravitazione universale di Newton.....	» 918
4 I pianeti del Sistema Solare	» 920
5 L'origine del Sistema Solare	» 922
6 Il pianeta Terra.....	» 922
a) Moti fondamentali	» 923
b) Moti secondari	» 926
7 La Luna	» 926
a) Moti della Luna	» 928
b) Eclissi	» 928
c) Teorie sull'origine della Luna.....	» 929
8 L'esplorazione dello spazio: l'ultima frontiera della scoperta	» 929
a) Agli albori.....	» 930
b) Animali e uomini in orbita: l'iniziale primato sovietico.....	» 930
c) La rivincita degli Stati Uniti.....	» 931
d) Dalla «competizione» alla «cooperazione»	» 931
e) Perché andare nello spazio.....	» 932
f) Dove siamo stati finora	» 932
g) Ma come si fa ad andare in orbita?	» 933
h) L'Italia, terzo Paese del mondo a lanciare un satellite	» 934
9 I propellenti dei vettori spaziali.....	» 934

Capitolo 3 L'atmosfera

1 La composizione	» 936
2 Le suddivisioni.....	» 936
3 La temperatura.....	» 937
4 Pressione e venti	» 938
5 Umidità e precipitazioni	» 939

Capitolo 4 L'idrosfera

1 Il ciclo idrologico	» 941
2 Gli oceani	» 941
a) Onde.....	» 942
b) Maree	» 943
c) Correnti	» 943

3 I ghiacciai.....	Pag.	943
4 Le acque continentali.....	»	944
a) Fiumi.....	»	944
b) Laghi.....	»	945

Capitolo 5 La struttura della Terra solida

1 La litosfera.....	»	946
2 Le rocce.....	»	946
a) Rocce ignee.....	»	947
b) Rocce sedimentarie.....	»	949
c) Rocce metamorfiche.....	»	950
3 L'interno della Terra e i diversi involucri che lo compongono.....	»	952

Capitolo 6 La geodinamica endogena

1 Il calore interno.....	»	954
2 Il magnetismo.....	»	954
3 La tettonica delle placche.....	»	955
4 I terremoti.....	»	956
5 I fenomeni endogeni: vulcanismo, sismi e orogenesi.....	»	958
6 I vulcani.....	»	959

Capitolo 7 La geodinamica esogena

1 Il modellamento del rilievo terrestre.....	»	961
2 I processi di degradazione delle rocce.....	»	961
3 Formazione e classificazione dei suoli.....	»	962
4 Modellamento dei versanti e dilavamento.....	»	964
5 Le frane.....	»	966
6 Morfologia fluviale ed erosione.....	»	967
7 Le alluvioni.....	»	969
8 Il carsismo.....	»	970
9 Morfologia glaciale e periglaciale.....	»	972
a) Azione erosiva dei ghiacciai.....	»	972
b) Funzione di trasporto e di deposito dei ghiacciai.....	»	973
c) Ambiente periglaciale.....	»	973
10 La morfologia eolica.....	»	974
11 La morfologia costiera.....	»	975
12 Ciclo d'erosione e superfici di spianamento.....	»	977

Capitolo 8 Elementi di stratigrafia e di tettonica

1 La stratigrafia.....	»	979
2 La tettonica.....	»	980

Capitolo 9 Paleogeografia ed evoluzione della vita

1 Le ere geologiche	Pag. 982
2 La paleogeografia.....	» 983
3 L'origine della vita	» 986

Capitolo 10 Il clima

1 Caratteri generali.....	» 987
2 I climi delle basse latitudini	» 989
3 I climi delle medie latitudini	» 990
4 I climi delle alte latitudini	» 992

Capitolo 11 Elementi di fitogeografia e di zoogeografia

1 Caratteri generali.....	» 994
2 Classi di formazioni vegetali e associazioni animali	» 996
a) Foresta	» 997
b) Savana	» 999
c) Prateria	» 1000
d) Deserto.....	» 1000

Capitolo 12 L'ecologia

1 L'equilibrio ecologico	» 1001
2 Uomo e ambiente.....	» 1002
3 Il dissesto idrogeologico: un problema anche, e soprattutto, italiano	» 1005

SEZIONE Educazione alla salute

Introduzione	» 1009

Capitolo 1 Sessualità e riproduzione umana

1 La riproduzione negli animali superiori.....	» 1010
2 Gonadi e sessualità	» 1010
3 Il processo di fecondazione dell'ovulo.....	» 1011
4 Lo sviluppo embrionale e fetale	» 1012
5 Nascita e sviluppo extrauterino.....	» 1013
6 Le malattie genetiche ed ereditarie	» 1014
a) Malattie cromosomiche	» 1014
b) Malattie monogeniche (o mendeliane)	» 1015
c) Malattie multifattoriali	» 1016
d) Prevenzione e diagnosi delle malattie genetiche ed ereditarie.....	» 1016
7 La contraccezione	» 1016

Capitolo 2 La gestione delle funzioni biologiche nell'uomo

1 Le funzioni biologiche	Pag. 1019
2 L'energia	» 1019
3 Nutrizione ed equilibrio idro-elettrolitico.....	» 1020
4 I disordini nutrizionali.....	» 1022
5 I disturbi alimentari.....	» 1023
a) Anoressia nervosa	» 1024
b) Bulimia.....	» 1025
c) Disturbi del comportamento alimentare non altrimenti specificati	» 1026
6 Igiene dell'apparato locomotore e importanza del movimento.....	» 1026
7 Fatica e riposo	» 1027

Capitolo 3 Consumi voluttuari e prevenzione delle dipendenze

1 La dipendenza	» 1029
2 Il tabagismo	» 1029
a) Il fumo come fenomeno sociale	» 1030
b) Normativa antifumo.....	» 1031
3 I rischi dell'alcolismo (o etilismo).....	» 1032
a) Cause dell'alcolismo	» 1033
b) Effetti dell'alcolismo.....	» 1033
4 La tossicodipendenza	» 1034
a) Processo di assuefazione	» 1034
b) Droghe.....	» 1034
c) Stupefacenti	» 1034
d) Stimolanti	» 1034
e) Sedativi o ipnotici.....	» 1035
f) Inebrianti	» 1035
g) Allucinogeni	» 1035
h) Nuove droghe	» 1035
i) Terapia e reinserimento sociale	» 1036
j) Doping.....	» 1036

Capitolo 4 La malattia come rottura dell'equilibrio biologico e psicologico

1 Il complesso equilibrio dell'organismo umano	» 1037
a) Concetto di salute	» 1037
b) Concetto di malattia.....	» 1038
2 Le malattie metaboliche	» 1038
a) Diabete mellito	» 1040
b) Iperlipoproteinemie	» 1040
c) Gotta e altri disordini del metabolismo purinico.....	» 1041
3 Le malattie degenerative	» 1042
4 Le malattie tumorali.....	» 1043
5 Le malattie infettive	» 1044
6 Le malattie a trasmissione sessuale	» 1046
7 Le malattie da lavoro	» 1048
8 L'educazione alla solidarietà nella sofferenza, nella vecchiaia e nell'inabilità temporanea o permanente.....	» 1049

Capitolo 5 La difesa della salute

1 L'educazione sanitaria	Pag. 1050
2 La medicina preventiva	» 1052
a) Prevenzione primaria	» 1053
b) Prevenzione secondaria	» 1053
c) Prevenzione terziaria	» 1053
d) Educazione ambientale ed ecologica	» 1053
3 Visite periodiche per la diagnosi precoce dei difetti sensoriali, delle malattie del ricambio e dei tumori	» 1054
4 Epidemiologia e profilassi delle malattie infettive	» 1055
a) Vaccinoprofilassi	» 1057
b) Vaccinazioni	» 1057
c) Calendario delle vaccinazioni	» 1058
d) Classificazione dei vaccini	» 1058
e) Vaccini per malattie batteriche	» 1060
f) Vaccini per malattie virali	» 1061
g) Esantemi del bambino	» 1063

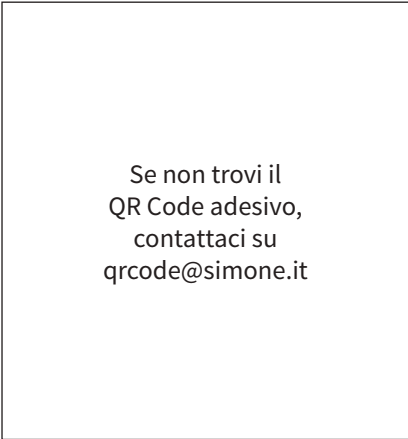
Capitolo 6 Conoscenza e prevenzione della patologia incidentale

1 Patologia incidentale e malattie traumatiche	» 1066
2 Educazione alla conoscenza delle cause e alla prevenzione della patologia incidentale	» 1066
a) Traumi	» 1067
b) Folgorazioni e lesioni da elettricità	» 1067
c) Ustioni	» 1068
d) Intossicazioni chimiche	» 1068
e) Incidenti stradali	» 1069

Capitolo 7 La salute mentale

1 Lo sviluppo neuromotorio e psichico	» 1072
a) <i>Screening</i> dello sviluppo psicomotorio	» 1074
b) Età scolare ed età mentale	» 1075
c) Principali cause di turbamento dello sviluppo neuromotorio e psichico	» 1075
2 La prevenzione delle nevrosi favorite dallo stile di vita	» 1076
a) Stress	» 1076
b) Nervosismo, ansia, depressione, nevrosi e psicosi	» 1077
3 I problemi relativi alla malattia mentale e alla disabilità psicofisica	» 1079
a) Disadattamenti scolastici e disabilità	» 1080
b) Instabilità psicomotoria	» 1081
c) Carenze nella coordinazione motoria, disgrafia e dislessia	» 1081
d) Disabilità sensoriali	» 1081
e) Disturbi emotivi e cognitivi	» 1081
f) Disabilità di tipo nevrotico e psicotico	» 1082

Il QR Code per accedere alle risorse online



Se non trovi il
QR Code adesivo,
contattaci su
qrcode@simone.it

Come accedere alle nostre risorse online?

Se hai uno smartphone, per utilizzare il QR Code devi utilizzare un lettore di QR Code.

Se non hai un lettore di QR Code sul tuo smartphone, puoi scaricarlo da iTunes (se hai un iPhone) o da Play Store (se hai uno smartphone Android).

Scaricato il lettore, utilizzalo per **inquadrare il QR Code che trovi su questa pagina** con la fotocamera del tuo smartphone. Una volta inquadrato e riconosciuto il QR Code, avrai automaticamente il link per accedere alle nostre risorse online.

Se, invece, non hai uno smartphone puoi accedere alle nostre risorse online direttamente utilizzando questo indirizzo nel tuo browser:

www.simone.it/d/526_12

In entrambi i casi, registrati sulla pagina che ti abbiamo indicato (in alto a destra, nel box, **Area Riservata**, clicca sul link **Registrati**). Terminata la registrazione dovrai attendere una mail con la password che ti è stata assegnata.

Ricevuta la password **accedi all'Area Riservata** e potrai scaricare i contenuti online.

Per accedere ad alcuni contenuti è necessario digitare il **Codice identificativo** che trovi sul QR Code.

Le espansioni online allegate al presente volume sono valide fino alla data di pubblicazione della eventuale successiva edizione del volume.

in **cattedra**

Concorsi Scuola & FIT

Questo manuale è indirizzato a quanti vogliono affrontare le prove scritte e orali dei nuovi **concorsi a cattedra e dei FIT**.

In particolare, il testo ripercorre per punti e snodi essenziali tutto il *programma d'esame delle discipline d'insegnamento* relative alla *classe di concorso A28 - Matematica e Scienze* (ex A059), al fine di consentire un *rapido ripasso delle materie, nonché approfondimenti sulla didattica disciplinare*. Completano l'opera le **Indicazioni nazionali** e le **Linee guida** (disponibili come **espansioni online** e accessibili tramite il **QR Code** posto all'interno del volume) riguardanti, appunto, le discipline oggetto d'insegnamento della classe suddetta, indispensabili per una *corretta progettazione di lezioni e unità d'apprendimento*.

Sempre in *espansione online* offriamo ulteriori approfondimenti su taluni degli argomenti trattati.

Per completare la preparazione sulle **AVVERTENZE GENERALI**



Vol. 526/B

MANUALE DELLE METODOLOGIE E TECNOLOGIE DIDATTICHE

Manuale di preparazione alle
prove metodologico-didattiche
dei concorsi a cattedra e dei FIT



Vol. 526/C

LEGISLAZIONE SCOLASTICA

Manuale di preparazione alle
prove dei concorsi a cattedra

**I due volumi disponibili anche in Kit
(Vol. 526/AG) AVVERTENZE GENERALI 2018**



www.simone.it/social
www.simone.it

