

## Nelle edizioni Zanichelli

### Opere di consultazione

Baruk, **Dizionario di matematica elementare**

### Analisi e matematica generale

Adams **Calcolo differenziale 1** (Cea)

Adams **Calcolo differenziale 1** (Cea)

Antognini, Barozzi **Matematica e Mathematica**

Avvantaggiati **Istituzioni di matematica** (Cea)

Barozzi **Primo corso di analisi matematica**

Barozzi, Dore, Obrecht **Elementi di analisi matematica 1**

Benedetto, Degli Esposti, Maffei **Matematica per le scienze della vita 2<sup>a</sup> ed** (Cea)

Binmore **Calcolo differenziale di più variabili** (Cea)

Boieri, Chiti **Percorso di matematica**

Bramanti, Pagani, Salsa, **Matematica** (2<sup>a</sup> edizione)

Bramanti, Pagani, Salsa **Analisi matematica 1**

Bramanti, Pagani, Salsa **Analisi matematica 1 con elementi di algebra lineare**

Bramanti, Pagani, Salsa **Analisi matematica 2**

Bramanti, Travaglini **Matematica. Questione di metodo**

Chiantini **Algebra lineare e geometria analitica** (Cea)

Davis **Il mondo dei grandi numeri**

De Marco **Analisi zero** (3<sup>a</sup> edizione)

De Marco **Analisi Uno** (2<sup>a</sup> edizione - Decibel)

De Marco **Analisi Due** (2<sup>a</sup> edizione - Decibel)

De Marco **Analisi Due. Secondo corso di analisi matematica per l'Università** (Decibel)

De Marco, Mariconda **Esercizi di analisi due** (Decibel)

De Marco, Mariconda **Esercizi di calcolo in una variabile** (Decibel)

Ghizzetti, Rosati **Analisi matematica** (Masson - 2<sup>a</sup> edizione)

Israel, Millan Gasca **Pensare in matematica**

Pagani, Salsa **Analisi matematica** (Masson - 2 volumi)

Pagani, Salsa **Matematica** (Masson)

Pagani, Salsa **Serie di funzioni ed equazioni differenziali**

Ritelli, Bergamini, Trifone **Fondamenti di matematica**

Salsa, Squellati **Esercizi di matematica** (2 volumi)

Salsa, Squellati **Esercizi di analisi matematica** (2 volumi)

Salsa, Squellati **Esercizi di analisi matematica II** (Masson - 3 volumi)

### Analisi numerica e Fisica matematica

Bagarello **Fisica matematica**

Barozzi **Matematica per l'ingegneria dell'informazione** (ristampa aggiornata)

Bevilacqua, Bini, Capovani, Menchi **Introduzione alla matematica computazionale**

Bevilacqua, Bini, Capovani, Menchi **Metodi numerici**

Cercignani **Spazio, tempo, movimento. Introduzione alla meccanica razionale**

Codegone **Metodi matematici per l'ingegneria**

Fabrizio **Elementi di meccanica classica**

Finzi **Meccanica razionale** (2 volumi)

Levi-Civita **Caratteristiche dei sistemi differenziali e propagazione ondosa**

Vivarelli **Appunti di meccanica razionale** (2<sup>a</sup> edizione ampliata)

**Probabilità, Statistica e Matematica finanziaria**

Bolzani, Canestrari **Logica del test statistico** (Cea)

Bolzani **Problema di statistica** (Cea)

Bruglieri, Colorni **Ricerca operativa**

Carbonaro, Vitale, **Fondamenti di probabilità e statistica** (Cea)

Contucci, Isola **Probabilità elementare**

Dall'Aglio **Calcolo delle probabilità** (3<sup>a</sup> edizione)

De Giulii; Giorgi, Maggi, Magnani **Matematica per l'economia e la finanza**

Invernizzi, Rinaldi, Sgarro **Moduli di matematica e statistica**

Letta **Probabilità elementare**

Montanari, Agati, Calò **Statistica** (Cea)

Ottaviani **Riassunto delle lezioni di matematica attuariale**

Scozzafava **Incertezza e probabilità**

Scozzafava **Primi passi in probabilità e statistica** (Decibel - 2<sup>a</sup> edizione)

Toscano **Training autogeno in probabilità** (Decibel/Zanichelli, rist. aggiornata)

### Algebra e Geometria

Abeasis **Complementi di algebra lineare e geometria**

Abeasis **Elementi di algebra lineare e geometria**

Abeasis **Geometria analitica del piano e dello spazio**

Broglia, Fortuna, Luminati **Problemi risolti di algebra lineare**

Carfagna, Piccolella **Complementi ed esercizi di geometria e algebra lineare** (2<sup>a</sup> edizione)

Cerasoli, Eugeni, Protasi **Elementi di matematica discreta**

Dedò **Trasformazioni geometriche**

Enriques **Lezioni di geometria proiettiva** (Ristampa anastatica)

**Questioni riguardanti le matematiche elementari** 2 voll. a cura di Enriques

Facchini **Algebra** (Decibel)

Kazarinoff **Disuguaglianze geometriche**

Kosniowski **Introduzione alla topologia algebrica**

Maroscia **Geometria e algebra lineare**

Maroscia **Introduzione alla geometria e all'algebra lineare**

Mauri, Schlesinger **Esercizi di algebra lineare e geometrica**

Piacentini Cattaneo **Algebra** (Decibel)

Piacentini Cattaneo **Matematica discreta**

Procesi Ciampi, Rota **Esercizi di geometria e algebra** (Decibel)

Procesi **Elementi di teoria dei gruppi** (Decibel)

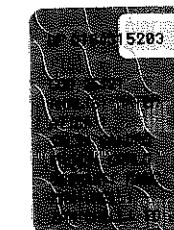
Procesi **Elementi di teoria di Galois** (Decibel)

Salce **Lezioni sulle matrici** (Decibel)

Schlesinger **Geometria e algebra lineare**

Serafini **Ottimizzazione**

Vaccaro, Carfagna, Piccolella **Lezioni di geometria e algebra lineare** (Masson)



Marco Bramanti Carlo D. Pagani Sandro Salsa

# Analisi matematica 1

con elementi di geometria e algebra lineare

ZANICHELLI

I diritti di elaborazione in qualsiasi forma o opera, di memorizzazione anche digitale su supporti di qualsiasi tipo (inclusi magnetici e ottici), di riproduzione e di adattamento totale o parziale con qualsiasi mezzo (compresi i microfilm e le copie fotostatiche), i diritti di noleggio, di prestito e di traduzione sono riservati per tutti i paesi. L'acquisto della presente copia dell'opera non implica il trasferimento dei suddetti diritti né li esaurisce.

Le fotocopie per uso personale (cioè privato e individuale, con esclusione quindi di strumenti di uso collettivo) possono essere effettuate, nei limiti del 15% di ciascun volume, dietro pagamento alla S.I.A.E del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633. Tali fotocopie possono essere effettuate negli esercizi commerciali convenzionati S.I.A.E. o con altre modalità indicate da S.I.A.E.

Per le riproduzioni ad uso non personale (ad esempio: professionale, economico, commerciale, strumenti di studio collettivi, come dispense e simili) l'editore potrà concedere a pagamento l'autorizzazione a riprodurre un numero di pagine non superiore al 15% delle pagine del presente volume. Le richieste per tale tipo di riproduzione vanno inoltrate a

Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali (CLEARedi)  
Corso di Porta Romana, n. 108  
20122 Milano  
e-mail autorizzazioni@clearedi.org e sito web www.clearedi.org

L'editore, per quanto di propria spettanza, considera rare le opere fuori del proprio catalogo editoriale, consultabile al sito [www.zanichelli.it/f\\_catalog.html](http://www.zanichelli.it/f_catalog.html). La fotocopia dei soli esemplari esistenti nelle biblioteche di tali opere è consentita, oltre il limite del 15%, non essendo concorrentiale all'opera. Non possono considerarsi rare le opere di cui esiste, nel catalogo dell'editore, una successiva edizione, le opere presenti in cataloghi di altri editori o le opere antologiche. Nei contratti di cessione è esclusa, per biblioteche, istituti di istruzione, musei ed archivi, la facoltà di cui all'art. 71 - ter legge diritto d'autore. Maggiori informazioni sul nostro sito: [www.zanichelli.it/fotocopie/](http://www.zanichelli.it/fotocopie/)

*Disegni e impaginazione:* Compomat, Configni (RI)

*Copertina:*  
- Progetto grafico: Miguel Sal & C., Bologna  
- Immagine di copertina: © Zygotehasnobrai/Shutterstock.

Prima edizione: luglio 2014

Ristampa

7 6 5 2017 2018 2019 2020

Realizzare un libro è un'operazione complessa, che richiede numerosi controlli: sul testo, sulle immagini e sulle relazioni che si stabiliscono tra essi.

L'esperienza suggerisce che è praticamente impossibile pubblicare un libro privo di errori. Saremo quindi grati ai lettori che vorranno segnalare gli errori.

Per segnalazioni o suggerimenti relativi a questo libro l'indirizzo a cui rivolgersi è:

Zanichelli editore S.p.A.  
piazza Castello 4  
20121 Milano  
e-mail: linea\_universitaria@zanichelli.it  
sito web: [www.zanichelli.it](http://www.zanichelli.it)  
fax 0272025050

Stampa: La Fotocromo Emiliana  
Via Sardegna 30, 40060 Osteria Grande (Bologna)  
per conto di Zanichelli editore S.p.A.  
Via Irnerio 34, 40126 Bologna

# Indice

## Prefazione

### 1 Numeri

1	Insiemi e logica	1
1.1	Concetti di base sugli insiemi	1
1.2	Un po' di logica elementare	9
2	Sommatorie e coefficienti binomiali	13
2.1	Il simbolo di sommatoria	13
2.2	Fattoriale di $n$	15
2.3	Coefficienti binomiali e formula di Newton	16
3	Campi ordinati	18
4	Numeri reali. Estremo superiore e assioma di continuità	20
4.1	Inadeguatezza dell'insieme dei razionali per misurare le lunghezze	20
4.2	Estremo superiore e assioma di continuità	21
4.3	Valore assoluto. Disuguaglianza triangolare	23
4.4	Intervalli	24
5	Radicali, potenze, logaritmi	25
5.1	Radici $n$ -esime aritmetiche	25
5.2	Potenze a esponente reale	26
5.3	Logaritmi	27
5.4	Approssimazioni	28
6	Insiemi infiniti	28
7	Il principio di induzione	32
8	Numeri complessi	35
8.1	Definizione di $\mathbb{C}$ e struttura di campo	35
8.2	Coniugato e modulo	37
8.3	Forma trigonometrica	40
8.4	Radici $n$ -esime	43

### 2 Funzioni di una variabile

1	Il concetto di funzione	49
2	Funzioni reali di variabile reale	52
2.1	Generalità	52
2.2	Funzioni limitate	53
2.3	Funzioni simmetriche	54
2.4	Funzioni monotone	55

	2.5 Funzioni periodiche	55
3	Funzioni elementari	56
	3.1 Funzioni potenza	56
	3.2 Funzioni esponenziali e logaritmiche	61
	3.3 Funzioni trigonometriche	63
	3.4 Fenomeni vibratori	64
	3.5 Funzioni parte intera e mantissa	68
	3.6 Funzioni iperboliche	69
	3.7 Operazioni sui grafici	70
	3.8 Funzioni definite a tratti	74
4	Funzioni composte e inverse	75
	4.1 Funzioni composte	75
	4.2 Funzioni invertibili; funzioni inverse	77
	4.3 Le funzioni trigonometriche inverse	80
	4.4 Le funzioni iperboliche inverse	82
<b>3</b>	<b>Limiti e continuità</b>	<b>87</b>
1	Successioni	87
	1.1 Definizione di successione. Definizione di limite	87
	1.2 Successioni monotone	93
	1.3 Il calcolo dei limiti	96
	1.4 Il numero $e$	101
	1.5 Confronti e stime asintotiche	103
2	Limiti di funzioni, continuità, asintoti	110
3	Il calcolo dei limiti	121
	3.1 Proprietà fondamentali di limiti e continuità	121
	3.2 Limiti notevoli	128
	3.3 Confronti e stime asintotiche	130
	3.4 Stime asintotiche e grafici	132
4	Proprietà globali delle funzioni continue o monotone su un intervallo	136
	4.1 Funzioni continue su un intervallo	136
	4.2 Funzioni monotone su un intervallo	141
	4.3 Continuità e invertibilità	143
<b>4</b>	<b>Calcolo differenziale per funzioni di una variabile</b>	<b>147</b>
1	Introduzione al calcolo differenziale	147
2	Derivata di una funzione	150
	2.1 Derivata e retta tangente	150
	2.2 Altre interpretazioni della derivata	153
	2.3 Derivate di funzioni elementari	154
	2.4 Punti angolosi, cuspidi, flessi a tangente verticale	157
3	Regole di calcolo delle derivate	160
	3.1 Algebra delle derivate	161
	3.2 Derivata di una funzione composta	162
	3.3 Derivata di funzione inversa	167
4	Il teorema del valor medio e le sue conseguenze	171
	4.1 Punti stazionari. Massimi e minimi locali	171

	4.2 Teorema del valor medio. Test di monotonía	174
	4.3 Soluzione di alcuni problemi di massimo e minimo	181
	4.4 Il teorema di de l'Hospital	187
	4.5 Limite della derivata e derivabilità	190
5	Derivata seconda	195
	5.1 Significato geometrico della derivata seconda	195
	5.2 Derivata seconda, concavità e convessità	196
6	Studio del grafico di una funzione	202
7	Calcolo differenziale e approssimazioni	208
	7.1 Differenziale e approssimazione lineare. Il simbolo di "o piccolo"	208
	7.2 Limiti notevoli e sviluppi	212
	7.3 Formula di Taylor-MacLaurin con resto secondo Peano	213
	7.4 Formula di Taylor-MacLaurin con resto secondo Lagrange	218
	7.5 Risoluzione approssimata di equazioni: il metodo di Newton	221
<b>5</b>	<b>Serie</b>	<b>229</b>
1	Serie numeriche	229
	1.1 Definizione e primi esempi	229
	1.2 Serie a termini non negativi	233
	1.3 Serie a termini di segno variabile	239
2	Serie di Taylor. Esponenziale complesso	245
	2.1 Serie di Taylor delle trascendenti elementari	245
	2.2 Serie nel campo complesso. Esponenziale complesso	249
<b>6</b>	<b>Calcolo integrale per funzioni di una variabile</b>	<b>257</b>
1	Introduzione al calcolo integrale	257
2	L'integrale come limite di somme	258
	2.1 La definizione di integrale	258
	2.2 Classi di funzioni integrabili	262
3	Proprietà dell'integrale	263
4	Il teorema fondamentale del calcolo integrale	266
5	Calcolo di integrali indefiniti e definiti	268
	5.1 Integrali immediati, per scomposizione, per sostituzione	268
	5.2 Integrazione delle funzioni razionali	273
	5.3 Integrazione per parti	277
	5.4 Integrazione delle funzioni trigonometriche	281
	5.5 Integrazione delle funzioni irrazionali	285
	5.6 Integrazione di funzioni discontinue	287
6	Alcune applicazioni fisiche e geometriche	289
7	Calcolo numerico approssimato di un integrale	294
8	Integrali generalizzati	296
	8.1 Integrazione di funzioni non limitate	296
	8.2 Criteri di integrabilità al finito	297
	8.3 Integrazione su intervalli illimitati	300
	8.4 Criteri di integrabilità all'infinito	303
9	Funzioni integrali	305
10	Convoluzione e sistemi fisici lineari	310

11	Teorema di Bolzano-Weierstrass, continuità uniforme e integrabilità delle funzioni continue	314
11.1	Alcuni risultati fondamentali per le successioni di numeri reali	314
11.2	Continuità uniforme	316
11.3	Integrabilità delle funzioni continue	318
<b>7</b>	<b>Modelli dinamici discreti</b>	<b>323</b>
1	Introduzione alla modellistica	323
1.1	Modello di Malthus	324
1.2	Modello logistico	326
1.3	Modello dell'acceleratore	327
2	Generalità sulle equazioni alle differenze	328
3	Equazioni lineari del prim'ordine a coefficienti costanti	329
4	Equazioni autonome non lineari	333
4.1	Orbite, diagrammi a gradino e punti fissi o d'equilibrio	333
4.2	Punti fissi e stabilità	335
4.3	Stabilità per linearizzazione	337
4.4	Orbite periodiche e stabilità	340
4.5	Esistenza di orbite periodiche	342
4.6	Comportamento caotico	343
4.7	Equazione logistica discreta	344
5	Equazioni lineari a coefficienti costanti del second'ordine	351
5.1	Equazioni omogenee	351
5.2	I numeri di Fibonacci	354
5.3	Equazioni non omogenee	356
5.4	Stabilità	358
<b>8</b>	<b>Elementi di geometria e algebra lineare</b>	<b>361</b>
1	Vettori nel piano e nello spazio	361
1.1	Operazioni fondamentali sui vettori	361
1.2	Prodotto scalare e vettoriale	369
2	Geometria analitica lineare nello spazio	374
3	Spazi vettoriali	385
3.1	Vettori $n$ -dimensionali: lo spazio $\mathbb{R}^n$ . Spazi vettoriali astratti	385
3.2	Prodotto scalare in $\mathbb{R}^n$	391
3.3	Spazi vettoriali con prodotto scalare	392
3.4	Il concetto di linearità	399
4	Matrici e trasformazioni lineari	402
4.1	L'algebra delle matrici	402
4.2	Rappresentazione matriciale delle trasformazioni lineari	409
4.3	Determinante	413
4.4	Caratteristica di una matrice	420
4.5	Matrice inversa	422
5	Sistemi lineari	426
5.1	Generalità. Metodo di Cramer	426
5.2	Immagine e nucleo di una trasformazione lineare da $\mathbb{R}^n$ a $\mathbb{R}^m$	433
5.3	Sistemi generali. Teorema di Rouché-Capelli	437

5.4	Metodo di eliminazione di Gauss	443
6	Autovettori ed autovalori. Diagonalizzazione	451
6.1	Matrici diagonalizzabili	451
6.2	Autovalori ed autovettori di una matrice	453
6.3	Condizioni di diagonalizzabilità	457
<b>9</b>	<b>Coniche e quadriche</b>	<b>467</b>
1	Le coniche	467
1.1	Le coniche dal punto di vista della geometria elementare e le loro equazioni canoniche	467
1.2	L'equazione generale delle coniche. Riduzione a forma canonica	472
1.3	Coniche ed eccentricità	485
2	Le superfici quadriche	487
2.1	Rototraslazioni nello spazio e invarianti di una quadrica	489
2.2	Riduzione a forma canonica delle quadriche: quadriche a centro	492
2.3	Riduzione a forma canonica delle quadriche: quadriche non a centro	497
2.4	Riconoscimento di una quadrica mediante gli invarianti	501
3	Esempi di applicazioni di coniche e quadriche	504
<b>A</b>	<b>Formule utili</b>	<b>507</b>
1	Costanti matematiche	507
2	Funzioni trigonometriche	507
3	Funzioni iperboliche	510
4	Derivate elementari	511
5	Regole di derivazione	511
6	Sviluppi di Mac Laurin delle principali funzioni	512
7	Tabella di primitive	513
<b>B</b>	<b>Grafici</b>	<b>515</b>
	<b>Indice analitico</b>	<b>523</b>

# Prefazione

Questa nuova edizione del testo di *Analisi matematica 1* affianca ai contenuti di analisi già presenti nell'edizione del 2008 una nuova sezione di *algebra lineare e geometria*. La ragione di questa scelta sta nella constatazione del fatto che in molti atenei esistono tuttora corsi che combinano in un unico esame contenuti di analisi matematica e di algebra lineare e geometria. Per questi corsi non è sempre facile individuare un unico libro di testo che tratti i contenuti richiesti con l'approfondimento desiderato ma anche con la necessaria sinteticità.

Riguardo alla parte di analisi matematica, i contenuti di questo testo sono quelli fondamentali di un corso di analisi per funzioni di una variabile: numeri reali e complessi, limiti e continuità, calcolo differenziale e integrale, serie numeriche e di Taylor. Merita un discorso a parte solo il capitolo 7, “Modelli dinamici discreti”, che presenta un argomento meno tradizionale con il quale si intende aprire per lo studente una finestra verso la modellistica, con la speranza di stimolarne la curiosità verso gli sviluppi successivi dell’analisi e delle sue applicazioni. La controparte continua dei modelli dinamici, ossia le equazioni differenziali, sono invece collocate nel volume di *Analisi matematica 2*.

Per quanto riguarda la parte di algebra lineare e geometria abbiamo presentato anzitutto, nel capitolo 8, un'esposizione dei contenuti fondamentali dell'algebra lineare e della geometria analitica, con un percorso graduale nel livello di astrazione introdotto: prima la geometria nel piano e nello spazio, poi lo studio dello spazio  $\mathbb{R}^n$ , quindi gli spazi vettoriali astratti. Si passa quindi al concetto di trasformazione lineare tra spazi vettoriali, ai primi elementi del calcolo matriciale e della teoria dei sistemi lineari, e si chiude il capitolo con i concetti fondamentali relativi agli autovекторi di una matrice. Nel capitolo 9 ci occupiamo invece della geometria delle coniche e delle quadriche. Si tratta di un argomento molto classico, che purtroppo ha sempre meno spazio nell'insegnamento universitario, mentre ci sembra istruttivo sia per le molte applicazioni che queste curve e superfici hanno, sia perché è un buon esempio di applicazione geometrica dei metodi dell'algebra lineare studiati in precedenza, di cui quindi viene illustrata la potenza.

Riteniamo utile esplicitare anche in questa nuova edizione i criteri didattici generali che ci hanno guidato fin qui e che ispirano questo testo:

1. Anzitutto, introdurre il minimo di astrazione necessaria per raggiungere l'obiettivo di conoscere, comprendere e saper utilizzare i contenuti di base dell'analisi matematica, l'algebra lineare e la geometria, con particolare riguardo agli aspetti effettivamente utilizzati negli altri corsi della laurea triennale.

2. Mantenere un equilibrio tra sinteticità e chiarezza: *Things should be made as simple as possible, but not any simpler*<sup>1</sup> (Einstein). L'eccessiva brevità oscura le idee. La giustificazione del risultato, la dimostrazione, quando non richieda un apparato formale troppo pesante, e quindi non sia incompatibile con la sinteticità, rende più consapevoli dei nessi e perciò aiuta a comprendere.
3. Motivazione. In uno studio impegnativo come quello della matematica, la motivazione gioca un ruolo fondamentale. D'altro canto, lo studente che affronta un corso di matematica di base, di solito sta iniziando lo studio di una disciplina tecnico-scientifica, che costituisce il suo interesse principale. Perciò si è cercato di presentare ogni nuovo concetto attraverso esempi tratti dalle applicazioni più comuni e di sviluppare la teoria accompagnandola costantemente con riferimenti a problemi tratti dalle varie scienze, evidenziando ove possibile il ruolo dello strumento matematico nella modellizzazione scientifica. I significati geometrici e fisici dei vari concetti introdotti sono stati costantemente evidenziati.
4. Nessuna separazione tra “teoria” e “pratica”. Non esiste sapere senza saper fare, e viceversa. Esempi, esercizi e applicazioni sono costantemente alternati alla presentazione teorica.
5. Modularità. I corsi di matematica di base sono variamente organizzati nei vari corsi di studio e nelle varie sedi. Inevitabilmente ogni docente dovrà scegliere quali parti del testo svolgere e quali no, nei propri corsi. Si è cercato di mantenere la massima modularità e indipendenza possibile, compatibilmente con la struttura logica del discorso matematico. In ogni capitolo la materia è stata organizzata raggruppando i concetti irrinunciabili in alcuni paragrafi. Tutto questo dovrebbe rendere agevole per il docente, e quindi per lo studente, un utilizzo parziale del libro.

In particolare, di quasi tutti i teoremi del calcolo differenziale e integrale in una variabile qui presentati è stata fornita una dimostrazione, ottenuta con un percorso logico che renda il più contenuto possibile l’investimento iniziale di tipo fondazionale, ridotto sostanzialmente al solo teorema di esistenza del limite per successioni monotone. Alla fine del capitolo sul calcolo integrale è stata inserita una sezione che, avendo come punto d’arrivo la dimostrazione dell’integrabilità delle funzioni continue, presenta sinteticamente i principali temi e risultati fondazionali tradizionali: teorema di Bolzano-Weierstrass, completezza di  $\mathbb{R}$ , teorema di Heine-Cantor. Svolgendo anche questa sezione si ha un quadro ragionevolmente completo dei fondamenti dell’analisi matematica di base, mentre omettendolo non è compromessa in alcun modo la comprensione delle restanti parti del testo. Per quanto riguarda l’algebra lineare, considerato il numero non piccolo di argomenti diversi trattati, abbiamo preferito omettere alcune dimostrazioni particolarmente impegnative, pensando che difficilmente in un corso di questo tipo potrebbero trovare spazio senza pregiudicare la trattazione di altri argomenti importanti. Con queste poche eccezioni, comunque, anche in questa parte abbiamo cercato di mantenere la massima continuità del discorso deduttivo.

Desideriamo anche in questa occasione ringraziare tutti i nostri colleghi e studenti che nell’arco di questi anni, con i loro suggerimenti, osservazioni e critiche, ci hanno stimolato a cercare di migliorare continuamente i nostri testi.

Giugno 2014

Gli Autori

<sup>1</sup> “Le cose andrebbero rese il più semplice possibile, ma non troppo semplici”

# 1

# Numeri

L'obiettivo principale di questo capitolo è introdurre gli *oggetti* più elementari del discorso matematico: i *numeri*. Sul concetto di numero, come vedremo, si basa quello, centrale in tutto questo corso, di *funzione*, che sarà introdotto nel prossimo capitolo e ampiamente ripreso in tutto il seguito.

Tuttavia, è necessario prima introdurre o puntualizzare alcuni concetti che hanno a che fare col *linguaggio matematico*. Come vedremo, non si tratta solo di questioni terminologiche, ma piuttosto di questioni logiche e di metodo: il linguaggio matematico infatti è strettamente legato alla *logica*. A sua volta, da più di cent'anni la logica matematica utilizza anche il linguaggio specifico degli *insiemi*.

Gli insiemi quindi si possono vedere per un verso come altri oggetti elementari del discorso matematico (come lo sono i numeri), per un altro verso come parte integrante del linguaggio logico con cui parliamo di matematica. Al linguaggio logico-insiemistico dedicheremo perciò il primo paragrafo.

## ■ 1 INSIEMI E LOGICA

### 1.1 Concetti di base sugli insiemi

La teoria degli insiemi può essere presentata e studiata come teoria assiomatica. Storicamente questo è stato fatto a partire dal 1908, con Zermelo<sup>1</sup>. In modo più informale, la teoria veniva utilizzata già dagli anni 1880<sup>2</sup>; noi la utilizzeremo in questo modo informale, o, come si usa dire, “ingenuo”.

#### Le parole chiave

Il linguaggio degli insiemi si basa su tre parole chiave:

1. Insieme
2. Elemento
3. Appartenenza

---

<sup>1</sup>Ernst Zermelo, 1871-1953, matematico tedesco.

<sup>2</sup>Ad esempio da Georg Cantor (1845-1918), normalmente considerato il padre del concetto di insieme.

### Significato dei termini

Non *definiamo* queste tre parole, ma possiamo spiegarle in modo informale ed implicito. Quando si dice: l'insieme dei punti di un piano, l'insieme degli iscritti a un'università, l'insieme delle stelle di una galassia... tutti comprendono il significato di queste frasi; la nozione di *insieme* è generalmente assunta come primitiva (cioè non riducibile a concetti più elementari). Useremo come sinonimi le espressioni: collezione, classe, aggregato, famiglia... Un insieme è determinato dai suoi *elementi*, nel senso che un insieme è definito quando abbiamo un criterio con cui stabilire se un dato oggetto è o non è elemento di questo insieme. Invece di dire "il tale oggetto è un elemento del tale insieme" si dice anche che quell'elemento *appartiene* all'insieme.

### Simboli

Per indicare gli insiemi si usano solitamente lettere maiuscole, come

$$A, B, X, Y \dots$$

Per indicare gli elementi di un insieme si usano solitamente lettere minuscole  $a, b, \dots, x, y \dots$ , e per indicare che un elemento  $x$  appartiene all'insieme  $A$  scriviamo

$$x \in A$$

o anche  $A \ni x$  ("Ad  $A$  appartiene  $x$ ").

### Come si specifica un insieme

La scrittura

$$A = \{1, 2, 5\}$$

significa che l'insieme  $A$  ha come elementi i numeri 1, 2, 5 (e nient'altro): l'insieme è quindi ben definito. In questo caso si dice che l'abbiamo definito *per tabulazione*, ossia elencando gli elementi che vi appartengono. Per questo scopo si usa sempre la scrittura  $A = \{\dots\}$  (cioè con le parentesi graffe che contengono l'elenco degli elementi, separati da virgolette). Definire un insieme per tabulazione presuppone che l'insieme abbia un numero *finito* di elementi. Come vedremo, molto spesso in matematica abbiamo a che fare con insiemi infiniti; per definire questi insiemi ci vorrà quindi un altro metodo: vedremo in seguito quale.

### Osservazioni sul concetto di insieme

Notiamo che

$$\{1, 2, 5\} = \{1, 5, 2\}$$

ossia: l'*ordine* in cui elenchiamo gli elementi è irrilevante, perché, come detto sopra, ciò che determina un insieme è sapere quali elementi appartengono o meno ad esso, e nient'altro. Il concetto di un *ordine* tra gli elementi è estraneo al concetto di insieme. Così pure gli è estraneo il concetto di *molteplicità* degli elementi: in altre parole, non ha senso dire che un certo elemento di un insieme è "contato due volte". Ad esempio, l'insieme delle soluzioni dell'equazione

$$(x - 1)^2 = 0$$

(che ha come unico elemento il numero 1) è uguale all'insieme delle soluzioni dell'equazione

$$(x - 1)^2 = 0.$$

Il fatto che la seconda equazione abbia la soluzione  $x = 1$  che ha molteplicità algebrica 2, non cambia l'insieme delle sue soluzioni.

### Relazioni tra insiemi: uguaglianza

Abbiamo già detto che due insiemi sono uguali quando possiedono gli stessi elementi. Detto più esplicitamente:

$$A = B \text{ significa:}$$

"Ogni elemento che appartiene ad  $A$  appartiene anche a  $B$  e ogni elemento che appartiene a  $B$  appartiene anche ad  $A$ ". Si noti che dimostrare l'uguaglianza tra due insiemi comporta quindi dimostrare *due* proposizioni distinte. Formalmente:

$$(1.1) \quad \forall x (x \in A \Rightarrow x \in B) \text{ e } \forall x (x \in B \Rightarrow x \in A).$$

Abbiamo introdotto due *simboli logici*:

il simbolo  $\forall$  si chiama *quantificatore universale* e si legge "per ogni", "per tutti", "per ciascuno";

il simbolo  $\Rightarrow$  si chiama *implicazione logica* e si legge "implica" o "se... allora" (ad esempio: "se  $x \in A$  allora  $x \in B$ ").

Riprenderemo entrambi più avanti.

### Relazioni tra insiemi: inclusione

Può accadere che valga solo una delle due richieste espresse dalla definizione di uguaglianza (1.1). Ad esempio, se sappiamo solo che ogni elemento di  $A$  è anche elemento di  $B$ , potremo dire che  $A$  è *contenuto* in  $B$ . Si scrive:

$$A \subseteq B$$

(e si legge "A è contenuto in B") per indicare che

$$\forall x (x \in A \Rightarrow x \in B).$$

Per quanto detto, quindi, dire che  $A = B$  equivale a dire che " $A \subseteq B$  e  $B \subseteq A$ ". Quando  $A \subseteq B$ , diciamo anche che  $A$  è un *sottoinsieme* di  $B$ .

Se si afferma che  $A \subseteq B$ , non si esclude che sia  $A = B$ . Se invece vogliamo proprio affermare che  $A$  è contenuto in  $B$  ma non coincide con  $B$ , diremo che  $A$  è *strettamente contenuto* in  $B$ , e scriveremo

$$A \subset B$$

(si scrive talvolta anche  $A \subsetneq B$ ). Si parla in tal caso di *inclusione stretta*. Esplicitamente, affermare che  $A \subset B$  significa affermare che

"Ogni elemento che appartiene ad  $A$  appartiene anche a  $B$  ed esiste un elemento di  $B$  che non appartiene ad  $A$ ".

Formalmente:

$$\forall x (x \in A \Rightarrow x \in B) \text{ e } \exists x \in B : x \notin A.$$

Così facendo abbiamo usato altri tre simboli logici:

- il simbolo  $\exists$  si chiama *quantificatore esistenziale* e si legge “esiste”, “esistono”;
- il simbolo  $:$  si legge “tale che” (e tipicamente accompagna il simbolo “esiste” nella costruzione “*esiste*  $x$  appartenente ad  $A$  tale che  $x$  ha una certa proprietà”);
- il simbolo  $\notin$  si legge “non appartiene”.

Si faccia attenzione a non confondere tra loro le relazioni “appartiene a” e “è contenuto in” (in simboli,  $\in$  e  $\subseteq$ ). In un certo senso, sono due modi per indicare che “qualcosa sta dentro qualcos’altro”, ma hanno una differenza logica fondamentale: un *insieme* è *contenuto* in un altro insieme; un *elemento appartiene* a un insieme. L’appartenenza è una relazione tra due oggetti matematici che stanno su due piani gerarchici diversi, mentre l’inclusione è una relazione tra due oggetti logici che stanno sullo stesso livello gerarchico. Ad esempio:

$$\begin{aligned} 3 &\in \{1, 3, 4\}; \\ \{3\} &\subseteq \{1, 3, 4\}; \\ \{1, 4\} &\subseteq \{1, 3, 4\}. \end{aligned}$$

In particolare, non si confonda 3 (che è un numero) con  $\{3\}$ , che è l’insieme che contiene come unico elemento il numero 3.

Talvolta si considerano insiemi che hanno per elementi altri insiemi; questo non contraddice l’osservazione appena fatta: in ogni caso, i simboli  $\in$  e  $\subseteq$  non sono intercambiabili, ma devono rispettare la distinzione dei livelli gerarchici che abbiamo spiegato. Ad esempio, se definiamo l’insieme

$$A = \{\{1\}, \{2\}, \{1, 2\}\},$$

allora è corretto affermare che  $\{2\} \in A$ , perché  $\{2\}$  è un insieme che svolge il ruolo di *elemento* di  $A$ , mentre sarebbe scorretto dire che  $\{2\} \subseteq A$ , perché questo significherebbe che  $2 \in A$ , mentre  $A$  ha come elemento  $\{2\}$ , ma non 2.

#### Insieme vuoto

Si definisce *insieme vuoto* l’insieme che non contiene alcun elemento, e si indica tale insieme con il simbolo

$$\emptyset.$$

L’insieme vuoto “ha zero elementi”, ma *non è* il numero zero! Non è neanche l’insieme  $\{0\}$  (cioè l’insieme che ha zero come unico elemento); infatti,  $\{0\}$  ha un elemento, mentre  $\emptyset$  non ne ha nessuno. (Il lettore quindi, d’ora in poi *non usi* il simbolo  $\emptyset$  per indicare il numero zero).

Per qualsiasi insieme  $A$ , possiamo affermare che

$$\emptyset \subseteq A.$$

Per dimostrarlo, dovremmo provare che ogni elemento che appartiene a  $\emptyset$ , appartiene anche ad  $A$ ; ma nessun elemento appartiene a  $\emptyset$ , per cui la dimostrazione è terminata.

#### Insieme delle parti

L’esempio più naturale di insieme che ha come elementi altri insiemi è il seguente. Fissiamo un insieme  $X$ ; consideriamo ora l’insieme che ha per elementi tutti i sottoinsiemi di  $X$ . Questo insieme si chiama *insieme delle parti di  $X$* , e si indica col simbolo

$$\mathcal{P}(X).$$

Ogni insieme  $X$  ha due *sottoinsiemi banali*, che sono  $X$  stesso e l’insieme vuoto  $\emptyset$  (potrebbero coincidere, se  $X$  è vuoto): non bisogna dimenticarseli, quando si elencano tutti i sottoinsiemi di  $X$ . Ad esempio, se

$$X = \{1, 2, 3\},$$

allora

$$\mathcal{P}(X) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, X\}.$$

Per esercizio, il lettore provi a dimostrare che se  $X$  ha  $n$  elementi, allora  $\mathcal{P}(X)$  ha  $2^n$  elementi. Se  $X$  è infinito, anche  $\mathcal{P}(X)$  è infinito, naturalmente. Si può dimostrare che, in un certo senso, anche in questo caso è vero che  $\mathcal{P}(X)$  è “più numeroso” di  $X$ . Approfondiremo quest’ultima osservazione nel par. 6 sugli insiemi infiniti.

#### Insiemi numerici

Prima di continuare il discorso, è bene introdurre i principali *insiemi numerici* che incontreremo nel resto del corso. Per il momento non diremo niente sulle proprietà di questi insiemi: ci basta introdurre il nome e il simbolo con cui si denotano e richiamare allo studente chi sono queste sue vecchie conoscenze.

$\mathbb{N}$  è l’insieme dei *numeri naturali*, ossia i numeri con cui si conta:  $0, 1, 2, 3, \dots$

$\mathbb{Z}$  è l’insieme dei *numeri interi* (detti anche numeri relativi), ossia:  $0, 1, -1, 2, -2, \dots$

$\mathbb{Q}$  è l’insieme dei *numeri razionali*, ossia delle frazioni

$$\frac{n}{m}, \text{ dove } n, m \in \mathbb{Z}, \text{ e } m \neq 0.$$

I numeri razionali possono essere scritti anche in forma decimale. Ad esempio:

$$\frac{3}{4} = 0,75, \text{ mentre } \frac{1}{3} = 0,33333\dots = 0,\overline{3}.$$

Gli esempi vogliono ricordare il seguente fatto: un numero razionale, scritto in forma decimale, dopo la virgola può presentare un numero finito di cifre (diverse da zero), oppure un numero infinito di cifre diverse da zero, che però si ripetono *periodicamente*. Si ricordi che  $0,\overline{9} = 1$ . (Per convincersene, il lettore si chieda quanto vale  $1 - 0,\overline{9}$ , e provi a rispondere ragionando).

$\mathbb{R}$  è l’insieme dei *numeri reali*, ossia quelli che, scritti in forma decimale, presentano dopo la virgola una successione qualsiasi di cifre diverse da zero, eventualmente anche infinita e non periodica. Che esistano numeri di quest’ultimo tipo (ossia reali ma non razionali), si capisce riflettendo su esempi come:

$$0,10110111011110\dots$$

Il numero precedente dopo la virgola ha: una cifra uguale a 1, poi 0, poi due cifre uguali a 1, poi 0, poi tre cifre uguali a 1... e così via. È chiaro che questo criterio

definisce con precisione un numero decimale. D'altro canto, l'allineamento di cifre diverse da zero dopo la virgola non è né finito né periodico: questo numero perciò è *irrazionale*.

Nel corso incontreremo anche:  $\mathbb{C}$ , l'insieme dei numeri complessi, ossia del tipo  $a + ib$ , dove  $a, b$  sono numeri reali, e  $i$  è l'unità immaginaria, ossia un numero il cui quadrato è  $-1$ . Tra gli insiemi numerici ora ricordati, questo è l'unico che probabilmente non è una vecchia conoscenza per lo studente: dedicheremo a questo insieme il par. 8, definendolo rigorosamente. Per ora ci accontentiamo di averne introdotto il nome e il simbolo.

Tra gli insiemi numerici introdotti, valgono le inclusioni:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Come indicato dai simboli, tutte le inclusioni sono strette: esistono numeri interi non naturali (i numeri negativi), esistono numeri razionali non interi (le frazioni proprie), esistono numeri reali non razionali (i numeri irrazionali), esistono numeri complessi non reali (i numeri immaginari).

### Operazioni su insiemi

Molto spesso, nel contesto di un certo discorso matematico, esiste un insieme  $X$  che svolge il ruolo di *universo*, nel senso che tutti gli insiemi di cui si parla in quel contesto sono sottoinsiemi di  $X$ . Ad esempio, in questioni di aritmetica potrebbe essere  $X = \mathbb{N}$ , mentre in questioni di analisi potrebbe essere  $X = \mathbb{R}$ . Perciò succede spesso di trovarsi a operare con insiemi che sono tutti sottoinsiemi di un certo universo comune,  $X$ . In queste condizioni, si possono definire certe *operazioni sugli insiemi*.

- *Intersezione*. L'intersezione di due insiemi  $A, B$  è definita da:

$$A \cap B = \{x : x \in A \text{ e } x \in B\}.$$

In altre parole, è l'insieme degli elementi che appartengono sia al primo sia al secondo insieme.

- *Unione*. L'unione di due insiemi  $A, B$  è definita da:

$$A \cup B = \{x : x \in A \text{ o } x \in B\}.$$

In altre parole, è l'insieme degli elementi che appartengono al primo o al secondo insieme, intendendo la "o" in modo non esclusivo (esplicitamente:  $A \cup B$  è l'insieme degli elementi che appartengono ad  $A$  o a  $B$  o a entrambi).

- *Differenza*. La differenza tra due insiemi  $A, B$  è definita da:

$$A \setminus B = \{x : x \in A \text{ e } x \notin B\}.$$

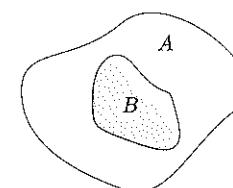
In altre parole, è l'insieme degli elementi che appartengono al primo ma non al secondo insieme.

- *Complementare*. È un tipo particolare di differenza. Se  $A \subseteq X$ , il complementare di  $A$  rispetto a  $X$  è per definizione:

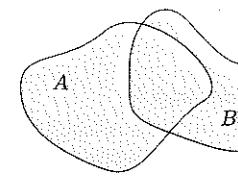
$$\complement_X A = X \setminus A.$$

Spesso  $X$  è sottointeso: si usa allora il simbolo  $A^c$ .

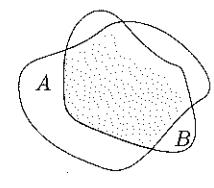
Le definizioni appena date delle operazioni tra insiemi sono visualizzate nei diagrammi in figura 1.1.



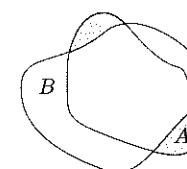
inclusione:  $B \subseteq A$



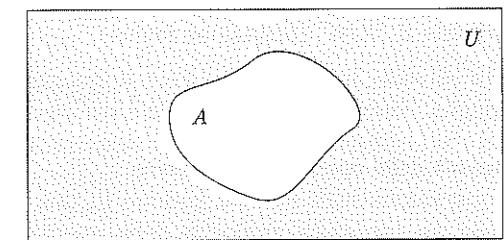
unione:  $A \cup B$



intersezione:  $A \cap B$



differenza:  $A \setminus B$



complementazione:  $\complement_U A$

Figura 1.1. Diagrammi di Venn.

C'è poi un'altra operazione sugli insiemi, che può essere eseguita su due insiemi qualsiasi (cioè due insiemi non necessariamente contenuti nel medesimo universo):

- *Prodotto cartesiano*. Dati due insiemi non necessariamente distinti  $A$  e  $B$ , possiamo considerare un nuovo insieme costituito da tutte le *coppie ordinate*  $(a, b)$  con  $a \in A$  e  $b \in B$ . Esso prende il nome di *prodotto cartesiano* di  $A$  per  $B$  e si indica col simbolo  $A \times B$ .

Una illustrazione è data nella figura 1.2, dove si può osservare anche che  $A \times B$  in generale è diverso da  $B \times A$ .

Un tipico uso di prodotto cartesiano si ha con  $\mathbb{R} \times \mathbb{R}$ , che si abbrevia col simbolo  $\mathbb{R}^2$ , e denota l'insieme delle *coppie ordinate di numeri reali*. Analogamente,  $\mathbb{R}^n$  (abbreviazione del prodotto cartesiano di  $n$  insiemi uguali ad  $\mathbb{R}$ ) è l'insieme delle *n-uple ordinate di numeri reali*:

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) : x_i \in \mathbb{R}\}.$$

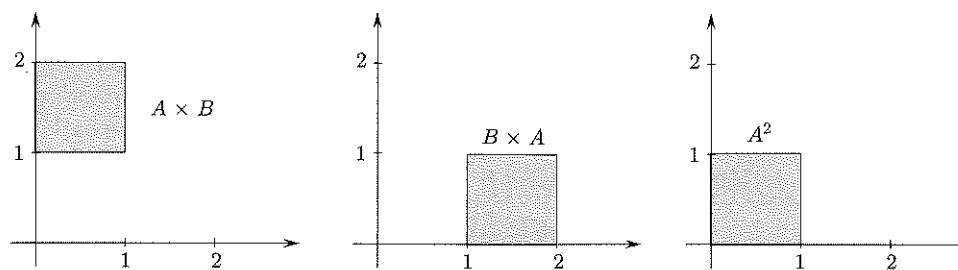


Figura 1.2.

### Operazioni tra insiemi e operazioni logiche

Nelle definizioni appena date, non sarà sfuggita al lettore la relazione tra *operazioni sugli insiemi* e *operazioni logiche*. Precisamente:

- L'intersezione insiemistica è definita mediante la “e” (congiunzione logica).
- L'unione insiemistica è definita mediante la “o” (disgiunzione logica).
- La differenza insiemistica e l'operazione di complementazione sono definite mediante il “non” (negazione logica).

Come si vede, linguaggio logico e linguaggio insiemistico sono due facce della stessa medaglia. Vedremo in seguito altre corrispondenze di questo tipo.

### Proprietà delle operazioni su insiemi

Le operazioni sugli insiemi godono di una serie di proprietà, che discendono in sostanza da proprietà analoghe delle operazioni logiche (congiunzione, disgiunzione, negazione)<sup>3</sup>. Queste proprietà sono utili quando si vuole trasformare un'espressione astratta che coinvolge operazioni su insiemi in un'altra equivalente, analogamente a quanto facciamo comunemente nel calcolo letterale, quando trasformiamo un'espressione in un'altra. Si osservi che alcune proprietà delle operazioni insiemistiche non assomigliano a quelle delle operazioni sui numeri:

- *Proprietà dell'intersezione:*  
commutativa:  $A \cap B = B \cap A$ ;  
associativa:  $A \cap (B \cap C) = (A \cap B) \cap C$   
idempotenza:  $A \cap A = A$
- *Proprietà dell'unione:*  
commutativa:  $A \cup B = B \cup A$ ;  
associativa:  $A \cup (B \cup C) = (A \cup B) \cup C$   
idempotenza:  $A \cup A = A$
- *Proprietà distributive* (legano unione e intersezione):  
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
  
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

<sup>3</sup>Si tratta delle regole logiche con cui utilizziamo “e”, “o”, “non” in qualsiasi ragionamento deduttivo. A questo livello informale del discorso non è il caso nemmeno di esplicitarle.

- *Leggi di De Morgan* (legano il complementare con le altre operazioni):  

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

$$(A^c)^c = A$$

Si noti come l'operazione di complementare scambi tra loro unione e intersezione.

- Operazioni con l'insieme vuoto e con l'insieme universo  $X$ :

$$\begin{aligned} A \cap \emptyset &= \emptyset \\ A \cap X &= A \\ A \cup \emptyset &= A \\ A \cup X &= X \\ X^c &= \emptyset \\ \emptyset^c &= X \end{aligned}$$

- Relazioni tra l'inclusione e le operazioni di unione e intersezione:

$$A \subseteq B \text{ se e solo se } A \cap B = A \text{ se e solo se } A \cup B = B.$$

## 1.2 Un po' di logica elementare

### Predicati (o proprietà) e proposizioni (o enunciati)

Consideriamo la seguente affermazione:

$$(1.2) \quad \text{"il numero naturale } n \text{ è dispari"}$$

e chiediamoci se è vera. Ovviamente l'unica risposta sensata è: “dipende da  $n$ ”. Infatti nella (1.2) il simbolo  $n$  rappresenta una *variabile* che può assumere valori diversi e che possono rendere l'affermazione vera o falsa.

Una frase di questo tipo si chiama *predicato* o *proprietà*: la sua verità o falsità dipende dai valori della o delle variabili che in essa compaiono.

Esaminiamo ora l'affermazione

$$\text{"Per ogni numero naturale } n, \text{ se } n \text{ è dispari allora } n^2 \text{ è dispari"}$$

che possiamo scrivere più formalmente nel modo seguente:

$$(1.3) \quad \forall n \in \mathbb{N} (n \text{ dispari} \implies n^2 \text{ dispari})$$

In questo caso non ha senso assegnare liberamente valori particolari ad  $n$ . Per esempio, per  $n = 3$  si otterebbe “Per ogni numero naturale 3, ...” che è priva di significato. Si dice che la variabile  $n$  non è libera, ma vincolata dal *quantificatore*  $\forall$ .

Come conseguenza si ha che la (1.3) è *vera* o *falsa* “una volta per tutte” e prende il nome di *enunciato* o *proposizione*. In particolare, i componenti della (1.3) sono l'insieme  $\mathbb{N}$ , due predicati definiti su  $\mathbb{N}$  dati da

$$p(n) : "n \text{ dispari}" \quad \text{e} \quad q(n) : "n^2 \text{ dispari}"$$

e l'implicazione  $p(n) \implies q(n)$ . In generale un enunciato che presenta un insieme  $A$ , due predicati  $p(x)$  e  $q(x)$  il cui argomento  $x$  varia in  $A$  e la struttura logica

$$(1.4) \quad \forall x \in A (p(x) \implies q(x))$$

prende il nome di *implicazione universale*.

**Dimostrazioni e controesempi**

La maggior parte dei teoremi è costituita da implicazioni universali, nelle quali il predicato  $p$  fa la parte dell'*ipotesi* e  $q$  quella della *tesi*. In particolare la (1.3) è un teorema. In questo caso ci si convince facilmente che l'asserto è vero, ma come si fa a dimostrarlo rigorosamente? Per esempio, è sufficiente osservare che

$$3 \text{ è dispari e } 3^2 \text{ è dispari}$$

per affermare che il teorema è vero?

Certamente no, poiché il teorema pretende che l'implicazione valga per *ogni* numero naturale. Tuttavia, i numeri dispari sono infiniti: come facciamo a provare un'implicazione per infiniti numeri? Il procedimento chiave è questo: si considera il generico  $n$  che soddisfa l'*ipotesi* (essere dispari) e si dimostra che  $n$  soddisfa la *tesi* (il suo quadrato è dispari). Vediamo come si opera.

**DIMOSTRAZIONE.** Sia  $n$  dispari. Osserviamo che qualunque numero dispari si può scrivere nella forma  $2k + 1$ , con  $k$  opportuno. Sia dunque  $n = 2k + 1$  il generico numero dispari. Per dimostrare il teorema occorre poter scrivere  $n^2$  come (intero pari +1). Si ha:

$$n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1.$$

Poiché  $2(2k^2 + 2k)$  è un intero pari  $n^2$  risulta dispari. ◆

Evidenziamo il procedimento logico usato nella dimostrazione: per dimostrare un'implicazione universale come (1.4), si considera il generico  $x$  che soddisfa l'ipotesi  $p(x)$  e si cerca di mostrare che  $q(x)$  è vera.

Veniamo ora ai *controesempi*, che costituiscono un'importante risorsa per dimostrare la... falsità di un'implicazione universale. Chiediamoci, per esempio, se il seguente teorema è vero o falso:

**TEOREMA 1.1** *Per ogni numero naturale  $n$ , se  $n$  è primo allora  $n$  è dispari.*

Un attimo di riflessione mostra che questo teorema è falso. Infatti, il numero 2 è primo ma è pari.

Si osservi che anche questo teorema esprime un'implicazione universale, ossia ha la stessa struttura logica di quello dell'esempio precedente. Riflettiamo: come mai per poter affermare che un'implicazione universale è vera è necessaria una dimostrazione (mentre un esempio non è sufficiente), mentre per mostrare che un'implicazione universale è falsa basta un esempio contrario?

Il punto è che l'implicazione universale pretende proprio che *ogni*  $x$  che soddisfa l'ipotesi soddisfi anche la tesi: perciò, se troviamo anche un solo esempio di  $x$  che soddisfa l'ipotesi ma non la tesi, questo significa che l'implicazione universale è falsa. Si badi: non “falsa in un caso”, ma semplicemente “falsa”, perché, come abbiamo già osservato, l'implicazione universale è vera o falsa una volta per tutte.

In generale, un esempio di oggetto  $x$  che soddisfa l'ipotesi ma non la tesi di una implicazione universale, e quindi ne mostra la falsità, si chiama *controesempio*. Il numero  $n = 2$  è un controesempio al fatto che “per ogni naturale  $n$ , se  $n$  è primo allora  $n$  è dispari”. Detto in un altro modo: la *negazione* della proposizione

“Per ogni  $x \in A$ , se vale  $p(x)$  allora vale  $q(x)$ ” è

“Esiste un  $x \in A$  per cui vale  $p(x)$  ma non vale  $q(x)$ ”. (Questo  $x$  particolare costituisce un controesempio).

**Implicazione logica e inclusione insiemistica. Insiemi definiti da proprietà**

Abbiamo visto in precedenza che esiste un parallelismo tra le operazioni di intersezione, unione, complementare e le operazioni logiche di congiunzione (“e”), disgiunzione (“o”), e negazione (“non”). Possiamo ora completare questo quadro osservando il parallelismo che esiste tra la relazione di inclusione insiemistica e l'implicazione logica. Per spiegarlo, consideriamo l'implicazione universale:

“Per ogni numero naturale  $n$ , se  $n$  è divisibile per 4 allora  $n$  è divisibile per 2”.

Se indichiamo con

$$D_4 = \{n \in \mathbb{N} : n \text{ è divisibile per } 4\} \text{ e } D_2 = \{n \in \mathbb{N} : n \text{ è divisibile per } 2\}$$

possiamo osservare che l'implicazione universale scritta sopra è equivalente all'affermazione:

$$“D_4 \subseteq D_2”.$$

Infatti, questa inclusione significa che ogni elemento appartenente a  $D_4$  appartiene anche a  $D_2$ , cioè che ogni numero naturale divisibile per 4 è anche divisibile per 2.

Astraendo dall'esempio, possiamo affermare che l'implicazione universale:

“Per ogni  $x \in A$ , se vale  $p(x)$  allora vale  $q(x)$ ” è equivalente all'inclusione insiemistica

$$\{x \in A : p(x) \text{ è vera}\} \subseteq \{x \in A : q(x) \text{ è vera}\}.$$

Questo ci porta anche a osservare esplicitamente il ruolo importante giocato dagli *insiemi definiti da proprietà*. Introducendo il concetto di insieme, abbiamo spiegato cosa significa definire un insieme per tabulazione dei suoi elementi, come quando scriviamo

$$A = \{1, 2, 5\}.$$

Abbiamo anche osservato che questo metodo ha senso solo quando l'insieme che si vuole definire possiede un numero finito di elementi (altrimenti, non possiamo elencarli tutti!). Ora possiamo esplicitare qual è l'altro modo di definire un insieme, modo che abbiamo già usato in vari esempi e che è tipico di quando si vuole definire un insieme infinito: definire un insieme mediante una *proprietà caratteristica*, come nell'esempio:

$$D_4 = \{n \in \mathbb{N} : n \text{ è divisibile per } 4\}$$

o, volendo enunciare in astratto questo schema logico:

$$E = \{x \in A : p(x) \text{ è vera}\}.$$

In queste definizioni, l'importante è che la proprietà  $p(x)$  che si utilizza abbia senso per ogni  $x$  dell'insieme  $A$ , e quindi risulti vera o falsa (senza ambiguità di significato) per ogni particolare  $x \in A$ ; l'insieme  $E$  consisterà allora di *tutti e soli* quegli  $x$  appartenenti ad  $A$  per cui la proprietà  $p(x)$  è vera. A questo proposito, è fondamentale che sia specificato l'insieme universo  $A$  in cui si ragiona: ad esempio, “L'insieme dei numeri naturali divisibili per 4” e non invece “L'insieme dei numeri divisibili per 4”, che avrebbe un significato ambiguo (il numero  $\sqrt{2}$  è divisibile per 4?).

**Negazioni e dimostrazioni indirette**

Terminiamo questa rapida panoramica sui concetti logici e insiemistici puntualizzando alcuni fatti relativi alla negazione di una proposizione, e il suo utilizzo nelle dimostrazioni indirette.

Cominciamo ad osservare che l'implicazione universale

$$(1.5) \quad \forall x \in A (p(x) \implies q(x))$$

è logicamente equivalente a

$$(1.6) \quad \forall x \in A (\text{non } q(x) \implies \text{non } p(x)) .$$

La seconda implicazione si dice la *controinversa* (o *contronomiale*) della prima. Ad esempio: poiché sappiamo che vale il teorema

$$\forall n \in \mathbb{N} (n \text{ dispari} \implies n^2 \text{ dispari}) ,$$

allora possiamo anche affermare che

$$(1.7) \quad \forall n \in \mathbb{N} (n^2 \text{ pari} \implies n \text{ pari}) .$$

Infatti: sia  $n^2$  pari, e proviamo che allora  $n$  è pari; se  $n$  non è pari, allora è dispari, e quindi per la prima implicazione anche  $n^2$  è dispari, contro l'ipotesi che  $n^2$  sia pari. Di conseguenza  $n$  è pari e la seconda implicazione è dimostrata.

Un attimo di riflessione mostra che il ragionamento appena fatto ha una validità generale, e mostra appunto che se è vera la (1.5) allora è vera la (1.6); di più, se è vera la seconda allora è vera la prima (perché “non non  $p$ ” è logicamente equivalente a  $p$ ), per cui le due sono logicamente equivalenti.

L'equivalenza tra (1.5) e (1.6) è detta *legge delle controinverse* e un tipico metodo di dimostrazione indiretta consiste appunto nel provare la (1.6) per mostrare che è vera la (1.5).

L'altra tipica dimostrazione indiretta è la *dimostrazione per assurdo*, che esemplifichiamo col seguente teorema, che ci sarà utile in seguito.

### TEOREMA 1.2

*Non esiste un numero razionale il cui quadrato è 2.*

La seguente dimostrazione si trova negli Elementi di Euclide (circa 300 a.C.).

**DIMOSTRAZIONE.** Supponiamo per assurdo che esista un numero  $r \in \mathbb{Q}$  tale che  $r^2 = 2$ . Possiamo scrivere  $r = \frac{n}{m}$  con  $n, m \in \mathbb{Z}$ ,  $m \neq 0$ . Inoltre, possiamo supporre che la frazione  $n/m$  sia già ridotta ai minimi termini, ossia “semplificata” (in altre parole:  $n, m$  non contengono fattori comuni). Abbiamo dunque la catena di implicazioni:

$$\begin{aligned} \left(\frac{n}{m}\right)^2 &= 2 \\ n^2 &= 2m^2 \end{aligned}$$

per cui  $n^2$  è pari; ma allora per la (1.7) anche  $n$  è pari e possiamo scrivere  $n = 2k$  per qualche  $k \in \mathbb{Z}$ . Allora la relazione  $n^2 = 2m^2$  si riscrive

$$\begin{aligned} (2k)^2 &= 2m^2 \\ 4k^2 &= 2m^2 \\ m^2 &= 2k^2 \end{aligned}$$

per cui  $m^2$  è pari; ma allora per la (1.7) anche  $m$  è pari; dunque sia  $n$  che  $m$  sono pari, e questo è assurdo, perché avevamo supposto che la frazione  $n/m$  fosse già stata semplificata. 

In generale, la dimostrazione per assurdo consiste nel supporre vera l'ipotesi del teorema e la negazione della tesi, e dedurre da questi fatti una contraddizione di qualsiasi tipo.

Sia nelle dimostrazioni per assurdo sia nell'usare la legge delle controinverse, occorre naturalmente saper costruire la corretta *negazione* di una proposizione o proprietà data. Si osservi che questa è un'operazione puramente formale, sintattica, che non ha a che fare col contenuto della proposizione, né col fatto che essa sia vera o falsa.

Riportiamo schematicamente alcune regole tipiche con cui si costruisce la negazione di una proposizione o proprietà. Il lettore è invitato a rendersi conto della validità di queste regole ragionando su esempi concreti.

Se  $p(x), q(x)$  sono due proprietà qualsiasi, allora:

- la negazione di “ $p(x) \wedge q(x)$ ” è “ $\text{non } p(x) \vee \text{non } q(x)$ ”;
- la negazione di “ $p(x) \vee q(x)$ ” è “ $\text{non } p(x) \wedge \text{non } q(x)$ ”;
- la negazione di “ $\text{non } p(x)$ ” è  $p(x)$ ;
- la negazione di “ $\forall x$  vale  $p(x)$ ” è “ $\exists x : \text{non } p(x)$ ”
- la negazione di “ $\exists x : \text{vale } p(x)$ ” è “ $\forall x \text{ non } p(x)$ ”
- la negazione di “ $\forall x (p(x) \implies q(x))$ ” è “ $\exists x : (p(x) \wedge \text{non } q(x))$ ”.

## 2 SOMMATORIE E COEFFICIENTI BINOMIALI

### 2.1 Il simbolo di sommatoria

**DEFINIZIONE 1.1** Siano  $a_1, a_2, \dots, a_n$ ,  $n$  numeri reali. La loro somma

$$a_1 + a_2 + \dots + a_n$$

si può indicare in forma compatta col simbolo di *sommatoria*:

$$\sum_{i=1}^n a_i$$

che si legge: “sommatoria per  $i$  da 1 a  $n$  di  $a_i$ ”. Il simbolo  $i$  si dice *indice di sommatoria*. 

Il simbolo di sommatoria è dunque una pura e semplice stenografia, che tuttavia risulta molto utile quando i termini  $a_i$  sono definiti esplicitamente in funzione dell'indice  $i$ , ad esempio:

$$\sum_{i=1}^{10} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{10}$$

$$\sum_{i=3}^n i^2 = 3^2 + 4^2 + 5^2 + \dots + n^2$$

Per usare agilmente il simbolo di sommatoria occorre capire bene l'utilizzo dell'indice di sommatoria. Anzitutto, l'indice di sommatoria è un *indice muto*. Questo vuol dire che se si sostituisce  $i$  con  $j, k$  o qualunque altro indice (in tutte le sue occorrenze) il senso dell'espressione non cambia:

$$\sum_{i=3}^n i^2 = \sum_{j=3}^n j^2$$

Invece,

$$\sum_{i=1}^n i^2 \neq \sum_{i=1}^m i^2$$

in quanto i due simboli indicano la somma, rispettivamente, dei primi  $n$  oppure dei primi  $m$  quadrati: se  $n \neq m$  il risultato sarà diverso.

Le seguenti proprietà formali delle sommatorie sono facilmente comprensibili se si pensa a ciò che esse affermano in termini di somme scritte per esteso.

#### PROPOSIZIONE 1.1 (PROPRIETÀ FORMALI DELLE SOMMATORIE)

1. *Prodotto per una costante (proprietà distributiva):*

$$\sum_{k=1}^n (c \cdot a_k) = c \cdot \sum_{k=1}^n a_k$$

2. *Sommatoria con termine costante:*

$$\sum_{k=1}^n c = c \cdot n = c \cdot (\text{numero di addendi della sommatoria})$$

3. *Somma di sommatorie:*

$$\sum_{k=1}^n a_k + \sum_{k=1}^n b_k = \sum_{k=1}^n (a_k + b_k)$$

4. *Scomposizione:*

$$\sum_{k=1}^{n+m} a_k = \sum_{k=1}^n a_k + \sum_{k=n+1}^{n+m} a_k$$

5. *Traslazione di indici:*

$$\sum_{k=1}^n a_k = \sum_{k=1+m}^{n+m} a_{k-m}$$

6. *Riflessione di indici:*

$$\sum_{k=1}^n a_k = \sum_{k=1}^n a_{n-k+1} = \sum_{k=0}^{n-1} a_{n-k}$$

La dimostrazione di queste proprietà è un esercizio di trascrizione: basta, cioè, scrivere per esteso che cosa indica ciascuna sommatoria ed eseguire qualche passaggio elementare.

#### Somma di una progressione geometrica

Si dice che  $n$  termini sono in *progressione geometrica* se il rapporto tra ogni termine (a partire dal secondo) e il precedente è costante. Tale costante si dice *ragione* della progressione. Se il primo termine è  $a$  e la ragione è  $q$ , i termini successivi saranno  $aq, aq^2, aq^3$ , e così via.

**PROPOSIZIONE 1.2** *Per la somma dei primi termini della progressione geometrica di ragione  $q \neq 1$  (e  $a = 1$ ), vale la formula:*

$$(2.1) \quad \sum_{k=0}^n q^k = \frac{1 - q^{n+1}}{1 - q}$$

*Se  $q = 1$  la sommatoria scritta vale  $n + 1$ , ovviamente.*

**DIMOSTRAZIONE.** Proviamo l'identità nella forma equivalente:

$$(1 - q) \sum_{k=0}^n q^k = 1 - q^{n+1}.$$

Infatti, applicando le proprietà delle sommatorie 1, 5, 4 (v. proposizione 1.1) si ha:

$$\begin{aligned} (1 - q) \sum_{k=0}^n q^k &= \sum_{k=0}^n q^k - q \sum_{k=0}^n q^k = \\ &= \sum_{k=0}^n q^k - \sum_{k=0}^n q^{k+1} = \sum_{k=0}^n q^k - \sum_{k=1}^{n+1} q^k = \\ &= 1 + \sum_{k=1}^n q^k - \left( \sum_{k=1}^n q^k + q^{n+1} \right) = 1 - q^{n+1} \end{aligned}$$



#### 2.2 Fattoriale di $n$

Un'espressione di uso frequente in Analisi è il *fattoriale* di  $n$ . Con ciò si intende il prodotto dei primi  $n$  interi: si indica con  $n!$  (e si legge “ $n$  fattoriale”)

$$n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$$

Si pone, per definizione,  $0! = 1$ . Il numero  $n!$  cresce molto rapidamente al crescere di  $n$ . Ecco i primi valori

$n$	0	1	2	3	4	5	6	7	8	9	10
$n!$	1	1	2	6	24	120	720	5040	40320	362880	3628800

Alcune proprietà del fattoriale, di verifica immediata, sono:

$$n! = n(n-1)!$$

$$\text{se } 0 < k < n \quad \frac{n!}{(n-k)!} = n(n-1)(n-2)\dots(n-k+1)$$

L'ultimo prodotto scritto è il *prodotto di k fattori, partendo da n e decrescendo*. Per esempio,

$$\frac{100!}{95!} = 100 \cdot 99 \cdot 98 \cdot 97 \cdot 96$$

Si noti che una calcolatrice tascabile riesce a calcolare  $100 \cdot 99 \cdot 98 \cdot 97 \cdot 96$  senza problemi, ma segnala "errore" non appena si digita  $100!$ . Conviene sempre semplificare il più possibile le espressioni che contengono il fattoriale, prima di calcolarle.

Una prima applicazione del fattoriale si ha nel problema seguente: in quanti modi è possibile ordinare  $n$  oggetti distinti?

Per esempio, se abbiamo tre oggetti:  $a, b, c$ , questi possono essere ordinati in  $3! = 6$  modi:

$$abc, \quad acb, \quad bac, \quad bca, \quad cab, \quad cba$$

Ognuno di questi ordinamenti si chiama *permutazione* (o *sostituzione*) degli  $n$  oggetti. Si ottiene facilmente che il numero delle permutazioni di  $n$  oggetti distinti è  $n!$

Basta infatti pensare di realizzare una permutazione collocando gli  $n$  oggetti in  $n$  scatole numerate da 1 a  $n$ . La prima scatola si può riempire in  $n$  modi. Riempita la prima scatola, si può riempire la seconda in  $(n-1)$  modi con i restanti oggetti. Così il complesso delle prime due scatole si può riempire in  $n(n-1)$  modi. Il complesso delle prime tre scatole si può riempire in  $n(n-1)(n-2)$  modi e così via fino a riempire tutte le scatole.

### 2.3 Coefficienti binomiali e formula di Newton

Sviluppando la potenza  $n$ -esima di un binomio  $(a+b)^n$  si trova

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

I numeri  $c_{n,k}$  sono detti *coefficienti binomiali* e intervengono in molte questioni non solo di Analisi, ma anche di Probabilità, Statistica, etc. Essi si indicano col simbolo  $\binom{n}{k}$  (si legga:  $n$  su  $k$ ); si può dimostrare che hanno l'espressione seguente:

$$c_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad 0 \leq k \leq n$$

Per quanto osservato poco sopra, si può scrivere anche:

$$\binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!}$$

espressione che è più maneggevole per il calcolo effettivo.

La formula di Newton si scrive dunque:

$$(2.2) \quad (a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

Valgono anche le seguenti proprietà, di facile verifica:

$$(2.3) \quad \binom{n}{k} = \binom{n}{n-k} \quad \binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

Una dimostrazione delle proprietà (2.2)-(2.3) sarà data nel paragrafo 7. Quest'ultima relazione permette di calcolare i numeri  $\binom{n}{k}$  per mezzo del cosiddetto *triangolo di Tartaglia*; in cima al triangolo si pone il numero  $\binom{0}{0} = 1$  (per definizione); ai lati si pongono i numeri  $\binom{n}{0} = \binom{n}{n} = 1$  per ogni  $n \geq 1$  e allora, per  $0 < k < n$ , il numero  $\binom{n}{k}$  viene scritto all'incrocio della  $n$ -esima riga e della  $k$ -esima colonna e risulta somma dei due numeri che si trovano nella riga precedente, quello sulla stessa colonna e quello sulla colonna precedente.

	potenza	coefficienti
$n = 0$		1
1	1	1
2	1	2 1
3	1	3 3 1
4	1	4 6 4 1
5	1	5 10 10 5 1
		.....

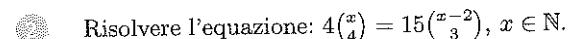
Figura 1.3.

Per esempio, abbiamo  $(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$ .



Usando il triangolo di Tartaglia scrivere esplicitamente

$$(1+a)^8 \quad (2a-3b)^7$$



Risolvere l'equazione:  $4\binom{x}{4} = 15\binom{x-2}{3}$ ,  $x \in \mathbb{N}$ .

Dimostrare le seguenti identità, sfruttando le proprietà delle sommatorie e i suggerimenti forniti:



$$\sum_{i=1}^n i = \frac{n(n+1)}{2}$$

Suggerimento: calcolare  $2 \sum_{i=1}^n i = \sum_{i=1}^n i + \sum_{i=1}^n i$  eseguendo nella seconda sommatoria una riflessione di indici.



$$\sum_{k=0}^{n-1} (2k+1) = n^2$$

Suggerimento: sfruttare il risultato dell'esercizio precedente.



$$\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$$

Suggerimento:  $\sum_{k=0}^n (k+1)^3 = \sum_{k=0}^n (k^3 + 3k^2 + 3k + 1) = \sum_{k=1}^n k^3 + 3 \sum_{k=1}^n k^2 + \dots$   
D'altra parte,  $\sum_{k=0}^n (k+1)^3 = \sum_{k=1}^{n+1} k^3 = \sum_{k=1}^n k^3 + (n+1)^3$ . Dal confronto tra le due espressioni...

Calcolare esplicitamente la seguente somma, semplificando opportunamente:

$$\sum_{k=1}^{100} \left( \frac{1}{k} - \frac{1}{k+1} \right)$$

Calcolare esplicitamente le seguenti somme, sfruttando la formula per una progressione geometrica:

$$\sum_{k=0}^{30} (-1)^k \frac{2^{3k+1}}{3^k} \quad \sum_{k=2}^{100} 3^{2-k}$$

### 3 CAMPI ORDINATI

Inizieremo ora a studiare più da vicino la struttura degli insiemi numerici che abbiamo introdotto in precedenza, ed in particolare l'insieme  $\mathbb{Q}$  dei numeri razionali e l'insieme  $\mathbb{R}$  dei numeri reali. Chiariamo subito che non intendiamo presentare una costruzione rigorosa di questi insiemi. Piuttosto, dando per presupposto che lo studente sappia già cosa sono i numeri razionali e reali (ovvero accontentandoci della descrizione ingenua di questi insiemi che abbiamo presentato nel paragrafo 1), puntualizzeremo con più precisione quali sono le *proprietà* di cui godono questi insiemi numerici.

Iniziamo quindi col richiamare le proprietà (ben note) dei numeri razionali; indichiamo con  $a, b, c \dots$  generici numeri razionali.

- $R_1$ . È definita in  $\mathbb{Q}$  un'operazione (detta *addizione* o *somma*) che ha le seguenti proprietà:

$$1. \forall a, b \quad a + b = b + a \quad (\text{proprietà commutativa})$$

$$2. \forall a, b, c \quad (a + b) + c = a + (b + c) \quad (\text{proprietà associativa})$$

3. Esiste un elemento (elemento neutro della somma, detto *zero* e indicato con 0) tale che  $\forall a, a + 0 = a$

4.  $\forall a$ , esiste un elemento (l'inverso di  $a$  rispetto alla somma, detto *opposto* di  $a$  e indicato con  $-a$ ) tale che  $a + (-a) = 0$ .

- $R_2$ . È definita in  $\mathbb{Q}$  un'operazione (detta *moltiplicazione* o *prodotto*) che ha le seguenti proprietà:

$$1. \forall a, b \quad a \cdot b = b \cdot a \quad (\text{proprietà commutativa}).$$

$$2. \forall a, b, c \quad (a \cdot b) \cdot c = a \cdot (b \cdot c) \quad (\text{proprietà associativa}).$$

3. Esiste un elemento (elemento neutro del prodotto, detto *unità* e indicato con 1) tale che  $\forall a, a \cdot 1 = a$ .

4.  $\forall a \neq 0$ , esiste un elemento (l'inverso di  $a$  rispetto al prodotto, detto *reciproco* di  $a$  e indicato con  $\frac{1}{a}$  oppure  $a^{-1}$ ) tale che  $a \cdot a^{-1} = 1$ .

Le operazioni di somma e prodotto sono legate dalla seguente proprietà:

$$5. \forall a, b, c \quad (a + b) \cdot c = a \cdot c + b \cdot c \quad (\text{proprietà distributiva}).$$

Si osservi che le proprietà delle due operazioni sono quasi le stesse; le uniche differenze sono: il fatto che, per il prodotto, non esiste l'inverso dell'elemento neutro della somma (cioè non esiste  $1/0$ ), e la proprietà distributiva (in cui somma e prodotto giocano un ruolo asimmetrico). Questi fatti fanno sì che le due operazioni siano sostanzialmente diverse.

Dalle proprietà  $R_1$  e  $R_2$  discende la possibilità di eseguire senza restrizioni le quattro operazioni fondamentali: addizione e moltiplicazione (sopra definite), sottrazione (ponendo  $a - b = a + (-b)$ ) e divisione (ponendo  $a : b = ab^{-1}$  purché sia  $b \neq 0$ ).

#### Rappresentazione geometrica

Una rappresentazione di tipo geometrico di  $\mathbb{Q}$  si può ottenere associando a ogni numero razionale un punto della retta euclidea. Ad un punto di questa, scelto arbitrariamente, si associa 0 e a un altro, distinto dal primo, si associa 1, individuando così il segmento orientato 01 che costituisce l'unità di misura. A questo punto si ha una corrispondenza biunivoca tra i numeri razionali e quei punti  $P$  della retta che sono estremi dei segmenti orientati  $OP$  *commensurabili* (nel senso classico della geometria elementare, cioè che hanno un multiplo comune) con 01.



Figura 1.4.

Per descrivere compiutamente la struttura dell'insieme  $\mathbb{Q}$  osserviamo anche che esso è un *insieme ordinato*; cioè che nell'insieme  $\mathbb{Q}$  è definita la relazione  $\leq$  (minore o uguale di ...) ; analogamente la relazione  $\geq$ : maggiore o uguale di ...), la quale è una *relazione d'ordine*, cioè verifica le tre proprietà seguenti:

1.  $\forall a \quad a \leq a$  riflessiva
2.  $\forall a, b \quad \text{se } a \leq b \text{ e } b \leq a \text{ allora } a = b$  antisimmetrica
3.  $\forall a, b, c \quad \text{se } a \leq b \text{ e } b \leq c \text{ allora } a \leq c$  transitiva

Inoltre, presi due qualsiasi numeri razionali  $a, b$ , è sempre possibile confrontarli per mezzo della relazione  $\leq$ , nel senso che vale necessariamente almeno una delle relazioni  $a \leq b$  o  $b \leq a$ . Si esprime questo fatto dicendo che l'ordinamento di  $\mathbb{R}$  è totale, o che  $\mathbb{R}$  è un *insieme totalmente ordinato*.

**OSSERVAZIONE** Si rifletta sulla natura generale astratta delle proprietà che definiscono una relazione d'ordine. Ad esempio, se  $X$  è un insieme qualsiasi (per esempio  $\mathbb{Q}$ ), nell'insieme  $\mathcal{P}(X)$  (insieme di tutti i sottoinsiemi di  $X$ ) possiamo considerare la relazione  $\subseteq$  di inclusione insiemistica; chiaramente, per ogni  $A, B, C \subseteq X$ , è vero che

1.  $A \subseteq A$  (proprietà riflessiva);
2.  $A \subseteq B$  e  $B \subseteq A$  implica  $A = B$  (proprietà antisimmetrica);
3.  $A \subseteq B$  e  $B \subseteq C$  implica  $A \subseteq C$  (proprietà transitiva).

Quindi  $\mathcal{P}(X)$  è un insieme ordinato, rispetto alla relazione d'ordine  $\subseteq$ . Si osservi tuttavia che non è un insieme *totalmente ordinato*. Infatti, dati due insiemi qualsiasi  $A, B \subseteq X$ , in generale può non valere né  $A \subseteq B$  né  $B \subseteq A$ . Quindi: non ogni insieme ordinato è totalmente ordinato.

Mettiamo in evidenza le proprietà dell'ordinamento dei numeri razionali:

- $R_3$ . È definita in  $\mathbb{Q}$  una relazione d'ordine totale  $\leq$ , compatibile con la struttura algebrica, cioè:

1.  $\forall a, b, c \quad \text{se } a \leq b \quad \text{allora } a + c \leq b + c$
2.  $\forall a, b, c \text{ con } c > 0 \quad \text{se } a \leq b \quad \text{allora } ac \leq bc.$

Osserviamo che tutte le regole ben note del calcolo algebrico derivano dalle proprietà  $R_1, R_2, R_3$ . A titolo di esempio, i “principi di identità” usati nel risolvere le disequazioni, non sono altro che le proprietà di compatibilità della relazione d’ordine con la somma e col prodotto: tutte le usuali procedure con cui si risolvono le disequazioni sono quindi conseguenza di questi assiomi e delle proprietà algebriche della somma e del prodotto, espresse da  $R_1, R_2$ .

Astraendo ora dal caso particolare dell’insieme dei numeri razionali, diamo la seguente

**DEFINIZIONE 1.2** Chiameremo *campo ordinato* un insieme in cui sono definite due operazioni (che chiamiamo somma e prodotto) e una relazione d’ordine, che soddisfano tutte le proprietà  $R_1, R_2, R_3$ . Un insieme con le proprietà  $R_1, R_2$  si dice *campo*.

Un momento di riflessione mostra che tutto ciò che abbiamo detto fin qui riguardo alle operazioni di somma e prodotto e alla relazione d’ordine  $\leq$  nell’insieme dei numeri razionali, vale anche per l’insieme dei numeri reali. Possiamo quindi sintetizzare il punto fondamentale di questo paragrafo nella seguente affermazione:

Sia  $\mathbb{Q}$  sia  $\mathbb{R}$  sono campi ordinati.

Da questo punto di vista, quindi,  $\mathbb{Q}$  ed  $\mathbb{R}$  appaiono dotati di proprietà simili. Nel prossimo paragrafo invece evidenzieremo qual è la proprietà che distingue sostanzialmente  $\mathbb{R}$  da  $\mathbb{Q}$  e che rende  $\mathbb{R}$  l’ambiente giusto per sviluppare l’analisi matematica.

## ■ 4 NUMERI REALI. ESTREMO SUPERIORE E ASSIOMA DI CONTINUITÀ

### 4.1 Inadeguatezza dell’insieme dei razionali per misurare le lunghezze

La struttura di campo ordinato dei razionali assolve alla maggior parte degli scopi pratici del calcolo, nel senso che si può esprimere con un numero razionale la misura di ogni grandezza con sufficiente precisione. Tuttavia è noto che ci sono grandezze che non sono commensurabili tra loro: l’esempio classico è dato dalla diagonale e dal lato di un quadrato. Con riferimento alla figura 1.5, se il lato misura 1, l’ascissa  $d$  che misura la diagonale non è un numero razionale. Infatti, per il Teorema di Pitagora,  $d^2 = 1^2 + 1^2 = 2$  ma, come abbiamo dimostrato nel par. 1.2 (Teorema 1.2), non esiste un numero razionale il cui quadrato è 2.

Dunque il punto  $d$  sulla retta non è il rappresentante di alcun numero razionale. Ciò significa che, dopo aver “occupato” i punti della retta con i numeri razionali, su

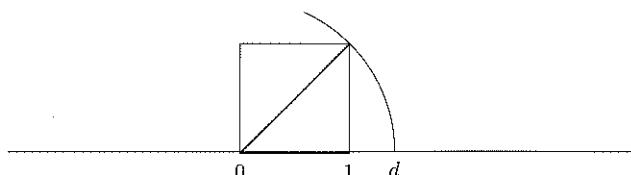


Figura 1.5.

di essa rimangono ancora dei posti vuoti. Significa anche che *l’insieme dei numeri razionali è inadeguato ad esprimere le lunghezze dei segmenti*. Ma la lunghezza è solo la più semplice delle grandezze che si incontrano in geometria e in fisica: un problema esattamente analogo si incontra nella misurazione di qualsiasi altra grandezza, come aree, volumi, tempi, velocità ecc.

Sorge spontanea allora la domanda: è possibile ampliare l’insieme dei razionali in modo da avere ancora un campo ordinato, i cui elementi (numeri) siano in *corrispondenza biunivoca* con i punti della retta euclidea? Il candidato naturale a questa estensione è l’insieme dei numeri reali. La domanda dunque diventa: l’insieme  $\mathbb{R}$  può essere messo in corrispondenza biunivoca con i punti della retta? Vedremo che la risposta a questa domanda (pur di riformularla in modo puramente analitico) è positiva. Per renderci conto di questo dobbiamo però introdurre un nuovo concetto, che ci porterà alla comprensione della fondamentale *proprietà di continuità di  $\mathbb{R}$* , che si rivelerà essenziale per l’analisi matematica.

### 4.2 Estremo superiore e assioma di continuità

Sia  $X$  un qualunque insieme totalmente ordinato. In pratica, ci interesserà il caso  $X = \mathbb{Q}$  o  $X = \mathbb{R}$ , ma è utile non specificare per ora quale dei due ambienti consideriamo.

Sia  $E$  un insieme contenuto in  $X$ . Tale insieme si dice *limitato superiormente* se esiste un numero  $M$  per cui risulti  $x \leq M$  per ogni elemento  $x \in E$ ; *limitato inferiormente* se esiste un numero  $m$  per cui risulta  $x \geq m$  per ogni elemento  $x \in E$ ; *limitato* se valgono entrambe le condizioni, ossia esistono due numeri,  $m$  e  $M$ , tali che

$$m \leq x \leq M \text{ per ogni } x \in E.$$

Introduciamo ora il concetto di elemento *massimo (minimo)* di un insieme. Diremo che un elemento  $\bar{x}$  è *massimo* per  $E$  se:

- i)  $\bar{x} \in E$ ;
- ii)  $x \leq \bar{x} \forall x \in E$ .

Analogia definizione per il minimo  $\underline{x}$ . È evidente che, affinché il massimo (minimo) esista, l’insieme deve essere superiormente (inferiormente) limitato.



	Insieme $E$	Max	Min
I)	$\mathbb{N}$	non esiste	0
II)	Numeri pari relativi	non esiste	non esiste
III)	$\left\{1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots\right\}$	1	non esiste
IV)	$\left\{n \in \mathbb{N} : \frac{n-1}{n+1}\right\}$	non esiste	-1
V)	$\left\{x \in \mathbb{R} : x^3 \geq 27\right\}$	non esiste	3
VI)	$\left\{x \in \mathbb{Q} : x \geq 0, x^2 < 2\right\}$	non esiste	0

Si osserva, negli esempi precedenti, che talvolta, pur essendo l'insieme  $E$  limitato, esso non possiede massimo o minimo.

Introduciamo allora un concetto, fondamentale per l'Analisi matematica, che è sostitutivo del concetto di massimo o minimo: quello di *estremo superiore* (sup) ed *estremo inferiore* (inf).

**DEFINIZIONE 1.3** Sia  $E \subseteq X$ . Un numero  $k \in X$  (non necessariamente  $\in E$ ) si dice un *maggiorante* (*minorante*) di  $E$  se  $k \geq x$  ( $k \leq x$ )  $\forall x \in E$ . Notiamo che un insieme superiormente (inferiormente) limitato ha molti maggioranti (minoranti). Chiameremo *estremo superiore* (*inferiore*) di  $E$ , e lo indicheremo con  $\sup E$  ( $\inf E$ ) il minimo (massimo) dei maggioranti (minoranti) di  $E$  (se esiste).

È evidente che, se l'insieme possiede massimo (minimo), questo coincide con l'estremo superiore (inferiore).

Riprendiamo gli esempi precedenti per vedere se, quando non esiste il massimo o il minimo, esiste però il sup o l'inf. Negli esempi I, II, V la risposta è negativa per l'evidente ragione che gli insiemi ivi considerati non sono limitati (da sopra e/o da sotto). Nell'esempio III è facile vedere che l'inf è 0 (il quale non è minimo poiché 0 non fa parte dell'insieme) e nell'esempio IV il sup è 1. Consideriamo l'esempio VI; si intuisce che il sup dovrebbe essere un numero il cui quadrato è 2; ma in  $\mathbb{Q}$  tale numero non esiste; esiste però in  $\mathbb{R}$  ed è  $\sqrt{2}$ . Questa circostanza non è casuale, ma illustra precisamente la differenza tra l'insieme  $\mathbb{Q}$  dei numeri razionali e l'insieme  $\mathbb{R}$  dei numeri reali. Per chiarirla, diamo la seguente definizione generale:

Si dice che l'insieme  $X$  (totalmente ordinato) possiede la *proprietà dell'estremo superiore*, se

- $R_4$ . Ogni insieme  $E \subset X$  non vuoto e limitato superiormente possiede estremo superiore in  $X$ .

Si prova facilmente che, se vale questa proprietà, allora è anche vero che ogni sottoinsieme di  $X$  non vuoto e inferiormente limitato ammette estremo inferiore.

Si faccia attenzione al contenuto della precedente definizione: non si richiede che  $X$  stesso abbia estremo superiore (ad esempio, nel caso  $X = \mathbb{Q}$  o  $\mathbb{R}$  questo è certamente falso!), ma che *ogni sottoinsieme non vuoto superiormente limitato di  $X$*  ne sia provvisto. L'esempio VI mostra che certamente  $\mathbb{Q}$  non ha la proprietà dell'estremo superiore. Invece,  $\mathbb{R}$  ha questa proprietà. Questo fatto non può essere dimostrato a partire dalla nostra "definizione ingenua" di  $\mathbb{R}$ , data in precedenza. Tuttavia nella presentazione assiomatica di  $\mathbb{R}$ , questa proprietà costituisce parte della definizione stessa di  $\mathbb{R}$ :

**DEFINIZIONE 1.4 (DEFINIZIONE ASSIOMATICA DI  $\mathbb{R}$ )** Chiamiamo  $\mathbb{R}$  un insieme che soddisfa le proprietà  $R_1, R_2, R_3, R_4$ , ossia un campo ordinato che ha la proprietà dell'estremo superiore.

Ricordiamo ancora, per confronto, che l'insieme  $\mathbb{Q}$  soddisfa  $R_1, R_2, R_3$ , ma non  $R_4$ . La proprietà dell'estremo superiore prende anche il nome di *assioma di Dedekind*, o *assioma di continuità*, o di *completezza* e si può enunciare anche nella seguente forma equivalente.

Sia  $\{A, B\}$  una partizione di  $\mathbb{R}$  (cioè  $A$  e  $B$  sono insiemi non vuoti e disgiunti la cui unione è  $\mathbb{R}$ ); essa si chiama *sezione* se:  $\forall a \in A$  e  $\forall b \in B$  risulta  $a < b$ . Allora si dimostra che:

- $R'_4$ . Per ogni sezione  $\{A, B\}$  di  $\mathbb{R}$  esiste un unico numero reale  $s$  (detto elemento separatore) tale che

$$a \leq s \leq b \quad \forall a \in A, \forall b \in B$$

(tale elemento separatore altro non è che  $\sup A = \inf B$ ).

Ritorniamo ora al problema dell'incommensurabilità di lato e diagonale del quadrato. Quando, con riga e compasso, fissato il lato unitario del quadrato, ne costruiamo la diagonale e la riportiamo sulla retta di partenza, l'arco tracciato dal compasso "spezza la retta in due", generando quella che abbiamo chiamato una *sezione*; l'elemento separatore, che esiste per l'assioma  $R_4$  rappresenta geometricamente il punto di intersezione, e quindi quel numero misura la diagonale. Quindi l'interpretazione geometrica della proprietà  $R_4$  ci mostra come l'insieme  $\mathbb{R}$  sia una rappresentazione adeguata della nostra idea intuitiva di retta, così adeguata che spesso in matematica si usa l'espressione "la retta reale" per indicare  $\mathbb{R}$ , confondendo l'insieme numerico con la sua rappresentazione geometrica naturale. Ricordiamo che l'assioma di continuità è stato messo in luce da R. Dedekind nel 1872, ossia circa 2400 anni dopo la scoperta pitagorica della incommensurabilità di lato e diagonale del quadrato.

Come vedremo nel resto del corso, le conseguenze *analitiche* dell'assioma di continuità sono rilevanti per l'Analisi matematica.

### 4.3 Valore assoluto. Diseguaglianza triangolare

Si dice *valore assoluto* del numero reale  $a$  (o *modulo* di  $a$ ) il numero non negativo così definito:

$$(4.1) \quad |a| = \begin{cases} a & \text{se } a \geq 0 \\ -a & \text{se } a < 0 \end{cases}$$

Dalla definizione di valore assoluto segue immediatamente che:

$$(4.2) \quad \forall a \geq 0 : |x| \leq a \iff -a \leq x \leq a$$

Dalla (4.2) segue l'importante *diseguaglianza triangolare*:

$$(4.3) \quad \forall x, y \in \mathbb{R} : |x + y| \leq |x| + |y|$$

Infatti, basta scrivere le relazioni:

$$-|x| \leq x \leq |x| \quad \text{e} \quad -|y| \leq y \leq |y|$$

quindi sommare membro a membro, ottenendo

$$-(|x| + |y|) \leq x + y \leq |x| + |y|$$

da cui, per la (4.2), segue la (4.3).

La diseguaglianza triangolare è spesso usata nella forma seguente:

$$(4.4) \quad |a - b| \leq |a - c| + |b - c| \quad \forall a, b, c \in \mathbb{R}$$

Basta porre nella (4.3)  $x = a - c$  e  $y = c - b$ . Se invece, sempre nella (4.3), si pone  $x = a - b$  e  $y = b$  si ottiene

$$|a| \leq |a - b| + |b| \quad \text{cioè} \quad |a| - |b| \leq |a - b|$$

Analogamente, scambiando  $a$  con  $b$  otteniamo:  $|b| - |a| \leq |a - b|$  e dunque, ricordando la (4.2),

$$(4.5) \quad ||a| - |b|| \leq |a - b|$$

L'espressione  $|a - b|$  rappresenta geometricamente la distanza (nel senso della geometria elementare) dei due punti  $a$  e  $b$  sulla retta.

La (4.3) può facilmente estendersi al caso di  $n$  addendi:

$$(4.6) \quad \left| \sum_{k=1}^n x_k \right| \leq \sum_{k=1}^n |x_k|$$

Valgono anche le seguenti proprietà immediate:

$$(4.7) \quad |ab| = |a| |b| \quad \left| \frac{a}{b} \right| = \frac{|a|}{|b|} \quad |-a| = |a|$$

#### 4.4 Intervalli

Dati due numeri reali  $a, b$ , si chiama *intervallo di estremi  $a$  e  $b$*  uno dei seguenti insiemi:

- $[a, b]$ , insieme dei numeri reali  $x$ ,  $a \leq x \leq b$
- $[a, b)$ , " " " " "  $x$ ,  $a \leq x < b$
- $(a, b]$ , " " " " "  $x$ ,  $a < x \leq b$
- $(a, b)$ , " " " " "  $x$ ,  $a < x < b$

Come si può notare la parentesi quadra (tonda) in corrispondenza di uno dei due estremi indica che quest'ultimo è incluso (escluso) nell'intervallo.

Gli intervalli  $[a, b]$  si dicono *chiusi*; quelli  $(a, b)$  si dicono *aperti*. Tutti gli intervalli indicati sopra sono *limitati*; si chiamano intervalli (illimitati) anche le semirette, per esempio:

- $(-\infty, b)$ , insieme dei numeri reali  $x$ ,  $x < b$
- $[a, +\infty)$ , " " " " "  $x$ ,  $x \geq a$

o l'intera retta  $\mathbb{R} = (-\infty, +\infty)$ .



Figura 1.6.

Si può dimostrare che gli intervalli, limitati o illimitati, sono tutti e soli i sottoinsiemi  $I$  di  $\mathbb{R}$  che soddisfano la seguente proprietà caratteristica (detta *connessione*): presi comunque tre numeri reali

$$x_1 < x_2 < x_3, \text{ se } x_1, x_3 \in I, \text{ allora } x_2 \in I.$$

Nel seguito ci capiterà di considerare il prodotto cartesiano di due (o più) intervalli, cui si può dare il significato geometrico di rettangolo (in due dimensioni) o parallelepipedo (in tre dimensioni).



Sia  $A = [0, 1]$ ,  $B = [1, 2]$ ; la figura 1.2 di pagina 8 illustra gli insiemi  $A \times B$ ,  $B \times A$ ,  $A \times A$  indicato anche con  $A^2$ .

## 5 RADICALI, POTENZE, LOGARITMI

In conseguenza della proprietà  $R_4$  possiamo eseguire, nel campo reale, operazioni che sono solo occasionalmente possibili nel campo razionale, come l'estrazione di radice o l'elevamento a potenza.

### 5.1 Radici $n$ -esime aritmetiche

**TEOREMA 1.3** Sia  $y \in \mathbb{R}, y > 0$  e  $n$  un intero  $\geq 1$ . Esiste un unico numero reale positivo  $x$  tale che  $x^n = y$ .

Tale numero si chiama *radice  $n$ -esima aritmetica di  $y$*  e si indica con uno dei simboli  $\sqrt[n]{y}$  oppure  $y^{1/n}$ . Si noti che la radice  $n$ -esima aritmetica è non negativa, per esempio  $\sqrt[4]{4} = 2$ ,  $\sqrt[3]{9} = 3$  e più in generale  $\sqrt{x^2} = |x|$ .

Una semplice dimostrazione di questo teorema potrà essere data più avanti nel corso, utilizzando le proprietà delle *funzioni continue* (si veda il cap. 3, par. 4.1). Per ora ci limitiamo a mostrare, con un esempio, come si può costruire la rappresentazione decimale della radice  $n$ -esima. Cerchiamo l'allineamento decimale di  $\sqrt{2}$ ; questo numero, non essendo razionale, sarà rappresentato da un allineamento infinito (non periodico). Si procede così: si costruisce una classe di numeri razionali della forma

$$\begin{aligned} 0 &< a_0, \\ a_0, a_1 \\ a_0, a_1 a_2 \\ a_0, a_1 a_2 a_3 \\ \dots \\ a_0, a_1 a_2 a_3 \dots a_n \\ \dots \end{aligned}$$

con la regola che: ognuno di questi numeri è il più grande tra quelli con lo stesso numero di decimali dopo la virgola il cui quadrato è minore di 2:

1	$1^2 = 1$	$2^2 = 4$
1,4	$(1,4)^2 = 1,96$	$(1,5)^2 = 2,25$
1,41	$(1,41)^2 = 1,9881$	$(1,42)^2 = 2,0164$
	.....	.....

Questo insieme di numeri (chiamiamolo  $E_-$ ) è limitato superiormente (ognuno è  $< 2$ ); per la proprietà  $R_4$  esso possiede estremo superiore; il numero  $\sqrt{2}$  si definisce precisamente come  $\sup E_-$ .

Si sarebbe potuto anche costruire una classe di numeri  $E_+$  come la precedente con la regola che ognuno di essi sia il più piccolo tra quelli con lo stesso numero di decimali dopo la virgola il cui quadrato è  $> 2$ ; avremmo ottenuto

$$2 \quad 1,5 \quad 1,42 \quad \dots$$

Questo insieme  $E_+$  è limitato inferiormente (ogni elemento è  $> 1$ ), perciò possiede estremo inferiore; si dimostra che  $\inf E_+ = \sup E_-$ . I numeri della classe  $E_-$  approssimano  $\sqrt{2}$  per difetto, quelli della classe  $E_+$  per eccesso.

## 5.2 Potenze a esponente reale

L'estrazione di radice  $n$ -esima è l'operazione inversa dell'elevamento a potenza intera. Si può estendere l'operazione di elevamento a potenza per *ogni* esponente razionale se la base è *positiva* (utilizzando il teorema precedente) ponendo

$$\text{se } r = \frac{m}{n} \quad a > 0 \quad a^r := (a^m)^{\frac{1}{n}} = \sqrt[n]{a^m}$$

(si assume:  $m \in \mathbb{Z}$ ,  $n$  positivo).

Se infine l'esponente è reale  $b = b_0.b_1b_2\dots b_n\dots$  il numero  $a^b$  ( $a > 0$ ) sarà individuato dalla classe di numeri

$$a^{b_0} \quad a^{b_0, b_1} \quad a^{b_0, b_1 b_2} \quad \dots$$

in un modo simile a quello del caso della radice.

Se la base  $a$  è negativa l'operazione di elevamento a potenza  $a^b$  è definita solo in certi casi: se l'esponente  $b$  è intero, oppure razionale  $n/m$  purché *non sia n* dispari ed  $m$  pari. Infatti: si pone  $a^{n/m} = \sqrt[m]{a^n}$  e inoltre, se  $c < 0$  e  $m$  dispari, si definisce  $\sqrt[m]{c} = -\sqrt[m]{-c}$ . Per esempio:

$$(-2)^{3/5} = \sqrt[5]{-8} = -\sqrt[5]{8}$$

$$(-2)^{2/5} = \sqrt[5]{4}$$

$$(-2)^{3/4} = \sqrt[4]{-8} \quad \text{non esiste nel campo reale!}$$

Quando si dice "non esiste in  $\mathbb{R}$ " si intende che non è possibile definire tale operazione in modo da mantenere valide le usuali regole di calcolo. Qualcosa di più sull'argomento sarà detto nel paragrafo 8, parlando dei numeri complessi.

Le proprietà principali dell'elevamento a potenza sono:

siano  $a, b$  reali positivi,  $c, d$  reali qualsiasi

- $E_0 \quad a^0 = 1 \quad \forall a \neq 0; \quad 1^c = 1 \quad \forall c$
- $E_1 \quad a^c > 0 \quad \forall c; \quad a^c \leq 1 \text{ se } a \leq 1 \text{ e } c > 0$
- $E_2 \quad a^{c+d} = a^c \cdot a^d$
- $E_3 \quad (ab)^c = a^c \cdot b^c$
- $E_4 \quad (a^b)^c = a^{bc}$
- $E_5 \quad c < d \implies a^c \leq a^d \text{ se } a \geq 1$
- $E_6 \quad 0 < a \leq b \implies a^c \leq b^c \quad \forall c > 0$

## 5.3 Logaritmi

Consideriamo l'equazione

$$a^x = y \quad a > 0$$

con  $y$  assegnato e  $x$  incognito.

Anzitutto, se  $a = 1$ , essa è risolubile solo se  $y = 1$  (e in tal caso ogni numero reale  $x$  è soluzione). Sia dunque  $a \neq 1$ . Se  $y \leq 0$  essa non ha alcuna soluzione (cfr.  $E_1$ ). Il seguente teorema ci dice che essa ha una sola soluzione per ogni  $y > 0$ .

**TEOREMA 1.4** *Sia  $a > 0$ ,  $a \neq 1$ ,  $y > 0$ . Esiste un unico numero reale  $x$  tale che  $a^x = y$ .*

Tale numero prende il nome di *logaritmo in base a di y* e si indica col simbolo  $\log_a y$ . Quindi, per definizione:  $a^{\log_a y} = y$ . Le proprietà dei logaritmi, che si deducono da quelle degli esponenziali, sono:

siano  $x, y, a$  reali positivi,  $a \neq 1$

- $L_1 \quad \log_a xy = \log_a x + \log_a y$
- $L_2 \quad \log_a \frac{x}{y} = \log_a x - \log_a y$
- $L_3 \quad \log_a x^\alpha = \alpha \log_a x \quad \forall \alpha \in \mathbb{R}$
- $L_4 \quad \log_a x = \frac{1}{\log_x a} = -\log_{\frac{1}{a}} x \quad (x \neq 1)$
- $L_5 \quad \log_b x = \frac{\log_a x}{\log_a b} \quad \forall b > 0, b \neq 1$

Nel capitolo 2 introdurremo il concetto di *funzione* e vedremo che le operazioni sopra definite consentono di introdurre tre importanti classi di funzioni: le *potenze*, gli *esponenziali* e i *logaritmi*.

#### 5.4 Approssimazioni

Concludiamo questo paragrafo con qualche osservazione sulle approssimazioni che si fanno lavorando coi numeri reali. Cominciamo col ricordare che un numero *razionale* può essere sempre espresso con precisione assoluta, sia ricorrendo alla scrittura frazionaria che a quella decimale (eventualmente con cifre periodiche). Invece non è possibile, naturalmente, scrivere *tutte* le cifre decimali di un numero irrazionale, dato che queste sono infinite e si susseguono senza periodicità. Cosa significa allora “conoscere” o “specificare” un numero irrazionale? Significa conoscere qualche algoritmo che ci consenta (almeno teoricamente) di scrivere tante cifre decimali esatte quante ne desideriamo. Nell’esempio del numero irrazionale

$$0,1010010001\dots$$

(formato in base alla regola: scrivere una cifra 1, una cifra 0, una cifra 1, due cifre 0, una cifra 1, tre cifre 0, e così via) è chiaro che potremmo scrivere tante cifre quante ne desideriamo. In altri casi, come  $\sqrt{2}$  o  $\log_2 3$ , le cose sono più laboriose, richiedendo l’esecuzione di calcoli iterativi per determinare ogni successiva cifra decimale; tuttavia, questi calcoli sono effettivamente eseguibili. Altre volte, in Analisi matematica, un numero viene specificato in modo non costruttivo, denotandolo come l’unico numero che risolve un determinato problema (una volta che si sia dimostrato appunto che tale soluzione esiste ed è unica). Si tratta di un modo operativamente meno soddisfacente, ma teoricamente ineccepibile.

Veniamo ora agli aspetti pratici. Ogni volta che eseguiamo calcoli con una calcolatrice tascabile, questa scriverà solo numeri con un numero fissato di cifre decimali dopo la virgola (tipicamente 9). Questo significa che stiamo lavorando solo con numeri razionali, anzi con un sottoinsieme *finito* dell’insieme  $\mathbb{Q}$ ! Lavorando con un computer le cose migliorano un po’, ma rimaniamo comunque nell’ambito di sottoinsiemi finiti di  $\mathbb{Q}$ . Occorre naturalmente esserne consapevoli.

Oltre all’approssimazione generata dagli strumenti di calcolo, a volte siamo noi stessi a non essere interessati a troppe cifre decimali; introduciamo allora volutamente delle approssimazioni, scrivendo ad esempio

$$\sqrt{2} \simeq 1,414.$$

Si ricordi, a tale proposito, la regola di *arrotondamento* espressa dai seguenti esempi:

$$2,4138 \simeq 2,41$$

$$2,4152 \simeq 2,42.$$

Esplicitamente: se la prima cifra che si trascura è da 0 a 4, l’ultima cifra che si scrive rimane inalterata; se la prima cifra che si trascura è da 5 a 9, l’ultima cifra che si scrive viene aumentata di un’unità.

## 6 INSIEMI INFINITI

Accenniamo ora ad una questione fondamentale che ha a che fare con gli insiemi, e precisamente il problema di come si possa confrontare la “numerosità” degli insiemi infiniti. Avvertiamo che ciò che diremo in questo paragrafo non sarà direttamente uti-

lizzato nel seguito del corso; tuttavia, ci sembra giusto dare spazio a questo problema, al tempo stesso così naturale e difficile<sup>4</sup>.

Affrontiamo il discorso a partire dagli insiemi numerici notevoli che abbiamo introdotto:  $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ . Chiediamoci: quanti sono gli elementi di ciascuno di questi insiemi?

Intuitivamente, la risposta sembra ovvia: ciascuno di questi insiemi ha *infiniti* elementi; tuttavia i numeri razionali sono più numerosi dei numeri interi, essendo  $\mathbb{Z} \subset \mathbb{Q}$ , e per lo stesso motivo i numeri reali sono di più dei numeri razionali. Queste risposte pongono però un problema: come possiamo affermare, al tempo stesso, che due insiemi sono *entrambi* infiniti, ma uno è *più numeroso* dell’altro? Com’è possibile, in generale, confrontare la numerosità degli insiemi infiniti? Ad esempio, ci sono più punti in una sfera o in un piano? (In questo caso, nessuno dei due insiemi è contenuto nell’altro).

Per dar senso a queste domande, prima ancora che per sapervi rispondere, occorre definire cosa si intenda per uguale numerosità di due insiemi. Astraendo dall’esperienza del *contare* gli elementi di un insieme finito, si è giunti a identificare l’idea di uguale numerosità con quella di *corrispondenza biunivoca*.

**DEFINIZIONE 1.5** Due insiemi  $A, B$  si dicono di uguale *cardinalità* o *potenza* (termini che traducono l’idea intuitiva di *numerosità*) se possono essere messi in corrispondenza biunivoca tra loro, cioè se esiste una legge che associa ad ogni elemento di  $A$  uno e un solo elemento di  $B$ , e viceversa.

Si osservi, ad esempio, la seguente corrispondenza tra gli insiemi  $\mathbb{Z}$  e  $\mathbb{N}$ :

$\mathbb{Z}$	0	1	-1	2	-2	...	$n$	- $n$	...
	↑	↓	↑	↓	↑	...	↓	↑	...
$\mathbb{N}$	0	1	2	3	4	...	$2n - 1$	$2n$	...

Come si vede, la legge ora definita realizza una corrispondenza biunivoca tra  $\mathbb{Z}$  e  $\mathbb{N}$ . Quindi, anche se dal punto di vista dell’inclusione  $\mathbb{Z}$  ha “più elementi” di  $\mathbb{N}$  (nel senso che ha tutti gli elementi di  $\mathbb{N}$  più altri), gli insiemi hanno la stessa cardinalità (si dice anche che sono *equipotenti*). In questo senso, quindi, vanno pensati come *ugualmente numerosi*.

**DEFINIZIONE 1.6** Si dice *numerabile* un insieme che ha la stessa cardinalità di  $\mathbb{N}$ .

Ad esempio,  $\mathbb{Z}$  è numerabile. Il termine numerabile indica che gli elementi dell’insieme si possono enumerare, ossia disporre in un elenco numerato (posto 1, posto 2, posto 3, ...).

Di più, possiamo dimostrare che:

**TEOREMA 1.5**  $\mathbb{Q}$  è *numerabile*.

<sup>4</sup>Le ricerche sulla numerosità degli insiemi infiniti iniziarono nel 19° secolo con Bolzano, Dedekind, e soprattutto Cantor, il cui nome è legato in maniera speciale a questo tema.

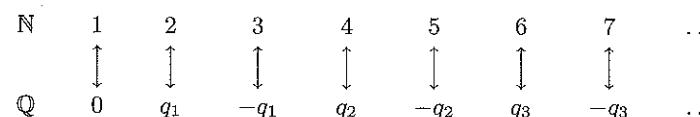
**DIMOSTRAZIONE.** Cominciamo a dimostrare che l'insieme dei numeri razionali *positivi* è numerabile. Per far questo, rappresentiamo i numeri razionali positivi come frazioni  $n/m$ , con  $n, m$  interi positivi, e disponiamo queste frazioni in una tabella triangolare infinita, al modo seguente:

$\frac{n}{m}$ , con $n + m = \dots$				
2	$\frac{1}{1}$			
3	$\frac{1}{2}$	$\frac{2}{1}$		
4	$\frac{1}{3}$	$\frac{2}{2}$	$\frac{3}{1}$	
5	$\frac{1}{4}$	$\frac{2}{3}$	$\frac{3}{2}$	$\frac{4}{1}$
...	...			

Si noti che: tutte le frazioni che rappresentano numeri razionali positivi compaiono in questa tabella almeno una volta; alcuni sono ripetuti più volte (ad es.  $1/1 = 2/2$ ); ogni riga ha lunghezza finita. Quindi possiamo mettere in corrispondenza biunivoca  $\mathbb{N}$  con l'insieme dei razionali positivi, percorrendo la tabella, riga dopo riga (e saltando un elemento quando è uguale ad uno già incontrato). Ad esempio:

1	2	3	4	5	6	7	8	9	...
$\frac{1}{1}$	$\frac{1}{2}$	$\frac{2}{1}$	$\frac{1}{3}$	$\frac{3}{1}$	$\frac{1}{4}$	$\frac{2}{3}$	$\frac{3}{2}$	$\frac{4}{1}$	...

(si noti che abbiamo saltato  $2/2$  perché uguale a  $1/1$ , già incontrato). Questo dimostra che l'insieme dei razionali positivi è numerabile; allora anche  $\mathbb{Q}$  è numerabile, e questo si può provare con una dimostrazione analoga a quella con cui abbiamo provato che  $\mathbb{Z}$  è numerabile: detti  $q_1, q_2, q_3, \dots$  i razionali positivi, si pone:



che realizza una corrispondenza biunivoca tra  $\mathbb{N}$  e  $\mathbb{Q}$ .  $\diamond$

Aver scoperto che diversi insiemi infiniti, uno propriamente contenuto nell'altro ( $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ ), hanno la stessa cardinalità, potrebbe far pensare che questo sia vero per tutti gli insiemi infiniti. Ciò non è vero. Infatti, si dimostra che:

#### TEOREMA 1.6 $\mathbb{R}$ non è numerabile.

**DIMOSTRAZIONE.** La tecnica di questa dimostrazione prende il nome di *procedimento diagonale di Cantor* ed è un'idea semplice ma profonda che si utilizza in varie parti della matematica.

Proveremo, per l'esattezza, che l'intervallo  $[0, 1]$  ha una cardinalità non numerabile, da cui seguirà, a maggior ragione, la non numerabilità di  $\mathbb{R}$ . Supponiamo dunque per assurdo che  $[0, 1]$  sia numerabile e disponiamo tutti i numeri reali dell'intervallo  $[0, 1]$  in un elenco  $r_1, r_2, r_3, \dots$ . Scriviamo ora ogni numero  $r_i$  in forma decimale (cioè nella forma  $0.a_1a_2a_3\dots$  dove gli  $a_i$  sono cifre da 0 a 9; in particolare, se le cifre sono tutte zero si ha  $r_i = 0$  e se sono tutte nove si ha  $r_i = 0.\overline{9} = 1$ ):

$$r_1 = 0.a_{11}a_{12}a_{13}\dots$$

$$r_2 = 0.a_{21}a_{22}a_{23}\dots$$

$$r_3 = 0.a_{31}a_{32}a_{33}\dots$$

.....

Definiamo ora il seguente numero decimale:

$$r = 0.b_1b_2b_3\dots$$

dove le cifre  $b_i$  sono definite con la seguente regola:

- se  $a_{ii}$  è una cifra 0, 1, 2, 3 o 4, poniamo  $b_i = 5$ ;
- se  $a_{ii}$  è una cifra 5, 6, 7, 8 o 9, poniamo  $b_i = 4$ .

Con questa definizione risulta  $b_i \neq a_{ii}$  per ogni  $i$ . Si noti che il numero  $r$  è stato costruito ragionando sulla *diagonale* della tabella infinita che ha per righe i numeri  $r_i$ , da cui il nome del procedimento.

Osserviamo ora che  $r$  è un numero reale appartenente all'intervallo  $[0, 1]$  (perché è del tipo  $0.b_1b_2b_3\dots$  con tutte le cifre uguali a 4 o 5) e d'altro canto non è uguale a nessuno dei numeri  $r_i$  dell'elenco, in quanto:

- $r \neq r_1$  perché la prima cifra di  $r$  è diversa dalla prima cifra di  $r_1$  (perché  $b_1 \neq a_{11}$ );
- $r \neq r_2$  perché la seconda cifra di  $r$  è diversa dalla seconda cifra di  $r_2$  (perché  $b_2 \neq a_{22}$ );
- $r \neq r_3$  perché...
- ...

Ma questo porta a una contraddizione, avevamo supposto che gli  $r_i$  esaurissero completamente l'insieme dei numeri reali dell'intervallo  $[0, 1]$ . Dunque  $[0, 1]$  non è numerabile.  $\diamond$

Poiché  $\mathbb{N}$  non può essere messo in corrispondenza biunivoca con  $\mathbb{R}$ , ma può esser messo in corrispondenza biunivoca con un sottoinsieme proprio di  $\mathbb{R}$  ( $\mathbb{N}$  stesso!), diciamo che  $\mathbb{N}$  ha cardinalità *minore* di  $\mathbb{R}$ . Ecco quindi com'è possibile dar senso al confronto tra la "numerosità" di due insiemi infiniti.

Notiamo anche che l'intervallo  $[0, 1]$ , anzi qualsiasi intervallo di  $\mathbb{R}$  (aperto o chiuso, limitato o illimitato) ha la stessa cardinalità di  $\mathbb{R}$ . In altre parole, la retta *non ha* una cardinalità maggiore del segmento, ma la stessa. Questo fatto si può dimostrare ad esempio con una costruzione geometrica elementare, che realizza una corrispondenza biunivoca tra punti della retta e di un segmento come quella in figura 1.7.

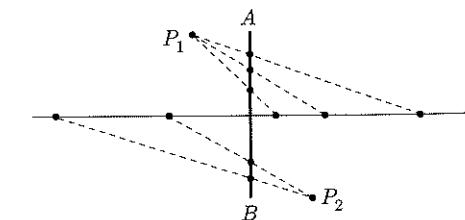


Figura 1.7. I punti del segmento (aperto)  $AB$  sono messi in corrispondenza biunivoca con i punti della retta, tracciando opportuni segmenti dai due punti fissati  $P_1, P_2$  alla retta stessa.

La cardinalità di  $\mathbb{R}$  prende il nome di *potenza del continuo*. Non solo ogni intervallo di  $\mathbb{R}$  ha questa stessa cardinalità. Si può dimostrare che lo stesso vale per il piano, per lo spazio tridimensionale e per i loro sottoinsiemi "continui": per esempio, un piano e una sfera hanno entrambi la potenza del continuo.

Abbiamo incontrato fin qui solo *due* livelli gerarchici di infinito: la potenza del numerabile e quella del continuo. Non si deve credere che esistano solo queste due: ad esempio, l'insieme di tutti i sottoinsiemi di  $\mathbb{R}$  è ancora più numeroso di  $\mathbb{R}$ ; anzi con questo procedimento si può sempre costruire un insieme più numeroso di un insieme dato:

**TEOREMA 1.7** *Sia  $A$  un insieme qualsiasi (finito o infinito!) e  $\mathcal{P}(A)$  il suo insieme delle parti. Allora la cardinalità di  $\mathcal{P}(A)$  è sempre maggiore della cardinalità di  $A$ .*

Una dimostrazione di questo teorema è suggerita nell'Esercizio 7 dei Complementi in fondo a questo capitolo.

Perciò i livelli gerarchici delle cardinalità infinite sono anch'essi infiniti.

## 7 IL PRINCIPIO DI INDUZIONE

Presentiamo ora un potente metodo dimostrativo, detto *dimostrazione per induzione*. Questo procedimento si può applicare a teoremi che abbiano la struttura seguente:

“Per ogni  $n \in \mathbb{N}$ ,  $n \geq n_0$ , vale la proprietà  $p(n)$ ”.

Il numero  $n_0$  è il più piccolo intero per cui si vuole che la proprietà sia vera; se  $n_0 = 0$  il teorema afferma semplicemente che la proprietà è vera per ogni  $n \in \mathbb{N}$ .

La dimostrazione per induzione consiste nei due passi seguenti:

1. Si dimostra che  $p(n)$  è vera per  $n = n_0$  (primo passo dell'induzione).
2. Si dimostra che, se  $n$  è un generico numero naturale  $\geq n_0$ , dal fatto che  $p(n)$  sia vera segue che  $p(n+1)$  è vera.

Si può allora concludere che per ogni  $n \geq n_0$ ,  $p(n)$  è vera. La validità di questo metodo dimostrativo, intuitivamente, si basa su questo fatto:

1. Per il punto 1, sappiamo che  $p(n_0)$  è vera. Supponiamo ad esempio  $n_0 = 1$ : sappiamo quindi che  $p(1)$  è vera (questo va dimostrato esplicitamente).
2. Per il punto 2, poiché è vera  $p(1)$ , sarà vera  $p(2)$ : infatti abbiamo dimostrato che qualunque sia  $n$ , se è vera  $p(n)$  è vera anche  $p(n+1)$ . Ma allora, poiché è vera  $p(2)$ , sarà vera  $p(3)$ ; ma allora è vera  $p(4)$ , ... e così via, dunque è vera  $p(n)$  per ogni  $n \geq 1$ .

Concretamente, la dimostrazione si articola in due fasi.

1. Dimostrare direttamente  $p(n_0)$ ;
2. Assumere come ipotesi  $p(n)$  (ipotesi induttiva) e provare  $p(n+1)$ .

È questo il punto delicato, spesso soggetto a equivoci. “Assumere per ipotesi”  $p(n)$  non significa assumere come ipotesi esattamente la nostra tesi? Non proprio. Quello che si deve provare è che:

per ogni  $n \geq n_0$ , se è vera  $p(n)$  allora è vera anche  $p(n+1)$

e non

se per ogni  $n$  è vera  $p(n)$ , allora è vera anche  $p(n+1)$ .

Lo studente riflette sulla differenza tra le due frasi appena scritte. La sottigliezza della dimostrazione per induzione è tutta lì. Diamo un paio di esempi significativi di applicazione di questo metodo:

**TEOREMA 1.8 (DISUGUAGLIANZA DI BERNOULLI)** *Per ogni intero  $n \geq 0$ ,  $x \in \mathbb{R}$ ,  $x > -1$*

$$(1+x)^n \geq 1 + nx.$$

DIMOSTRAZIONE. Per induzione su  $n$ . Sia  $n = 0$ . Allora l'asserto diventa

$$(1+x)^0 \geq 1 + 0 \cdot x,$$

cioè  $1 \geq 1$ , evidentemente vero. Supponiamo che sia vero per  $n$ , e proviamolo per  $(n+1)$ .

$$(1+x)^{n+1} = (1+x) \cdot (1+x)^n \geq$$

(per ipotesi induttiva  $(1+x)^n \geq 1 + nx$ ; inoltre per ipotesi  $(1+x) \geq 0$  perché  $x \geq -1$ )

$$\geq (1+x)(1+nx) = 1 + (n+1)x + nx^2 \geq 1 + (n+1)x$$

dove nell'ultima diseguaglianza si è sfruttato il fatto che  $nx^2 \geq 0$ . La catena di diseguaglianze mostra che  $(1+x)^{n+1} \geq 1 + (n+1)x$ , pertanto la tesi è dimostrata. 

**TEOREMA 1.9 (FORMULA DEL BINOMIO DI NEWTON)** *Per ogni intero  $n \geq 0$ , con  $a, b \in \mathbb{R}$ ,*

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

DIMOSTRAZIONE. (Questo risultato è stato già enunciato nel paragrafo 2.3; ora ne forniamo una dimostrazione). Per induzione su  $n$ . Per  $n = 0$  l'asserto è

$$(a+b)^0 = \binom{0}{0} a^0 b^0, \text{ cioè } 1 = 1,$$

vero. Sia vero per  $n$ , e dimostriamolo per  $(n+1)$ .

$$(a+b)^{n+1} = (a+b)(a+b)^n =$$

per l'ipotesi induttiva

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

eseguendo nella prima sommatoria una traslazione di indici

$$= \sum_{k=1}^{n+1} \binom{n}{k-1} a^k b^{n-k+1} + \sum_{k=0}^n \binom{n}{k} a^k b^{n+1-k} =$$

scorporando l'addendo per  $k = 0$  dalla seconda sommatoria e quello per  $k = n+1$  dalla prima

$$= a^{n+1} + \sum_{k=1}^n \binom{n}{k-1} a^k b^{n-k+1} + b^{n+1} + \sum_{k=1}^n \binom{n}{k} a^k b^{n+1-k} =$$

sommmando le due sommatorie

$$= a^{n+1} + b^{n+1} + \sum_{k=1}^n \left[ \binom{n}{k-1} + \binom{n}{k} \right] a^k b^{n-k+1} =$$

applicando la formula di ricorrenza  $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$

$$\begin{aligned} &= a^{n+1} + b^{n+1} + \sum_{k=1}^n \binom{n+1}{k} a^k b^{n-k+1} = \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n-k+1}, \end{aligned}$$

che è esattamente l'asserto voluto, per  $n+1$ . 

Chiediamoci ora che cosa giustifichi la validità del procedimento di dimostrazione per induzione dal punto di vista formale, e non puramente intuitivo. Per rispondere, si consideri anzitutto la seguente riformulazione del principio di induzione nel linguaggio degli insiemi:

“Sia  $S$  un sottoinsieme di  $\mathbb{N}$ . Se  $0 \in S$  e, per ogni  $n \in \mathbb{N}$ , se  $n \in S$  anche  $n+1 \in S$ , allora  $S = \mathbb{N}$ ”.

Formulato in questo modo, il principio di induzione viene ad essere uno degli assiomi del sistema dei numeri naturali; in altre parole, questo significa che la validità di questo principio è parte della definizione dell'insieme  $\mathbb{N}$ . La situazione è analoga a quella che abbiamo incontrato parlando della proprietà dell'estremo superiore: la validità della proprietà dell'estremo superiore (assioma di continuità) è parte della definizione dell'insieme  $\mathbb{R}$  dei numeri reali.



Lo studente dimostri per induzione le seguenti identità, alcune delle quali si sono già incontrate nel paragrafo 1.2:

La formula per la somma dei primi  $n$  interi:

$$\sum_{k=1}^n k = \frac{n(n+1)}{2} \text{ per ogni intero } n \geq 1.$$

La formula per la somma geometrica:

$$\sum_{k=0}^n q^k = \frac{1-q^{n+1}}{1-q} \text{ per ogni intero } n \geq 0 \text{ e numero reale } q \neq 1.$$

Ricavare la formula per la somma dei primi  $n$  numeri pari:

$$\sum_{k=1}^n (2k) \text{ per ogni intero } n \geq 1.$$

Ricavare la formula per la somma dei primi  $n$  numeri dispari:

$$\sum_{k=0}^{n-1} (2k+1) \text{ per ogni intero } n \geq 1.$$

Dimostrare che se  $A$  è un insieme costituito da  $n$  elementi, l'insieme delle parti di  $A$  ha  $2^n$  elementi.

Dimostrare l'identità dei coefficienti binomiali (utilizzata nella dimostrazione della formula del binomio di Newton):

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k} \text{ per } n, k \text{ interi, } 1 \leq k \leq n.$$

## 8 NUMERI COMPLESSI

Ci sono ragioni, principalmente di natura algebrica, che ci spingono ad ampliare ulteriormente il campo numerico; si tratta, in sostanza, di definire in tutta la sua generalità l'operazione di elevamento a potenza  $a^b$  (e la sua “inversa”), operazione che, nel campo reale, è definita per ogni esponente  $b$  solo se  $a$  è positivo, e, se  $a$  è negativo, soltanto per alcuni esponenti (interi o razionali  $\frac{m}{n}$  con  $n$  dispari).

In questo paragrafo introduciamo quindi il campo dei numeri complessi e definiamo su di essi le operazioni algebriche elementari e la radice  $n$ -esima. Definiremo in seguito (cap. 5, par. 2.2) l'operazione di elevamento a potenza qualunque, così come i logaritmi e gli esponenziali nel campo complesso. Molte applicazioni che lo studente potrà incontrare nel corso dei suoi studi richiedono l'uso di questi strumenti.

### 8.1 Definizione di $\mathbb{C}$ e struttura di campo

Abbiamo indicato con  $\mathbb{R}^2$  (abbreviazione di  $\mathbb{R} \times \mathbb{R}$ ) l'insieme delle coppie ordinate  $(a, b)$  di numeri reali. Su queste definiamo direttamente le operazioni di somma e prodotto con le seguenti regole:

$$(8.1) \quad (a, b) + (c, d) := (a+c, b+d)$$

$$(8.2) \quad (a, b) \cdot (c, d) := (ac - bd, ad + bc)$$

Questa “somma” e questo “prodotto” verificano le proprietà commutativa, associativa e distributiva enunciate nel paragrafo 3. Osserviamo anche che,  $\forall (a, b)$ :

$$(a, b) + (0, 0) = (0, 0) + (a, b) = (a, b)$$

dunque la coppia  $(0, 0)$  è *elemento neutro* per la somma.

$$(a, b) \cdot (1, 0) = (1, 0) \cdot (a, b) = (a, b)$$

dunque la coppia  $(1, 0)$  è *elemento neutro* per il prodotto.

$$(a, b) + (-a, -b) = (0, 0)$$

dunque  $(-a, -b)$  è l'*opposto* di  $(a, b)$ .

Se

$$(a, b) \neq (0, 0) \text{ allora } (a, b) \cdot \left( \frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right) = (1, 0)$$

dunque la coppia  $(a/(a^2 + b^2), -b/(a^2 + b^2))$  è il *reciproco* di  $(a, b)$ .

Dunque le proprietà  $R_1, R_2$  (vedi par. 3) sono verificate per la somma e il prodotto così definiti e perciò l'insieme  $\mathbb{R}^2$  così strutturato è un campo, che chiameremo *campo dei numeri complessi* e indicheremo con  $\mathbb{C}$ .

Osserviamo ora che  $\mathbb{C}$  contiene il sottoinsieme  $\mathbb{C}_0$  delle coppie del tipo  $(a, 0)$ ; esso è un *sottocampo* di  $\mathbb{C}$ , poiché somma e prodotto di coppie di questo tipo sono ancora coppie dello stesso tipo; infatti si ha:

$$(a, 0) + (b, 0) = (a + b, 0)$$

$$(a, 0) \cdot (b, 0) = (ab, 0)$$

Inoltre  $\mathbb{C}_0$  può essere ordinato ponendo  $(a, 0) < (b, 0)$  se  $a < b$ . Se allora mettiamo in corrispondenza biunivoca l'insieme dei numeri reali  $\mathbb{R}$  con  $\mathbb{C}_0$ , ponendo

$$(a, 0) \longleftrightarrow a$$

possiamo *identificare* i numeri reali  $a$  con i numeri complessi del tipo  $(a, 0)$ .

In questo senso il campo dei numeri complessi  $\mathbb{C}$  è un ampliamento di quello dei numeri reali  $\mathbb{R}$ .

Consideriamo ora il numero  $(0, 1)$ . Esso ha la singolare proprietà che:

$$(0, 1) \cdot (0, 1) = (-1, 0)$$

cioè *il suo quadrato coincide col numero reale  $-1$* ! Per questa ragione la coppia  $(0, 1)$  merita di essere indicata con un simbolo speciale: la indicheremo con " $i$ " e la chiameremo *unità immaginaria*.

### Forma algebrica dei numeri complessi

A questo punto conviene semplificare la notazione. Osserviamo che, se scriviamo semplicemente  $a$  invece di  $(a, 0)$  ecc. abbiamo

$$(a, b) = (a, 0) + (0, 1) \cdot (b, 0) = a + ib$$

Con questa notazione le regole (8.1) e (8.2) sono le ordinarie regole del calcolo letterale, ove si tenga conto che  $i^2 = -1$ :

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

$$(a + ib) \cdot (c + id) = (ac - bd) + i(ad + bc)$$

La scrittura

$$(8.3) \quad z = a + ib$$

è detta *forma algebrica* dei numeri complessi;  $a$  si chiama *parte reale* di  $z$  e si indica con  $\operatorname{Re}(z)$ ,  $b$  si chiama *parte immaginaria* e si indica con  $\operatorname{Im}(z)$ .

### Piano complesso

In un piano cartesiano, rappresentiamo i numeri complessi  $a + ib$  come punti di coordinate  $(a, b)$ : ecco una semplice e comoda immagine geometrica del campo complesso. In questo contesto, il piano viene detto *piano complesso* o *piano di Gauss*; gli assi  $x, y$  si dicono *asse reale* e *asse immaginario*: i punti sull'asse reale sono i numeri reali, i punti sull'asse immaginario sono i numeri immaginari puri (cioè del tipo  $ib$ ).

La somma di due numeri complessi è il numero complesso che ha per coordinate la somma delle coordinate: il significato geometrico di questo fatto è che il punto  $z + t$  si costruisce a partire dai punti  $z, t$  in base alla "regola del parallelogramma" (vedi fig. 1.8).

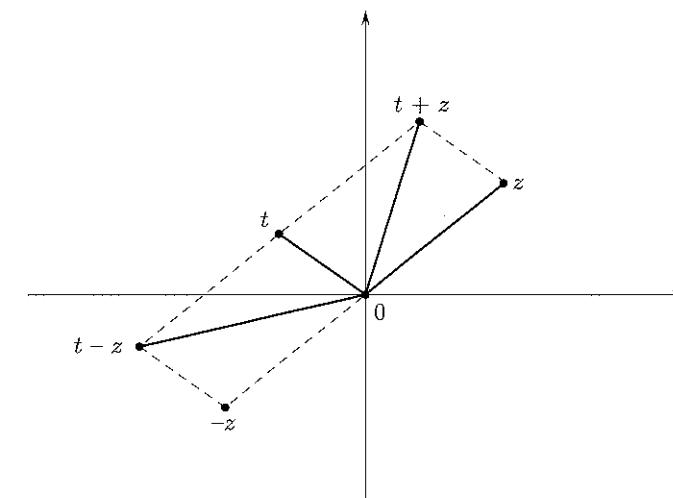


Figura 1.8.

Abbiamo visto che  $\mathbb{C}$  soddisfa gli assiomi di campo; *non* soddisfa però quelli di *campo ordinato*, ovvero non è possibile definire una relazione  $\leq$  tra i numeri complessi, in modo che valgano le proprietà elencate nel paragrafo 3 (in particolare, la  $R_3$ ). Infatti si può dimostrare che da quelle proprietà segue che *il quadrato di un numero qualsiasi non è mai negativo*, e d'altra parte, *se un numero è positivo il suo opposto è negativo*. Ora, in  $\mathbb{C}$  si ha:

$$1^2 = 1 \quad \text{e} \quad i^2 = -1$$

Abbiamo quindi due quadrati che sono l'uno l'opposto dell'altro. Nessuno dei due però può essere negativo (perché sono quadrati), e questo è assurdo (perché tra  $a$  e  $-a$  uno dev'essere negativo, se  $a \neq 0$ ). Concludiamo che  $\mathbb{C}$  *non è un campo ordinato*.

### 8.2 Coniugato e modulo

Il numero complesso  $a - ib$  si dice il *compleSSO coniugato* di  $z = a + ib$  e si indica con  $\bar{z}$ . Evidentemente si ha:

$$z + \bar{z} = 2a = 2\operatorname{Re}(z)$$

$$z - \bar{z} = 2bi = 2i\operatorname{Im}(z)$$

L'operazione di *coniugio* ha le seguenti elementari proprietà rispetto alla somma e al prodotto:

$$\overline{(z_1 + z_2)} = \bar{z}_1 + \bar{z}_2$$

$$\overline{(z_1 z_2)} = \bar{z}_1 \cdot \bar{z}_2$$

$$\overline{\left(\frac{1}{z}\right)} = \frac{1}{\bar{z}}$$

Osserviamo ora che

$$z\bar{z} = (a+ib)(a-ib) = a^2 + b^2 \geq 0$$

Si chiama *modulo* di  $z = a+ib$  il numero reale non negativo  $\sqrt{a^2 + b^2}$ , che si indica con  $|z|$ . Se  $z = a$  è reale, il suo modulo si chiama *valore assoluto* e si indica sempre con  $|a|$ , come detto nel paragrafo 4. Valgono le seguenti proprietà:

- a)  $|z| \geq 0$  e  $|z| = 0 \iff z = 0$
- b)  $|z| = |\bar{z}|$
- c)  $|\operatorname{Re}(z)| \leq |z| \quad |\operatorname{Im}(z)| \leq |z| \quad |z| \leq |\operatorname{Re}(z)| + |\operatorname{Im}(z)|$
- (8.4) d)  $|z_1 + z_2| \leq |z_1| + |z_2| \quad (\text{diseguaglianza triangolare})$
- (8.5) e)  $|z_1 + z_2| \geq ||z_1| - |z_2||$

Le proprietà a), b), c) si verificano immediatamente. Proviamo d), e).

Esse sono equivalenti alla seguente:

$$(|z_1| - |z_2|)^2 \leq |z_1 + z_2|^2 \leq (|z_1| + |z_2|)^2$$

Ponendo  $z_1 = a+ib$ ,  $z_2 = c+id$  otteniamo:

$$(\sqrt{a^2 + b^2} - \sqrt{c^2 + d^2})^2 \leq (a+c)^2 + (b+d)^2 \leq (\sqrt{a^2 + b^2} + \sqrt{c^2 + d^2})^2$$

Con calcoli elementari questa doppia diseguaglianza si riduce a

$$-\sqrt{a^2 + b^2} \cdot \sqrt{c^2 + d^2} \leq ac + bd \leq \sqrt{a^2 + b^2} \cdot \sqrt{c^2 + d^2}$$

che è equivalente alla seguente:

$$|ac + bd| \leq \sqrt{a^2 + b^2} \cdot \sqrt{c^2 + d^2}$$

Elevando al quadrato entrambi i membri si arriva a:

$$(ac + bd)^2 \leq (a^2 + b^2)(c^2 + d^2)$$

ovvero a

$$0 \leq -2acbd + a^2d^2 + b^2c^2 = (ad - bc)^2$$

che è vera per ogni  $a, b, c, d \in \mathbb{R}$ .

Geometricamente,  $|z|$  rappresenta la *distanza* del punto (o numero complesso)  $z$  dall'origine;  $|z_1 - z_2|$  rappresenta la distanza dei due punti  $z_1$  e  $z_2$ ; le diseguaglianze (8.4) e (8.5) traducono il noto teorema sulle lunghezze dei lati di un triangolo (vedi fig. 1.9).

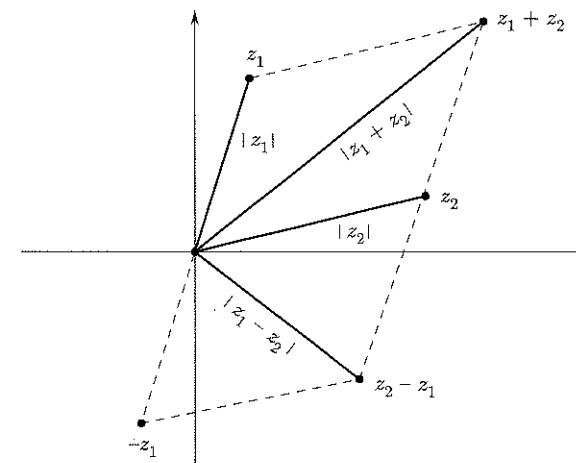


Figura 1.9.

Utilizzando i concetti ora introdotti, possiamo rappresentare in forma algebrica il rapporto di due numeri complessi  $\frac{a+ib}{c+id}$ ; basta moltiplicare numeratore e denominatore per  $c-id$ ; abbiamo:

$$\frac{a+ib}{c+id} = \frac{(a+ib)(c-id)}{|c+id|^2} = \frac{ac+bd}{c^2+d^2} + i \frac{bc-ad}{c^2+d^2}$$

Vediamo come si può risolvere un'equazione nel campo complesso, quando questa coinvolge l'incognita  $z = x+iy$  anche attraverso  $\operatorname{Re} z$ ,  $\operatorname{Im} z$ ,  $\bar{z}$ ,  $|z|$ .



$$z^2 + i\operatorname{Im} z + 2\bar{z} = 0$$

Poniamo  $z = x+iy$ , con  $x, y$  incognite reali, e trascriviamo l'equazione a questo modo

$$z^2 = (x+iy)^2 = x^2 - y^2 + 2ixy$$

$$i\operatorname{Im} z = iy$$

$$2\bar{z} = 2(x-iy) = 2x - 2iy$$

$$(x^2 - y^2 + 2ixy) + (iy) + (2x - 2iy) = 0$$

Ora, un numero complesso è zero se e solo se la sua parte reale e parte immaginaria sono zero. Perciò mettiamo in evidenza la parte reale e la parte immaginaria del primo membro

ed uguagliamo entrambe a zero:

$$(x^2 - y^2 + 2x) + i(2xy + y - 2y) = 0$$

$$\begin{cases} x^2 - y^2 + 2x = 0 \\ 2xy - y = 0 \end{cases}$$

Si è così trasformata l'equazione in una incognita complessa in un sistema di due equazioni in due incognite reali. Risolviamo il sistema. La seconda equazione dà:

$$y = 0 \quad \text{o} \quad x = \frac{1}{2}$$

Per  $y = 0$  la prima equazione diventa

$$x^2 + 2x = 0$$

che dà

$$x = 0 \quad \text{o} \quad x = -2$$

Per  $x = \frac{1}{2}$  la prima equazione diventa

$$-y^2 + \frac{5}{4} = 0$$

che ha soluzioni

$$y = \pm \frac{\sqrt{5}}{2}$$

Quindi le soluzioni sono:

$$z = 0 \quad z = -2 \quad z = \frac{1}{2} + i\frac{\sqrt{5}}{2} \quad z = \frac{1}{2} - i\frac{\sqrt{5}}{2}$$

L'equazione ha 4 soluzioni.

*Il metodo visto in quest'esempio* (passare alla parte reale e immaginaria dell'equazione) è applicabile in linea di principio ad ogni equazione in  $\mathbb{C}$ . In pratica, un generico sistema di due equazioni in due incognite è quasi sempre insolubile per via algebrica. Perciò prima di mettersi su questa strada è bene osservare se non ce ne sia una più semplice.

### 8.3 Forma trigonometrica

Come è noto dalla Geometria, i punti del piano possono essere individuati, oltre che dalle loro coordinate cartesiane, anche dalle coordinate polari:  $\rho$  (*raggio polare*, cioè distanza del punto dall'origine) e  $\theta$  (*angolo polare*, cioè l'angolo che la retta congiungente il punto con l'origine forma con l'asse delle ascisse positive, misurato in senso antiorario). È chiaro che una coppia  $\rho, \theta$ , con  $\rho > 0$ , individua un ben determinato punto del piano; invece un punto del piano individua univocamente la coordinata  $\rho$ , ma l'angolo  $\theta$ , misurato in radianti, è determinato solo a meno di multipli di  $2\pi$  (fig. 1.10).

Dato un numero complesso  $z$ , il suo modulo  $|z|$  coincide col raggio polare del punto che ne è l'immagine sul piano complesso. Chiamiamo *argomento* di  $z$ , e lo indicheremo con  $\arg(z)$ , uno qualsiasi degli angoli  $\theta$  relativi al punto  $z$ . In questo modo l'argomento

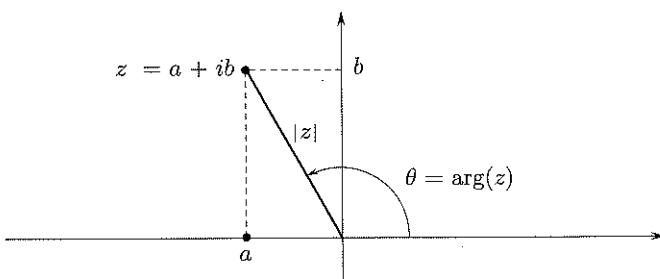


Figura 1.10.

di  $z$  non è univocamente determinato. Spesso questa indeterminatezza non porta alcun inconveniente. Altre volte invece è preferibile assegnare un ben determinato argomento a un numero complesso. Ciò può ottenersi in infiniti modi, fissando un qualsiasi intervallo, di ampiezza  $2\pi$ , entro il quale far variare l'angolo  $\theta$ . Gli intervalli più comunemente usati a questo scopo sono  $[0, 2\pi)$  e  $(-\pi, \pi]$ ; allora l'argomento di  $z$  viene detto *argomento principale*.

Per esempio, il numero  $-i$  ha come argomento  $-\pi/2$  oppure  $3\pi/2$  oppure qualsiasi altro valore della forma  $-\pi/2 + 2k\pi$  con  $k \in \mathbb{Z}$ . Il suo argomento principale sarà  $3\pi/2$  se si adotta la convenzione che  $\theta \in [0, 2\pi)$ ,  $-\pi/2$  con la convenzione che  $\theta \in (-\pi, \pi]$ . I numeri reali positivi hanno argomento principale 0 e quelli negativi  $\pi$  con entrambe le convenzioni. Per il numero 0 l'argomento non viene definito.

Dato il numero  $z = a + ib$ , dalla trigonometria ricaviamo immediatamente le relazioni tra le coordinate cartesiane  $a, b$  e quelle polari  $\rho, \theta$ :

$$(8.6) \quad a = \rho \cos \theta \quad b = \rho \sin \theta$$

cosicché il numero complesso  $z$  può anche scriversi nella forma

$$(8.7) \quad z = \rho(\cos \theta + i \sin \theta)$$

che è detta *forma trigonometrica* dei numeri complessi.

Le relazioni inverse di (8.6) sono:

$$(8.8) \quad \rho = \sqrt{a^2 + b^2} \quad \cos \theta = \frac{a}{\sqrt{a^2 + b^2}} \quad \sin \theta = \frac{b}{\sqrt{a^2 + b^2}}$$



Scriviamo in forma trigonometrica il numero complesso:

$$\sqrt{3} + i$$

Abbiamo  $\rho = \sqrt{a^2 + b^2} = \sqrt{3 + 1} = 2$ ;

$$\sqrt{3} + i = 2 \left( \frac{\sqrt{3}}{2} + i \frac{1}{2} \right) = 2 \left( \cos \frac{\pi}{6} + i \sin \frac{\pi}{6} \right)$$

**Formule di De Moivre**

La forma trigonometrica è comoda per esprimere prodotti e quozienti di numeri complessi. Se abbiamo infatti

$$z_1 = \rho_1(\cos \theta_1 + i \sin \theta_1) \quad z_2 = \rho_2(\cos \theta_2 + i \sin \theta_2)$$

otteniamo

$$\begin{aligned} z_1 z_2 &= \rho_1 \rho_2 \{ \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 + i(\sin \theta_1 \cos \theta_2 + \cos \theta_1 \sin \theta_2) \} = \\ (8.9) \quad &= \rho_1 \rho_2 \{ \cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2) \} \end{aligned}$$

Se  $z_2 \neq 0$ , abbiamo

$$\frac{z_1}{z_2} = \frac{\rho_1}{\rho_2} \cdot \frac{\cos \theta_1 + i \sin \theta_1}{\cos \theta_2 + i \sin \theta_2}$$

Moltiplicando numeratore e denominatore per  $\cos \theta_2 - i \sin \theta_2$  e tenendo conto che  $(\cos \theta_2)^2 + (\sin \theta_2)^2 = 1$  abbiamo:

$$(8.10) \quad \frac{z_1}{z_2} = \frac{\rho_1}{\rho_2} (\cos \theta_1 + i \sin \theta_1)(\cos \theta_2 - i \sin \theta_2) = \frac{\rho_1}{\rho_2} \{ \cos(\theta_1 - \theta_2) + i \sin(\theta_1 - \theta_2) \}$$

Pertanto il modulo del prodotto e del quoziente di due numeri complessi è, rispettivamente, il prodotto e il quoziente dei moduli; l'argomento è, rispettivamente, la somma e la differenza degli argomenti:

$$\begin{aligned} (8.11) \quad |z_1 \cdot z_2| &= |z_1| \cdot |z_2| \quad \arg(z_1 \cdot z_2) = \arg z_1 + \arg z_2 \\ |z_1/z_2| &= |z_1| / |z_2| \quad \arg(z_1/z_2) = \arg z_1 - \arg z_2 \end{aligned}$$

La (8.9) si generalizza al caso di un numero qualsiasi di fattori  $z_1, z_2, \dots, z_n$ .

$$(8.12) \quad z_1 z_2 \dots z_n = \rho_1 \rho_2 \dots \rho_n \{ \cos(\theta_1 + \theta_2 + \dots + \theta_n) + i \sin(\theta_1 + \theta_2 + \dots + \theta_n) \}$$

Se poi i fattori sono tutti uguali, otteniamo:

$$(8.13) \quad z^n = \rho^n (\cos(n\theta) + i \sin(n\theta))$$

Le relazioni (8.9)-(8.13) vanno sotto il nome di *formule di De Moivre*. Vediamone alcune applicazioni.



**ESEMPIO** Scrivere in forma algebrica  $(1+i)^7$ .

Determiniamo modulo e argomento di  $(1+i)$  e poi applichiamo la formula di De Moivre.

$$|1+i| = \sqrt{2} \quad \arg(1+i) = \frac{\pi}{4}$$

Quindi

$$|(1+i)^7| = (\sqrt{2})^7 = 8\sqrt{2} \quad \arg(1+i)^7 = \frac{7}{4}\pi$$

Di conseguenza:

$$(1+i)^7 = 8\sqrt{2} \left( \frac{1}{\sqrt{2}} - i \frac{1}{\sqrt{2}} \right) = 8 - 8i$$

**Risolvere l'equazione**

$$z^3 - |z| = 0$$

Sostituire  $z = x + iy$  e separare parte reale e parte immaginaria è possibile, ma porta a calcoli algebrici un po' pesanti. Se riscriviamo l'equazione nella forma

$$z^3 = |z|$$

possiamo invece notare che ambo i membri si esprimono facilmente se si pone  $z = \rho(\cos \theta + i \sin \theta)$ , ovvero se si usa la *forma trigonometrica*. Infatti:

$$z^3 = \rho^3 (\cos 3\theta + i \sin 3\theta) \quad |z| = \rho$$

e l'equazione è soddisfatta se e solo se i due membri hanno moduli uguali e argomenti che differiscono per multipli di  $2\pi$ , ovvero (il secondo membro ha argomento 0):

$$\begin{cases} \rho^3 = \rho \\ 3\theta = 2k\pi \quad \text{con } k \in \mathbb{Z} \end{cases}$$

La prima equazione dà  $\rho = 0$  e  $\rho = 1$  (attenzione:  $\rho$  dev'essere  $\geq 0$  perché è il modulo del numero complesso; perciò  $\rho = -1$  non è accettabile); la seconda dà:  $\theta = \frac{2k\pi}{3}$ . Si trova pertanto:

$$z = 0 \quad z = \cos \frac{2k\pi}{3} + i \sin \frac{2k\pi}{3} \quad \text{per } k = 0, 1, 2$$

Esplicitamente:

$$z = 0 \quad z = 1 \quad z = -\frac{1}{2} + i \frac{\sqrt{3}}{2} \quad z = -\frac{1}{2} - i \frac{\sqrt{3}}{2}$$

Le formule di De Moivre permettono di dare un'*interpretazione geometrica al prodotto di numeri complessi*. Sia  $z$ , per cominciare, un numero complesso di modulo 1, quindi del tipo  $(\cos \theta + i \sin \theta)$ . Allora, moltiplicare un numero per  $z$  significa sommare  $\theta$  al suo argomento, cioè eseguire una *rotazione di angolo  $\theta$* . Se  $z$  ha modulo  $\rho$  anziché 1, oltre ad eseguire una rotazione si esegue una *dilatazione di coefficiente  $\rho$* . Ad esempio:

- moltiplicare per  $i$  significa eseguire una rotazione di  $\frac{\pi}{2}$ ;
- moltiplicare per  $-1$  significa eseguire una rotazione di  $\pi$ ;
- moltiplicare per  $(1+i)$  significa eseguire una dilatazione di coefficiente  $\sqrt{2}$  e una rotazione di  $\pi/4$ .

**8.4 Radici  $n$ -esime**

Dato un numero complesso  $w$ , diremo che  $z$  è una *radice  $n$ -esima (complessa)* di  $w$  se risulta  $z^n = w$ .

**TEOREMA 1.10** Sia  $w \in \mathbb{C}$ ,  $w \neq 0$ , e  $n$  intero  $\geq 1$ . Esistono precisamente  $n$  radici  $n$ -esime complesse  $z_0, z_1, \dots, z_{n-1}$  di  $w$ ; posto  $w = r(\cos \varphi + i \sin \varphi)$  e  $z_k = \rho_k(\cos \theta_k + i \sin \theta_k)$  abbiamo

$$\begin{aligned} (8.14) \quad \rho_k &= r^{1/n} & k = 0, 1, \dots, n-1 \\ \theta_k &= \frac{\varphi + 2k\pi}{n} \end{aligned}$$

**DIMOSTRAZIONE.** I numeri  $z_k$  sono evidentemente radici di  $w$ , come risulta calcolando  $z_k^n$  mediante la formula di De Moivre. Mostriamo che non ve ne sono altre. Infatti, perché un numero  $R(\cos \psi + i \sin \psi)$  sia radice  $n$ -esima di  $w$ , dovrebbe risultare

$$R^n = r \quad \text{e} \quad n\psi = \varphi + 2h\pi \quad \text{con} \quad h \in \mathbb{Z}$$

e cioè

$$R = r^{1/n} \quad \text{e} \quad \psi = \varphi/n + 2h\pi/n$$

Dando a  $h$  i valori  $0, 1, \dots, n-1$  troviamo appunto i numeri  $z_k$ . Dando a  $h$  un qualsiasi altro valore  $\bar{h}$  diverso dai precedenti, questo può scriversi nella forma  $\bar{h} = k + mn$  ( $m \in \mathbb{Z}$  è il quoziente e  $k$  è il resto della divisione di  $\bar{h}$  per  $n$ ) per cui sarebbe

$$\psi = \frac{\varphi}{n} + \frac{2k\pi}{n} + 2m\pi = \theta_k + 2m\pi$$

e ritroveremmo ancora gli stessi  $z_k$  precedenti.  $\diamond$

Per le radici complesse si usa purtroppo una notazione un po' ambigua, la stessa in uso per indicare la radice aritmetica; si indica cioè con  $\sqrt[n]{z}$  o  $z^{1/n}$  l'insieme delle  $n$  radici complesse di  $z$ . Ciò può creare confusione quando  $z$  è reale. Infatti il simbolo  $\sqrt{4}$ , inteso come radice aritmetica di 4, è 2; inteso come radice complessa di 4 è l'insieme dei due numeri  $+2$  e  $-2$ .

### ESEMPIO

Le tre radici cubiche di  $-1$  sono i numeri  $z_k$  della forma:  $\cos \theta_k + i \sin \theta_k$  con  $\theta_k = \pi/3 + 2k\pi/3$ ,  $k = 0, 1, 2$ . Esplicitamente abbiamo:

$$z_0 = \frac{1}{2}(1 + \sqrt{3}i)$$

$$z_1 = -1$$

$$z_2 = \frac{1}{2}(1 - \sqrt{3}i)$$

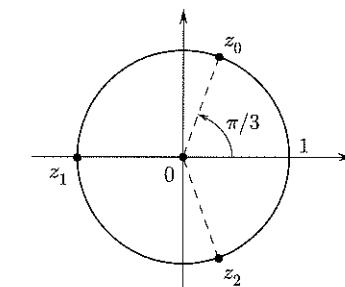


Figura 1.11.

La disposizione delle radici cubiche dell'esempio precedente nel piano di Gauss non è casuale. Infatti se  $w = r(\cos \varphi + i \sin \varphi)$  le radici  $n$ -esime  $z_0, z_1, \dots, z_{n-1}$  di  $w$  si trovano ai vertici del poligono regolare di  $n$  lati inscritto nella circonferenza di centro 0 e raggio  $r^{1/n}$ , con il vertice  $z_0$  posto nel punto di argomento  $\theta = \varphi/n$ . Nella figura 1.12 sono rappresentate le radici seste di  $i$ :  $z_0, z_1, \dots, z_5$ .

### ESEMPIO

Calcolare  $\sqrt[5]{1+i}$ . Il numero  $(1+i)$  ha modulo  $\sqrt{2}$  e argomento  $\pi/4$ . Le radici quinte saranno quindi:

$$\begin{aligned} \sqrt[5]{\sqrt{2}} \left[ \cos \left( \frac{\frac{\pi}{4} + 2k\pi}{5} \right) + i \sin \left( \frac{\frac{\pi}{4} + 2k\pi}{5} \right) \right] &= \\ = \sqrt[10]{2} \left[ \cos \left( \frac{\pi}{20} + \frac{k}{5}\pi \right) + i \sin \left( \frac{\pi}{20} + \frac{k}{5}\pi \right) \right] \quad \text{con} \quad k = 0, 1, 2, 3, 4 \end{aligned}$$

Gli angoli trovati non sono notevoli, ma volendo si possono calcolare seni e coseni in modo approssimato, usando una calcolatrice.

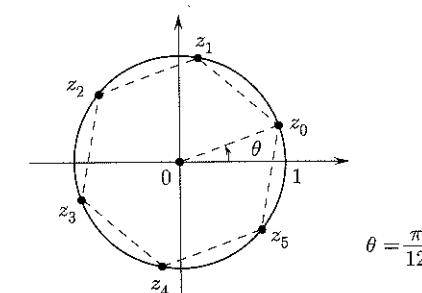


Figura 1.12.

### Equazioni di secondo grado

Un'equazione di secondo grado

$$az^2 + bz + c = 0$$

con coefficienti  $a, b, c \in \mathbb{C}$ , si risolve con la "solita" formula

$$z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

a patto di intendere la radice quadrata in senso complesso. Il segno  $\pm$  davanti al simbolo di radice è in realtà superfluo, perché il simbolo stesso  $\sqrt{\cdot}$  nel campo complesso denota *due numeri*, uno opposto dell'altro.

### ESEMPIO

$$z^2 + 2iz - \sqrt{3}i = 0$$

$$\begin{aligned} z &= -i \pm \sqrt{-1 + \sqrt{3}i} = -i \pm \sqrt{2 \left( \cos \frac{2}{3}\pi + i \sin \frac{2}{3}\pi \right)} = \\ &= -i \pm \sqrt{2} \left( \cos \frac{\pi}{3} + i \sin \frac{\pi}{3} \right) = \pm \frac{\sqrt{2}}{2} + i \left( -1 \pm \frac{\sqrt{6}}{2} \right) \end{aligned}$$

Il teorema 1.10 ci dice che un polinomio del tipo  $z^n + a$  (con  $a$  complesso) ha in  $\mathbb{C}$  esattamente  $n$  radici; nel campo reale invece l'equazione  $x^n + a = 0$  può avere due, una, o nessuna radice (esempi:  $x^2 - 1 = 0$ ,  $x^3 - 1 = 0$ ,  $x^2 + 1 = 0$ ); ora sappiamo che tale equazione ha sempre  $n$  radici in  $\mathbb{C}$ , ma solo occasionalmente una o due di esse stanno in  $\mathbb{R}$ .

Il risultato è di portata ben più generale, come afferma il seguente teorema, di cui non riportiamo la dimostrazione.

**TEOREMA 1.11 (TEOREMA FONDAMENTALE DELL'ALGEBRA)** *Un'equazione polinomiale della forma*

$$a_0 + a_1 z + \dots + a_n z^n = 0 \quad (a_n \neq 0)$$

con coefficienti complessi qualsiasi ha precisamente  $n$  radici in  $\mathbb{C}$ , se ognuna di esse viene contata con la sua molteplicità.<sup>5</sup>



- Scrivere in forma algebrica ( $z = a + ib$ ,  $a, b \in \mathbb{R}$ ) i seguenti numeri complessi:

$$\frac{(2+i)(1-i)}{3-2i} \quad \frac{1}{i(3+2i)^2} \quad \frac{(\sqrt{3}+i\sqrt{2})^3}{\sqrt{2}-\sqrt{3}i}$$

- Calcolare modulo e argomento dei seguenti numeri complessi:

$$-3i; -5; -\sqrt{3} + i$$

- Calcolare modulo e argomento dei numeri:

$$\frac{1+i\sqrt{3}}{1-i} \quad \frac{1+i}{1-i} \quad \frac{1+i}{\sqrt{3}+i}$$

- Disegnare nel piano complesso il luogo dei punti  $z$  tali che:

$$\text{i) } |z| = |z + i| \quad \text{ii) } \operatorname{Re}(z^2) > 2 \quad \text{iii) } \operatorname{Im} \frac{1}{z} = -1$$

- Verificare che,  $\forall z \in \mathbb{C}$ :

$$|z| \geq \operatorname{Re}(z) \quad |z| \geq \operatorname{Im}(z) \quad |z| \leq |\operatorname{Re}(z)| + |\operatorname{Im}(z)|$$

- Scrivere in forma algebrica  $a + ib$  i numeri

$$\frac{1-i\sqrt{3}}{1+i} \quad \left(\frac{1-i}{1+i}\right)^3 \quad (1+i)^{20} \quad (1-i)^{11} \quad (1+i\sqrt{3})^n - (1-i\sqrt{3})^n$$

- Calcolare le radici seste di  $-1$  e rappresentarle sul piano di Gauss.

- Risolvere le equazioni

$$\text{a) } (z-2)^3 = -i \quad \text{b) } z^2 - (4+i)z + 4 + 2i = 0$$

- Esprimere  $\cos 5\theta$  come combinazione lineare di potenze di  $\cos \theta$  e  $\sin \theta$ .

- Esprimere  $(\sin \theta)^4$  come combinazione lineare di termini del tipo  $\sin(k\theta)$  e  $\cos(k\theta)$ .

- Calcolare le seguenti radici  $n$ -esime nel campo complesso. Quando l'argomento di  $\sqrt[n]{z}$  è un valore notevole, riscrivere in forma algebrica le radici (ad esempio:  $2(\cos \frac{\pi}{6} + i \sin \frac{\pi}{6}) = \sqrt{3} + i$ ).

$$\sqrt[3]{i-1} \quad \sqrt{-1+\sqrt{3}i} \quad \sqrt[4]{-2-2\sqrt{3}i}$$

- Disegnare nel piano complesso i seguenti insiemi:

$$A = \left\{ z : 1 < |z| < 2; \frac{\pi}{6} < \arg z < \frac{\pi}{3} \right\} \quad B = \{iz : z \in A\}$$

<sup>5</sup>Se  $P(z)$  è un polinomio in  $z$  di grado  $n$  e  $z_0$  una sua radice, si dice che  $z_0$  è di molteplicità  $k$  ( $k$  intero,  $\geq 1$ ) se vale la formula  $P(z) = (z - z_0)^k Q(z)$ , dove  $Q$  è un polinomio tale che  $Q(z_0) \neq 0$ .

Determinare tutte le soluzioni delle seguenti equazioni nel campo complesso:

1) $z^2 + 2z + 3 = 0$	1) $z + 3i + (\operatorname{Re} z)(i + (\operatorname{Im} z)^2) = 0$
2) $z^2 + 2iz - 3 = 0$	2) $iz^3 = \bar{z}$
3) $iz^2 + (1+i)z + 1 = 0$	3) $z^6 + 2z^3 - 3 = 0$
4) $z^2 + \bar{z} = 0$	4) $(\bar{z})^4 =  z $
5) $i \operatorname{Re} z + z^2 =  z ^2 - 1$	5) $2 z ^2 = z^3$

## COMPLEMENTI

1 Nel paragrafo 1.1 si sono definite l'unione e l'intersezione di due (e quindi di un numero finito) di insiemi, e si è mostrata l'analogia tra unione e "o", intersezione ed "e". Consideriamo ora una famiglia infinita  $A_1, A_2, A_3, \dots$  di sottoinsiemi di un insieme fissato  $X$ . Come si potrebbero definire rigorosamente la loro unione e intersezione infinita, cioè

$$\bigcup_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} A_n ?$$

*Suggerimento:* non si possono usare infinite "e" od "o", bisognerà invece usare opportunamente i quantificatori.

- 2 Dimostrare che  $\log_2 3$  è irrazionale.

- 3 Dimostrare le proprietà dei coefficienti binomiali (2.3) enunciate nel paragrafo 2.3 mediante un calcolo algebrico diretto (e non per induzione).

- 4 Dimostrare la diseguaglianza di Bernoulli  $(1+x)^n \geq 1+nx$  nel caso particolare  $x \geq 0$ , usando lo sviluppo del binomio di Newton (invece che per induzione, come fatto nel paragrafo 7).

- 5 Dimostrare che l'intervallo  $[0, 1]$  e il quadrato  $[0, 1] \times [0, 1]$  hanno la stessa cardinalità. *Suggerimento:* rappresentare in forma decimale  $0.a_1a_2a_3\dots$  gli elementi di  $[0, 1]$ , e trovare un algoritmo esplicito che associa, biunivocamente, ad ogni elemento di questo tipo una coppia ordinata di elementi di questo tipo.

- 6 Dimostrare che  $\mathcal{P}(\mathbb{N})$  ha la stessa cardinalità dell'intervallo  $[0, 1]$ .

*Suggerimento:* rappresentare i numeri dell'intervallo  $[0, 1]$  in forma decimale binaria, cioè del tipo  $0.a_1a_2a_3\dots$  con  $a_i$  cifre 0 o 1; quindi, trovare una corrispondenza biunivoca tra l'insieme dei numeri così scritti e  $\mathcal{P}(\mathbb{N})$ .

- 7 Si studi e si comprenda in dettaglio la seguente dimostrazione del teorema, enunciato nel paragrafo 6: "Per ogni insieme  $A$ ,  $A$  e  $\mathcal{P}(A)$  non hanno la stessa cardinalità".

**DIMOSTRAZIONE.** Per assurdo, supponiamo che esista una corrispondenza biunivoca  $f$  tra  $A$  e  $\mathcal{P}(A)$ , e definiamo

$$X = \{a \in A : a \notin f(a)\}$$

dove  $f(a)$  indica l'elemento di  $A$  che corrisponde ad  $a$ . Sia ora  $x \in A$  l'elemento che corrisponde a  $X$ . Allora si ha:  $x \in f(x) \iff x \in X \iff x \notin f(x)$ , assurdo.

# 2

# Funzioni di una variabile

## 1 IL CONCETTO DI FUNZIONE

Il concetto di *funzione* è centrale in matematica, ed estremamente generale. Dal punto di vista fisico, si può dire che questo concetto nasca per descrivere matematicamente quello di *grandezza variabile*. Esempi di grandezze variabili sono pressoché ovunque attorno a noi: la posizione di un oggetto mobile, o la sua velocità, la temperatura in una stanza, la densità di un oggetto, i prezzi delle merci ecc. Si potrebbe pensare che, dopo tutto, queste grandezze sono espresse da *numeri*, per cui non c'è bisogno di introdurre alcun nuovo concetto per descriverle. Tuttavia, dire che, ad esempio, la velocità di un'automobile è espressa da un numero è un'affermazione piuttosto imprecisa, perché un numero può indicare solo la velocità *in un istante fissato*. Più in generale: i numeri, di cui finora ci siamo prevalentemente occupati, sono adatti a esprimere *grandezze costanti*; per esprimere grandezze variabili ci vuole un'idea diversa. Ora, un'analisi logica del concetto di grandezza variabile mostra che questa coinvolge sempre una relazione tra *due* grandezze: la velocità di un'automobile varia *al variare del tempo*; la densità di un oggetto può variare *da punto a punto*; in altre parole:

l'esistenza di una grandezza variabile sottointende l'esistenza di una relazione tra *due* grandezze (tempo e velocità; posizione e densità ecc.), ovvero la *dipendenza* di una grandezza da un'altra.

Questa dipendenza segue, di volta in volta, una certa *legge*. Vediamo alcuni esempi di questo tipo, che ci permetteranno di precisare una caratteristica fondamentale del concetto di funzione:

- Se si lascia cadere un oggetto pesante da una certa altezza, lo spazio percorso dall'oggetto varia col tempo secondo la formula

$$s(t) = \frac{1}{2}gt^2 \quad (g \text{ accelerazione di gravità})$$

Cioè a  $t$  viene associato  $s(t)$ ; in simboli:

$$t \mapsto s(t) = \text{spazio percorso al tempo } t.$$

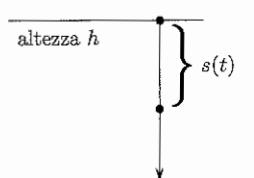


Figura 2.1.

- Se un capitale unitario viene investito per un anno al tasso mensile  $i$ , il capitale finale dipende da  $i$  secondo la formula

$$k(i) = \left(1 + \frac{i}{12}\right)^{12}$$

Cioè:

$$i \mapsto k(i) = \text{capitale alla fine dell'anno}$$

- La potenza  $P$  di una lente dipende dalla lunghezza focale  $f$  secondo la formula

$$P = \frac{1}{f}$$

(se  $f$  è misurata in metri,  $P$  è espressa in diottrie).

In ciascuno degli esempi precedenti, a un numero reale (ingresso) viene associato *univocamente* un altro numero reale (uscita). È proprio l'univocità della corrispondenza a caratterizzare una funzione. In generale, gli ingressi ammissibili per una data corrispondenza sono soggetti a restrizioni naturali, legate alla natura stessa della corrispondenza.

Nel primo esempio il numero reale "di partenza" ha il significato di tempo; immaginando di lasciar cadere l'oggetto a un tempo iniziale  $t = 0$  è evidente che ci si dovrà limitare a tempi  $t \geq 0$ .

Nel secondo esempio, evidentemente dovrà essere  $0 \leq i \leq 1$ . Infine, nel terzo esempio, dal significato di  $f$ , deve essere  $f \geq 0$ , mentre la formula impone ulteriormente  $f > 0$ .

L'insieme dei valori "di partenza" ammissibili per una data funzione prende il nome di *dominio*.

Spesso si usano le locuzioni *variabile indipendente* per indicare un ingresso generico e *variabile dipendente* per indicare l'uscita.

I prossimi esempi sono di tipo un po' diverso:

- Consideriamo una sbarra metallica sottile di lunghezza  $L$ ; possiamo rappresentarla con l'intervallo  $[0, L]$ . La temperatura  $T$  nel punto  $x$  della sbarra all'istante  $t$ , è espressa da un numero che dipende dalla *coppia* di numeri  $(x, t)$ , secondo una certa legge. In simboli:

$$(x, t) \mapsto T(x, t).$$

Questa volta l'*ingresso* non è un numero, ma piuttosto una coppia ordinata di numeri  $(x, t)$ , mentre l'*uscita* è ancora un numero. D'altro canto, se  $x \in [0, L]$  e  $t \geq 0$ , la coppia  $(x, t)$  è un elemento dell'insieme  $A = [0, L] \times [0, +\infty)$ . Si può ancora dire, quindi, che ad ogni elemento di un certo *insieme*  $A$  viene associato univocamente un elemento di un altro insieme  $B$  (qui  $B = \mathbb{R}$ ).

- Se  $A$  è l'insieme degli studenti di un certo corso e  $B$  è l'insieme dei loro nomi, la corrispondenza che associa ad ogni studente il suo nome è univoca (mentre il viceversa non è necessariamente vero: potrebbero esserci due studenti con lo stesso nome). In questo caso né  $A$  né  $B$  sono insiemi numerici, e tuttavia risulta ben definita una corrispondenza univoca tra questi due insiemi.

Si arriva così al concetto generale di funzione<sup>1</sup>:

<sup>1</sup>Questa definizione moderna di funzione è dovuta a Dirichlet, in una memoria del 1837.

**DEFINIZIONE 2.1** Dati due insiemi  $A, B$  qualsiasi, una *funzione f di dominio A a valori in B* (o anche "di codominio  $B$ ") è una qualsiasi legge che ad ogni elemento di  $A$  associa uno e un solo elemento di  $B$ . Scrivereemo:

$$f : A \rightarrow B$$

(che si legge "f definita da  $A$  a  $B$ "). La scrittura

$$f : x \mapsto f(x)$$

(che si legge "f associa  $f(x)$  ad  $x$ ") indica come la funzione  $f$  agisce sugli elementi. Il simbolo  $f(x)$  indica il valore che la funzione  $f$  associa ad  $x$ , e non va confuso col simbolo  $f$ , che denota la funzione stessa.

Ad esempio, per la funzione  $k$  sopra definita si avrebbe:

$$k : [0, 1] \rightarrow \mathbb{R}$$

$$k : i \mapsto \left(1 + \frac{i}{12}\right)^{12}$$

Si usa anche la scrittura

$$y = f(x)$$

per indicare che la variabile  $y$  è funzione di  $x$ , ossia che  $y$  è il valore che  $f$  associa all'ingresso  $x$ . Di nuovo, la proprietà caratteristica di  $f$ , affinché la si possa chiamare funzione, è l'*univocità* della corrispondenza: assegnato l'elemento di "ingresso"  $a \in A$ , l'elemento di "uscita"  $b = f(a)$  dev'essere univocamente determinato.

Si può pensare a una funzione come a una scatola nera che a ogni ingresso ammissibile  $x$  associa un'unica uscita  $f(x)$ :

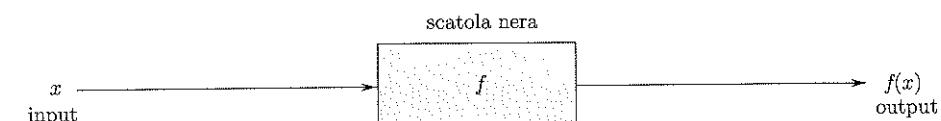


Figura 2.2.

L'uscita corrispondente a  $x$  si chiama *immagine* di  $x$ ; l'insieme delle possibili uscite si chiama *immagine* di  $D$  tramite  $f$  e si indica con il simbolo  $f(D)$  o  $\text{Im } f$ . Si noti che nella scrittura  $f : A \rightarrow B$ , il codominio  $B$  può essere più grande dell'immagine  $f(A)$  (ossia in generale è  $f(A) \subseteq B$ ). Ad esempio, se  $f$  ha valori reali, solitamente si scrive  $f : A \rightarrow \mathbb{R}$  senza precisare quale sia l'effettiva immagine di  $f$ .

Alcune classi di funzioni di cui ci occuperemo in questo corso sono le seguenti:

- le funzioni  $f : \mathbb{N} \rightarrow \mathbb{R}$ , che si dicono *successioni*: le studieremo nel capitolo 3, paragrafo 1;
- le funzioni  $f : \mathbb{R} \rightarrow \mathbb{R}$  (*funzioni reali di variabile reale*), di cui ci occuperemo a partire da questo capitolo;

- le funzioni  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  di tipo lineare, dette anche *trasformazioni lineari*, di cui si occupa l'algebra lineare;
- le funzioni  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  (non necessariamente lineari), di cui ci occuperemo nel secondo volume.

Infine, nello studio del calcolo differenziale e integrale, incontreremo anche alcuni esempi di funzioni definite tra insiemi che, a loro volta, hanno per elementi altre funzioni. Come si vede, il concetto di *funzione* (come quello di *insieme*, introdotto nel cap. 1) fa attualmente parte del linguaggio di base della matematica, che unifica tra loro concetti e oggetti molto diversi.

## 2 FUNZIONI REALI DI VARIABILE REALE

### 2.1 Generalità

Nel paragrafo precedente è stato introdotto il concetto di funzione nella sua generalità. Le funzioni di cui ci occuperemo a partire da questo capitolo, sono le *funzioni reali di variabile reale*, ossia le funzioni per cui sia la variabile di “ingresso” che quella di “uscita” sono numeri reali. La funzione ha dunque per *dominio* un sottoinsieme di  $\mathbb{R}$  e per *codominio*  $\mathbb{R}$ ; l'*immagine* di  $f$  sarà anch'essa un sottoinsieme di  $\mathbb{R}$ :

$$f : D \rightarrow \mathbb{R} \text{ con } D \subseteq \mathbb{R}$$

$$f : x \mapsto f(x)$$

Le funzioni più comuni, come vedremo, hanno come dominio e come immagine un *intervallo* (che eventualmente è tutto  $\mathbb{R}$ ) o l'*unione di un numero finito di intervalli* (cfr. par. 4.4 del cap. 1).

La dipendenza di  $f(x)$  da  $x$  si visualizza efficacemente disegnando il *grafico* di  $f$ , ossia l'insieme dei punti del piano di coordinate  $(x, y)$  con

$$y = f(x)$$

e  $x$  variabile nel dominio  $D$  (v. fig. 2.3). Per le funzioni più comuni, il grafico è una curva, nel senso intuitivo del termine.

La proprietà fondamentale che fa di  $f$  una funzione, ossia il fatto che ad ogni ingresso  $x \in D$  faccia corrispondere *una e una sola* uscita  $f(x) \in \mathbb{R}$ , ha allora il seguente significato geometrico:

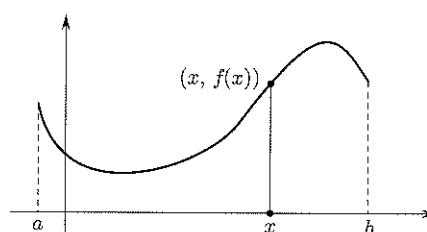


Figura 2.3. Grafico di una funzione di dominio  $D = [a, b]$ .

ogni retta parallela all'asse delle ordinate che taglia l'asse delle ascisse in un punto  $x$  del dominio  $D$ , interseca il grafico di  $f$  in uno e un sol punto.

Infatti, se la retta non intersecasse il grafico significherebbe che all'ingresso  $x$  non corrisponde alcuna uscita, mentre se lo intersecasse in più di un punto significherebbe che all'ingresso  $x$  corrispondono più uscite distinte, cadendo quindi l'univocità della funzione (v. fig. 2.4).

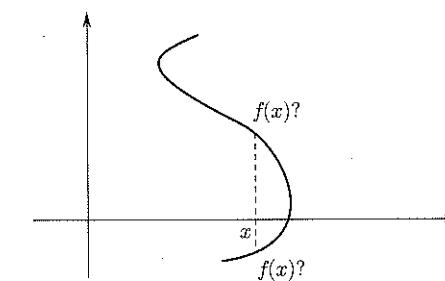


Figura 2.4. Questa curva non è il grafico di una funzione: all'ingresso  $x$  quale uscita  $f(x)$  corrisponde?

Si noti che, invece, nulla impedisce che una retta parallela all'asse delle ascisse tagli il grafico di  $f$  in più punti (o in nessun punto).

Per le funzioni reali di variabile reale si possono introdurre alcune importanti nozioni, che non sempre hanno senso per funzioni di tipo più generale.

### 2.2 Funzioni limitate

Se il grafico di una funzione  $f : D \rightarrow \mathbb{R}$  è contenuto nel semipiano inferiore delimitato da una retta parallela all'asse delle ascisse, per esempio di equazione  $y = M$ , la funzione si dice *limitata superiormente*. Analiticamente significa che

$$f(x) \leq M \quad \text{per ogni } x \in D$$

Analogamente,  $f$  si dirà *limitata inferiormente* se il suo grafico è contenuto nel semipiano superiore

$$f(x) \geq m \quad \text{per ogni } x \in D$$

Una funzione si dirà *limitata* se è limitata sia inferiormente che superiormente. Il grafico di una funzione limitata è contenuto in una striscia orizzontale del piano  $xy$ .

Per esempio (v. fig. 2.5):

- $x \mapsto x^3, x \in \mathbb{R}$ , non è limitata né superiormente, né inferiormente.
- $x \mapsto x^2, x \in \mathbb{R}$ , è limitata inferiormente; infatti  $x^2 \geq 0, \forall x \in \mathbb{R}$ .
- $x \mapsto \frac{1}{1+x^2}, x \in \mathbb{R}$ , è limitata, poiché  $0 < \frac{1}{1+x^2} \leq 1, \forall x \in \mathbb{R}$ .

Equivalentemente, si può dire che una funzione è limitata superiormente (limitata inferiormente, limitata) se, rispettivamente, la sua *immagine* è un sottoinsieme di  $\mathbb{R}$  limitato superiormente (inferiormente, limitato).

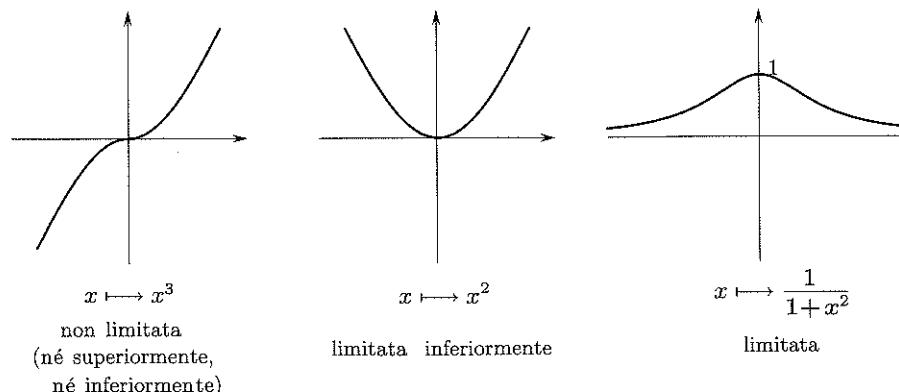


Figura 2.5.

### 2.3 Funzioni simmetriche

I grafici di alcune funzioni presentano particolari proprietà di simmetria. Per esempio vi sono funzioni il cui grafico è simmetrico rispetto all'asse delle ordinate. Queste funzioni si chiamano funzioni *pari*, hanno un dominio simmetrico rispetto a  $x = 0$  (tipicamente un intervallo del tipo  $(-a, a)$ ) e sono caratterizzate dalla relazione

$$(2.1) \quad f(-x) = f(x)$$

che esprime l'uguaglianza delle ordinate corrispondenti ai punti  $x$  e  $-x$ , simmetrici rispetto a  $x = 0$ .

Funzioni che hanno il grafico simmetrico rispetto all'origine si chiamano *dispari*; anch'esse hanno dominio simmetrico rispetto a  $x = 0$  e sono caratterizzate dalla relazione

$$(2.2) \quad f(-x) = -f(x)$$

Per esempio, la funzione  $x \mapsto x^2$  è pari, mentre  $x \mapsto x^3$  è dispari. Più in generale, le potenze a esponente intero sono funzioni pari (dispari) se l'esponente è pari (dispari).

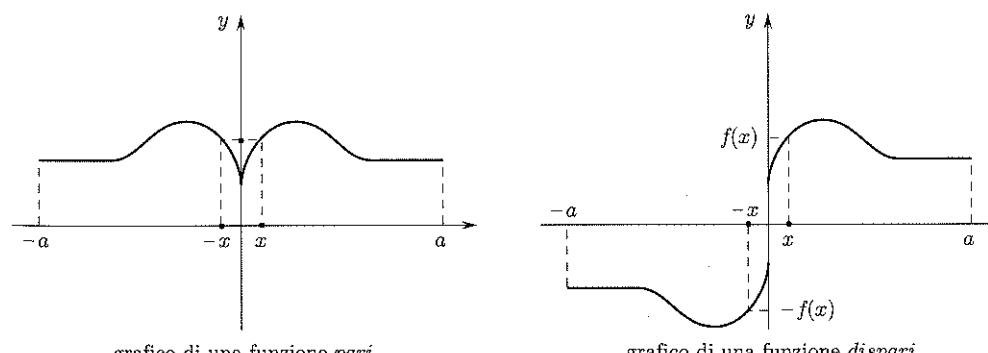


Figura 2.6.

Notiamo che una funzione *non può*, invece, avere il grafico simmetrico *rispetto all'asse x*: per quanto osservato nel paragrafo 2.1, una curva con tale simmetria non sarebbe il grafico di una funzione, venendo meno l'univocità della corrispondenza.

### 2.4 Funzioni monotone

Una funzione si dice *monotona crescente* o *non decrescente* se per ogni coppia di punti  $x_1, x_2$  nel dominio di  $f$  si ha:

$$(2.3) \quad x_1 > x_2 \text{ implica } f(x_1) \geq f(x_2)$$

Se nella (2.3) vale il maggiore anziché il maggiore o uguale si dice che  $f$  è *strettamente crescente*.

Se invece nella (2.3) si sostituisce " $\geq$ " con " $\leq$ "  $f$  si dice *decrescente* o *non crescente*; se poi la (2.3) vale con " $<$ " anziché con " $\leq$ "  $f$  si dirà *strettamente decrescente*.

In altri termini:  $f$  è crescente (strettamente crescente) se, all'aumentare di  $x$ , l'ordinata corrispondente sul grafico non diminuisce (aumenta);  $f$  è decrescente (strettamente decrescente) se, all'aumentare di  $x$ , l'ordinata corrispondente non aumenta (diminuisce).

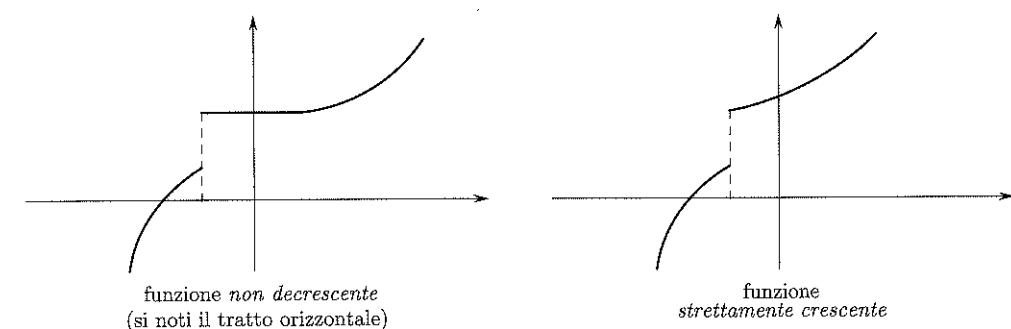


Figura 2.7.

Per esempio,  $x \mapsto x^3$  è strettamente crescente; la funzione costante  $x \mapsto k$  (che ha come grafico la retta di equazione  $y = k$ ) è sia non decrescente sia non crescente. Tutte queste funzioni (crescenti o decrescenti, strettamente o non) si dicono *monotone*.

### 2.5 Funzioni periodiche

La funzione  $f : D \rightarrow \mathbb{R}$  (non costante) è *periodica* di periodo  $T$ ,  $T > 0$ , se  $T$  è il più piccolo numero reale positivo tale che

$$f(x+T) = f(x) \text{ per ogni } x \in D$$

Ogni intervallo di lunghezza  $T$ , contenuto in  $D$ , si chiama *intervallo di periodicità*. Basterà disegnare il grafico di  $f$  su un qualunque intervallo di periodicità, per conoscere il grafico di  $f$  su tutto il dominio (v. fig. 2.8).

Tipici esempi di funzioni periodiche sono le funzioni trigonometriche  $x \mapsto \sin x$  ( $T = 2\pi$ ),  $x \mapsto \cos x$  ( $T = 2\pi$ ),  $x \mapsto \operatorname{tg} x$  ( $T = \pi$ ) sulle quali torneremo in un prossimo paragrafo.

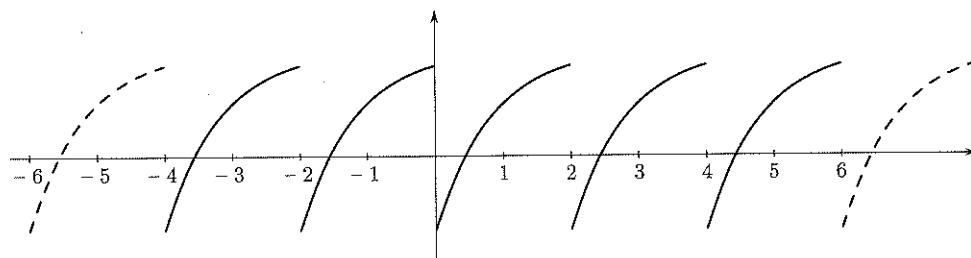


Figura 2.8. Funzione periodica di periodo 2.

### 3 FUNZIONI ELEMENTARI

Esaminiamo in questo paragrafo alcune tra le più comuni funzioni numeriche, evidenziandone le proprietà principali e segnalando uso e interpretazione in qualche contesto applicativo.

Lo studente conoscerà già, probabilmente, la maggior parte di queste funzioni. Si tenga presente che le proprietà che qui richiameremo brevemente (proprietà algebriche e loro utilizzo, grafici, ...) devono essere possedute con padronanza, per poter procedere con efficacia nello studio del calcolo infinitesimale. Lo studente che trovasse difficoltà, ad esempio, nello svolgere gli esercizi riportati alla fine di questo capitolo, è invitato a dedicare qualche tempo allo studio di questi argomenti elementari.<sup>2</sup>

#### 3.1 Funzioni potenza

Riserviamo la denominazione di *funzioni potenza* a funzioni del tipo seguente:

$$(3.1) \quad f(x) = kx^\alpha \quad (\alpha \neq 0)$$

dove  $k$  e  $\alpha$  sono numeri reali. In generale, queste funzioni sono definite solo per  $x \geq 0$ , se  $\alpha > 0$ , e per  $x > 0$ , se  $\alpha < 0$ .

La (3.1) indica che l'uscita  $f(x)$  è proporzionale, secondo la costante di proporzionalità  $k$ , a  $x^\alpha$ . Cominciamo a passare in rassegna alcuni esempi di funzioni potenza che si incontrano in applicazioni di vario tipo, quindi faremo alcune puntualizzazioni generali.

Negli esempi sottostanti usiamo lettere più appropriate alla situazione, per l'ingresso e l'uscita.

##### Moto rettilineo uniforme

La nota legge cinematica

$$\begin{array}{c} s(t) = v \cdot t \\ \downarrow \quad \downarrow \\ \text{spazio} \quad \text{tempo} \end{array}$$

esprime la proporzionalità dello spazio percorso rispetto al tempo impiegato a percorrerlo. La costante di proporzionalità è la velocità.

<sup>2</sup>Si rimanda, ad esempio, al testo: M. Bramanti: *PreCalculus*. Ed. Esculapio.

#### Capacità di un condensatore

In un comune condensatore, la quantità di carica  $Q$  presente sulle armature e la differenza di potenziale  $\Delta V$  esistente tra queste sono legate dalla formula

$$Q = C \cdot \Delta V$$

La costante di proporzionalità  $C$  è la capacità del condensatore.

Nei due esempi precedenti abbiamo dunque leggi del tipo

$$f(x) = kx \quad (\alpha = 1 \text{ nella (3.1)})$$

Il grafico al variare di  $k$  di  $f$  è ben noto:

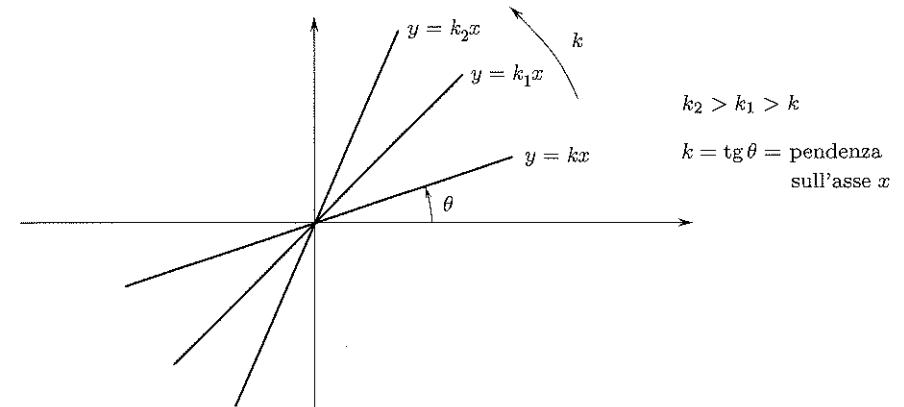


Figura 2.9.

#### Trasformazioni isotermiche

In una trasformazione a temperatura costante di una mole di un gas ideale, pressione e volume sono legati dalla legge

$$p = \frac{RT}{V}$$

Cioè,  $p$  è inversamente proporzionale a  $V$  secondo la costante  $RT$ , dove  $R = 1,986 \text{ cal/K}$  e  $T$  è la temperatura assoluta.

#### Potenziale elettrico generato da una carica puntiforme $q$

Sia  $q$  una carica elettrica concentrata in un punto  $Q$  dello spazio. Se  $P$  è un altro punto a distanza  $r$  da  $Q$ , il potenziale in  $P$  prodotto dalla carica  $q$  è dato da

$$V = h \frac{q}{r}$$

dove  $h$  è una costante che dipende dalle unità di misura usate.

Ancora,  $V$  è inversamente proporzionale alla distanza tra  $P$  e  $Q$ .

Nei due esempi precedenti, abbiamo leggi del tipo "proporzionalità inversa":

$$f(x) = \frac{k}{x} \quad (\alpha = -1 \text{ nella (3.1)})$$

Al variare di  $k$  (che qui consideriamo  $> 0$ ) si ottiene una famiglia di iperboli equilatere.

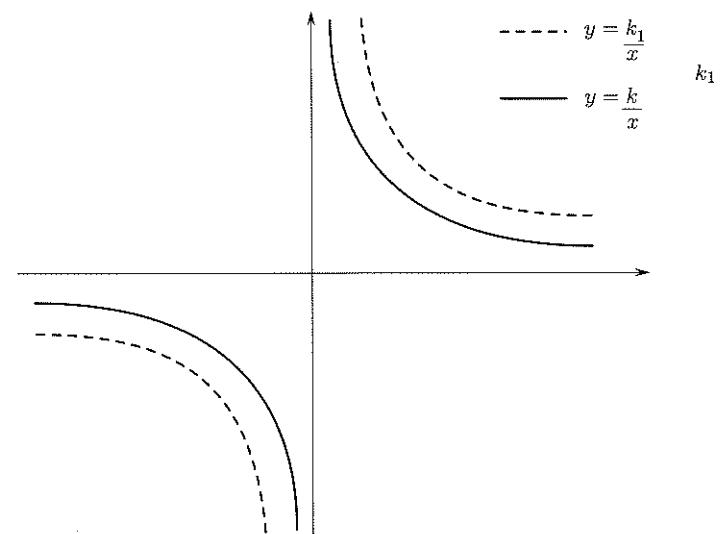


Figura 2.10.

**Superficie e volume di una sfera in funzione del raggio**

Per una sfera di raggio  $r$ , superficie e volume sono assegnati dalle formule seguenti:

$$S = 4\pi r^2 \quad V = \frac{4}{3}\pi r^3$$

Cioè, la superficie è proporzionale al quadrato del raggio, mentre il volume è proporzionale al suo cubo.

Abbiamo dunque leggi del tipo

$$f(x) = kx^n \quad \text{con } n = 2, n = 3$$

i cui grafici al variare di  $k$  sono rappresentati qui di seguito:

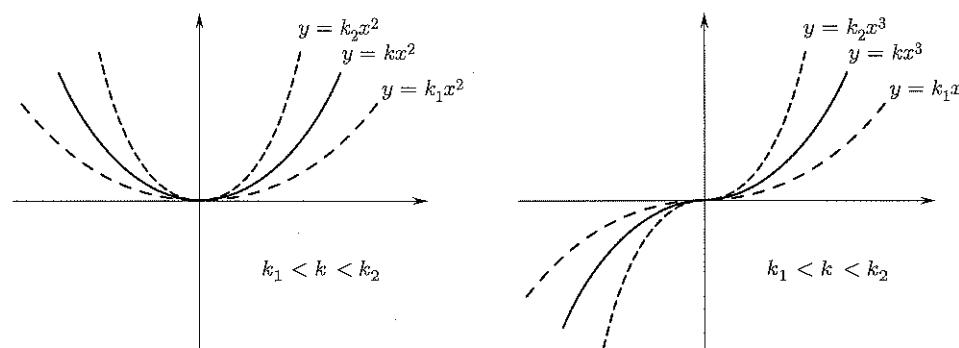


Figura 2.11.

**Moto lungo un piano inclinato**

Se una sferetta d'acciaio rotola senza attrito lungo un piano inclinato dal punto  $B$  al punto  $A$  come in figura, sotto la sola azione della gravità, giunge al punto  $A$  con

velocità data dalla formula

$$v = \sqrt{2gh} \quad (g = \text{accelerazione di gravità}).$$

Si tratta di una legge del tipo

$$f(x) = k\sqrt{x}$$

( $\alpha = \frac{1}{2}$  nella (3.1)), i cui grafici al variare di  $k$  sono illustrati sotto.

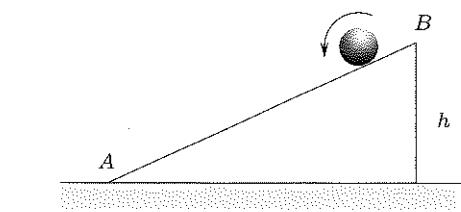


Figura 2.12.

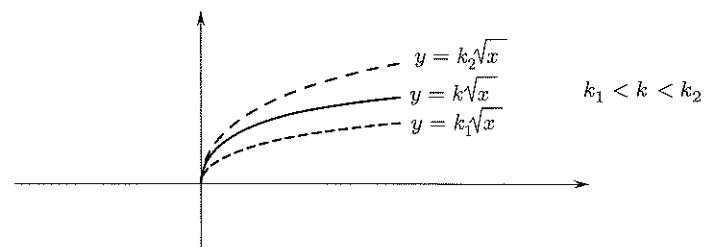


Figura 2.13.

**Trasformazioni adiabatiche**

In una trasformazione adiabatica di una mole di un gas ideale, pressione e volume sono legati dalla formula

$$p = \frac{c}{V^\gamma} \quad (c \text{ dipende da } R \text{ e } T)$$

dove  $\gamma = \frac{5}{3}$  per un gas monoatomico e  $\gamma = \frac{7}{5}$  per un gas diatomico.

Si tratta di una legge del tipo

$$f(x) = kx^{-\gamma} \quad (\alpha = -\gamma \text{ nella (3.1)})$$

Al variare di  $k$ , i grafici sono qualitativamente simili alle iperboli della figura 2.10 (limitandosi al primo quadrante se  $\gamma$  è irrazionale o se è una frazione ridotta ai minimi termini con denominatore pari).

Facciamo ora alcune osservazioni generali. Ricordiamo anzitutto che l'operazione di elevamento a potenza è ben definita per qualunque esponente se la base è positiva, ma può essere definita anche con base negativa se l'esponente è un intero oppure un razionale (frazione) con denominatore dispari.

Precisamente: la funzione  $x^{m/n}$  con  $n$  intero dispari è definita anche per  $x < 0$  ed è pari se  $m$  è un intero pari, dispari se  $m$  è un intero dispari. Ricapitoliamo i vari casi possibili dal punto di vista dei grafici di queste funzioni.

Cominciamo dalle *potenze a esponente razionale*:

$$f(x) = x^{m/n}$$

con  $m, n$  interi ridotti ai minimi termini. Le situazioni qualitativamente diverse, nel caso in cui l'esponente è positivo, sono schematizzate dagli esempi seguenti, che raccolgono tutti i casi possibili (a parte il caso banale  $f(x) = x$ ):

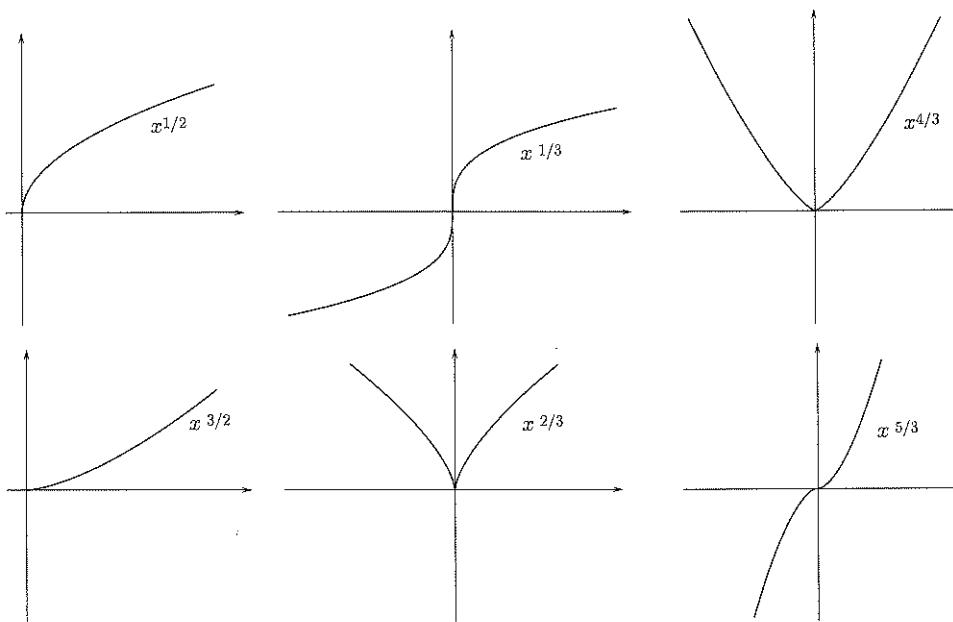


Figura 2.14.

A commento dei grafici precedenti, aggiungiamo le seguenti osservazioni, di cui lo studente è invitato a rendersi conto:

- in  $(0, +\infty)$  la funzione è strettamente crescente, mentre in  $(-\infty, 0)$ , se è definita, basta tener conto della simmetria di  $f$  per decidere se cresce o decresce;
- in  $(0, +\infty)$  la funzione presenta un andamento simile a quello delle funzioni  $x^2$  o  $\sqrt{x}$ , rispettivamente (e cioè tangente all'asse  $x$  o all'asse  $y$ , nell'origine), a seconda che l'esponente sia maggiore o minore di 1. Quest'ultima affermazione potrà essere dimostrata molto semplicemente una volta introdotto il concetto di derivata, nel capitolo 4.

Se l'esponente è negativo, le situazioni qualitativamente diverse sono schematizzate dagli esempi seguenti:

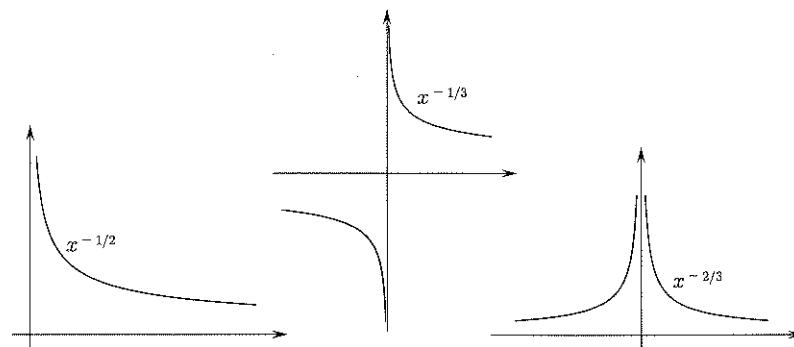


Figura 2.15.

In questi casi la funzione è strettamente decrescente in  $(0, +\infty)$ , mentre in  $(-\infty, 0)$ , se è definita, basta tener conto della simmetria di  $f$  per decidere se cresce o decresce.

Per le funzioni *potenza a esponente reale* (ma non razionale) la casistica si semplifica, perché  $f(x)$  non è definita per  $x < 0$ . Precisamente,  $f(x) = kx^\alpha$  è definita per  $x \geq 0$  se  $\alpha > 0$ , per  $x > 0$  se  $\alpha < 0$ . Le situazioni possibili sono le seguenti:

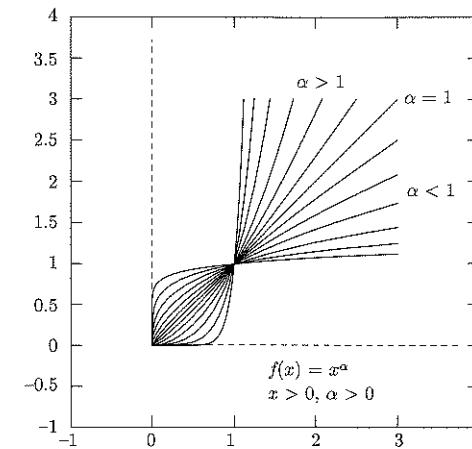


Figura 2.16. Grafico delle potenze a esponente reale.

### 3.2 Funzioni esponenziali e logaritmiche

In questo e nel prossimo paragrafo descriveremo due classi di funzioni notevoli: le funzioni esponenziali (e logaritmiche) e le funzioni circolari o trigonometriche.

L'importanza di queste due classi di funzioni deriva dal fatto che esse sono in un certo senso i prototipi utili a descrivere due corrispondenti gruppi di fenomeni frequentissimi in natura: i fenomeni di decadimento (o, al contrario, di crescita) e i fenomeni periodici. Esempi di fenomeni del primo tipo sono: il decadimento radioattivo, il processo di raffreddamento di un corpo, il diffondersi di un'infezione o il moltiplicarsi di una colonia di batteri. Esempi di fenomeni del secondo tipo sono: il moto dei pianeti, la propagazione di onde (meccaniche o elettromagnetiche), il moto di un pendolo, certi andamenti di malattie influenzali di carattere stagionale ecc. Ora, accade che spesso le funzioni esponenziali servano a descrivere i fenomeni del primo tipo mentre quelle trigonometriche siano utili a descrivere quelli del secondo tipo. Il motivo profondo di questo fatto sarà messo in luce nello studio delle equazioni differenziali di cui ci occuperemo nel secondo volume.

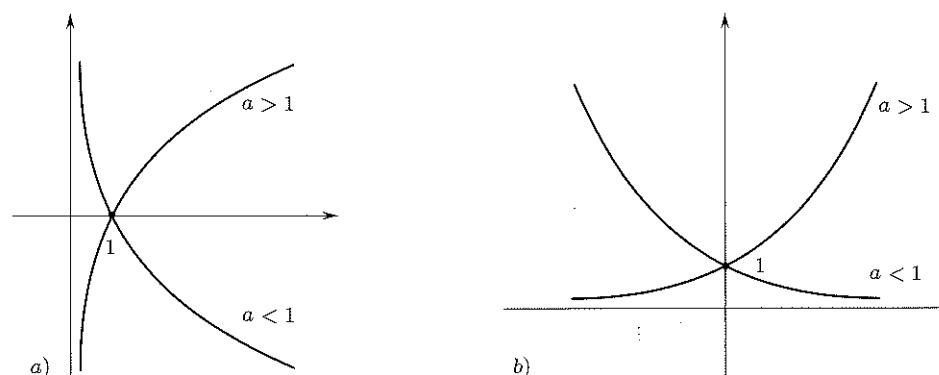
Se  $a$  è un numero reale *positivo e diverso da 1*, la funzione

$$f : (0, +\infty) \rightarrow \mathbb{R} \quad f(x) = \log_a x$$

si chiama funzione *logaritmo in base a*, mentre la funzione

$$g : \mathbb{R} \rightarrow \mathbb{R} \quad g(x) = a^x$$

si chiama funzione *esponenziale in base a*.

Figura 2.17. a) Grafico di  $f(x) = \log_a x$ ; b) grafico di  $g(x) = a^x$ .

Le funzioni  $f$  e  $g$  sono legate dalla relazione fondamentale

$$(3.2) \quad x = a^{\log_a x} \quad (x > 0, a > 0, a \neq 1)$$

ossia,

$$(3.3) \quad y = \log_a x \quad \text{equivale a} \quad x = a^y$$

Confrontando invece le funzioni potenza (introdotte nel paragrafo 3.1) con le funzioni esponenziali, notiamo che la stessa operazione algebrica di elevamento a potenza nel campo reale,  $a^b$ , è alla base della definizione di queste due classi di funzioni:

- se l'esponente  $b$  è fissato e la base è variabile abbiamo le funzioni potenza:

$$x \mapsto x^b$$

- se la base  $a$  è fissata e l'esponente è variabile abbiamo le funzioni esponenziali:

$$x \mapsto a^x$$

#### La base naturale

Tra le funzioni esponenziali e logaritmiche, si incontrano con particolare frequenza quelle la cui base è *il numero e di Nepero*. Si tratta di una costante molto importante in matematica, che sarà definita e trattata ampiamente nel prossimo capitolo 3. Per ora è sufficiente sapere che  $e$  vale circa 2,7, perciò le funzioni  $e^x$  e  $\log_e x$  (che scriviamo semplicemente  $\log x$  sottointendendo la base  $e$ ) hanno i grafici caratteristici delle funzioni esponenziali e logaritmiche di base maggiore di 1.

Notiamo che una funzione esponenziale  $a^x$  (di base qualsiasi, positiva e diversa da 1) si può sempre esprimere nella forma  $e^{bx}$ , scegliendo  $b = \log a$ .

**DEFINIZIONE** Se  $[H^+]$  indica la concentrazione di ioni idrogeno (= numero di moli/cm<sup>3</sup>) in una soluzione, si definisce come misura della sua acidità la quantità

$$\text{pH} = -\log_{10}[H^+]$$

Poiché per l'acqua pura  $[H^+] = 10^{-7}$ , il corrispondente pH è pari a  $-\log_{10} 10^{-7} = 7$ , equivalente a soluzione neutra. Se  $\text{pH} < 7$  la soluzione è acida, se  $\text{pH} > 7$  la soluzione è basica.

**■ Scarica di un condensatore.** In un circuito con capacità e resistenza, come quello indicato in figura 2.18a, la carica  $q(t)$  è assegnata dalla formula

$$(3.4) \quad q(t) = Qe^{-t/\tau}$$

dove  $Q$  è la carica iniziale (al tempo  $t = 0$ ) e  $\tau = RC$ .

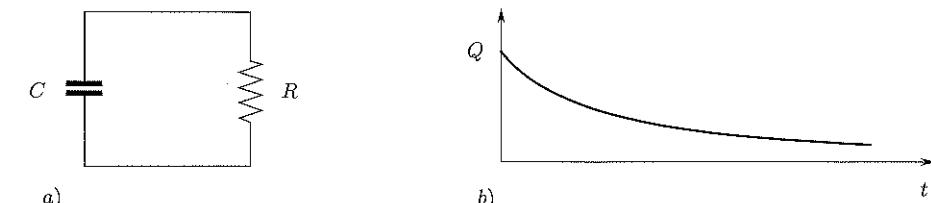


Figura 2.18. La carica decresce secondo la legge esponenziale (3.4) come indicato in b).

La funzione in (3.4) si può scrivere nella forma  $q(t) = Qa^t$  con  $a = e^{-1/\tau} < 1$ .

Nella tabella seguente sono raccolte le principali proprietà delle funzioni in oggetto. Le altre proprietà algebriche dei logaritmi sono state richiamate nel capitolo 1.

$f(x) = \log_a x$	$g(x) = a^x$
dominio: $(0, +\infty)$ ; immagine: $\mathbb{R}$ strettamente $\begin{cases} \text{crescente} & \text{se } a > 1 \\ \text{decrescente} & \text{se } a < 1 \end{cases}$ $\log_a x \cdot y = \log_a x + \log_a y$	dominio: $\mathbb{R}$ ; immagine: $(0, +\infty)$ strettamente $\begin{cases} \text{crescente} & \text{se } a > 1 \\ \text{decrescente} & \text{se } a < 1 \end{cases}$ $a^{x+y} = a^x \cdot a^y$

### 3.3 Funzioni trigonometriche

Le funzioni trigonometriche elementari sono:

$$\begin{array}{llll} x \mapsto \cos x & x \mapsto \sin x & x \mapsto \tan x & x \mapsto \cot x \\ (\text{coseno di } x) & (\text{seno di } x) & (\text{tangente di } x) & (\text{cotangente di } x) \end{array}$$

La variabile  $x$  ha il significato di *misura in radianti* di un angolo. I significati geometrici delle quattro funzioni sono illustrati nella figura 2.19.

Dalla figura si vede che  $\cos x$  = ascissa di  $P$ ,  $\sin x$  = ordinata di  $P$ ,  $\tan x$  = lunghezza con segno del segmento  $AT$ ,  $\cot x$  = lunghezza con segno del segmento  $BS$ .

Valgono le relazioni fondamentali seguenti

$$(\sin x)^2 + (\cos x)^2 = 1 \quad \tan x = \frac{\sin x}{\cos x} \quad \cot x = \frac{\cos x}{\sin x} = \frac{1}{\tan x}$$

La prima si ricava dal fatto che  $P$  si trova sulla circonferenza di equazione  $u^2 + v^2 = 1$ ; la seconda segue dalla similitudine dei triangoli  $OPQ$  e  $OTA$ ; la terza segue dalla similitudine dei triangoli  $OPQ$  e  $OBS$ .

Nella seguente tabella sono raggruppate le principali proprietà delle funzioni seno, coseno, tangente.

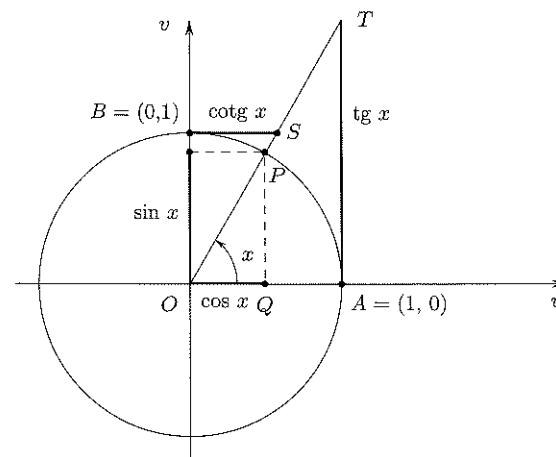


Figura 2.19. Definizione di seno, coseno, tangente e cotangente.

$x \mapsto \sin x, x \mapsto \cos x$	$x \mapsto \tan x$	$x \mapsto \cot x$
dominio: $\mathbb{R}$ ; immagine: $[-1, 1]$	dominio: $\mathbb{R} \setminus \{\frac{\pi}{2} + k\pi\}$ ; immagine: $\mathbb{R}$	dominio: $\mathbb{R} \setminus \{k\pi\}$ ; immagine: $\mathbb{R}$
periodiche di periodo $T = 2\pi$	periodica di periodo $T = \pi$	periodica di periodo $T = \pi$
$\sin(-x) = -\sin x$ (funzione dispari)	$\tan(-x) = -\tan x$ (funzione dispari)	$\cot(-x) = -\cot x$ (funzione dispari)
$\cos(-x) = \cos x$ (funzione pari)		

Poiché  $\sin x = \cos(x - \frac{\pi}{2})$  ricaviamo che  $\sin x$  si ottiene da  $\cos x$  con uno *sfasamento* di  $x$  pari a  $\frac{\pi}{2}$ , ossia, il grafico di  $\sin x$  si ottiene da quello di  $\cos x$  traslandolo a destra di  $\frac{\pi}{2}$  (v. fig. 2.20).

### 3.4 Fenomeni vibratori

Abbiamo visto che le funzioni seno e coseno sono periodiche di periodo  $2\pi$ . Più in generale le funzioni

$$(3.5) \quad t \mapsto a \sin \omega t \quad t \mapsto b \cos \omega t$$

dove  $a, b, \omega$  sono numeri reali positivi, sono periodiche di periodo  $T = \frac{2\pi}{\omega}$ . Infatti, per esempio

$$\sin \left[ \omega \left( t + \frac{2\pi}{\omega} \right) \right] = \sin(\omega t + 2\pi) = \sin \omega t$$

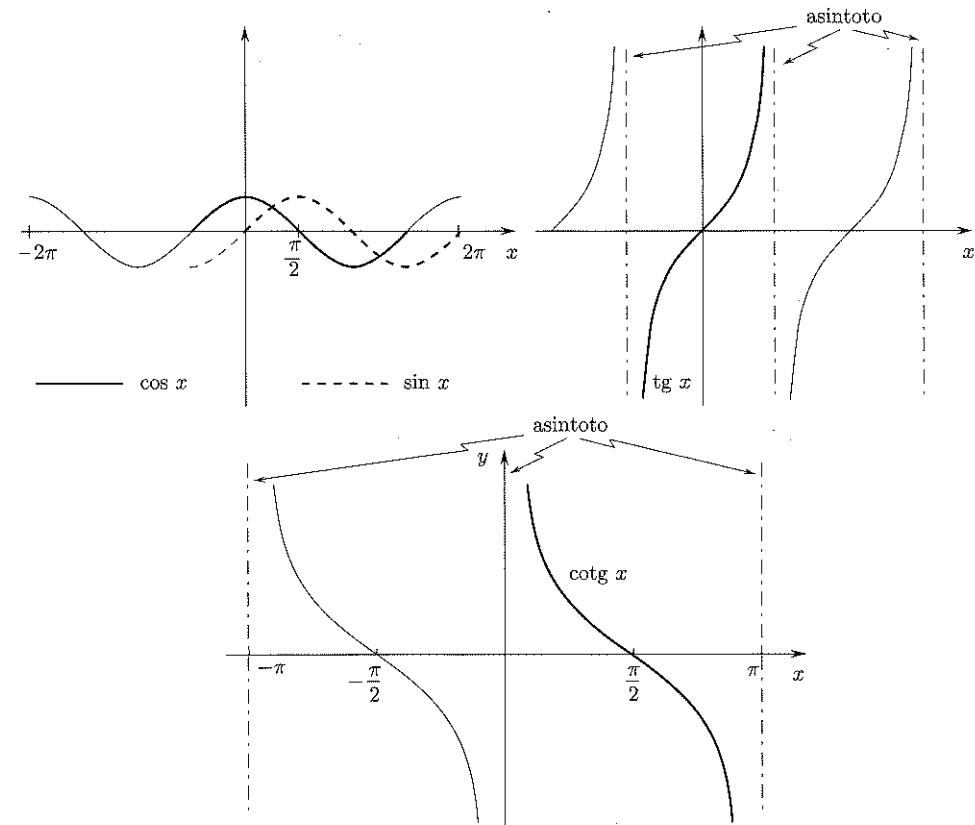
Inoltre, essendo  $|\sin \omega t| \leq 1$ ,  $|\cos \omega t| \leq 1$  si ha

$$|a \sin \omega t| \leq a, \quad |b \cos \omega t| \leq b$$

Le funzioni (3.5) descrivono vibrazioni elementari di *ampiezza*  $a$  e  $b$ , rispettivamente, di *pulsazione*  $\omega = \frac{2\pi}{T}$  e *frequenza*  $\nu = \frac{1}{T}$ .

La funzione

$$h(t) = a \sin \omega t + b \cos \omega t$$

Figura 2.20. Grafici di  $y = \cos x$ ,  $y = \sin x$ ,  $y = \tan x$  e  $y = \cot x$ .

descrive la sovrapposizione delle due vibrazioni elementari di uguale pulsazione  $\omega$ . Quest'ultima è ancora una vibrazione elementare *sfasata* rispetto alle precedenti. Infatti, posto  $A = \sqrt{a^2 + b^2}$ , si può scrivere

$$(3.6) \quad h(t) = A \frac{a}{\sqrt{a^2 + b^2}} \sin \omega t + A \frac{b}{\sqrt{a^2 + b^2}} \cos \omega t$$

Osserviamo ora che i numeri  $\alpha = a/\sqrt{a^2 + b^2}$  e  $\beta = b/\sqrt{a^2 + b^2}$  soddisfano le condizioni

$$-1 \leq \alpha \leq 1 \quad -1 \leq \beta \leq 1 \quad \alpha^2 + \beta^2 = 1$$

Esiste quindi un unico angolo  $\varphi$  tale che

$$\cos \varphi = \alpha, \quad \sin \varphi = \beta$$

cosicché la (3.6) si può riscrivere nella forma seguente

$$(3.7) \quad h(t) = A \cos \varphi \sin \omega t + A \sin \varphi \cos \omega t = A \sin(\omega t + \varphi)$$

Dunque  $h(t)$  rappresenta una vibrazione elementare di ampiezza  $A$ , pulsazione  $\omega = \frac{2\pi}{T}$ , sfasata di un angolo  $\varphi$ .

In sintesi

$$\begin{aligned} a \sin \omega t + b \cos \omega t &= A \sin(\omega t + \varphi) \\ A &= \sqrt{a^2 + b^2} \quad \begin{cases} a = A \cos \varphi \\ b = A \sin \varphi \end{cases} \end{aligned}$$

Sotto condizioni abbastanza generali, un fenomeno naturale periodico si potrà scrivere come sovrapposizione di un numero finito o infinito di vibrazioni elementari di frequenza diversa; ciò conduce al concetto di serie di Fourier, ovvero a somme del tipo

$$\sum a_n \cos n\omega t + b_n \sin n\omega t$$

che tratteremo nel secondo volume.

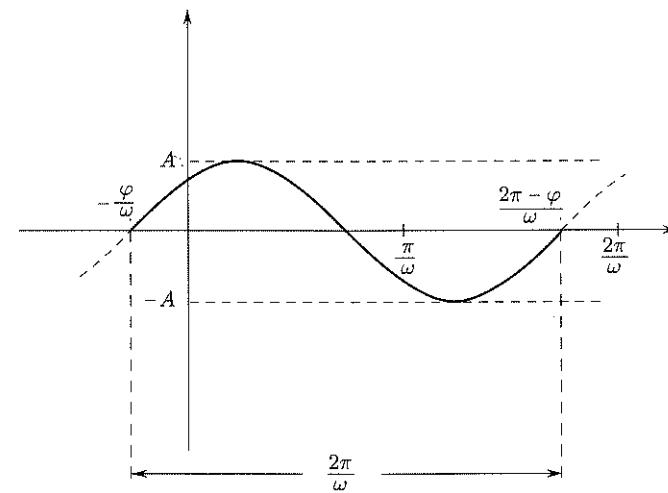


Figura 2.21. La vibrazione  $h(t) = A \sin(\omega t + \varphi)$ .

Moltiplicando una vibrazione elementare per potenze o esponenziali si possono modellizzare effetti di smorzamento o di amplificazione.

Per esempio, consideriamo la funzione

$$h(t) = t \sin \omega t \quad (t \geq 0)$$

essendo  $-1 \leq \sin \omega t \leq 1$ , si ha

$$-t \leq t \sin \omega t \leq t$$

e quindi il grafico di  $h$  si trova tra i grafici delle rette di equazione  $y = -t$ ,  $y = t$ . Nei punti in cui  $\sin \omega t = 1$ , cioè  $t = \frac{\pi}{2\omega} + k \frac{2\pi}{\omega}$  ( $k = 0, 1, 2, \dots$ ), il grafico di  $h$  tocca quello di  $y = t$ , mentre nei punti in cui  $\sin \omega t = -1$ , cioè  $t = \frac{3\pi}{2\omega} + k \frac{2\pi}{\omega}$ , il grafico di  $h$  tocca quello di  $y = -t$ .

Il grafico di  $h$  è perciò il seguente:

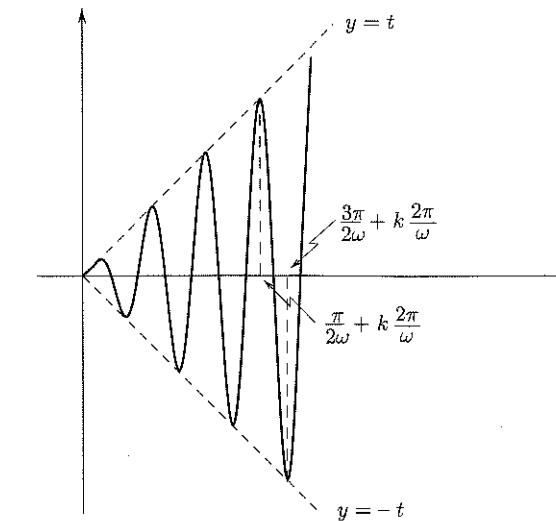


Figura 2.22. Grafico di  $h(t) = t \sin \omega t$  (oscillazioni amplificate).

Dal grafico si vede come la moltiplicazione per  $t$  abbia l'effetto di una amplificazione della vibrazione, all'aumentare di  $t$ .

Analogamente, la funzione

$$k(t) = e^{-\alpha t} \sin \omega t \quad (\alpha > 0)$$

modellizza una *vibrazione smorzata*.

Ricordando che  $e^{-\alpha t} = \left(\frac{1}{e^\alpha}\right)^t$  è un'esponenziale con base minore di 1, considerazioni analoghe a quelle svolte per la funzione  $h$ , indicano che il grafico di  $k$  è compreso tra i grafici delle funzioni  $y_1 = e^{-\alpha t}$  e  $y_2 = -e^{-\alpha t}$ , come in figura:

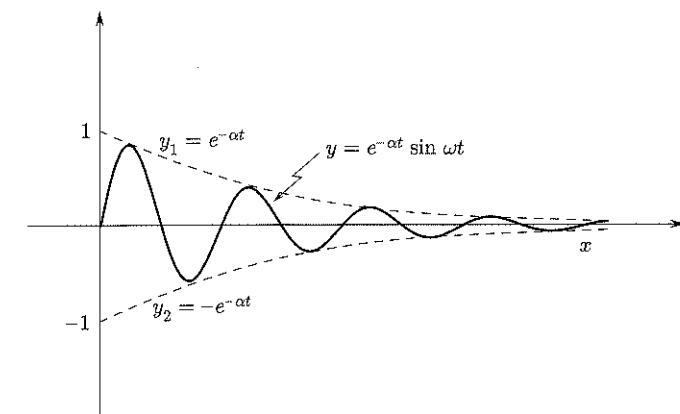


Figura 2.23. Grafico di  $k(t) = e^{-\alpha t} \sin \omega t$  (oscillazioni smorzate).

### 3.5 Funzioni parte intera e mantissa

Due funzioni che tipicamente si incontrano nella scrittura di algoritmi sono la funzione *parte intera* e la funzione *mantissa* o *parte decimale*. La funzione “parte intera di  $x$ ”, talvolta indicata con  $[x]$ , è definita da:

$$[x] = \text{quell'intero } n \text{ tale che } n \leq x < n + 1.$$

Ad esempio,

$$[2,38] = 2;$$

$$[3] = 3;$$

$$[-1,8] = -2.$$

Mentre per i numeri positivi, quindi, la parte intera si ottiene semplicemente “buttando via le cifre dopo la virgola”, per i numeri negativi occorre prendere il massimo intero  $\leq x$ , che è diverso da quello che si ottiene buttando via le cifre dopo la virgola (tranne nel caso in cui  $x$  sia già un intero). Il grafico della funzione parte intera è:

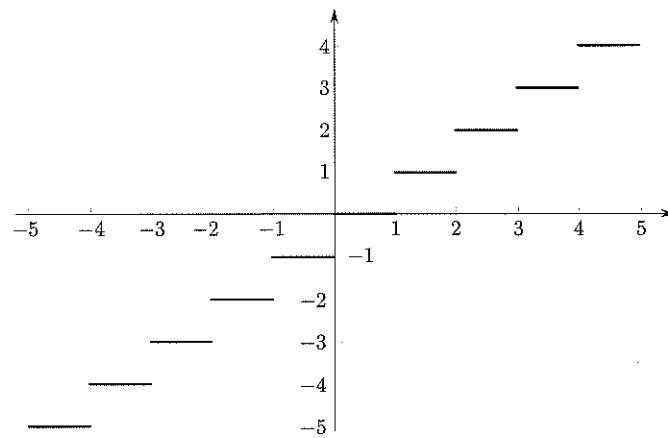


Figura 2.24. Grafico della funzione parte intera.

La mantissa, o parte decimale di  $x$ , indicata con  $(x)$  oppure  $\text{mant}(x)$ , è definita da:

$$(x) = x - [x].$$

Ad esempio, si ha:

$$(2,38) = 0,38;$$

$$(3) = 0;$$

$$(-1,8) = 0,2.$$

La mantissa quindi non è un intero ma un numero reale, compreso in  $[0, 1)$ . Per i numeri positivi, si ottiene semplicemente “buttando via le cifre prima della virgola”, per i numeri negativi la mantissa è il complemento a uno del numero che si ottiene buttando via le cifre prima della virgola.

Il grafico della funzione mantissa è:

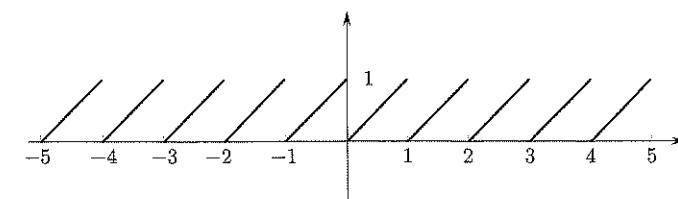


Figura 2.25. Grafico della funzione mantissa.

Si noti che la mantissa è una funzione periodica di periodo 1; si tratta di una semplice funzione periodica di tipo diverso da quelle trigonometriche.

### 3.6 Funzioni iperboliche

Nelle applicazioni della matematica sono importanti certe combinazioni delle funzioni  $e^x$  ed  $e^{-x}$  che ora definiamo:

$$\begin{aligned} \text{Sh : seno iperbolico} \quad \text{Sh } x &:= \frac{e^x - e^{-x}}{2} \\ \text{Ch : coseno iperbolico} \quad \text{Ch } x &:= \frac{e^x + e^{-x}}{2} \\ \text{Th : tangente iperbolica} \quad \text{Th } x &:= \frac{\text{Sh } x}{\text{Ch } x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \end{aligned}$$

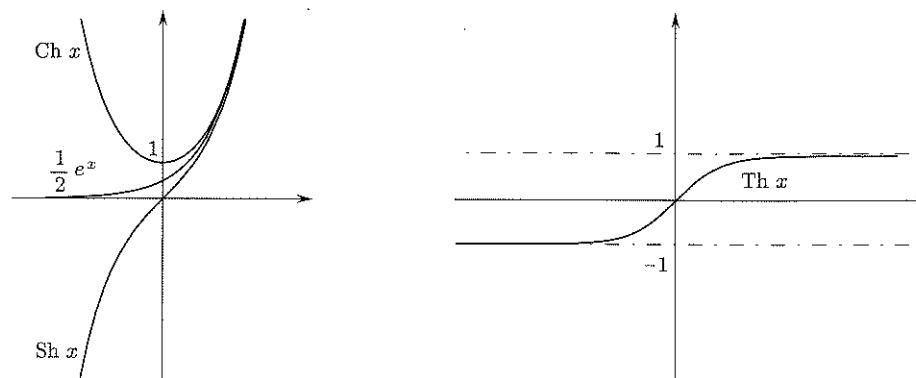
Esse sono evidentemente definite su tutto  $\mathbb{R}$ . Le seguenti proprietà discendono immediatamente dalla definizione:

$$\begin{aligned} 1) \quad \text{Sh}(-x) &= -\text{Sh } x && (\text{funzione dispari}) \\ \text{Ch}(-x) &= \text{Ch } x && (\text{funzione pari}) \\ \text{Th}(-x) &= -\text{Th } x && (\text{funzione dispari}) \end{aligned}$$

$$\begin{aligned} 2) \quad \text{Sh}(0) &= 0 & \text{Ch}(0) &= 1 & \text{Th}(0) &= 0 \\ \text{Sh } x &\leq \frac{1}{2}e^x \leq \text{Ch } x & \forall x \in \mathbb{R} \end{aligned}$$

$$\begin{aligned} 3) \quad (\text{Ch } x)^2 - (\text{Sh } x)^2 &= 1 \\ \text{Sh}(x+y) &= \text{Sh } x \text{Ch } y + \text{Sh } y \text{Ch } x \\ \text{Ch}(x+y) &= \text{Ch } x \text{Ch } y + \text{Sh } y \text{Sh } x \\ \text{Th}(x+y) &= \frac{\text{Th } x + \text{Th } y}{1 + \text{Th } x \text{Th } y} \\ \text{Sh}(2x) &= 2 \text{Sh } x \text{Ch } x \\ \text{Ch}(2x) &= (\text{Ch } x)^2 + (\text{Sh } x)^2 \end{aligned}$$

I grafici di queste funzioni sono illustrati nella figura 2.26.

Figura 2.26. Grafici di  $\text{Sh } x = \frac{e^x - e^{-x}}{2}$ ,  $\text{Ch } x = \frac{e^x + e^{-x}}{2}$ ,  $\text{Th } x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ .

La funzione  $\text{Ch } x$  compare nella soluzione di un semplice problema fisico: trovare la sagoma lungo la quale si dispone un filo pesante, omogeneo, fissato per le due estremità. In queste ipotesi il filo descriverà, in un piano verticale, il grafico della funzione

$$f(x) = \text{Ch } x$$

(a patto di scegliere opportunamente il sistema di riferimento e le unità di misura sui due assi). Per questo motivo la curva grafico di  $\text{Ch } x$  è anche detta *catenaria*.

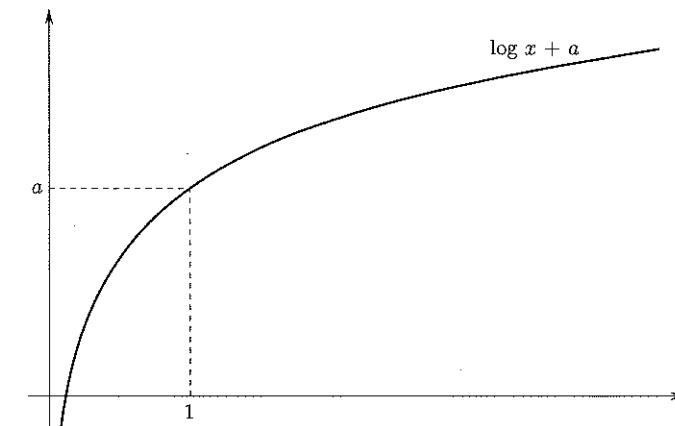
### 3.7 Operazioni sui grafici

Conoscendo il grafico di una funzione  $y = f(x)$ , mediante semplici trasformazioni geometriche è possibile disegnare il grafico delle seguenti funzioni:

$$\begin{aligned} y_1 &= f(x) + a, \quad a \in \mathbb{R}; \\ y_2 &= f(x + a), \quad a \in \mathbb{R}; \\ y_3 &= k \cdot f(x), \quad k \in \mathbb{R}; \\ y_4 &= f(kx), \quad k \in \mathbb{R}; \\ y_5 &= |f(x)|; \\ y_6 &= f(|x|). \end{aligned}$$

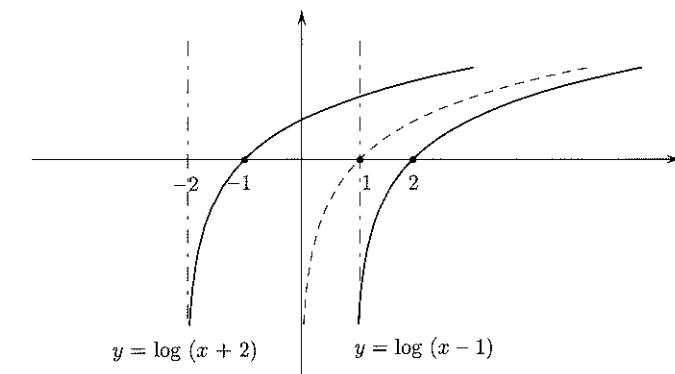
Inoltre, più operazioni di questi tipi possono essere effettuate in sequenza. Pertanto, a partire dalle funzioni elementari precedentemente introdotte, è possibile costruire mediante queste operazioni i grafici di una grande varietà di nuove funzioni. Illustriamo su esempi il significato e l'azione di queste trasformazioni.

- Consideriamo per esempio  $f(x) = \log x$ . Allora  $y_1 = f(x) + a = (\log x) + a$ . La funzione  $y_1$  è definita per  $x > 0$ , come  $\log x$ , e il grafico di  $y_1$  si ottiene da quello di  $y$  con una traslazione di  $a$  unità verso l'alto se  $a > 0$ , verso il basso se  $a < 0$ . Infatti, la quantità  $a$  viene aggiunta al valore della funzione  $f(x)$  (non alla variabile  $x$ ), dopo che questa è stata calcolata.

Figura 2.27. Grafico di  $\log x + a$ ,  $a > 0$ .

- Consideriamo ancora  $f(x) = \log x$ . Allora  $y_2 = f(x + a) = \log(x + a)$ . Essendo  $\log x$  definito per  $x > 0$ ,  $\log(x + a)$  sarà definito per  $x + a > 0$  e cioè per  $x > -a$ . Poiché  $\log x = 0$  se  $x = 1$ , si ha  $\log(x + a) = 0$  per  $x + a = 1$  e cioè  $x = 1 - a$ .

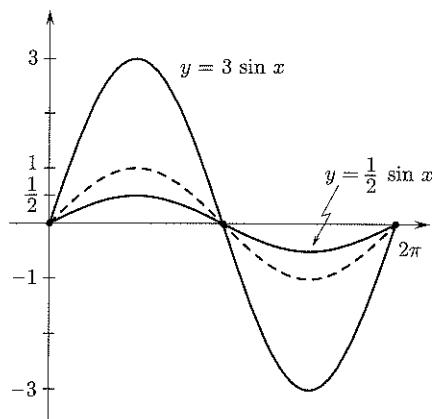
In altri termini il grafico di  $y_2$  si ottiene da quello di  $y$  con una traslazione di  $a$  unità a sinistra se  $a > 0$ , a destra se  $a < 0$ .

Figura 2.28. Grafici di  $y = \log(x - 1)$  e  $y = \log(x + 2)$ .

- Il grafico di  $y_3 = kf(x)$  si ottiene da quello di  $f$  moltiplicando per  $k$  tutte le ordinate  $f(x)$ . In particolare se  $k = -1$  le ordinate sono semplicemente cambiate di segno, cosicché il grafico di  $y_2$  è simmetrico, rispetto all'asse  $x$ , a quello di  $f$ .

Per esempio, sia  $f(x) = \sin x$ . I grafici di  $y = 3 \sin x$ ,  $y = \frac{1}{2} \sin x$ ,  $y = -\sin x$  sono riportati in figura 2.29.

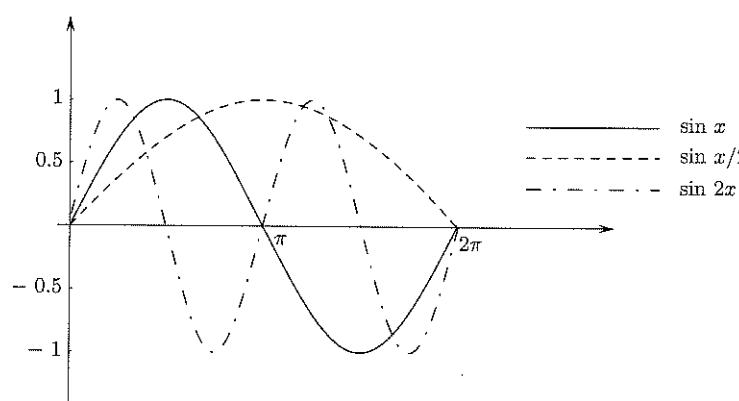
Osserviamo che se  $k > 1$ , il grafico si "stira" nella direzione verticale, dilatando verso l'alto le ordinate positive e verso il basso quelle negative. Al contrario, se  $0 < k < 1$  il grafico si contrae, sempre in direzione verticale. Perciò l'operazione di

Figura 2.29. Grafici di  $y = 3 \sin x$ ,  $y = \frac{1}{2} \sin x$  e  $y = -\sin x$ .

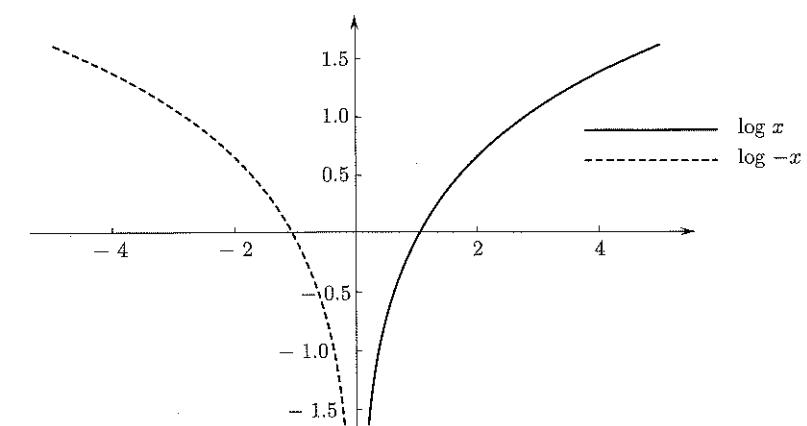
moltiplicazione di  $f(x)$  per  $k$  ha il significato geometrico di *dilatazione* (se  $|k| > 1$ ) o *contrazione* (se  $|k| < 1$ ) sull'asse delle  $y$ , eventualmente accompagnata da una *riflessione rispetto all'asse  $x$* , se  $k < 0$ .

- Il grafico di  $y_4 = f(kx)$  si ottiene da quello di  $f(x)$  con un cambiamento di scala sull'asse  $x$ . Se  $k > 1$ ,  $kx$  cresce più rapidamente di  $x$  e perciò il grafico di  $y_4$  sarà simile a quello di  $f$  ma con oscillazioni più rapide, ovvero sarà “compresso” in direzione orizzontale, di un fattore  $1/k$ . Analogamente se  $0 < k < 1$  il grafico apparirà “dilatato” in direzione orizzontale, con oscillazioni più dolci. Se  $k < 0$ , oltre ad una compressione (se  $|k| > 1$ ) o dilatazione (se  $|k| < 1$ ) sull'asse delle  $x$ , ci sarà una riflessione rispetto all'asse delle  $y$ .

Ad esempio, si confrontino i grafici di  $\sin x$ ,  $\sin 2x$ ,  $\sin \frac{x}{2}$  su  $[0, 2\pi]$ :

Figura 2.30. Grafici di  $y = \sin x$ ,  $y = \sin \frac{x}{2}$  e  $y = \sin 2x$ .

Per quanto riguarda la riflessione per  $k < 0$ , si confrontino anche i grafici di  $\log x$  e  $\log(-x)$ :

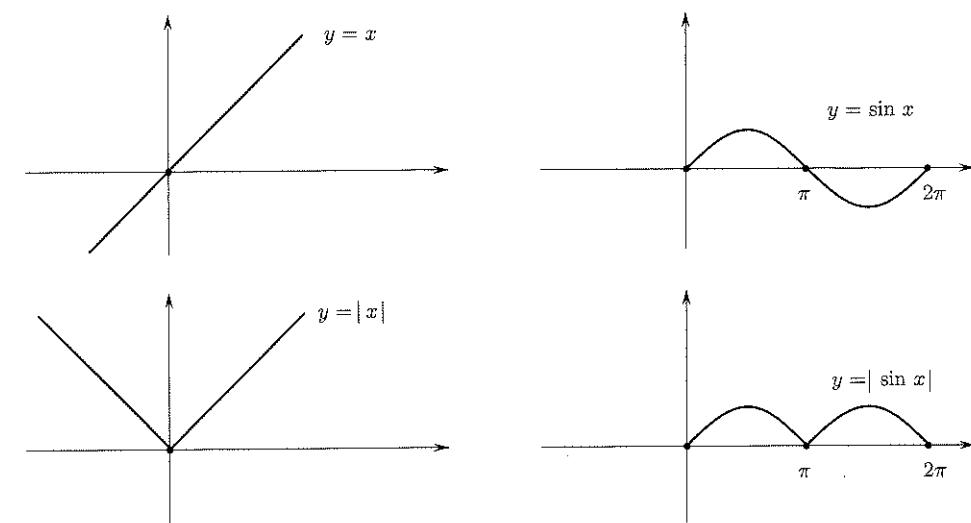
Figura 2.31. Grafici di  $y = \log x$  e  $y = \log(-x)$ .

- Per disegnare il grafico di  $y_5 = |f(x)|$  ricordiamo che

$$|f(x)| = \begin{cases} f(x) & \text{se } f(x) \geq 0 \\ -f(x) & \text{se } f(x) < 0 \end{cases}$$

Dunque nel passare dal grafico di  $f$  a quello di  $|f|$  i punti a ordinata non negativa rimangono inalterati mentre quelli a ordinata negativa vengono trasformati nei loro simmetrici rispetto all'asse  $x$ .

Il grafico di  $|f|$  si ottiene perciò da quello di  $f$  “ribaltando” simmetricamente rispetto all'asse delle ascisse la parte del grafico di  $f$  che si trova nel semipiano inferiore e lasciando inalterato il resto. Per esempio, per le funzioni  $y = x$  e  $y = \sin x$  si ha:

Figura 2.32. Dal grafico di  $f$  a quello di  $|f|$ .

- Infine, per tracciare il grafico di  $y_6 = f(|x|)$ , osserviamo che:  $|x| = x$  per  $x \geq 0$ , e quindi nel semipiano destro i due grafici coincidono;  $|-x| = |x|$ , e quindi  $y_6$  è una funzione pari, perciò simmetrica rispetto all'asse delle ordinate.

Di conseguenza il grafico di  $y_6 = f(|x|)$  verrà tracciato lasciando inalterato il grafico di  $f$  nel semipiano destro e ribaltandolo simmetricamente rispetto all'asse delle ordinate. Per esempio, se  $y = e^x$  si ha:

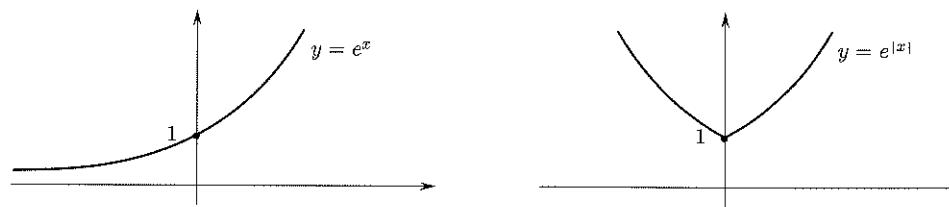


Figura 2.33. Dal grafico di  $y = e^x$  a quello di  $y = e^{|x|}$ .

### 3.8 Funzioni definite a tratti

Abbiamo visto come traslazioni, dilatazioni e riflessioni di vario tipo consentano di costruire funzioni più complesse a partire da quelle elementari. Un altro modo tipico con cui a partire dalle funzioni elementari se ne costruiscono di nuove, è quello di usare definizioni analitiche diverse su intervalli diversi. Si consideri il prossimo:

**ESEMPIO**

Sia

$$f(x) = \begin{cases} \log x & \text{se } x > 1 \\ x^2 & \text{se } 0 < x \leq 1 \\ x & \text{se } x \leq 0 \end{cases}$$

Un attimo di riflessione mostra che la funzione  $f$  è definita su tutto  $\mathbb{R}$  e ha il grafico seguente:

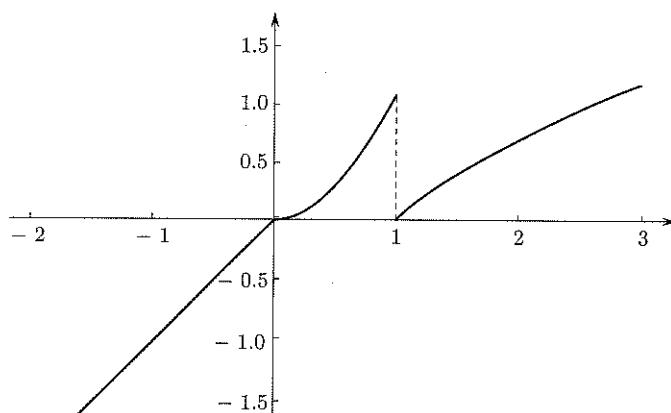


Figura 2.34.

Il valore  $f(x)$  è calcolato mediante “istruzioni” diverse a seconda dell’intervallo in cui cade la  $x$ . Una funzione di questo tipo si chiama solitamente “funzione definita a tratti”. Dal punto di vista logico / algoritmico, si può dire che una funzione di questo tipo ha la particolarità di essere costruita utilizzando, come “ingredienti”, non solo le funzioni matematiche elementari discusse fin qui, ma anche la *funzione logica* “se... allora”.<sup>3</sup>

Si presti attenzione all’insieme di definizione di una funzione definita a tratti: la funzione  $f(x)$  dell’esempio è definita anche per  $x < 0$  nonostante la presenza di  $\log x$ . Infatti, l’istruzione “calcola  $\log x$ ” interviene solo se  $x > 1$ .

## 4 FUNZIONI COMPOSTE E INVERSE

### 4.1 Funzioni composte

Nel paragrafo 3.7 abbiamo illustrato come si può costruire il grafico di  $|f|$  a partire da quello di  $f$ . La funzione  $x \mapsto |f(x)|$  è in realtà “composta” di due funzioni: dato  $x$ , si calcola  $f(x)$ ; calcolato  $f(x)$ , si calcola  $|f(x)|$ .

Si tratta di operare in serie con due scatole nere, la prima corrispondente a  $f$ , la seconda al suo valore assoluto, secondo lo schema seguente:

$$x \mapsto \boxed{f} \mapsto f(x) \mapsto \boxed{|\cdot|} \mapsto |f(x)|$$

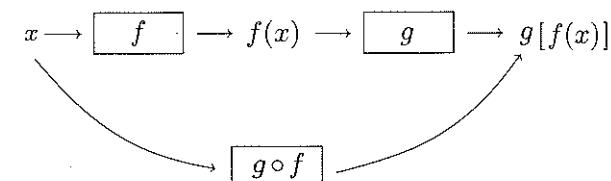
In generale, date due funzioni

$$f : E \rightarrow \mathbb{R} \quad g : F \rightarrow \mathbb{R}$$

se  $f(E) \subseteq F$  (cioè se per ogni  $x \in E$  si ha che  $f(x) \in F$ ) si può definire la funzione  $h : E \rightarrow \mathbb{R}$  composta di  $f$  e  $g$  (nell’ordine), denotata col simbolo  $g \circ f$ , mediante la formula

$$(4.1) \quad h(x) = (g \circ f)(x) = g[f(x)]$$

ossia



Può accadere che risultino ben definite sia la composizione  $g \circ f$  che  $f \circ g$ ; ma in generale sarà

$$f \circ g \neq g \circ f$$

<sup>3</sup>Per implementare la funzione dell’esempio in qualche linguaggio di programmazione, occorrerebbe annidare due funzioni “se... allora” una dentro l’altra, secondo uno schema del tipo:  
“se  $x > 1$  allora  $\log x$ , altrimenti (se  $x > 0$  allora  $x^2$ , altrimenti  $x$ )”

(il prodotto di composizione non è commutativo!) L'operazione di composizione si può estendere a tre o più fattori. Si verifica che se la composizione  $(f \circ g) \circ r$  esiste, allora esiste anche  $f \circ (g \circ r)$  e sono uguali (proprietà associativa)

$$(f \circ g) \circ r = f \circ (g \circ r)$$



$f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^2$  e  $g : \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(x) = \cos x$ .

Poiché  $g$  è definita su tutto  $\mathbb{R}$ ,  $h = g \circ f$  è ben definita su  $\mathbb{R}$  e vale la formula

$$h(x) = g[f(x)] = \cos(x^2)$$

ossia

$$x \mapsto \boxed{(\cdot)^2} \mapsto x^2 \mapsto \boxed{\cos(\cdot)} \mapsto \cos(x^2)$$

Si noti che anche  $k = f \circ g$  è ben definita ed è data dalla formula

$$k(x) = (\cos x)^2$$

Si osservi che queste due funzioni sono ben diverse tra loro: ad esempio,  $(\cos x)^2$  è periodica e mai negativa,  $\cos(x^2)$  non è periodica e ha segno oscillante. Si vedano i prossimi grafici:

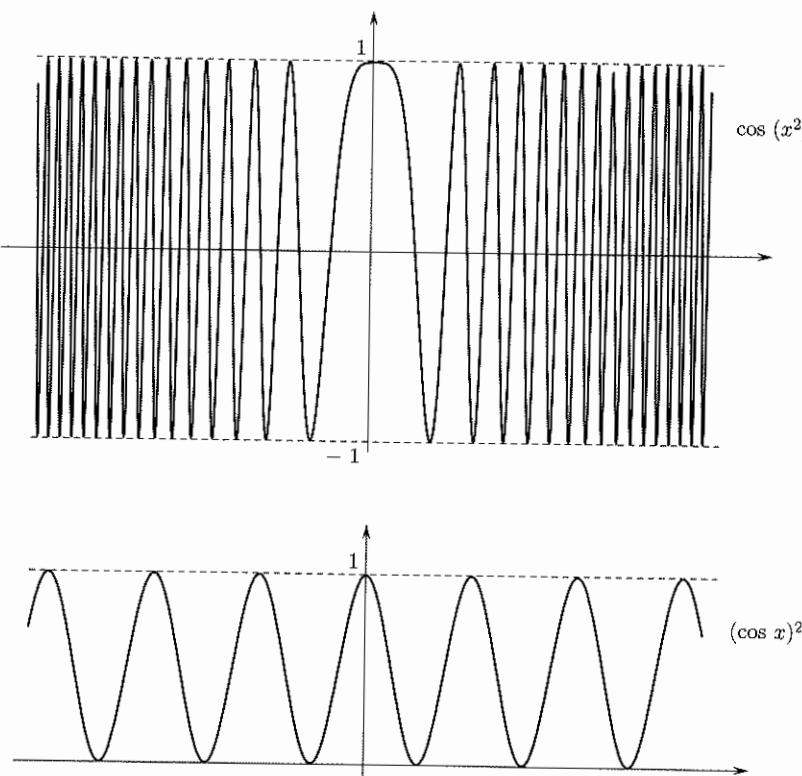


Figura 2.35. Grafico di  $y = \cos(x^2)$  e di  $(\cos x)^2$ .

$f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = e^x$  e  $g : [0, +\infty) \rightarrow \mathbb{R}$ ,  $g(x) = \sqrt{x}$ .

Poiché  $f(x) > 0$  per ogni  $x$  in  $\mathbb{R}$ ,  $h = g \circ f$  è ben definita in  $\mathbb{R}$  e si ha  $h(x) = \sqrt{e^x}$ .

Anche  $k = f \circ g$  è ben definita in  $\mathbb{R}_+$  e si ha  $k(x) = e^{\sqrt{x}}$ .

Se una funzione  $f : D \rightarrow \mathbb{R}$  è tale che  $f(D) \subseteq D$ , allora si può comporre con sé stessa; la funzione

$$(4.2) \quad f^2 = f \circ f \quad \text{ossia} \quad f^2(x) = f[f(x)]$$

viene detta *iterata seconda* di  $f$ . Analogamente,  $f^n(x) = f[f[\dots[f(x)]\dots]]$  ( $n$  volte) si dice *iterata n-esima* di  $f$ .

## 4.2 Funzioni invertibili; funzioni inverse

Supponiamo che  $f$  abbia come dominio un insieme  $D \subseteq \mathbb{R}$ . Per ogni ingresso  $x \in D$  esiste un'unica uscita  $f(x)$ . Se succede che per ogni uscita  $y \in f(D)$  esiste un solo ingresso  $x \in D$  tale che  $f(x) = y$ , allora  $f$  si dice *invertibile*, e realizza una corrispondenza biunivoca tra  $D$  e  $f(D)$ .

Più formalmente, si dice che  $f : D \rightarrow \mathbb{R}$  è invertibile in  $D$  se vale una delle seguenti condizioni (equivalenti):

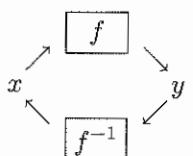
- (i)  $\forall x_1, x_2 \in D$ ,  $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$
- (ii)  $\forall x_1, x_2 \in D$ ,  $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$
- (iii)  $\forall y \in f(D)$   $\exists! x \in D$  tale che  $f(x) = y$ .

La funzione che associa a ogni uscita  $y \in f(D)$  l'unico ingresso  $x \in D$  tale che  $f(x) = y$  si chiama *funzione inversa di  $f$*  e si indica con il simbolo  $f^{-1}$ .

In sintesi

$$(4.3) \quad \begin{cases} y = f(x) \\ x \in D \end{cases} \quad \text{equivale a} \quad \begin{cases} x = f^{-1}(y) \\ y \in f(D) \end{cases}$$

La scatola nera di  $f^{-1}$  lavora a ritroso rispetto a quella di  $f$ , secondo lo schema seguente

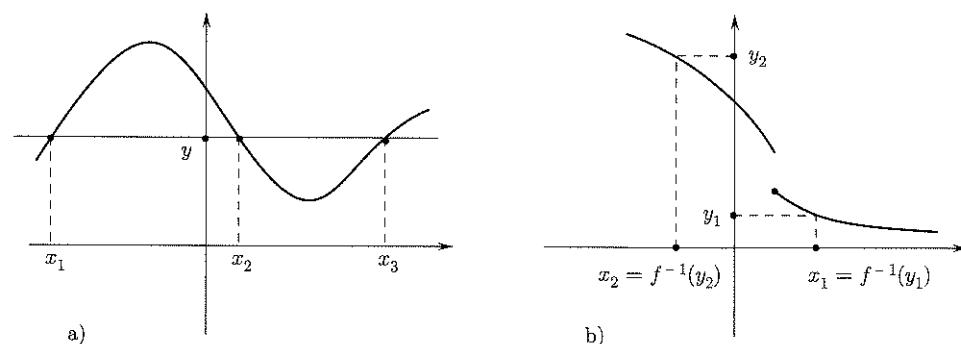


Partire da  $x$  e ritornare dopo un giro in  $x$  equivale a  $f^{-1}[f(x)] = x$  per ogni  $x \in I$ ; partire da  $y$  e tornare in  $y$  equivale a  $f[f^{-1}(y)] = y$ , per ogni  $y \in f(I)$ .

La condizione di invertibilità equivale a richiedere che il grafico di  $f$  sia intersecato al massimo in un punto da *ogni* retta parallela all'asse delle ascisse.

La funzione in figura 2.36a non è invertibile in quanto per il valore di  $y$  indicato esistono tre punti  $x_1, x_2, x_3$  che hanno immagine  $y$ . La funzione in figura 2.36b è invece invertibile in quanto ogni retta parallela all'asse delle ascisse o non interseca il grafico di  $f$  o lo interseca esattamente in un punto.

Classi di funzioni che sicuramente *non risultano* invertibili sono: le funzioni simmetriche pari ( $f(-x) = f(x)$ ), le funzioni periodiche ( $f(x+T) = f(x)$ ).

Figura 2.36. a) Funzione non invertibile: all'uscita  $y$  corrispondono 3 ingressi. b) Funzione invertibile.

Osserviamo invece che:

**TEOREMA 2.1** Una funzione  $f : D \rightarrow \mathbb{R}$  strettamente monotona in  $D$  è invertibile in  $D$ . Inoltre, la sua inversa è ancora strettamente monotona.

**DIMOSTRAZIONE.** Supponiamo per fissare le idee che  $f$  sia strettamente crescente in  $D$ . Siano  $x_1, x_2 \in D$  e proviamo che

$$x_1 \neq x_2 \implies f(x_1) \neq f(x_2).$$

Se  $x_1 \neq x_2$ , allora o  $x_1 < x_2$ , oppure  $x_1 > x_2$ . Per la monotonia stretta di  $f$ , nel primo caso si ha  $f(x_1) < f(x_2)$ , nel secondo  $f(x_1) > f(x_2)$ ; in entrambi i casi  $f(x_1) \neq f(x_2)$ , perciò  $f$  è invertibile. Sia ora  $x = f^{-1}(y)$  la sua funzione inversa, e proviamo che  $f^{-1}$  è strettamente crescente. Sia dunque  $y_1 < y_2$ . Se fosse  $x_1 \geq x_2$  (dove  $x_i = f^{-1}(y_i)$ ), poiché  $f$  è crescente avremmo  $y_1 \geq y_2$ , assurdo; quindi  $x_1 < x_2$ , ossia  $f^{-1}(y_1) < f^{-1}(y_2)$ , e  $f^{-1}$  è strettamente crescente.  $\diamond$

Si noti, ad ogni modo, che una funzione può essere invertibile anche senza essere strettamente monotona, come mostra il prossimo esempio:

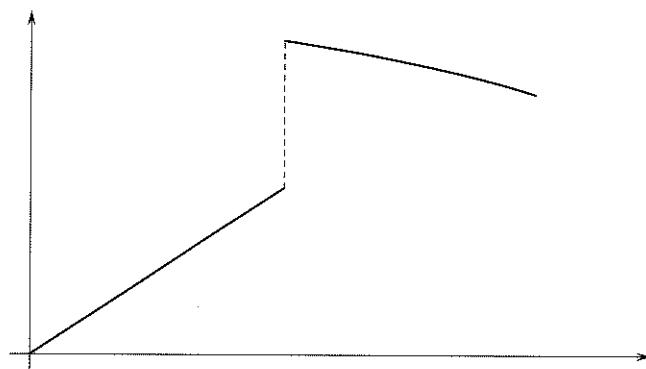
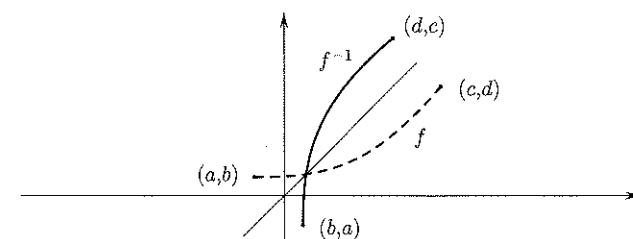


Figura 2.37.

Nel prossimo capitolo, parlando di funzioni continue, diremo qualcosa di più sulle condizioni di invertibilità per una funzione definita su un intervallo.

#### Grafico di $f^{-1}$

Le relazioni (4.3) indicano che se il punto  $(x_0, y_0)$  è sul grafico di  $f$  allora il punto  $(y_0, x_0)$  sta sul grafico di  $f^{-1}$ . Essendo i punti  $(x_0, y_0)$  e  $(y_0, x_0)$  simmetrici rispetto alla bisettrice di equazione  $y = x$ , si deduce che il grafico di  $f^{-1}$  si ricava da quello di  $f$  per simmetria rispetto alla bisettrice  $y = x$ .

Figura 2.38. Il grafico di  $f$  e della sua inversa.

Se è nota l'espressione analitica di  $f$ , invertibile, l'espressione analitica di  $f^{-1}$  si trova cercando di risolvere (quasi sempre però è impossibile!) rispetto a  $x$  l'equazione

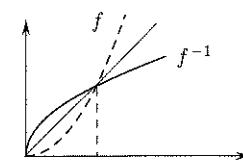
$$f(x) = y \quad (\text{Trovare l'ingresso } x \text{ che produce l'uscita } y)$$

Per esempio, se  $f(x) = 2x + 3$ ,  $f$  è invertibile su  $\mathbb{R}$  e l'equazione  $2x + 3 = y$  dà per  $f^{-1}$  l'espressione analitica

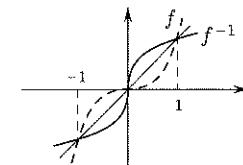
$$f^{-1}(y) = \frac{y - 3}{2}$$

Copie  $f, f^{-1}$  sono le seguenti:

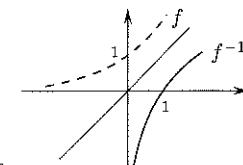
$$f : \begin{cases} y = x^2 \\ x \geq 0 \end{cases} \quad f^{-1} : \begin{cases} x = \sqrt{y} \\ y \geq 0 \end{cases}$$



$$f : \begin{cases} y = x^3 \\ x \in \mathbb{R} \end{cases} \quad f^{-1} : \begin{cases} x = \sqrt[3]{y} \\ y \in \mathbb{R} \end{cases}$$



$$f : \begin{cases} y = a^x \\ x \in \mathbb{R} \end{cases} \quad f^{-1} : \begin{cases} x = \log_a y \\ y > 0 \end{cases}$$



Si consideri anche il prossimo esempio:  $f(x) = x + e^x$ .

Essendo somma di due funzioni strettamente crescenti in tutto  $\mathbb{R}$ ,  $f(x)$  è strettamente crescente e quindi invertibile su tutto  $\mathbb{R}$ . Tuttavia, cercheremmo inutilmente di risolvere rispetto a  $x$  l'equazione  $x + e^x = y$ . In altre parole,  $f^{-1}$  esiste, ma non si sa scrivere esplicitamente.

Si noti che *tutte* le funzioni potenza:  $y = x^\alpha$ , per  $x > 0$ , sono invertibili e le funzioni inverse  $x = y^{1/\alpha}$  sono ancora potenze (con esponente reciproco di quello della funzione data). Lo studente controlli questa affermazione sulla famiglia dei grafici di  $x^\alpha$ .

Osserviamo anche che le potenze pari:  $x^{2n}$  ( $n = 1, 2, \dots$ ) *non* sono invertibili su tutta la retta, ma solo sulla semiretta  $x \geq 0$ , mentre le potenze dispari:  $x^{2n+1}$  e le potenze  $x^{p/q}$  con  $q$  dispari essendo monotone strettamente crescenti sono invertibili da  $-\infty$  a  $+\infty$ .

### 4.3 Le funzioni trigonometriche inverse

Essendo periodiche, le funzioni trigonometriche non possono essere invertibili. Infatti, per esempio, l'equazione

$$\sin x = y$$

ha infinite soluzioni se  $-1 \leq y \leq 1$  (l'uscita  $y$  corrisponde a infiniti ingressi) oppure non ha soluzioni reali se  $|y| > 1$ .

Per parlare di funzioni inverse di seno, coseno e tangente occorrerà restringersi a intervalli nei quali queste funzioni siano strettamente monotone e perciò invertibili.

Un intervallo nel quale la funzione seno è invertibile è  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ . La funzione inversa si chiama *arcoseno* (arcsin)

Dunque:

$$\begin{cases} y = \sin x \\ x \in [-\frac{\pi}{2}, \frac{\pi}{2}] \end{cases} \quad \text{equivale a} \quad \begin{cases} x = \arcsin y \\ y \in [-1, 1] \end{cases}$$

Il grafico dell'arcoseno si ottiene dall'arco di sinusoide ristretto all'intervallo  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ , per simmetria rispetto alla bisettrice  $y = x$ , come mostrato in figura 2.39.

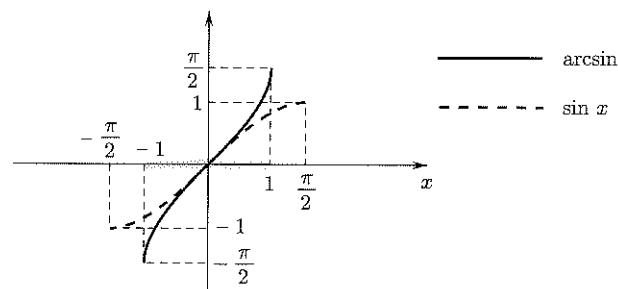


Figura 2.39. Seno e arcoseno.

Analogamente, un intervallo di monotonia del coseno è  $[0, \pi]$ , nel quale dunque esso è invertibile. La funzione inversa si chiama *arcocoseno* (arcos)

Pertanto:

$$\begin{cases} y = \cos x \\ x \in [0, \pi] \end{cases} \quad \text{equivale a} \quad \begin{cases} x = \arccos y \\ y \in [-1, 1] \end{cases}$$

Il grafico dell'arcocoseno è illustrato in figura 2.40.

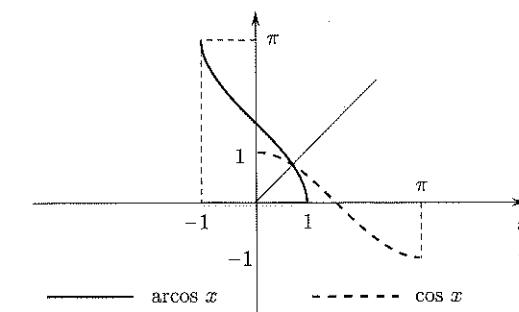


Figura 2.40. Coseno e arcocoseno.

Infine, osserviamo che nell'intervallo  $(-\frac{\pi}{2}, \frac{\pi}{2})$  la tangente è strettamente monotona e quindi invertibile. La funzione inversa si chiama *arcotangente* (arctg), ed è definita in  $(-\infty, +\infty)$ .

Dunque:

$$\begin{cases} y = \tan x \\ x \in (-\frac{\pi}{2}, \frac{\pi}{2}) \end{cases} \quad \text{equivale a} \quad \begin{cases} x = \operatorname{arctg} y \\ y \in (-\infty, +\infty) \end{cases}$$

Il grafico dell'arcotangente è mostrato in figura 2.41.

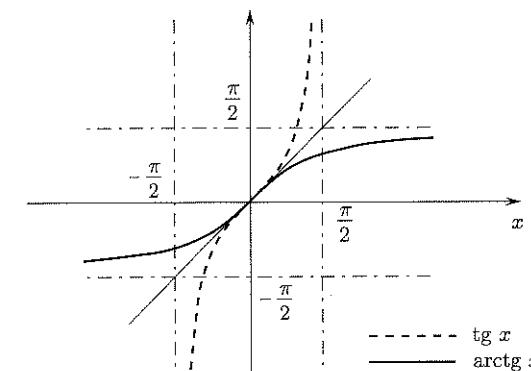


Figura 2.41. Tangente e arcotangente.

Per mezzo delle funzioni trigonometriche inverse, possiamo esprimere le soluzioni di un'equazione o disequazione trigonometrica, quando questa coinvolge angoli non notevoli. Lo studente rifletta sui prossimi esempi:

- Le soluzioni di  $\sin x = \frac{1}{4}$  sono  $x = \arcsin \frac{1}{4} + 2k\pi; x = \pi - \arcsin \frac{1}{4} + 2k\pi$
- Le soluzioni di  $\cos x < \frac{1}{5}$  sono  $\arccos \frac{1}{5} + 2k\pi < x < 2\pi - \arccos \frac{1}{5} + 2k\pi$
- Le soluzioni di  $\tan x \geq 3$  sono  $\operatorname{arctg} 3 + k\pi \leq x < \frac{\pi}{2} + k\pi$

(Per convincersi delle affermazioni fatte, ragionare sulla circonferenza trigonometrica e sulla definizione delle funzioni trigonometriche inverse).

#### 4.4 Le funzioni iperboliche inverse

Consideriamo la funzione  $y = \text{Sh}x$ . È definita e strettamente crescente in tutto  $\mathbb{R}$ , perciò è invertibile. Inaspettatamente, la sua funzione inversa è facile da scrivere esplicitamente. Infatti, poniamo

$$x = \text{Sh}y = \frac{e^y - e^{-y}}{2}$$

e risolviamo rispetto ad  $y$ . Moltiplicando ambo i membri per  $e^y$  l'equazione si può riscrivere come:

$$e^{2y} - 2xe^y - 1 = 0$$

che è un'equazione di secondo grado nell'incognita  $e^y$ . Ricaviamo:

$$e^y = x \pm \sqrt{x^2 + 1}$$

poiché  $e^y > 0$ , la soluzione col segno  $-$  va scartata. Rimane dunque:

$$y = \log(x + \sqrt{x^2 + 1})$$

Questa è l'espressione analitica della funzione inversa di  $\text{Sh}x$ , che prende il nome di *settore seno iperbolico*, e si indica anche con  $\text{SettSh}x$ . È definita per ogni  $x$  reale.

La funzione  $y = \text{Ch}x$  è definita in  $\mathbb{R}$ , strettamente crescente per  $x \geq 0$ , decrescente per  $x \leq 0$ . Perciò non è invertibile su tutto  $\mathbb{R}$ . La sua restrizione a  $x \geq 0$  però lo è. Vogliamo determinare l'espressione analitica della funzione inversa di questa restrizione.

Con calcoli analoghi ai precedenti, scriviamo

$$x = \text{Ch}y = \frac{e^y + e^{-y}}{2}$$

e risolviamo rispetto ad  $y$ :

$$e^{2y} - 2xe^y + 1 = 0$$

$$e^y = x \pm \sqrt{x^2 - 1}$$

Questa volta entrambi i numeri  $x \pm \sqrt{x^2 - 1}$  sono positivi; ricordiamo però che stiamo ragionando solo per  $y \geq 0$ , che equivale a scegliere il segno  $+$ . Pertanto:

$$y = \log(x + \sqrt{x^2 - 1})$$

Questa è l'espressione analitica della funzione inversa di  $\text{Ch}x$ , che prende il nome di *settore coseno iperbolico*, e si indica anche con  $\text{SettCh}x$ . Si noti che è definita per  $x \geq 1$ . Si osservi anche che, mentre ad esempio l'equazione

$$\text{Sh}x = 2 \text{ ha l'unica soluzione } x = \text{SettSh}2 = \log(2 + \sqrt{5})$$

l'equazione

$$\text{Ch}x = 3 \text{ ha le due soluzioni } x = \pm \text{SettCh}3 = \pm \log(3 + 2\sqrt{2})$$

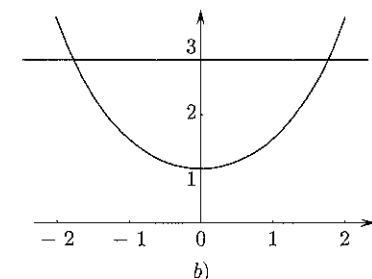
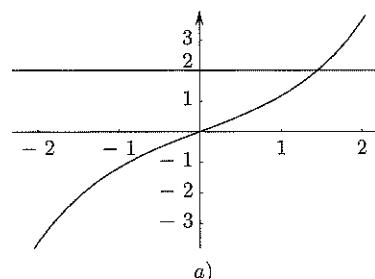


Figura 2.42. a) Grafico di  $y = \text{Sh}x$  e  $y = 2$ ; b) grafico di  $y = \text{Ch}x$  e  $y = 3$ .

Le funzioni iperboliche inverse saranno utilizzate nell'*integrazione delle funzioni irrazionali* (cap. 6).



Determinare l'insieme di definizione delle seguenti funzioni:

a)  $f(x) = \sqrt{e^{2x} - e^x}$     b)  $f(x) = \log(1 - x^3)$     c)  $f(x) = \arcsin(x^2 - 2)$

Utilizzando il grafico di  $f(x) = e^x$  e i metodi del paragrafo 3.7, disegnare i grafici di

$$y_1 = e^{|x|} \quad y_2 = 3e^x \quad y_3 = e^x - 1 \quad y_4 = |e^x - 1|$$

Verificare i risultati al computer.

Utilizzando il grafico di  $f(x) = \log x$  e i metodi del paragrafo 3.7, disegnare i grafici di

$$y_1 = \log|x| \quad y_2 = 3 \log x \quad y_3 = \log(x - 2) \quad y_4 = |\log x| \quad y_5 = |\log|x||$$

Verificare i risultati al computer.

Una vibrazione elementare sia descritta dalla funzione

$$f(x) = A \cos \omega x \quad A = \text{ampiezza} \quad \omega/2\pi = \text{frequenza}$$

Tracciare il grafico di  $f$  per  $A = 1, -1, 3, -3$  e  $\omega = 1, 2, 3$ .

Visualizzare al computer le variazioni del grafico al variare di  $A$  e di  $\omega$ .

Sia  $f(x) = \sin x$ . Tracciare il grafico e studiare, aiutandosi col computer, come esso si modifichi moltiplicando  $f$  per  $x, \frac{1}{x}, e^x, e^{-x}$ .

Determinare dominio e immagine delle seguenti funzioni. Nel caso in cui la funzione sia invertibile, determinare la funzione inversa

a)  $y_1 = x^3 + 2$     b)  $y_2 = e^{-|x|}$     c)  $y_3 = x|x|$

Sia  $f(x) = \frac{1}{x}$ . Calcolare la funzione composta  $f^2 = f \circ f$ .

Determinare l'espressione analitica della funzione composta  $f \circ g$  dove

$$f(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ -1 & \text{se } x < 0 \end{cases} \quad g(x) = \sin x$$

Disegnare il grafico di  $f \circ g$ .

Disegnare al computer le funzioni seguenti

$$y_1 = [x^2] = \text{parte intera di } x^2 \quad \text{e} \quad y_2 = (x^2) = \text{mantissa di } x^2$$

Dire se le seguenti funzioni sono simmetriche (pari o dispari):

$$\text{a)} \frac{x}{1+x^2} \quad \text{b)} x \sin 2x \quad \text{c)} \cos(x^3) \quad \text{d)} 2^{x^3} \quad \text{e)} \frac{x^2+1}{x^4+2} \quad \text{f)} \frac{x^3+1}{x+4}$$

Stabilire se le seguenti funzioni sono periodiche, e in caso affermativo determinarne il periodo:

$$\text{a)} \sin 2x + \cos 3x \quad \text{b)} (\sin x)^2 \quad \text{c)} \cos(2x+1) \quad \text{d)} \sin(x^3) \quad \text{e)} 2^{\sin x} \quad \text{f)} \operatorname{tg} \frac{x}{2} + (\cos x)^2$$

Stabilire con ragionamenti elementari se le seguenti funzioni sono monotone sul loro insieme di definizione:

$$\text{a)} \frac{1}{x^2} \quad \text{b)} e^{x^3+1} \quad \text{c)} \sin 2x \quad \text{d)} \log(1+2^{-x}) \quad \text{e)} \frac{1}{1+x^2}$$

Tracciare il grafico qualitativo delle seguenti funzioni potenza:

$$\text{a)} x^{-2/7} \quad \text{b)} x^{4/5} \quad \text{c)} x^{6/5} \quad \text{d)} x^\pi \quad \text{e)} x^{1/\sqrt{2}}$$

Risolvere le seguenti disequazioni:

$$\text{i)} |x-4| \geq x+2 \quad \text{ii)} \frac{(2x-1)(x+1)}{x} \geq 0 \quad \text{iii)} \frac{x}{|x|}(1-x) \leq 1+x$$

Risolvere le seguenti disequazioni:

$$\text{i)} \sqrt{|x-1|} < 2-x \quad \text{ii)} \sqrt[3]{1+x^3} > x-4 \quad \text{iii)} \frac{\sqrt{2x+1}-3}{x^2-4} \leq 0$$

Risolvere le disequazioni:

$$\text{i)} \log_2(x-4) - \log_2(x-1) > 1 \quad \text{ii)} (\log_2 x)^2 - 3 \log_2 x - 5 \geq 0$$

Trovare le soluzioni dell'equazione:

$$\log_{\frac{1}{2}}(1+x) = \log_2(2-x)$$

Risolvere le seguenti disequazioni:

$$\text{i)} 2(\sin x)^2 + \sin x - 3 < 0 \quad \text{ii)} (\cos x)^2 + 2 \cos x - 1 \geq 0$$

Risolvere le seguenti disequazioni, eseguendo anche, ove necessario, un confronto grafico, e localizzando i valori che non si possono determinare esattamente (Esempio: soluzioni  $\alpha \leq x \leq 5$ , con  $3 < \alpha < 4$ ):

$$\text{i)} e^{\sqrt{x}-x+1} > 1 \quad \text{ii)} \frac{e^x}{1-x^2} < 1$$

Dire quali sono il dominio e l'immagine delle seguenti funzioni:

$$f(x) = 2 \arcsin(x-1); \quad g(x) = \frac{1}{2} \arccos(2x+1)$$

Con lo stesso metodo usato per invertire la funzione  $\operatorname{Sh} x$ , scrivere la funzione inversa di  $\operatorname{Th} x$ .

## COMPLEMENTI

Dimostrare che la funzione  $f(x) = x^2$  è strettamente crescente per  $x > 0$ . (Suggerimento: moltiplicare ogni membro della diseguaglianza  $0 < x_1 < x_2$  una volta per  $x_1$  e una volta per  $x_2$ , e confrontare le diseguaglianze ottenute).

Utilizzando un'idea simile a quella suggerita, provare per induzione su  $n$  che  $x^n$  è strettamente crescente in  $(0, +\infty)$ .

Utilizzare l'identità

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2}$$

per dimostrare che ogni funzione  $f : \mathbb{R} \rightarrow \mathbb{R}$  si può scrivere come somma di una funzione pari  $p$  e una dispari  $d$ . Se si applica questo procedimento a  $f(x) = e^x$ , chi sono  $p(x)$  e  $d(x)$ ?

Siano  $A$  e  $B$  due intervalli. Una funzione crescente (o decrescente) sia in  $A$  che in  $B$ , è crescente (o decrescente) in  $A \cup B$ ? Rispondere ragionando sull'esempio  $f(x) = 1/x$ .

Siano  $f, g$  funzioni monotone per cui è definita  $f \circ g$ . Allora  $f \circ g$  è monotona, e precisamente:

	se $f$ è crescente...	se $f$ è decrescente...
... e $g$ è crescente	allora $f \circ g$ è...	allora $f \circ g$ è...
... e $g$ è decrescente	allora $f \circ g$ è...	allora $f \circ g$ è...

Siano  $f, g$  funzioni simmetriche (pari o dispari) per cui è definita  $f \circ g$ . Allora  $f \circ g$  è simmetrica, e precisamente:

	se $f$ è pari...	se $f$ è dispari...
... e $g$ è pari	allora $f \circ g$ ...	allora $f \circ g$ ...
... e $g$ è dispari	allora $f \circ g$ è...	allora $f \circ g$ è...

6 Siano  $f, g$  funzioni per cui è definita  $f \circ g$ . Chiediamoci se  $f \circ g$  è periodica, nei seguenti casi:

	se $f$ è periodica...	se $f$ non è periodica...
... e $g$ è periodica	allora $f \circ g$ ...	allora $f \circ g$ ...
$g$ non è periodica	allora $f \circ g$ ...	allora $f \circ g$ ...

7 Sia  $f : \mathbb{R} \rightarrow \mathbb{R}$  una funzione periodica di periodo  $T$ . Si dica se le seguenti funzioni sono periodiche, indicandone il periodo, oppure non lo sono, esibendo in tal caso un contesempio esplicito:

$$\begin{aligned} f_1(x) &= f(\omega x) \quad (\text{con } \omega > 0); & f_2(x) &= f(x + c) \quad (\text{con } c \in \mathbb{R}); \\ f_3(x) &= |f(x)|; & f_4(x) &= f(|x|); \\ f_5(x) &= f(x^2) \end{aligned}$$

8 Siano  $f_1, f_2 : \mathbb{R} \rightarrow \mathbb{R}$  funzioni periodiche di periodo  $T_1, T_2$ , rispettivamente. Ci chiediamo se la funzione  $f_1 + f_2$  è periodica o no, e in caso affermativo qual è il suo periodo. Se vogliamo che risulti

$$(f_1 + f_2)(x + T) = (f_1 + f_2)(x) \quad \forall x,$$

dovrà essere:

$$T = n_1 T_1 = n_2 T_2$$

per due opportuni interi positivi  $n_1, n_2$ . Ne segue che  $T$  dev'essere un multiplo comune di  $T_1$  e  $T_2$ . In particolare:

i) condizione necessaria affinché  $f_1 + f_2$  sia periodica è che il rapporto tra i due periodi  $T_1/T_2$  sia un numero *razionale*;

ii) se questa condizione è soddisfatta, un numero  $T$  che soddisfa la condizione è il *minimo comune multiplo* di  $T_1, T_2$ . Questo numero è il periodo di  $f$  (oppure un multiplo del periodo di  $f$ , che in tal caso è più piccolo).

Usando di questi fatti:

a) Si determinino i periodi delle seguenti funzioni:

$$\sin 4x + 3 \cos 6x; \quad \operatorname{tg} 6x - 2 \cos 8x; \quad 5 \sin \frac{x}{3} + \cos \frac{x}{4}$$

e si controlli il risultato previsto tracciando il grafico di queste funzioni col computer.

b) Si dimostri che la funzione

$$\sin x + \sin(\pi x)$$

non è periodica (pur essendo somma di due funzioni periodiche), e ci si renda conto di questo fatto tracciando il grafico di questa funzione col computer, su intervalli di varia ampiezza.

9 Dimostrare, in base alla definizione, le proprietà 3) delle funzioni iperboliche enunciate nel paragrafo 3.6.

# 3

## Limiti e continuità

In questo corso ci occuperemo prevalentemente del *calcolo infinitesimale*, disciplina matematica che affonda le sue radici nella Grecia del III secolo a.C. (Euclide, Archimede), ha un grande sviluppo a partire dal Seicento, parallelamente al nascere della scienza moderna, in particolare ad opera di Newton e Leibniz, tra il 1670 e il 1710 circa; viene quindi sottoposta a revisione critica e fondata rigorosamente nell'Ottocento, prima da Cauchy, nel 1821<sup>1</sup>, poi da Weierstrass e da vari altri matematici (Heine, Cantor, Méray,...) intorno al 1870. Le idee e le tecniche di calcolo proprie di questa disciplina fanno oggi parte del bagaglio essenziale con cui scienza e tecnologia si esprimono e procedono.

Il fondamento concettuale del calcolo infinitesimale sta nella nozione di *limite* (D'Alembert 1765, Cauchy 1821), che quindi può a buon diritto considerarsi una pietra miliare nella storia del pensiero scientifico. Noi introdurremo questo concetto gradualmente, prima nel caso discreto (par. 1) e poi in quello continuo, in cui storicamente è nato (par. 2).

Nel contesto discreto, il limite si può vedere come un'*operazione* che, a differenza delle operazioni algebriche elementari (somma, prodotto), viene eseguita non su una coppia di numeri ma su una *successione* di infiniti numeri. Per prima cosa introdurremo quindi il concetto di successione. Questo argomento, oltre a servirci immediatamente per introdurre il concetto di limite di funzione, sarà ripreso nel capitolo 5 parlando di *serie numeriche*, e nel capitolo 7 discutendo i *modelli dinamici discreti*.

### 1 SUCCESSIONI

#### 1.1 Definizione di successione. Definizione di limite

Consideriamo l'insieme  $\mathbb{N}$  degli interi non negativi ordinato secondo l'ordine naturale

$$\mathbb{N} : 0, 1, 2, 3, \dots, n, \dots$$

<sup>1</sup>Corso di Analisi per l'Ecole Polytechnique di Parigi.

Questo è l'esempio canonico di successione. Stabiliamo ora una legge che associa, a ogni elemento di  $\mathbb{N}$  (o da un certo intero in poi) un numero (reale):

$$n \mapsto a_n$$

Chiameremo *successione* una tale corrispondenza.

Una successione può dunque vedersi come una funzione

$$f : \mathbb{N} \rightarrow \mathbb{R}$$

$$f : n \mapsto a_n$$

(o eventualmente,  $f : \{n \in \mathbb{N} : n \geq n_0\} \rightarrow \mathbb{R}$ , per un certo  $n_0$  fissato). Il fatto che il dominio della funzione  $f$  sia l'insieme dei naturali, rende possibile visualizzare la successione enumerando i suoi valori, nell'ordine in cui essi si succedono al crescere di  $n$ :<sup>2</sup>

$$a_0, a_1, a_2, \dots, a_n, \dots$$



$n \geq 0$	$n \mapsto n^2$	$0, 1, 4, 9, 16, \dots$
$n \geq 0$	$n \mapsto (-1)^n$	$1, -1, 1, -1, 1, -1, \dots$
$n \geq 1$	$n \mapsto 2^{1/n}$	$2, \sqrt[3]{2}, \sqrt[4]{2}, \dots$
$n \geq 1$	$n \mapsto \frac{1}{n}$	$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$
$n \geq 2$	$n \mapsto \frac{n+1}{n-1}$	$3, 2, \frac{5}{3}, \frac{6}{4}, \frac{7}{5}, \dots$
$n \geq 0$	$n \mapsto 4$	$4, 4, 4, \dots$ (successione costante)

Possiamo rappresentare graficamente questa corrispondenza con i punti del piano cartesiano di coordinate  $(n, a_n)$  (fig.3.1).

Sottolineiamo che la successione è nota quando è nota la legge che, dato l'intero  $n$ , determina il numero  $a_n$  associato a quell'intero. Per indicare una successione useremo i simboli

$$n \mapsto a_n \quad \text{oppure} \quad \{a_n\}$$

precisando l'insieme in cui varia l'indice  $n$  (tutto  $\mathbb{N}$  o da un certo intero in poi).

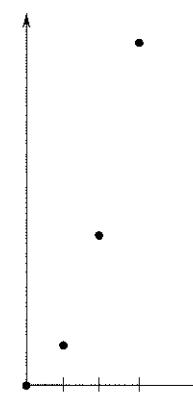
Una successione  $\{a_n\}$  si dirà

*limitata inferiormente* se esiste un numero  $m$  tale che  $a_n \geq m \forall n$

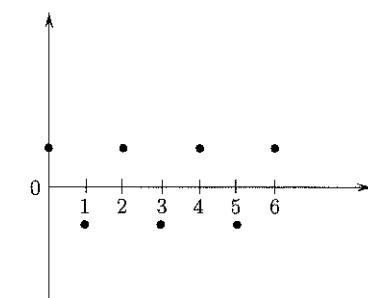
*limitata superiormente* se esiste un numero  $M$  tale che  $a_n \leq M \forall n$

*limitata* se esistono due numeri  $m$  e  $M$  tali che  $m \leq a_n \leq M \forall n$

<sup>2</sup>I puntini di sospensione... scritti nella formula seguente dopo  $a_n$  sono fondamentali: significano che non stiamo considerando soltanto i primi  $n$  termini della successione (cioè un insieme finito di numeri), ma l'intera successione di infiniti termini (in cui  $n$  gioca il ruolo di indice muto).



a)  $n \mapsto n^2$



b)  $n \mapsto (-1)^n$

Figura 3.1.

Per esempio, la successione  $\{(-1)^n\}$  è limitata;  $\{n^2\}$  è limitata solo inferiormente; la successione  $\{(-2)^n\}$  non è limitata (né inferiormente, né superiormente).

L'operazione che vogliamo definire (il *limite*) consente di rispondere in forma rigorosa alla domanda: *come si comportano i numeri  $\{a_n\}$  quando  $n$  diventa sempre più grande?*

Cominciamo con l'introdurre un modo di dire molto utile.

**DEFINIZIONE 3.1** Diciamo che una successione  $\{a_n\}$  possiede (o acquista) *definitivamente* una certa proprietà se esiste un  $N \in \mathbb{N}$  tale che  $a_n$  soddisfa quella proprietà per ogni intero  $n \geq N$ .



La successione  $\{n - 10\sqrt{n}\}$  è definitivamente positiva; la successione  $\{\frac{1}{n}\}$  ( $n = 1, 2, \dots$ ) è definitivamente minore di  $10^{-10^{10}}$ .

### Successioni convergenti

**DEFINIZIONE 3.2** Una successione  $\{a_n\}$  si dice *convergente* se esiste un numero  $l \in \mathbb{R}$  con questa proprietà: qualunque sia  $\varepsilon > 0$  risulta definitivamente

$$(1.1) \quad |a_n - l| < \varepsilon.$$

In altre parole: per ogni  $\varepsilon > 0$  si può trovare un intero  $N$  (che naturalmente dipenderà in generale da questo  $\varepsilon$ ) tale che

$$|a_n - l| < \varepsilon \quad \text{per ogni } n \geq N$$

Se la successione  $\{a_n\}$  è convergente, ad essa è associato perciò il numero  $l$ . Si osservi che tale numero è unico, poiché, se ve ne fossero due,  $l_1$  e  $l_2$ , associati alla medesi-

ma successione, risulterebbe definitivamente (applicando la diseguaglianza triangolare (4.4), cap. 1)

$$|l_1 - l_2| = |l_1 - a_n + a_n - l_2| \leq |l_1 - a_n| + |a_n - l_2| < 2\varepsilon$$

ma tale diseguaglianza, potendo noi scegliere  $\varepsilon$  come vogliamo, può sussistere solo se  $l_1 = l_2$ .

**DEFINIZIONE 3.3** Il numero  $l$  che compare nella (1.1) si chiama *limite* della successione  $\{a_n\}$ , e si scrive

$$\lim_{n \rightarrow +\infty} a_n = l \quad \text{oppure} \quad a_n \rightarrow l \quad \text{per } n \rightarrow +\infty$$

(si legge, rispettivamente: il limite, per  $n$  tendente all'infinito, di  $a_n$  è  $l$ , oppure:  $a_n$  tende a  $l$  per  $n$  tendente a infinito).

Si noti che la diseguaglianza (1.1) corrisponde, più esplicitamente, alle seguenti due:

$$(1.2) \quad l - \varepsilon < a_n < l + \varepsilon$$

Rappresentando graficamente i punti della successione

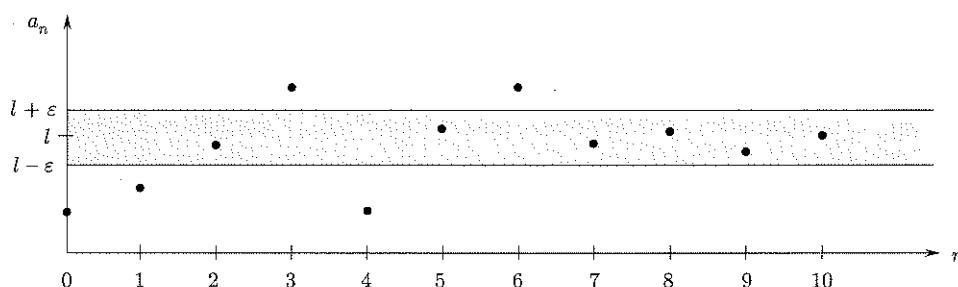


Figura 3.2.

la condizione di convergenza significa che, fissata una striscia orizzontale  $[l - \varepsilon, l + \varepsilon]$  “comunque stretta”, da un certo indice in poi i punti della successione non escono più da questa striscia (v. fig. 3.2). Da questa osservazione risulta chiaramente che: *ogni successione convergente è limitata*.



**■■■** Mostriamo che  $\lim_{n \rightarrow +\infty} \frac{n+1}{n-1} = 1$  (cosa che si può facilmente sospettare osservando l'andamento della successione). Delle due diseguaglianze

$$1 - \varepsilon < \frac{n+1}{n-1} < 1 + \varepsilon$$

quella di sinistra è sempre soddisfatta, mentre quella di destra è soddisfatta per

$$n > \frac{2+\varepsilon}{\varepsilon}$$

Fissato  $\varepsilon > 0$ , basterà scegliere  $N = (2 + \varepsilon)/\varepsilon$  (o uguale al primo intero  $> (2 + \varepsilon)/\varepsilon$ ) per soddisfare la condizione richiesta dalla definizione di limite.

**■■■** Per mostrare che  $2^{1/n} \rightarrow 1$  per  $n \rightarrow \infty$ , si studiano le diseguaglianze

$$1 - \varepsilon < 2^{1/n} < 1 + \varepsilon$$

Quella di sinistra è sempre soddisfatta; quella di destra, prendendo il logaritmo (in base 2) di ambo i membri, si scrive

$$\frac{1}{n} < \log_2(1 + \varepsilon)$$

ed è soddisfatta se  $n > 1/\log_2(1 + \varepsilon)$ . Si sceglie perciò  $N = \dots$

Non risultano convergenti invece le prime due successioni dell'esempio 1.1. Esse sono però molto diverse tra loro e conviene introdurre definizioni che ne mettano in risalto la differenza.

#### Successioni divergenti. Successioni irregolari

**DEFINIZIONE 3.4** Quando, al crescere di  $n$ , una successione supera definitivamente qualunque numero  $M > 0$  fissato, diremo che *diverge a  $+\infty$* ; se invece scende al di sotto di  $-M$ , diremo che *diverge a  $-\infty$* . (Il simbolo  $\infty$  si legge “infinito”).

Diremo nei due casi, rispettivamente, che  $+\infty$  e  $-\infty$  sono i limiti della successione e scriveremo, rispettivamente:

$$\lim_{n \rightarrow +\infty} a_n = +\infty \quad \text{oppure} \quad \lim_{n \rightarrow +\infty} a_n = -\infty.$$

Questi simboli,  $+\infty$  e  $-\infty$ , non sono numeri. Se rappresentiamo i numeri reali sulla retta euclidea, ogni numero corrisponde a un punto e ogni punto a un numero. Con i simboli  $+\infty$  e  $-\infty$  conveniamo di indicare due “punti”, uno ( $+\infty$ ) sta alla destra di ogni punto di  $\mathbb{R}$  e l’altro ( $-\infty$ ) alla sinistra; a questi due punti non corrisponde però alcun numero (in altre parole, non possiamo definire sui simboli  $+\infty$  e  $-\infty$  le operazioni di somma e prodotto con le proprietà indicate in  $R_1$  e  $R_2$ , anche se, come vedremo, potremo fare “parzialmente” queste operazioni).

L’insieme dei numeri reali  $\mathbb{R}$  con l’aggiunta dei due elementi  $\{+\infty\}$  e  $\{-\infty\}$  sarà indicato con  $\mathbb{R}^*$

$$\mathbb{R}^* = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$$

Possiamo rappresentare “visivamente” l’insieme  $\mathbb{R}^*$  mettendo in corrispondenza biunivoca (fig. 3.3) i punti della retta con quelli di una semicirconferenza, proiettando questi ultimi dal centro  $C$  sulla retta  $\mathbb{R}$ :

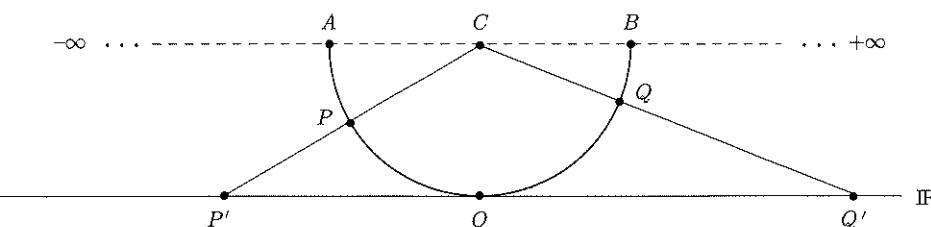


Figura 3.3.

Ai punti  $A$  e  $B$  non corrisponde su  $\mathbb{R}$  alcun punto; diremo che  $\{-\infty\}$  è il corrispondente del punto  $A$  e  $\{+\infty\}$  il corrispondente di  $B$ .

L'operazione di limite risulta completamente significativa se ambientata in  $\mathbb{R}^*$  invece che in  $\mathbb{R}$ ; cioè il limite di una successione può essere un numero reale  $l$  oppure  $+\infty$  oppure  $-\infty$ ; le successioni il cui limite è un numero reale si dicono *convergenti*, quelle il cui limite è  $+\infty$  oppure  $-\infty$  si dicono *divergenti*. La successione canonica  $\{n\}$  degli interi naturali evidentemente diverge a  $+\infty$ ; così pure la successione  $\{2^n\}$ .

Infine osserviamo che ci sono successioni che non ricadono in nessuna delle categorie precedenti, cioè non sono convergenti né divergenti; per esempio la successione  $\{(-1)^n\}$  oppure  $\{(-2)^n\}$  (si noti che la prima è limitata e la seconda no). Tali successioni si diranno *irregolari* o *indeterminate*. Per esse l'operazione di limite non è definita, ovvero *il loro limite non esiste*.

#### Insiemi non limitati

È comodo adottare la convenzione introdotta per i limiti anche per il sup e per l'inf, estendendo la definizione di queste quantità nel modo seguente:

**DEFINIZIONE 3.5** Se l'insieme  $E \subseteq \mathbb{R}$  non è limitato superiormente (inferiormente) diremo che

$$\sup E = +\infty \quad (\inf E = -\infty).$$

In questo modo la proprietà  $R_4$  dei numeri reali può essere enunciata così:

ogni insieme  $E \subseteq \mathbb{R}$  non vuoto è dotato di estremo superiore e inferiore;  $\sup E$  ( $\inf E$ ) è un numero se  $E$  è limitato superiormente (inferiormente), altrimenti è  $+\infty$  ( $-\infty$ ).

#### Infinitesimi e infiniti

Una successione  $a_n$  tendente a zero si dice *infinitesima*. Ad esempio, sono infinitesime le successioni  $\{\frac{1}{n}\}$ ,  $\{\frac{1}{n^2}\}, \dots$

Il concetto di infinitesimo gioca un ruolo centrale ed è fondamentale anche per avere un'*immagine intuitiva* corretta ed efficace dei concetti del calcolo infinitesimale. Vedremo nel par. 2 che il concetto di infinitesimo nel continuo (cioè parlando di funzioni) sarà perfettamente analogo. L'idea chiave a cui prestare attenzione è la seguente:

“infinitesimo” non è un “numero infinitamente piccolo” (concetto privo di senso, se non si vuole che denoti semplicemente il numero 0) ma una *quantità variabile* (successione o, come vedremo, funzione), che *diviene infinitamente piccola*.

Analogamente, una successione  $a_n$  tendente a  $\pm\infty$  si dice *infinita*. Ad esempio,  $\{n^2\}$ ,  $\{n!\}$  sono infiniti.

Talvolta è possibile precisare se una successione convergente si avvicina al suo limite *per eccesso* o *per difetto*. Questo concetto è precisato dalla prossima

**DEFINIZIONE 3.6** Si dice che la successione  $\{a_n\}$  tende a  $l \in \mathbb{R}$  *per eccesso* (*per difetto*), e si scrive

$$\lim_{n \rightarrow +\infty} a_n = l^+ \text{ oppure } a_n \rightarrow l^+ \text{ per } n \rightarrow +\infty$$

(rispettivamente:

$$\lim_{n \rightarrow +\infty} a_n = l^- \text{ oppure } a_n \rightarrow l^- \text{ per } n \rightarrow +\infty)$$

se per ogni  $\varepsilon > 0$  si ha che

$$0 \leq a_n - l < \varepsilon \text{ definitivamente}$$

(rispettivamente:

$$0 \leq l - a_n < \varepsilon \text{ definitivamente}).$$

In sostanza, dire che  $a_n \rightarrow l^+$  significa affermare che  $a_n \rightarrow l$  e inoltre  $a_n \geq l$  definitivamente; dunque  $a_n$  si avvicina ad  $l$  “da sopra”, ossia approssima  $l$  per eccesso.

Si rifletta sui prossimi esempi:



$$\lim_{n \rightarrow +\infty} \frac{1}{n} = 0^+;$$



$$\lim_{n \rightarrow +\infty} \frac{n}{n+1} = 1^-;$$



$$\lim_{n \rightarrow +\infty} \frac{(-1)^n}{n} = 0,$$

ma in questo caso non si può affermare né che  $a_n \rightarrow 0^+$  né che  $a_n \rightarrow 0^-$ .

## 1.2 Successioni monotone

**DEFINIZIONE 3.7** Una successione  $\{a_n\}$  si dirà:

monotona crescente se  $a_n \leq a_{n+1}$ ; strettamente crescente se  $a_n < a_{n+1} \forall n$ ,  
monotona decrescente se  $a_n \geq a_{n+1}$ ; strettamente decrescente se  $a_n > a_{n+1} \forall n$ .

Per esempio, la successione  $\{n^2\}$  è monotona strettamente crescente, la successione  $\{\frac{1}{n}\}$  è monotona strettamente decrescente, la successione  $\{(-1)^n\}$  non è monotona; ogni successione costante è monotona (crescente o decrescente, non strettamente). Riguardo all'operazione di limite, queste successioni hanno una importanza particolare; infatti esse non sono mai irregolari, ma sono convergenti oppure divergenti a seconda che siano limitate oppure no. Il risultato fondamentale è il seguente:

**TEOREMA 3.1 (DI MONOTONIA)** Sia  $\{a_n\}$  una successione monotona crescente e superiormente limitata. Allora  $\{a_n\}$  è convergente, e il suo limite è uguale a  $\sup \{a_n : n \in \mathbb{N}\}$ .

Analogamente, se  $\{a_n\}$  è una successione monotona decrescente e inferiormente limitata, allora  $\{a_n\}$  è convergente, e il suo limite è uguale a  $\inf \{a_n : n \in \mathbb{N}\}$ .

Ricordiamo che il simbolo  $\sup \{a_n : n \in \mathbb{N}\}$  denota l'estremo superiore dell'insieme dei valori  $a_n$  assunti dalla successione.

**DIMOSTRAZIONE.** Consideriamo l'insieme dei valori assunti dalla successione,  $\{a_n : n \in \mathbb{N}\}$ . Poiché la successione è limitata superiormente, questo insieme è limitato superiormente, quindi (per la proprietà dell'estremo superiore di cui gode  $\mathbb{R}$ ) esiste il suo estremo superiore,

$$\Lambda = \sup \{a_n : n \in \mathbb{N}\}, \Lambda \in \mathbb{R}.$$

Proviamo ora che

$$\lim_{n \rightarrow \infty} a_n = \Lambda.$$

Occorre mostrare che per ogni  $\varepsilon > 0$  si ha

$$\Lambda - \varepsilon < a_n < \Lambda + \varepsilon \text{ definitivamente.}$$

La seconda disegualanza è ovvia: per ogni  $n$  è  $a_n \leq \Lambda$  (e quindi  $a_n < \Lambda + \varepsilon$ ), perché  $\Lambda$  è l'estremo superiore degli  $a_n$ , quindi in particolare è un maggiorante dell'insieme dei valori  $a_n$ . Per provare la prima disegualanza, consideriamo il numero  $\Lambda - \varepsilon$ . Ricordiamo che per definizione di estremo superiore,  $\Lambda$  è il *minimo* dei maggioranti dell'insieme  $\{a_n : n \in \mathbb{N}\}$ . Perciò, essendo  $\Lambda - \varepsilon < \Lambda$ , certamente  $\Lambda - \varepsilon$  non è un maggiorante dell'insieme  $\{a_n : n \in \mathbb{N}\}$ . Questo significa che esiste un  $n_0$  per cui

$$a_{n_0} > \Lambda - \varepsilon.$$

D'altra canto la successione è monotona crescente, perciò per ogni  $n \geq n_0$  risulta  $a_n \geq a_{n_0}$ . Abbiamo quindi provato che

$$a_n \geq a_{n_0} > \Lambda - \varepsilon$$

per ogni  $n \geq n_0$ , ossia definitivamente. Quindi  $\lim_{n \rightarrow \infty} a_n = \Lambda$  (fig. 3.4).

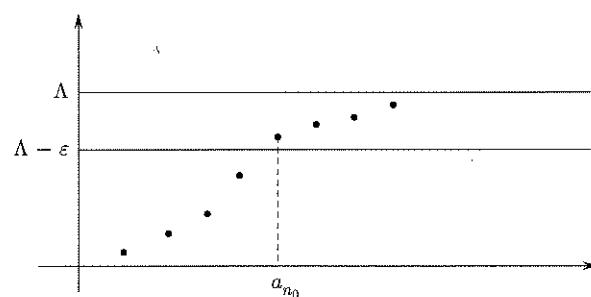


Figura 3.4.

Per esprimere anche simbolicamente che il limite è il sup (o l'inf) di una successione crescente (o decrescente) si usa la notazione (di evidente significato):

$$a_n \uparrow l \quad \text{oppure} \quad a_n \downarrow l.$$

Questo in particolare implica che  $a_n \rightarrow l^-$  (rispettivamente,  $l^+$ ), ma contiene una ulteriore informazione: la monotonia della successione.

Questo teorema è una conseguenza dell'assioma di continuità  $R_4$  dei numeri reali (v. cap. 1, par. 4.2), anzi è equivalente a questo. Pertanto vale se l'ambiente che

consideriamo è  $\mathbb{R}$ . Ad esempio, non è vero che una successione crescente e limitata di numeri razionali ammette sempre limite razionale, cioè in  $\mathbb{Q}$ .



Sia  $\{a_n\}$  la successione così definita:

$$\begin{aligned} a_0 &= 0 \\ a_1 &= 0,1 \\ a_2 &= 0,1011 \\ a_3 &= 0,10110111 \\ &\dots \end{aligned}$$

e così via. (Al passo  $n$  si aggiunge al numero decimale ottenuto al passo precedente una cifra zero seguita da  $n$  cifre uguali a 1).

La successione  $\{a_n\}$  è evidentemente crescente, e superiormente limitata (ad esempio,  $a_n \leq 1$ ). In  $\mathbb{R}$ , la successione converge al numero  $\sup \{a_n : n \in \mathbb{N}\}$ , che dopo la virgola presenta un allineamento decimale illimitato e non periodico di cifre (una cifra 1, una cifra 0, due cifre 1, una cifra 0, tre cifre 1, una cifra 0, e così via all'infinito). Quindi il limite della successione è un numero irrazionale. Quest'esempio mostra che nell'insieme  $\mathbb{Q}$  il teorema di monotonia è falso.

Vedremo in seguito che il teorema di monotonia sarà utilizzato per dimostrare importanti proprietà delle funzioni continue. Questo teorema quindi costituisce una delle motivazioni per cui è utile studiare l'analisi matematica nell'ambiente dei numeri reali, anziché in quello dei numeri razionali.

Il teorema di monotonia si può completare con il prossimo enunciato, che considera successioni limitate o illimitate:

**COROLLARIO 3.2** Sia  $\{a_n\}$  una successione monotona crescente. Allora esiste

$$\lim_{n \rightarrow \infty} a_n = \sup \{a_n : n \in \mathbb{N}\}.$$

Esplicitamente: se  $\{a_n\}$  è superiormente limitata, allora converge (e il suo limite è uguale all'estremo superiore dei suoi valori, che in questo caso è un numero reale); se invece  $\{a_n\}$  è superiormente illimitata, allora  $a_n$  tende a  $+\infty$  (che in questo caso è pari all'estremo superiore dei suoi valori).

Si può anche dire, sinteticamente: una successione monotona converge o diverge (non può essere irregolare).

**DIMOSTRAZIONE.** Se  $\{a_n\}$  è superiormente limitata, l'enunciato è contenuto nel teorema di monotonia. Se invece  $\{a_n\}$  è superiormente illimitata, questo significa che per ogni  $K > 0$  esiste un  $n_0$  tale che  $a_{n_0} > K$ . D'altra canto la successione è crescente, perciò per ogni  $n \geq n_0$  si ha  $a_n \geq a_{n_0} > K$ . Abbiamo quindi provato che per ogni  $K > 0$  è  $a_n > K$  definitivamente. Questo significa che  $a_n \rightarrow +\infty$ .



**(Successione geometrica)** Si consideri la progressione geometrica di ragione  $q$ ,  $\{q^n\}$  (cfr. cap. 1, paragrafo 2):

$$1, q, q^2, q^3, \dots, q^n, \dots$$

Se  $q > 1$  la successione è monotona crescente, illimitata superiormente.

Se  $q = 1$  la successione è costante. Se  $0 < q < 1$ , la successione è monotona decrescente; è facile mostrare che tende a zero. Se  $q$  è negativo la successione non è più monotona. Lo studente verifichi le seguenti affermazioni:

$$\lim_{n \rightarrow +\infty} q^n = \begin{cases} +\infty & \text{se } q > 1 \\ 1 & \text{se } q = 1 \\ 0 & \text{se } |q| < 1 \\ \text{non esiste} & \text{se } q \leq -1 \end{cases}$$

### 1.3 Il calcolo dei limiti

In questo paragrafo passeremo in rassegna i teoremi basilari sul calcolo dei limiti. Le dimostrazioni si basano sulla definizione di limite, sull'uso di *disuguaglianze*, e sull'uso di proprietà *definitivamente vere*. In particolare, questi teoremi illustrano la relazione tra l'operazione di limite e le strutture algebriche e d'ordine presenti in  $\mathbb{R}$ .

Cominciamo ad esaminare le proprietà dell'operazione di limite rispetto alle *operazioni algebriche*. Quando il limite esiste finito, si dimostra un risultato semplice e naturale: l'operazione di limite commuta con queste operazioni, cioè:

**TEOREMA 3.3 (ALGEBRA DEI LIMITI)** *Se  $a_n \rightarrow a$  e  $b_n \rightarrow b$  allora*

$$\begin{aligned} a_n \pm b_n &\rightarrow a \pm b \\ a_n b_n &\rightarrow ab \\ \frac{a_n}{b_n} &\rightarrow \frac{a}{b} \quad (b_n, b \neq 0) \\ a_n^{b_n} &\rightarrow a^b \quad (a_n, a > 0) \end{aligned}$$

**DIMOSTRAZIONE.** Proviamo ad esempio che:

$$a_n \rightarrow a, b_n \rightarrow b \Rightarrow a_n + b_n \rightarrow a + b.$$

Fissiamo  $\varepsilon > 0$  e consideriamo:

$$\begin{aligned} |(a_n + b_n) - (a + b)| &= |(a_n - a) + (b_n - b)| \leq \\ &\leq |a_n - a| + |b_n - b| \end{aligned}$$

(per la diseguaglianza triangolare, cap.1, par. 4.3). Poiché per ipotesi  $a_n \rightarrow a$  e  $b_n \rightarrow b$ , si ha che

$$|a_n - a| < \varepsilon \text{ e } |b_n - b| < \varepsilon \text{ definitivamente,}$$

perciò concludiamo che

$$|(a_n + b_n) - (a + b)| < 2\varepsilon \text{ definitivamente,}$$

quindi  $a_n + b_n \rightarrow a + b$ .

Proviamo ora:

$$a_n \rightarrow a, b_n \rightarrow b \Rightarrow a_n b_n \rightarrow ab$$

Fissiamo  $\varepsilon > 0$ . Usando ancora la diseguaglianza triangolare si ha:

$$\begin{aligned} |a_n b_n - ab| &= |a_n(b_n - b) + b(a_n - a)| \leq \\ &\leq |a_n(b_n - b)| + |b(a_n - a)| = \\ &= |a_n||b_n - b| + |b||a_n - a|. \end{aligned}$$

Poiché  $a_n \rightarrow a$ ,  $|a_n - a| < \varepsilon$  definitivamente; inoltre  $|a_n| < |a| + \varepsilon$  definitivamente; poiché  $b_n \rightarrow b$ ,  $|b_n - b| < \varepsilon$  definitivamente. Quindi:

$$|a_n b_n - ab| < (|a| + \varepsilon)\varepsilon + |b|\varepsilon < \varepsilon \cdot \text{cost.}$$

definitivamente. Per l'arbitrarietà di  $\varepsilon$  segue la tesi.

Tralasciamo le dimostrazioni degli altri due casi.

L'operazione di limite mantiene inoltre l'*ordinamento* cioè:

**TEOREMA 3.4 (DI PERMANENZA DEL SEGNO, 1<sup>a</sup> FORMA)** *Se  $a_n \rightarrow a$  e  $a > 0$  ( $a < 0$ ) allora  $a_n > 0$  definitivamente ( $a_n < 0$  definitivamente).*

**DIMOSTRAZIONE.** Fissato  $\varepsilon > 0$ , per definizione di limite abbiamo che

$$|a_n - a| < \varepsilon \text{ definitivamente,}$$

che riscriviamo nella forma:

$$a - \varepsilon < a_n < a + \varepsilon \text{ definitivamente.}$$

Poiché  $a > 0$ , possiamo scegliere  $\varepsilon > 0$  in modo che sia anche  $a - \varepsilon > 0$ , allora la diseguaglianza  $a - \varepsilon < a_n$  mostra che  $a_n > 0$  definitivamente. In modo analogo si dimostra il caso  $a < 0$ .

**TEOREMA 3.5 (DI PERMANENZA DEL SEGNO, 2<sup>a</sup> FORMA)** *Se  $a_n \rightarrow a \in \mathbb{R}$ , e  $a_n \geq 0$  definitivamente, allora risulta  $a \geq 0$ . Più in generale:*

$$\text{se } a_n \rightarrow a, b_n \rightarrow b \text{ e } a_n \geq b_n \text{ definitivamente, allora } a \geq b.$$

**DIMOSTRAZIONE.** Segue dal teorema precedente. Infatti, se per assurdo fosse  $a < 0$ , dal teorema precedente si avrebbe  $a_n < 0$  definitivamente, il che è incompatibile con l'ipotesi che sia  $a_n \geq 0$  definitivamente. Questo dimostra la prima affermazione del teorema. Applicando ora questa proprietà alla differenza  $a_n - b_n$  si ottiene anche la seconda affermazione. Infatti:

$$a_n \geq b_n \text{ definitivamente} \implies a_n - b_n \geq 0 \text{ definitivamente,}$$

e poiché  $a_n - b_n \rightarrow a - b$  (teorema sull'algebra dei limiti), si conclude  $a - b \geq 0$ , ovvero  $a \geq b$ .

L'ultima relazione significa che in una disegualanza tra due successioni si può passare al limite in ambo i membri, mantenendo il segno  $\leq$ . Si noti che in generale, invece, nel passaggio al limite non si conserva il segno di disegualanza *stretta*: ad esempio, anche se gli  $a_n$  sono strettamente positivi, il loro limite  $a$  è positivo o *nullo*, come mostra il semplice esempio di  $\frac{1}{n} \rightarrow 0$ .

È utile anche la seguente proposizione:

**TEOREMA 3.6 (DEL CONFRONTO)** Se  $a_n \leq b_n \leq c_n$  definitivamente e

$$a_n \rightarrow l, c_n \rightarrow l \in \mathbb{R},$$

allora anche  $b_n \rightarrow l$ .

**DIMOSTRAZIONE.** Fissiamo  $\varepsilon > 0$ . Allora definitivamente si ha

$$l - \varepsilon < a_n < l + \varepsilon; \quad l - \varepsilon < c_n < l + \varepsilon$$

da cui segue (definitivamente)

$$l - \varepsilon < a_n \leq b_n \leq c_n < l + \varepsilon$$

e quindi, definitivamente,

$$l - \varepsilon < b_n < l + \varepsilon$$

Dunque  $b_n \rightarrow l$ .

Casi particolari di questo teorema che si usano frequentemente sono espressi dal prossimo corollario, molto utile quando si studia il prodotto tra una successione oscillante (ma limitata) e una che tende a zero:

**COROLLARIO 3.7** 1. Se  $|b_n| \leq c_n$  definitivamente e  $c_n \rightarrow 0$ , allora anche  $b_n \rightarrow 0$ .  
2. Se  $c_n \rightarrow 0$  e  $b_n$  è limitata (ma non necessariamente convergente), allora  $c_n b_n \rightarrow 0$ . Detto a parole: il prodotto di una successione infinitesima e una limitata è infinitesimo.

**DIMOSTRAZIONE.**

1. Sappiamo che definitivamente è  $-c_n \leq b_n \leq c_n$ ; d'altro canto se  $c_n \rightarrow 0$  anche  $-c_n \rightarrow 0$ , quindi per il teorema precedente (con  $a_n = -c_n$  e  $l = 0$ ) si ha che  $b_n \rightarrow 0$ .
2. Se  $b_n$  è limitata, ossia  $|b_n| \leq K$  per un certo  $K > 0$  e per ogni  $n$ , possiamo scrivere

$$|b_n c_n| \leq K |c_n|.$$

Poiché  $c_n \rightarrow 0$ , anche  $K |c_n| \rightarrow 0$ , e per il punto 1 si conclude che  $b_n c_n \rightarrow 0$ .



Applicando la definizione si dimostri che

$$n^\alpha \rightarrow \begin{cases} +\infty & \text{se } \alpha > 0 \\ 1 & \text{se } \alpha = 0 \\ 0 & \text{se } \alpha < 0 \end{cases}$$

Infatti, se  $\alpha > 0$ , fissato  $M > 0$  risulta  $n^\alpha > M$  per  $n > M^{1/\alpha}$ , perciò  $n^\alpha \rightarrow +\infty$ ; se  $\alpha < 0$ , scrivendo  $n^\alpha = 1/n^{|\alpha|}$ , si osserva che  $1/n^{|\alpha|} < \varepsilon$  per  $n^{|\alpha|} > 1/\varepsilon$  ossia per  $n > 1/\varepsilon^{1/|\alpha|}$ , cioè definitivamente.

Limiti che si presentano nella forma di rapporto di due espressioni, ognuna costituita dalla somma di potenze di  $n$ :

$$\frac{n^{5/2} - 3n + 7}{n^3 + \sqrt{n} - 3n^2}$$

Si mette in evidenza a numeratore come a denominatore la potenza maggiore

$$\dots = \frac{n^{5/2} \left(1 - \frac{3}{n^{3/2}} + \frac{7}{n^{5/2}}\right)}{n^3 \left(1 + \frac{1}{n^{5/2}} - \frac{3}{n}\right)} = \frac{1}{\sqrt{n}} \cdot \left( \frac{1 - \frac{3}{n^{3/2}} + \frac{7}{n^{5/2}}}{1 + \frac{1}{n^{5/2}} - \frac{3}{n}} \right).$$

Ora per il teorema sull'algebra dei limiti e sapendo che potenze negative di  $n$  tendono a zero possiamo affermare che:

$$1 - \frac{3}{n^{3/2}} + \frac{7}{n^{5/2}} \rightarrow 1; \quad 1 + \frac{1}{n^{5/2}} - \frac{3}{n} \rightarrow 1.$$

Pertanto la successione tra parentesi tende a 1; d'altro canto  $\frac{1}{\sqrt{n}} \rightarrow 0$ , perciò la successione di partenza tende a zero.

La successione  $\frac{\sin n}{n}$  è il prodotto di  $\frac{1}{n}$  (convergente) e  $\sin n$  (irregolare); quindi il teorema sull'algebra dei limiti non è applicabile. Tuttavia è applicabile l'ultimo corollario: la successione  $\frac{1}{n}$  è infinitesima mentre  $|\sin n| \leq 1$ , perciò  $\frac{\sin n}{n} \rightarrow 0$ .

Fin qui abbiamo visto teoremi che operano su coppie di successioni entrambe convergenti o comunque limitate. Consideriamo ora il caso in cui i limiti sono  $+\infty$  o  $-\infty$ . Supponiamo, per esempio, che  $a_n \rightarrow a$  e  $b_n \rightarrow +\infty$ ; allora è facile (e intuitivo) vedere che  $a_n + b_n \rightarrow +\infty$ . Abbrevieremo questa scrittura così:  $a + \infty = +\infty$ . Ragionando in maniera analoga possiamo compendiare le regole per il limite della somma (o differenza) di due successioni delle quali una o entrambe sono divergenti con le scritture seguenti:

$a + \infty = +\infty$
$a - \infty = -\infty$
$+\infty + \infty = +\infty$
$-\infty - \infty = -\infty$

Analogamente per il prodotto (o il rapporto)abbiamo le regole seguenti; il segno di  $\infty$  va determinato con la usuale regola dei segni:

$a \cdot \infty = \infty \quad (a \neq 0)$
$\frac{a}{0} = \infty \quad (a \neq 0)$
$\frac{a}{\infty} = 0$

Esplicitamente, la seconda regola scritta significa, ad esempio, quanto segue:

$$\text{se } a_n \rightarrow a > 0 \text{ e } b_n \rightarrow 0^+, \text{ allora } \frac{a_n}{b_n} \rightarrow +\infty.$$

In modo analogo si procede, in base alla regola dei segni, se  $a < 0$  o  $b_n \rightarrow 0^-$ ; è comunque necessario, per applicare questo teorema, sapere che  $b_n$  tende a zero per eccesso o per difetto.

Le precedenti regole prendono il nome di “aritmetizzazione parziale del simbolo di infinito”.

**DIMOSTRAZIONE.** Proviamo ad esempio che:

$$a_n \rightarrow +\infty, b_n \rightarrow b \in \mathbb{R} \Rightarrow \frac{b_n}{a_n} \rightarrow 0.$$

Fissiamo  $\varepsilon > 0$ . Poiché  $a_n \rightarrow +\infty$ , definitivamente si ha

$$a_n > \frac{1}{\varepsilon}$$

Poiché  $b_n \rightarrow b$ , definitivamente si ha

$$|b_n| < |b| + \varepsilon$$

Ne segue che definitivamente

$$\left| \frac{b_n}{a_n} \right| < \varepsilon (|b| + \varepsilon) < \varepsilon \cdot \text{cost.}$$

Per l’arbitrarietà di  $\varepsilon$ , segue la tesi.

◆ Lo studente noterà che mancano le regole

relative a quattro operazioni, e cioè

$$+\infty - \infty, \quad 0 \cdot \infty, \quad \frac{0}{0}, \quad \frac{\infty}{\infty}$$

Queste espressioni si chiamano *forme di indecisione*, poiché nessuna regola può essere stabilita a priori per determinare il risultato, come vedremo negli esempi sotto illustrati.

Le regole sopra elencate (e la mancanza di regole per le forme di indecisione) confermano la natura particolare dei “punti”  $+\infty$  e  $-\infty$ , che non possono essere considerati “numeri” poiché non rispettano le proprietà  $R_1$  e  $R_2$  del capitolo 1 paragrafo 3.



◆  $\sqrt{n+1} - \sqrt{n-1}$ : forma di indecisione del tipo  $+\infty - \infty$ .

Moltiplicando e dividendo per  $\sqrt{n+1} + \sqrt{n-1}$  si trova

$$\dots = \frac{(\sqrt{n+1} - \sqrt{n-1})(\sqrt{n+1} + \sqrt{n-1})}{\sqrt{n+1} + \sqrt{n-1}} = \frac{2}{\sqrt{n+1} + \sqrt{n-1}} \rightarrow 0$$

◆ Limiti di successioni che si presentano nella forma  $a_n^{b_n}$  si trattano più facilmente considerando la successione dei logaritmi (per fissare le idee, in base 10)

$$\log_{10} a_n^{b_n} = b_n \log_{10} a_n$$

Si mostra che (cfr. anche più avanti, al paragrafo 3) se questa successione converge a  $l$ , diverge a  $+\infty$ ,  $-\infty$ , o è indeterminata, la successione  $a_n^{b_n}$ , rispettivamente, converge a  $10^l$ ,  $+\infty$ , 0, è indeterminata.

Si può così osservare che le espressioni

$$1^{\pm\infty} \quad 0^0 \quad (+\infty)^0$$

sono altrettante forme di indecisione, corrispondenti (passando al logaritmo) alla forma di indecisione  $0 \cdot \infty$ .

#### 1.4 Il numero e

Introdurremo ora un numero molto importante per l’Analisi, che sarà definito come limite di una particolare successione. Cominciamo a dimostrare il seguente risultato:

**TEOREMA 3.8** *La successione*

$$a_n = \left(1 + \frac{1}{n}\right)^n$$

è convergente.

Si osservi che questa successione presenta una forma di indecisione  $1^\infty$ .

**DIMOSTRAZIONE.** Si proverà che la successione è monotona crescente

$$a_{n+1} > a_n$$

e limitata:

$$2 \leq a_n < 4$$

e perciò è convergente, per il Teorema di monotonia (par. 1.2).

Per provare che  $a_n$  è monotona crescente, studiamo, per  $n \geq 2$ , il rapporto:

$$\begin{aligned} \frac{a_n}{a_{n-1}} &= \frac{\left(1 + \frac{1}{n}\right)^n}{\left(1 + \frac{1}{n-1}\right)^{n-1}} = \frac{\left(\frac{n+1}{n}\right)^n}{\left(\frac{n}{n-1}\right)^{n-1}} = \\ &= \left(\frac{\frac{n+1}{n}}{\frac{n}{n-1}}\right)^n \cdot \frac{1}{\left(\frac{n}{n-1}\right)^{-1}} = \left(\frac{n^2-1}{n^2}\right)^n \cdot \frac{1}{\left(\frac{n-1}{n}\right)} = \\ &= \frac{\left(1 - \frac{1}{n^2}\right)^n}{1 - \frac{1}{n}} \geq \frac{1 - n \cdot \frac{1}{n^2}}{1 - \frac{1}{n}} = 1, \end{aligned}$$

dove nell’ultima diseguaglianza scritta si è applicata la *diseguaglianza di Bernoulli* (v. cap. 1, par. 7)

$$(1+x)^n \geq 1 + nx$$

con  $x = -\frac{1}{n^2}$ . Questo mostra che  $\frac{a_n}{a_{n-1}} \geq 1$ , ossia  $a_n \geq a_{n-1}$ , e la successione è monotona crescente. In particolare, essendo  $a_1 = 2$ , segue  $a_n \geq 2 \forall n \geq 1$ .

Consideriamo ora la successione

$$b_n = \left(1 + \frac{1}{n}\right)^{n+1}.$$

Si noti che

$$b_n = a_n \cdot \left(1 + \frac{1}{n}\right),$$

perciò  $b_n > a_n$ . Con calcoli simili a quelli appena svolti (lasciamo i dettagli al lettore), si mostra che  $b_n \leq b_{n-1}$ . Poiché  $b_1 = 4$ , risulta quindi

$$a_n < b_n \leq b_1 = 4, \forall n \geq 1,$$

e  $a_n$  è limitata.



Il limite della successione  $a_n$  appena studiata è un numero (irrazionale) molto importante in matematica, per varie ragioni che vedremo; esso viene indicato con la lettera  $e$  (*numero di Nepero*) e la sua rappresentazione decimale inizia così:

$$2.7182818284\dots$$

Dunque, *per definizione*,

$$e = \lim_{n \rightarrow +\infty} \left(1 + \frac{1}{n}\right)^n.$$

Questo numero viene molto spesso usato come base dei logaritmi, i quali, quando si usa questa base, vengono detti *naturali* o *neperiani* (dal nome del matematico John Napier) e indicati semplicemente col simbolo  $\log$  (oppure  $\ln$ ) senza indicazione della base.



**Il numero di Nepero e un problema... finanziario.** Supponiamo di possedere un capitale (per esempio, 1 milione di euro) e di investirlo al tasso di interesse annuale  $t$  (cioè con una rendita di  $t$  milioni all'anno).

Se l'interesse viene pagato annualmente, dopo un anno il capitale posseduto sarà  $1+t$ . Se l'interesse viene calcolato mensilmente avremo:

- dopo il primo mese un capitale pari a  $1 + \frac{t}{12}$ ;
- dopo il secondo mese, pari a  $1 + \frac{t}{12} + \frac{t}{12}(1 + \frac{t}{12}) = (1 + \frac{t}{12})^2$ ;
- dopo il terzo mese, pari a  $(1 + \frac{t}{12})^2 \frac{t}{12} + (1 + \frac{t}{12})^2 = (1 + \frac{t}{12})^3$ ;
- alla fine dell'anno avremo un capitale pari a  $(1 + \frac{t}{12})^{12}$ .

Se l'interesse viene calcolato ogni  $n$ -esimo di anno, avremo alla fine un capitale pari a

$$\left(1 + \frac{t}{n}\right)^n$$

Per  $t = 1$  (rendita del 100%) otteniamo esattamente la successione che definisce  $e$ .

Far tendere  $n$  a  $\infty$  significa calcolare l'interesse dopo frazioni di anno sempre più piccole fino ad arrivare a calcolarlo con continuità. Il capitale che si ottiene alla fine dell'anno in quest'ultimo caso è esattamente pari a  $e$ .

Una volta *definito* il numero  $e$  come sopra, si può *dimostrare* che

$$(1.3) \quad e = \lim_{n \rightarrow +\infty} \left(1 - \frac{1}{n}\right)^{-n}.$$

Inoltre, a partire dalle due relazioni precedenti, si può provare il seguente

**TEOREMA 3.9** *Sia  $\{a_n\}$  una qualsiasi successione divergente ( $a \rightarrow +\infty$  o  $a \rightarrow -\infty$ ). Allora esiste*

$$\lim_{n \rightarrow +\infty} \left(1 + \frac{1}{a_n}\right)^{a_n} = e.$$

(Una traccia per la dimostrazione della (1.3) e di questo teorema sarà fornita nei Complementi alla fine del par. 1). Quest'ultimo teorema si rivela estremamente utile nel calcolo di limiti che coinvolgono la forma di indeterminazione  $[1^\infty]$ .



**Calcoliamo**

$$\lim_{n \rightarrow +\infty} \left(\frac{n}{3+n}\right)^{5n+1}.$$

Si tratta di una forma di indeterminazione  $[1^\infty]$ . Scriviamo:

$$\left(\frac{n}{3+n}\right)^{5n+1} = \frac{1}{\left(1 + \frac{3}{n}\right)^{5n+1}} = \frac{1}{\left[\left(1 + \frac{3}{n}\right)^{\frac{n}{3}}\right]^{\frac{3(5n+1)}{n}}} = \frac{1}{\left[\left(1 + \frac{1}{a_n}\right)^{a_n}\right]^{b_n}}$$

con  $a_n = \frac{n}{3}$ . Per il teorema precedente, la successione entro parentesi quadre tende ad  $e$ , mentre l'esponente

$$b_n = \frac{3(5n+1)}{n} \rightarrow 15$$

perciò il limite cercato è  $1/e^{15}$ .

Altre situazioni di questo tipo saranno illustrate negli esercizi alla fine del par. 1.

## 1.5 Confronti e stime asintotiche

Abbiamo visto che una successione che tende a 0 è un *infinitesimo*; una successione che diverge ( $a \rightarrow +\infty$ ,  $a \rightarrow -\infty$ ) si dice *infinito*. Quando due successioni sono entrambe infinitesimi o entrambe infiniti, è utile poter stabilire un confronto tra di esse, per capire quale delle due tenda “più rapidamente” a 0 o all’infinito.

Esempi di infiniti sono le successioni seguenti:

$$\{\log n\} \quad \{\sqrt{n}\} \quad \{n^2\} \quad \{2^n\}$$

Esempi di infinitesimi si ottengono dalle successioni precedenti considerando gli elementi reciproci.

Siano  $\{a_n\}$  e  $\{b_n\}$  due infiniti. Consideriamo il limite del rapporto  $a_n/b_n$ ; si hanno quattro possibilità:

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \begin{cases} 0 & \text{i)} \\ l \text{ finito e } l \neq 0 & \text{ii)} \\ \pm\infty & \text{iii)} \\ \text{inesistente} & \text{iv)} \end{cases}$$

- Diciamo che:
- $\{a_n\}$  è un infinito di *ordine inferiore* a  $\{b_n\}$
  - $\{a_n\}$  e  $\{b_n\}$  sono infiniti dello stesso *ordine*
  - $\{a_n\}$  è un infinito di *ordine superiore* a  $\{b_n\}$
  - $\{a_n\}$  e  $\{b_n\}$  non sono confrontabili.

Se  $\{a_n\}$  e  $\{b_n\}$  sono due infinitesimi (e  $b_n \neq 0$ ), ancora si considera il limite del rapporto  $a_n/b_n$  e, in corrispondenza delle 4 possibilità sopra elencate, diremo che:

- $\{a_n\}$  è un infinitesimo di *ordine superiore* a  $\{b_n\}$
- $\{a_n\}$  e  $\{b_n\}$  sono infinitesimi dello stesso *ordine*
- $\{a_n\}$  è un infinitesimo di *ordine inferiore* a  $\{b_n\}$
- $\{a_n\}$  e  $\{b_n\}$  non sono confrontabili.

Il caso  $\frac{a_n}{b_n} \rightarrow 1$  è particolarmente importante: si usa dire, in tal caso, che le due successioni  $\{a_n\}$  e  $\{b_n\}$  sono *asintotiche* e, per indicare questa circostanza, si scrive

$$a_n \sim b_n$$

(si legge:  $a_n$  è *asintotico* a  $b_n$ ).

Il simbolo di asintotico è molto utile nel calcolo dei limiti per le seguenti proprietà:

### PROPOSIZIONE 3.1

- Se  $a_n \sim b_n$ , le due successioni hanno lo stesso comportamento: convergono allo stesso limite, o divergono entrambe a  $\pm\infty$ , o entrambe non hanno limite.
- Si possono scrivere catene di relazioni asintotiche, cioè:

$$\text{se } a_n \sim b_n \sim \dots \sim c_n, \text{ allora } a_n \sim c_n$$

- Un'espressione composta da prodotto o quoziente di più fattori può essere stimata fattore per fattore:

$$\text{se } a_n \sim a'_n, b_n \sim b'_n, c_n \sim c'_n, \text{ allora } \frac{a_n b_n}{c_n} \sim \frac{a'_n b'_n}{c'_n}$$

Attenzione: lo stesso *non vale* per le somme o per l'esponenziale.

#### DIMOSTRAZIONE.

- Dimostriamo la prima affermazione: se  $a_n \sim b_n$ , le due successioni hanno lo stesso comportamento. Sia  $a_n \rightarrow l \in \mathbb{R}$ ; poiché  $b_n = \frac{b_n}{a_n} \cdot a_n$  e  $\frac{b_n}{a_n} \rightarrow 1$  (per definizione di asintotico), allora per il teorema sul limite del prodotto

$$b_n \rightarrow 1 \cdot l = l.$$

Lo stesso teorema (nel caso di limite infinito) permette di concludere che se  $a_n \rightarrow \pm\infty$  e  $b_n \sim a_n$ , anche  $b_n \rightarrow \pm\infty$ . Osserviamo ora che la relazione di asintotico è simmetrica,

quindi quanto appena provato mostra anche che se  $b_n$  converge (diverge), anche  $a_n$  converge (diverge). Ne concludiamo che se  $a_n$  è irregolare, anche  $b_n$  è irregolare, perché se per assurdo non lo fosse, per quanto appena affermato anche  $a_n$  sarebbe convergente o divergente.

2. Proviamo la transitività della relazione di asintotico:

$$\text{se } a_n \sim b_n \text{ e } b_n \sim c_n \text{ allora } a_n \sim c_n.$$

Le nostre ipotesi significano che

$$\frac{a_n}{b_n} \rightarrow 1 \text{ e } \frac{b_n}{c_n} \rightarrow 1.$$

Allora per il teorema sul limite del prodotto,

$$\frac{a_n}{c_n} = \frac{a_n}{b_n} \cdot \frac{b_n}{c_n} \rightarrow 1.$$

Analogamente si prova la terza proprietà. ♦ Un tipico modo per mostrare che  $a_n \sim b_n$  consiste nello scrivere  $a_n = b_n c_n$  con  $c_n \rightarrow 1$ . Ad esempio:

$$2n^2 + 3n + 1 = 2n^2 \left(1 + \frac{3}{n} + \frac{1}{2n^2}\right) \sim 2n^2$$

$$\text{perché } \left(1 + \frac{3}{n} + \frac{1}{2n^2}\right) \rightarrow 1.$$

**OSSERVAZIONE** Il fatto che la relazione di asintotico soddisfi le 3 proprietà:

- Riflessiva:  $a_n \sim a_n$ ;
- Simmetrica: se  $a_n \sim b_n$  allora  $b_n \sim a_n$ ;
- Transitiva: se  $a_n \sim b_n$  e  $b_n \sim c_n$  allora  $a_n \sim c_n$

si esprime dicendo che “asintotico” è una *relazione di equivalenza*. (Si noti che le prime due proprietà sono immediate, mentre la terza è stata dimostrata sopra). Più in generale, in matematica si chiama relazione d’equivalenza una relazione che soddisfa i 3 assiomi ora enunciati.

Mostriamo ora il seguente:

### TEOREMA 3.10 (GERARCHIA DEGLI INFINITI)

$$\lim_{n \rightarrow +\infty} \frac{\log_a n}{n^\alpha} = 0$$

$$\lim_{n \rightarrow +\infty} \frac{n^\alpha}{a^n} = 0$$

per ogni  $a > 1, \alpha > 0$ .

Questi limiti descrivono la “velocità” con cui i logaritmi (con base  $> 1$ ), le potenze, gli esponenziali (con base  $> 1$ ) vanno all’ $\infty$ : i logaritmi vanno più lentamente di qualsiasi potenza, le potenze più lentamente di qualsiasi esponenziale a base  $> 1$ .

**DIMOSTRAZIONE.** Per dimostrare la prima relazione, iniziamo a stabilire un'utile diseguaglianza tra un qualsiasi numero reale positivo e il suo logaritmo. Per  $x \in \mathbb{R}, x > 0$ , sia  $k$  la parte intera di  $x$  (v. cap. 2, par. 3.6), ossia l'intero  $k \geq 0$  per cui è  $k \leq x < k + 1$ . Si ha:

$$2^x \geq 2^k = (1+1)^k \geq 1+k > x$$

dove: la prima diseguaglianza segue dalla monotonia della funzione esponenziale, la seconda dallo sviluppo del binomio di Newton (o dalla diseguaglianza di Bernoulli). Passando ai logaritmi in base  $a$ , otteniamo

$$\log_a x < x \log_a 2$$

per ogni  $x > 0$ . Applichiamo ora questa diseguaglianza al numero  $x = n^{\alpha/2}$  e abbiamo:

$$\frac{\alpha}{2} \log_a n < n^{\alpha/2} \log_a 2$$

e quindi

$$\frac{\log_a n}{n^{\alpha/2}} < \frac{2}{\alpha} \log_a 2$$

$$\frac{\log_a n}{n^\alpha} = \frac{\log_a n}{n^{\alpha/2}} \cdot \frac{1}{n^{\alpha/2}} \leq \frac{2}{\alpha} \log_a 2 \cdot \frac{1}{n^{\alpha/2}}.$$

Per il teorema del confronto,  $\frac{\log_a n}{n^{\alpha/2}} \rightarrow 0$ , e la prima relazione è dimostrata.

Applichiamo ora questo risultato sostituendo all'intero  $n$  l'intero  $2^n$ . Avremo:

$$0 = \lim_{n \rightarrow \infty} \frac{\log_a (2^n)}{(2^n)^\alpha} = \lim_{n \rightarrow \infty} \frac{n \log_a 2}{(2^\alpha)^n}.$$

Se ora  $a > 1$  è fissato, scegliendo  $\alpha > 0$  in modo che sia  $2^\alpha = a$  otteniamo che  $\frac{n}{a^n} \rightarrow 0$ , che è la seconda relazione nel caso particolare in cui  $n$  è elevato ad esponente 1. Il caso generale segue dall'identità:

$$\frac{n^\alpha}{a^n} = \left(\frac{n}{a^{n/\alpha}}\right)^\alpha = \left(\frac{n}{(a^{1/\alpha})^n}\right)^\alpha.$$

Infatti per il risultato precedente  $\frac{n}{(a^{1/\alpha})^n} \rightarrow 0$  (la base  $a^{1/\alpha}$  è ancora un numero  $> 1$ ), quindi anche  $\left(\frac{n}{(a^{1/\alpha})^n}\right)^\alpha \rightarrow 0$ .

Altri confronti tra infiniti si possono risolvere grazie al seguente criterio:

**TEOREMA 3.11 (CRITERIO DEL RAPPORTO)** Sia  $a_n$  una successione positiva (cioè  $a_n > 0$  per ogni  $n$ ). Se esiste

$$\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = l$$

e  $l < 1$ , allora  $a_n \rightarrow 0$ ; se  $l > 1$  (ed eventualmente  $l = +\infty$ ), allora  $a_n \rightarrow +\infty$ .

Il teorema precedente riconduce lo studio del carattere di una successione positiva,  $a_n$ , al calcolo del limite di un'altra successione, la successione (dei rapporti),  $\frac{a_{n+1}}{a_n}$ . In certi casi quest'ultima è più semplice da studiare di quella di partenza, come vedremo negli esempi. Si osservi che nel caso  $l = 1$  il teorema non permette di concludere nulla.

**DIMOSTRAZIONE.** Supponiamo che

$$\frac{a_{n+1}}{a_n} \rightarrow l < 1.$$

Fissato  $\varepsilon > 0$ , definitivamente, ossia per ogni  $n \geq n_0$  (con  $n_0$  opportuno) si ha  $\frac{a_{n+1}}{a_n} < l + \varepsilon$ . Scegliamo  $\varepsilon$  abbastanza piccolo perché si abbia  $l + \varepsilon < 1$ . Possiamo scrivere la catena di diseguaglianze:

$$\begin{aligned} a_{n_0+1} &< (l + \varepsilon) a_{n_0}; \\ a_{n_0+2} &< (l + \varepsilon) a_{n_0+1} < (l + \varepsilon)^2 a_{n_0} \\ &\dots \\ a_{n_0+k} &< (l + \varepsilon)^k a_{n_0}. \end{aligned}$$

Poiché  $(l + \varepsilon) < 1$ , si ha  $(l + \varepsilon)^k \rightarrow 0$  per  $k \rightarrow \infty$ . D'altro canto  $a_{n_0}$  è fissato; dunque per  $k$  abbastanza grande il secondo membro (e quindi il primo) è piccolo quanto si vuole, il che dimostra che  $a_n \rightarrow 0$ .

Supponiamo ora che  $\frac{a_{n+1}}{a_n} \rightarrow l > 1$ . Fissato  $\varepsilon > 0$ , definitivamente si ha  $\frac{a_{n+1}}{a_n} > l - \varepsilon$ . Scegliamo  $\varepsilon$  abbastanza piccolo perché si abbia  $l - \varepsilon > 1$ . Analogamente a prima, possiamo ottenere la diseguaglianza:

$$a_{n_0+k} > (l - \varepsilon)^k a_{n_0}$$

per un certo  $n_0$  fissato e qualsiasi  $k$ . Poiché  $(l - \varepsilon) > 1$  e quindi  $(l - \varepsilon)^k \rightarrow +\infty$ , si conclude che  $a_n \rightarrow +\infty$ .



**Dimostriamo che**

$$\lim_{n \rightarrow \infty} \frac{b^n}{n!} = 0 \text{ per ogni } b > 0.$$

Applichiamo il criterio del rapporto alla successione  $a_n = \frac{b^n}{n!}$ . Si ha:

$$\frac{a_{n+1}}{a_n} = \frac{b^{n+1}}{(n+1)!} \cdot \frac{n!}{b^n} = \frac{b}{n+1} \rightarrow 0.$$

Per il criterio del rapporto allora,  $\frac{b^n}{n!} \rightarrow 0$ . Abbiamo quindi ottenuto un nuovo caso nella gerarchia degli infiniti.



**Si provi a calcolare**

$$\lim_{n \rightarrow \infty} \frac{\log n}{n}$$

col criterio del rapporto, osservando che il metodo fallisce.

Ecco alcuni esempi di come si applicano tutte le osservazioni precedenti per risolvere alcune forme di indeterminazione. Quando avremo studiato un certo numero di *limiti notevoli* (par. 3.2) potremo risolvere mediante stime asintotiche situazioni più complesse di queste.



$$\lim_{n \rightarrow +\infty} \frac{2n^3 + 4n + 1}{5(n+1)^3} = \left[ \frac{\infty}{\infty} \right]$$

Considerando solo le potenze di grado massimo a numeratore e denominatore, possiamo scrivere:

$$\frac{2n^3 + 4n + 1}{5(n+1)^3} \sim \frac{2n^3}{5n^3} = \frac{2}{5}$$

e pertanto la successione tende a  $2/5$ .

$\lim_{n \rightarrow +\infty} \frac{2^n + n}{2^{n+1}} = \left[ \frac{\infty}{\infty} \right].$

$2^n$  è un infinito di ordine superiore rispetto ad  $n$ ; possiamo scrivere quindi:

$$2^n + n = 2^n \left(1 + \frac{n}{2^n}\right) \sim 2^n$$

perché  $\frac{n}{2^n} \rightarrow 0$ , e quindi  $\left(1 + \frac{n}{2^n}\right) \rightarrow 1$ . Pertanto

$$\frac{2^n + n}{2^{n+1}} \sim \frac{2^n}{2^{n+1}} = \frac{1}{2}$$

e il limite è  $\frac{1}{2}$ .

$\lim_{n \rightarrow +\infty} \sqrt[n]{n} = [\infty^0]$

Scriviamo

$$\sqrt[n]{n} = n^{1/n} = e^{\log n^{1/n}} = e^{\frac{\log n}{n}}.$$

Usando il confronto di infiniti,  $\frac{\log n}{n} \rightarrow 0$ , e quindi si deduce

$$\lim_{n \rightarrow +\infty} \sqrt[n]{n} = e^0 = 1.$$

$\lim_{n \rightarrow \infty} \frac{n!}{n^n} = \left[ \frac{\infty}{\infty} \right].$

Applicando il criterio del rapporto, consideriamo:

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)!}{(n+1)^{n+1}} \cdot \frac{n^n}{n!} = \frac{(n+1)n^n}{(n+1)^n \cdot (n+1)} = \left(\frac{n}{n+1}\right)^n = \frac{1}{\left(1 + \frac{1}{n}\right)^n} \rightarrow \frac{1}{e} < 1,$$

dove nell'ultimo passaggio si è usata la definizione di  $e$ ; perciò  $a_n \rightarrow 0$ . Abbiamo quindi provato che  $n!$  è un infinito di ordine inferiore rispetto a  $n^n$ .



1 Dare esempi di “infiniti” di ordine inferiore a  $\{\log n\}$  e di ordine superiore a  $\{2^n\}$ .

2 Provare che

$$\log(n+1) \sim \log n$$

3 Calcolare

$$\lim_{n \rightarrow +\infty} \frac{a_{n+1}}{a_n}$$

per le seguenti successioni:

$$a_n = n^n \quad a_n = n! \quad a_n = 2^n \quad a_n = n^2$$

4 Dare una stima asintotica delle seguenti successioni, mediante una successione “più semplice”, e calcolare quindi il limite:

$$\frac{n^3 + 2n^2 + \sin n}{n + \log n} \quad \frac{n^2 \log n + n}{\log 3n} \quad \frac{n! - (n-1)!}{n + (n-2)!}$$

Lo studente, utilizzando una normale calcolatrice tascabile o un PC, calcoli  $\left(1 + \frac{1}{n}\right)^n$ . Troverà, per esempio

$n$	$(1 + 1/n)^n$
$10^3$	2,7169239
$10^4$	2,7181459
$\vdots$	$\vdots$
$10^{11}$	1
$10^{12}$	1

Cerchi di spiegarne la ragione.

Un altro esempio: il limite

$$\lim_{n \rightarrow +\infty} (\sqrt{n^2 + 1} - n)n$$

si presenta sotto la forma di indecisione  $\infty - \infty$ . Moltiplicando e dividendo l'espressione per  $\sqrt{n^2 + 1} + n$  tale limite prende la forma

$$\lim_{n \rightarrow +\infty} \frac{n}{\sqrt{n^2 + 1} + n} = \frac{1}{2}$$

Se si usa una calcolatrice per il calcolo del limite si ottiene (per esempio)

$n$	$(\sqrt{n^2 + 1} - n)n$	$n/(\sqrt{n^2 + 1} + n)$
$10^2$	0,4999875	0,4999875
$10^3$	0,5	0,499999
$\vdots$	$\vdots$	$\vdots$
$10^7$	0	0,5
$10^8$	0	0,5

Spiegare la ragione del diverso risultato.

Calcolare i limiti, per  $n \rightarrow +\infty$ , delle successioni seguenti:

$$\begin{array}{lll} \sqrt{n+1} - \sqrt{n} & \sqrt{n^2 + n} - n & (\sqrt{n^4 + 1} - n^2)n \\ \log(n+1) - \log n & \left(\frac{n}{n+1}\right)^n & \left(1 + \frac{1}{n^2}\right)^n \quad \left(1 + \frac{1}{n^7}\right)^{n^9} \end{array}$$

## COMPLEMENTI

1 Dimostrare il teorema sul limite del quoziente: se  $a_n \rightarrow a, b_n \rightarrow b$ , allora  $\frac{a_n}{b_n} \rightarrow \frac{a}{b}$  purché  $b_n, b \neq 0$ .

Suggerimento: per maggiorare  $\left|\frac{a_n}{b_n} - \frac{a}{b}\right|$  fare denominatore comune; ora il numeratore si maggiorna in modo simile a quello visto nella dimostrazione del teorema sul limite del prodotto; per il denominatore occorre invece dimostrare che, ad esempio,  $|b_n| \geq |b|/2$  definitivamente, e poi...

**2** Dimostrare le relazioni che abbiamo chiamato “aritmetizzazione parziale di infinito”, nei casi che non sono stati dimostrati.

**3** Dimostrare che i valori assunti dalla successione  $\sin n$  sono tutti diversi tra loro, ossia che  $n, m \in \mathbb{N}, n \neq m$  implica  $\sin n \neq \sin m$ .

**4** Provare che

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{1}{n}\right)^{-n} = e$$

*Suggerimento:*  $(1 - \frac{1}{n})^{-n} = \left(\frac{n}{n-1}\right)^n = \dots$ ; ricondursi al limite che definisce  $e$ .

**5** Provare il Teorema 3.9 enunciato nel par. 1.4, ad esempio nel caso  $a_n \rightarrow +\infty$ .

*Suggerimento:* detta  $[a_n]$  la parte intera di  $a_n$ , provare anzitutto le disuguaglianze:

$$\left(1 + \frac{1}{[a_n] + 1}\right)^{[a_n]} \leq \left(1 + \frac{1}{a_n}\right)^{a_n} \leq \left(1 + \frac{1}{[a_n]}\right)^{[a_n] + 1}$$

Quindi sfruttando opportunamente il fatto che  $[a_n]$  è intero, ricondursi al limite già noto di  $\left(1 + \frac{1}{n}\right)^n$ .

**6** Dimostrare la proprietà 3 del simbolo di  $\sim$  enunciata nel paragrafo 1.5 (Proposizione 3.1), usando la definizione di “asintotico” e di limite.

**7** Si faccia un esempio di due successioni  $a_n, b_n$  tendenti a  $+\infty$ , per cui si ha:

$$a_n \sim b_n \text{ ma } e^{a_n} \text{ non asintotico a } e^{b_n}.$$

Dunque il simbolo di asintotico non si può usare con gli esponenziali come si userebbe nei prodotti o quozienti.

*Suggerimento:* scegliere come  $a_n$  la somma di due infiniti di tipo diverso.

## 2 LIMITI DI FUNZIONI, CONTINUITÀ, ASINTOTI

L’operazione di limite si può estendere dalle successioni alle funzioni. Potremo così precisare il comportamento di una funzione quando la variabile indipendente si muove vicino a un determinato punto oppure diventa molto grande (in valore assoluto).

In questo paragrafo introdurremo i concetti fondamentali riguardanti i limiti e la continuità di funzioni; nel prossimo paragrafo 3 svilupperemo i teoremi e gli strumenti che permetteranno il calcolo effettivo dei limiti; infine, nel paragrafo 4 approfondiremo lo studio delle proprietà delle funzioni continue, incontrando alcuni dei più significativi teoremi dell’analisi delle funzioni di una variabile. Nei capitoli successivi, mediante l’operazione di limite introdurremo i concetti di derivata, di differenziale e di integrale per una funzione reale di variabile reale.

Consideriamo come caso tipico un intervallo  $I$ , un punto  $c \in I$  e una funzione  $f$  a valori reali, definita in  $I$ , salvo al più nel punto  $c$ . L’intervallo  $I$  può essere limitato o illimitato, chiuso o aperto; il punto  $c$  può essere interno all’intervallo oppure uno dei suoi estremi (eventualmente  $+\infty$  o  $-\infty$ ).

Prendiamo ora una qualunque successione di punti  $x_n$  ( $n = 1, 2, \dots$ ), nell’intervallo  $I$  e diversi da  $c$ , che tenda a  $c$ , per  $n \rightarrow +\infty$ .

In corrispondenza alla successione di ingressi  $x_n$ , consideriamo la successione delle uscite  $f(x_n)$ .

Se, qualunque sia la successione scelta, si ha che  $f(x_n)$  tende al limite  $\ell$  (finito o infinito) si dice che il *limite di  $f(x)$  per  $x$  che tende a  $c$  è  $\ell$*  e si scrive

$$\lim_{x \rightarrow c} f(x) = \ell \text{ oppure } f(x) \rightarrow \ell \text{ per } x \rightarrow c$$

(che si leggono: “il limite per  $x$  tendente a  $c$  di  $f(x)$  è uguale a  $\ell$ ” oppure “ $f(x)$  tende a  $\ell$  per  $x$  tendente a  $c$ ”).

Sinteticamente:

**DEFINIZIONE 3.8 (SUCCESSIONALE DI LIMITE)** Si dice che

$$\lim_{x \rightarrow c} f(x) = \ell$$

(dove  $c, \ell \in \mathbb{R}^*$ ) se per ogni successione  $\{x_n\}$  di punti di  $I$  diversi da  $c$ , tale che  $x_n \rightarrow c$ , si ha che  $f(x_n) \rightarrow \ell$  per  $n \rightarrow \infty$ .

Una funzione che per  $x \rightarrow c$  tende a 0 si dice *infinitesima* per  $x \rightarrow c$ ; analogamente una funzione che tende a  $\pm\infty$  si dice *infinita*.

La definizione appena data di *limite di funzione* si dice definizione *successionale* di limite, in quanto riconduce il concetto di limite di funzione a quello di limite di successione. Non è l’unica definizione possibile, come vedremo, e a prima vista può sembrare una definizione poco operativa. Tuttavia, come dimostreranno sia il successivo sviluppo della teoria, sia gli esempi che faremo di calcolo di limiti, il fatto di avere già sviluppato le basi del calcolo dei limiti per le successioni, renderà molto vantaggiosa questa definizione, proprio dal punto di vista operativo. Ad esempio, possiamo subito affermare che vale il:

**TEOREMA 3.12 (DI UNICITÀ DEL LIMITE)** Se esiste  $\lim_{x \rightarrow c} f(x) = \ell$ , tale limite  $\ell$  è unico.

Infatti, se esistessero due limiti  $\ell_1, \ell_2$  diversi tra loro, presa una qualsiasi successione  $x_n \rightarrow c$  si avrebbe

$$f(x_n) \rightarrow \ell_1 \text{ e } f(x_n) \rightarrow \ell_2$$

dunque la successione  $f(x_n)$  avrebbe due limiti distinti, assurdo.

Osserveremo ora un po’ più da vicino le varie situazioni che si possono verificare nella definizione di limite di funzione. È subito chiaro che la casistica sarà più ricca e complessa rispetto al caso delle successioni, perché mentre per una successione  $a_n$  il limite si calcola necessariamente per  $n \rightarrow +\infty$ , per una funzione  $f(x)$  il limite si può calcolare per  $x \rightarrow c$  dove  $c$  può essere qualsiasi elemento di  $\mathbb{R}^*$ .

Anzitutto chiariamo la terminologia: nella scrittura

$$\lim_{x \rightarrow c} f(x) = \ell$$

parleremo di

$$\text{limite } \begin{cases} \text{finito} & \text{se } \begin{cases} \ell \in \mathbb{R} \\ l = +\infty, l = -\infty \end{cases} \\ \text{infinito} & \text{se } \begin{cases} \ell \in \mathbb{R} \\ l = +\infty, l = -\infty \end{cases} \end{cases}$$

e parleremo di

$$\limite \begin{cases} \text{al finito} & \text{se } \begin{cases} c \in \mathbb{R} \\ c = +\infty, \quad c = -\infty \end{cases} \\ \text{all'infinito} & \end{cases}$$

I prossimi esempi hanno lo scopo sia di mostrare come si utilizza concretamente la definizione successionale di limite, sia di illustrare le varie situazioni tipiche che si possono realizzare.

### Limite finito all'infinito



Dimostriamo che:

$$\lim_{x \rightarrow -\infty} e^x = 0.$$

Applicando la definizione, si tratta di provare che per ogni successione  $\{x_n\}$  tale che  $x_n \rightarrow -\infty$ , si ha

$$\lim_{n \rightarrow +\infty} e^{x_n} = 0.$$

A sua volta, per definizione di limite di successione, questo significa provare che fissato  $\varepsilon > 0$  qualsiasi, risulti

$$|e^{x_n}| < \varepsilon \text{ definitivamente,}$$

ossia (essendo l'esponenziale sempre positivo),

$$e^{x_n} < \varepsilon \text{ definitivamente.}$$

L'ultima diseguaglianza è equivalente a:

$$x_n < \log \varepsilon.$$

Se  $\varepsilon > 0$  è un numero piccolo ( $< 1$ ),  $\log \varepsilon$  sarà un numero negativo (e grande in valore assoluto); poniamo  $K = -\log \varepsilon > 0$ . Dobbiamo provare che, fissato questo numero  $K > 0$ , risulta  $x_n < -K$  definitivamente; ma questo è proprio ciò che vale per ipotesi, perché  $x_n \rightarrow -\infty$ . Questo conclude la dimostrazione.

### Asintoto orizzontale

Si dice che  $f$  ha un *asintoto orizzontale* di equazione  $y = \ell$  ( $\ell \in \mathbb{R}$ ) per  $x \rightarrow +\infty$  oppure per  $x \rightarrow -\infty$  se  $\lim_{x \rightarrow +\infty} f(x) = \ell$  oppure  $\lim_{x \rightarrow -\infty} f(x) = \ell$ , rispettivamente.

Ogni situazione di limite finito all'infinito, quindi, corrisponde graficamente alla presenza di un asintoto orizzontale, ossia di una retta orizzontale a cui il grafico della funzione si avvicina sempre più (come nell'esempio precedente).

### Limite per eccesso e per difetto

Come abbiamo visto per le successioni (par. 1.1), quando una funzione ha limite finito  $\ell$ , talvolta è possibile precisare se questo limite viene raggiunto *per eccesso* ( $\ell^+$ ) o *per difetto* ( $\ell^-$ ). Graficamente, questo significa che il grafico della funzione si avvicina alla quota  $y = \ell$  dall'alto o dal basso. La definizione precisa è la seguente:

**DEFINIZIONE 3.9 (LIMITE PER ECCESSO O PER DIFETTO)** Se  $\ell \in \mathbb{R}$  e  $c \in \mathbb{R}^*$ , si dice che

$$\lim_{x \rightarrow c} f(x) = \ell^+ \text{ (rispettivamente, } \ell^-),$$

e in tal caso si dice che  $f(x)$  tende a  $\ell$  per eccesso (per difetto) per  $x$  tendente a  $c$ , se per ogni successione  $\{x_n\}$  di punti di  $I$  diversi da  $c$ , tale che  $x_n \rightarrow c$ , si ha che  $f(x_n) \rightarrow \ell^+$  (rispettivamente  $\ell^-$ ) per  $n \rightarrow \infty$ . Ricordiamo che, come visto nel paragrafo 1.1, affermare che  $f(x_n) \rightarrow \ell^+$  significa che  $f(x_n) \rightarrow \ell$  e inoltre  $f(x_n) \geq \ell$  definitivamente.

Ad esempio,

$$\lim_{x \rightarrow -\infty} e^x = 0^+.$$

Un errore comune consiste nel pensare che il simbolo  $0^+$  denoti un (misterioso) numero diverso da zero e "poco più grande di 0". Sbagliato: il limite della funzione è il "solito" numero zero che conosciamo; semplicemente, la scrittura  $0^+$  aggiunge una informazione: la funzione tende a zero per eccesso, ossia i valori di  $f(x)$  tendono a zero mantenendosi non negativi.

Si osservi che non ogni limite finito è necessariamente assunto per eccesso o per difetto, come mostra la figura 3.5. Questa funzione tende a 1 per  $x \rightarrow +\infty$ , ma non si può affermare che  $f(x) \rightarrow 1^+$ , né che  $f(x) \rightarrow 1^-$ .

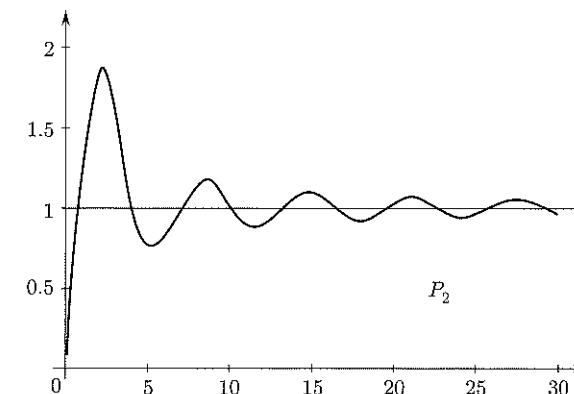


Figura 3.5.

### Limite infinito all'infinito



Dimostriamo che:

$$\lim_{x \rightarrow +\infty} \log_{1/2} x = -\infty.$$

Dobbiamo provare che per ogni successione  $\{x_n\}$  tale che  $x_n \rightarrow +\infty$ , si ha

$$\lim_{n \rightarrow +\infty} \log_{1/2} x_n = -\infty.$$

A sua volta, questo significa provare che fissato  $K > 0$  qualsiasi, risulti

$$\log_{1/2} x_n < -K \text{ definitivamente,}$$

il che equivale a:

$$x_n > \left(\frac{1}{2}\right)^{-K} = 2^K \text{ definitivamente.}$$

Ma per ipotesi  $x_n \rightarrow +\infty$ , quindi fissata la quantità positiva  $2^K$ , certamente risulta  $x_n > 2^K$  definitivamente, e l'asserto è dimostrato.

### Asintoto obliquo

Nei casi in cui una funzione presenta limite infinito all'infinito, può accadere (ma non sempre accade) che esista una retta, obliqua, a cui il grafico della funzione si avvicina indefinitamente. Si parla in tal caso di *asintoto obliquo*. Precisamente: si dice che una funzione  $f(x)$  ha *asintoto obliquo*  $y = mx + q$  ( $m \neq 0, q \in \mathbb{R}$ ) per  $x \rightarrow +\infty$  (o per  $x \rightarrow -\infty$ ) se

$$\lim_{x \rightarrow +\infty (-\infty)} [f(x) - (mx + q)] = 0.$$



Sia  $f(x) = 2x + 1 + e^x$ . Studiamo la funzione per  $x \rightarrow -\infty$ . Si vede facilmente che  $f(x) \rightarrow -\infty$ ; inoltre,

$$\lim_{x \rightarrow -\infty} [f(x) - (2x + 1)] = \lim_{x \rightarrow -\infty} e^x = 0.$$

Perciò, per definizione di asintoto obliquo, si riconosce che la retta  $y = 2x + 1$  è asintoto obliquo per  $f$ , per  $x \rightarrow -\infty$ .

In casi meno elementari, anziché dover "indovinare" qual è l'asintoto obliquo è utile avere un criterio operativo per cercarlo. Vale in proposito la seguente:

**PROPOSIZIONE 3.2** La funzione  $f(x)$  ammette asintoto obliquo per  $x \rightarrow +\infty$  se e solo se valgono le seguenti due condizioni:

1. Esiste finito

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{x} = m \neq 0;$$

2. Esiste finito

$$\lim_{x \rightarrow +\infty} [f(x) - mx] = q$$

(dove  $m$  è il numero calcolato al punto 1). In tal caso l'asintoto è  $y = mx + q$ . Analogico criterio vale per  $x \rightarrow -\infty$ .

Si lascia per esercizio la facile dimostrazione. Vediamo un semplice esempio di applicazione; altri, più elaborati, si potranno affrontare dopo aver approfondito, nella prossima sezione, i metodi di calcolo dei limiti.



Studiamo la funzione  $f(x) = 3x + \sqrt{x}$  per  $x \rightarrow +\infty$ . Si vede subito che  $f(x) \rightarrow +\infty$  per  $x \rightarrow +\infty$ . Vediamo se presenta un asintoto obliquo. Calcoliamo perciò (in base alla Proposizione precedente):

$$\lim_{x \rightarrow +\infty} \frac{3x + \sqrt{x}}{x} = \lim_{x \rightarrow +\infty} \left(3 + \frac{1}{\sqrt{x}}\right) = 3 \equiv m$$

(come si verifica facilmente applicando la definizione di limite). Calcoliamo ora

$$\lim_{x \rightarrow +\infty} [f(x) - mx] = \lim_{x \rightarrow +\infty} [(3x + \sqrt{x}) - 3x] = \lim_{x \rightarrow +\infty} \sqrt{x} = +\infty.$$

Poiché questo limite è infinito (non esiste  $q$ ), la funzione *non ammette* asintoto obliquo.

### Limite infinito al finito



Dimostriamo che:

$$\lim_{x \rightarrow 0} \frac{1}{x^2} = +\infty.$$

Dobbiamo provare che per ogni successione  $\{x_n\}$  tale che  $x_n \rightarrow 0$  e  $x_n \neq 0 \forall n$  si ha

$$\lim_{n \rightarrow +\infty} \frac{1}{x_n^2} = +\infty.$$

(Si noti che la precisazione " $x_n \neq 0 \forall n$ ", che è contenuta nella definizione di limite, era irrilevante negli esempi precedenti in cui  $c = \pm\infty$ , ma diventa ora importante). A sua volta, questo significa provare che fissato  $K > 0$  qualsiasi, risulti

$$\frac{1}{x_n^2} > K \text{ definitivamente,}$$

il che equivale a:

$$|x_n| < \frac{1}{\sqrt{K}} \text{ definitivamente.}$$

Ma per ipotesi  $x_n \rightarrow 0$ , quindi fissata la quantità positiva  $\varepsilon = \frac{1}{\sqrt{K}}$ , certamente risulta  $|x_n| < \varepsilon$  definitivamente, e l'asserto è dimostrato.

### Limite destro e sinistro

Talvolta una funzione si comporta diversamente (dal punto di vista del suo limite) a seconda che  $x$  si avvicini a  $c$  da destra o da sinistra. Ad esempio, per  $x \rightarrow 0$  la funzione  $1/x$  diventa sempre più grande in valore assoluto, ma con un segno diverso a seconda che sia  $x \geq 0$ . Per descrivere questo tipo di situazioni, si introduce il concetto di limite destro e sinistro:

**DEFINIZIONE 3.10 (LIMITE DESTRO E SINISTRO)** Se  $c \in \mathbb{R}$  e  $\ell \in \mathbb{R}^*$ , si dice che

$$\lim_{x \rightarrow c^+} f(x) = \ell \text{ (rispettivamente, } \lim_{x \rightarrow c^-} f(x) = \ell\text{),}$$

e in tal caso si dice che il limite destro (sinistro) di  $f(x)$  per  $x$  tendente a  $c$  è  $\ell$ , se per ogni successione  $\{x_n\}$  di punti di  $I$  tali che  $x_n \rightarrow c^+$  ( $x_n \rightarrow c^-$ ) per  $n \rightarrow +\infty$  e  $x_n \neq c \forall n$ , si ha che  $f(x_n) \rightarrow \ell$  per  $n \rightarrow \infty$ .

Ad esempio,

$$\lim_{x \rightarrow 0^+} \frac{1}{x} = +\infty \text{ e } \lim_{x \rightarrow 0^-} \frac{1}{x} = -\infty,$$

come si verifica facilmente. Si noti che, invece,

$$\lim_{x \rightarrow 0} \frac{1}{x} \text{ non esiste!}$$

Più in generale:

il limite  $\lim_{x \rightarrow c} f(x) = \ell$  esiste se e solo se esistono, e sono entrambi uguali a  $\ell$ , i limiti destro e sinistro  $\lim_{x \rightarrow c^+} f(x)$ ,  $\lim_{x \rightarrow c^-} f(x)$ ; tuttavia, può accadere che i limiti destro e sinistro esistano ma siano diversi tra loro, oppure uno solo dei due esista: in questi casi  $\lim_{x \rightarrow c} f(x)$  non esiste.

Se  $f$  è definita in  $(a, b)$ , il limite per  $x \rightarrow a$  (rispettivamente  $x \rightarrow b$ ) è automaticamente un limite destro (rispettivamente sinistro).

#### Asintoto verticale

Si dice che  $f$  ha un *asintoto verticale* di equazione  $x = c$  ( $c \in \mathbb{R}$ ) per  $x \rightarrow c$  (oppure per  $x \rightarrow c^+$  o  $x \rightarrow c^-$ ) se  $\lim_{x \rightarrow c} f(x) = +\infty$  o  $-\infty$  (oppure questo accade per  $x \rightarrow c^+$  o  $x \rightarrow c^-$ , rispettivamente).

Ogni situazione di limite infinito al finito, quindi, corrisponde graficamente alla presenza di un asintoto verticale, ossia di una retta verticale a cui il grafico della funzione si avvicina sempre più. Ad esempio, diciamo che

$x = 0$  è asintoto verticale per  $\frac{1}{x^2}$  (per  $x \rightarrow 0$ );

$x = 0$  è asintoto verticale per  $\frac{1}{x}$  (per  $x \rightarrow 0^+$  e per  $x \rightarrow 0^-$ );

$x = 0$  è asintoto verticale per  $\log x$  (per  $x \rightarrow 0^+$ ).

Il lettore verifichi per esercizio l'ultima affermazione.

#### Limite finito al finito

Questo è, in un certo senso, il caso meno intuitivo dei 4 finora trattati; vediamo perché.



Chiediamoci quanto vale:

$$\lim_{x \rightarrow 0} \sin x.$$

Chiunque, probabilmente, risponderebbe: zero. Dopotutto,  $\sin 0 = 0$ . Occorre rendersi conto, però, che anche se il risultato è proprio questo, la *motivazione* è errata. Infatti, la definizione di limite richiede che:

$\forall$  successione  $\{x_n\}$  tale che  $x_n \neq 0 \ \forall n$  e  $x_n \rightarrow 0$  per  $n \rightarrow \infty$ , si abbia  $\sin x_n \rightarrow 0$  per  $n \rightarrow \infty$ .

Quindi, per mostrare che  $\lim_{x \rightarrow 0} \sin x = 0$  non si deve calcolare  $\sin 0$ ! Si può procedere così, invece: dalla diseguaglianza elementare

$$|\sin x| \leq |x|, \text{ valida } \forall x \in \mathbb{R}$$

(si ragioni sulla circonferenza trigonometrica per rendersene conto), leggiamo che  $|\sin x_n| \leq |x_n|$  e quindi, per il teorema del confronto, se  $x_n \rightarrow 0$  anche  $\sin x_n \rightarrow 0$ . Pertanto

$$\lim_{x \rightarrow 0} \sin x = 0.$$

Si confronti col prossimo:

Sia

$$f(x) = \begin{cases} 1 & \text{se } x \neq 0 \\ 0 & \text{se } x = 0 \end{cases}$$

e si voglia calcolare

$$\lim_{x \rightarrow 0} f(x).$$

Applicando la definizione di limite si vede subito che il limite vale 1. Infatti, se  $x_n \rightarrow 0$  per  $n \rightarrow \infty$  ma  $x_n \neq 0 \ \forall n$ , risulta  $f(x_n) = 1 \ \forall n$ , e quindi  $f(x_n) \rightarrow 1$  per  $n \rightarrow \infty$ . In questo caso, dunque (a differenza dell'esempio precedente), si ha:

$$\lim_{x \rightarrow 0} f(x) \neq f(0).$$

#### Continuità

Si rifletta sui due esempi appena visti. In entrambi i casi il limite al finito di una certa funzione esiste ed è finito. Nel primo caso, tale limite coincide col valore della funzione nel punto considerato, nel secondo caso no. Il buon senso suggerisce che la prima situazione è quella "normale", mentre la seconda è un po' "patologica". Questa idea è chiarita dalla prossima definizione, che introduce un nuovo concetto fondamentale.

**DEFINIZIONE 3.11 (CONTINUITÀ)** Se  $f : I \rightarrow \mathbb{R}$  ( $I$  intervallo) e  $c \in I$ , si dice che  $f$  è *continua* in  $c$  se esiste

$$\lim_{x \rightarrow c} f(x) = f(c).$$

Si dice che  $f$  è continua in  $I$  se è continua in ciascun punto di  $I$ . Una funzione *non continua* in un punto  $c$  si dice *discontinua* in  $c$ .

Gli esempi precedenti si possono leggere dicendo che la funzione  $\sin x$  è continua in  $x = 0$ , mentre la funzione  $f(x)$  dell'esempio 2.7 è discontinua in  $x = 0$ . Si consideri anche il prossimo



$$\lim_{x \rightarrow 0^\pm} \frac{x}{|x|} = \pm 1.$$

Questo esempio è interessante per due motivi. Anzitutto, in questo caso non sarebbe neppure possibile calcolare  $f(0)$ , quindi è chiaro a priori che il limite non si potrà calcolare semplicemente valutando la funzione nel punto 0. Inoltre, in questo caso i limiti destro e sinistro esistono finiti e sono diversi tra loro. Questo fatto merita una definizione:

**DEFINIZIONE 3.12** Si dice che  $c$  è un *punto di discontinuità a salto* per  $f(x)$  quando i limiti destro e sinistro in  $c$  esistono finiti, ma diversi tra loro. Il salto è costituito dalla differenza dei limiti e precisamente:

$$\text{salto in } c = \lim_{x \rightarrow c^+} f(x) - \lim_{x \rightarrow c^-} f(x).$$

Se uno dei due limiti destro o sinistro per  $x \rightarrow c$  coincide con  $f(c)$ , si dice che  $f$  è *continua da destra o da sinistra*, rispettivamente.



Nell'esempio precedente, la funzione  $\frac{x}{|x|}$  ha un punto di discontinuità a salto in  $x = 0$ , con salto 2.

Le funzioni che presentano discontinuità a salto si prestano bene a modellizzare fenomeni che registrano bruschi cambiamenti.

Le funzioni continue, d'altro canto, devono la loro importanza al fatto che  $\lim_{x \rightarrow x_0} f(x) = f(x_0)$  può essere interpretato dicendo che "se  $x$  è vicino a  $x_0$ " allora " $f(x)$  è vicino a  $f(x_0)$ ", ossia

$$\text{piccole variazioni di } x \mapsto \boxed{f} \longrightarrow \text{piccole variazioni di } f(x)$$

Per coglierne la rilevanza, pensiamo al seguente problema: noto  $x_0$  e nota  $f$ , mediante una calcolatrice tascabile, calcolare  $f(x_0)$ .

Se  $x_0$  è un numero irrazionale, la calcolatrice approssima  $x_0$  con  $x_0 + \Delta x$  ( $\Delta x$  è l'errore commesso nell'approssimazione) e calcola di conseguenza  $f(x_0 + \Delta x)$ , approssimando ulteriormente il risultato finale.

Se  $f$  è continua in  $x_0$  e  $|\Delta x| \ll 1$  (cioè l'errore assoluto  $|\Delta x|$  è molto piccolo) è ragionevole considerare il risultato ottenuto come una buona approssimazione del risultato finale. Ma se  $f$  è discontinua, in corrispondenza di uno spostamento, anche piccolissimo, a destra o a sinistra di  $x_0$ , si trova un valore  $f(x_0 + \Delta x)$  molto diverso da  $f(x_0)$ .

#### Non esistenza del limite

Il limite di una funzione può anche non esistere, naturalmente:



$$\lim_{x \rightarrow +\infty} \sin x \text{ non esiste}$$

Per dimostrarlo è sufficiente trovare due successioni  $\{x_n\}$  e  $\{y_n\}$  divergenti a  $+\infty$  tali che  $\sin x_n$  e  $\sin y_n$  tendano a due limiti diversi. Si può scegliere  $x_n = n\pi$  e  $y_n = \frac{\pi}{2} + 2n\pi$ ; in corrispondenza di tali successioni si ha  $\sin x_n = 0$  e  $\sin y_n = 1$  e quindi la definizione di limite non è soddisfatta.

Si riflette sul fatto che, in base alle definizioni che abbiamo dato di limite di successione e limite di funzione, dimostrare rigorosamente che una funzione non ha limite è molto più facile che dimostrare che una *successione* non ha limite. Ad esempio, non è così banale dimostrare che non esiste  $\lim_{n \rightarrow +\infty} \sin n$ . Analogamente:



$$\lim_{x \rightarrow 0} \sin \frac{1}{x} \text{ non esiste}$$

Infatti: sia  $x_n = 1/(2n\pi)$ ,  $y_n = 1/(\frac{\pi}{2} + 2n\pi)$ . Entrambe le successioni tendono a 0, ma  $\sin \frac{1}{x_n} = 0$ ,  $\sin \frac{1}{y_n} = 1$ . Poiché lungo successioni diverse che tendono a 0 la funzione ha limiti diversi, il limite della funzione non esiste.

La situazione è analoga all'esempio precedente, ma questa volta la funzione ha infinite oscillazioni in uno spazio finito (perciò non è possibile materialmente disegnarne il grafico in tutto l'intervallo  $(0, 1)$ , ad esempio; si veda la figura 3.6).

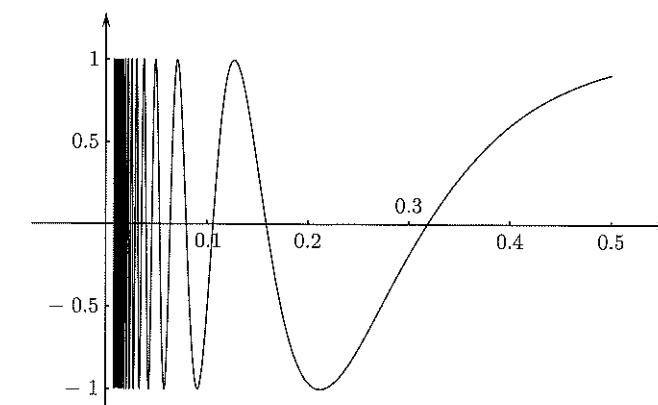


Figura 3.6.

#### Definizione topologica di limite

Concludiamo questa sezione presentando una diversa (ma equivalente) definizione di limite di funzione, che ha la caratteristica di essere indipendente dal concetto di successione e limite di successione, e che talvolta ci sarà utile.

Per precisare l'idea di "vicinanza a un punto" è comoda la nozione di *intorno*:

**DEFINIZIONE 3.13 (INTORNI)** Un *intorno* di un punto  $x_0 \in \mathbb{R}$  è un intervallo aperto che contiene  $x_0$ . Useremo intorni del tipo  $(x_0 - \delta, x_0 + \delta)$ , centrati cioè in  $x_0$ . Dire che " $x$  si muove in un intorno di  $x_0$ " significa affermare che  $x \in (x_0 - \delta, x_0 + \delta)$ , dove si pensa  $0 < \delta \ll 1$ . Si può parlare anche di intorni di  $+\infty$  o di  $-\infty$ : sono intervalli della forma  $(a, +\infty)$  o  $(-\infty, b)$ , rispettivamente.

Sarà utile anche introdurre la locuzione "definitivamente per  $x \rightarrow c$ ", analogamente a quanto abbiamo fatto nel caso delle successioni:

**DEFINIZIONE 3.14 (DEFINITIVAMENTE)** Diremo che una funzione  $f(x)$  ha una certa proprietà *definitivamente* per  $x \rightarrow c$  se esiste un intorno  $U$  di  $c$  tale che la proprietà vale per  $f(x)$  per ogni  $x \in U$ ,  $x \neq c$ .

Veniamo ora alla definizione topologica di limite.

**DEFINIZIONE 3.15** Sia  $c \in \mathbb{R}^*$  e sia  $f$  una funzione definita almeno definitivamente per  $x \rightarrow c$ . Si dice che

$$\lim_{x \rightarrow c} f(x) = \ell \text{ (con } \ell \in \mathbb{R}^*)$$

se per ogni intorno  $U_\ell$  di  $\ell$  esiste un intorno  $V_c$  di  $c$  tale che

$$\forall x \in V_c, x \neq c \text{ si ha } f(x) \in U_\ell.$$

Si noti che, come anticipato, questa definizione di limite di funzione non fa ricorso al concetto di limite di successione. Anche questa definizione, come quella successionale, è una *definizione unitaria*, che contiene come casi particolari i 4 tipi di limiti che abbiamo studiato analiticamente (limite finito o infinito, al finito o all'infinito). Il concetto centrale che permette questa sintesi è in questo caso il concetto di *intorno*.

Poiché questo è un concetto base di quella branca della matematica che prende il nome di *topologia*, questa definizione di limite “mediante intorni” si dice anche “definizione topologica di limite”.

Il lettore, a titolo di esercizio, verifichi che, particolarizzando il concetto di intorno di  $c$  ai casi  $c \in \mathbb{R}$  o  $c = \pm\infty$ , la definizione topologica di limite si traduce nelle seguenti 4 definizioni:

Si dice che

$$\lim_{x \rightarrow c} f(x) = \ell$$

se:

1. (Limite finito al finito:  $c, \ell \in \mathbb{R}$ )

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall x \neq c, |x - c| < \delta \implies |f(x) - \ell| < \varepsilon.$$

2. (Limite infinito al finito:  $c \in \mathbb{R}, \ell = +\infty$ )

$$\forall K > 0 \exists \delta > 0 : \forall x \neq c, |x - c| < \delta \implies f(x) > K.$$

3. (Limite finito all’infinito:  $c = +\infty, \ell \in \mathbb{R}$ )

$$\forall \varepsilon > 0 \exists K > 0 : \forall x > K \implies |f(x) - \ell| < \varepsilon.$$

4. (Limite infinito all’infinito:  $c = +\infty, \ell = +\infty$ )

$$\forall K > 0 \exists H > 0 : \forall x > H \implies f(x) > K.$$



Dimostrare in base alla definizione successionale di limite, i seguenti limiti di funzioni elementari:

$$\lim_{x \rightarrow +\infty} x^\alpha = \begin{cases} +\infty & \text{se } \alpha > 0 \\ 0^+ & \text{se } \alpha < 0 \end{cases}$$

$$\lim_{x \rightarrow +\infty} a^x = \begin{cases} +\infty & \text{se } a > 1 \\ 0^+ & \text{se } 0 < a < 1 \end{cases}$$

$$\lim_{x \rightarrow 0^+} \log_a x = \begin{cases} -\infty & \text{se } a > 1 \\ +\infty & \text{se } 0 < a < 1 \end{cases}$$

$$\lim_{x \rightarrow (\frac{\pi}{2})^\pm} \tan x = \mp\infty$$

Dimostrare in base alla definizione successionale di limite che non esiste

$$\lim_{x \rightarrow +\infty} \tan x.$$

Più in generale, dimostrare che non esiste

$$\lim_{x \rightarrow +\infty} f(x)$$

se  $f : \mathbb{R} \rightarrow \mathbb{R}$  è una funzione periodica non costante.

Dimostrare il criterio per la ricerca dell’asintoto obliqua contenuto nella Proposizione 3.2

Dimostrare in base alla definizione successionale di limite che non esiste (né finito, né infinito)

$$\lim_{x \rightarrow +\infty} x \sin x.$$

Sia

$$f(x) = \begin{cases} 1 & \text{se } x \text{ è razionale} \\ 0 & \text{se } x \text{ è irrazionale} \end{cases}$$

Dimostrare che  $f$  è discontinua in ogni punto.

Per ciascuna delle due funzioni parte intera di  $x$  e mantissa di  $x$ , si calcoli il limite per  $x \rightarrow +\infty$ , e si dica in quali punti la funzione è discontinua, precisando se in tali punti è (almeno) continua da destra o da sinistra.

Si determinino tutti i punti di discontinuità delle seguenti funzioni:

$$\arctg \frac{1}{x}; \quad \sin \left( \frac{1}{\sin \frac{1}{x}} \right); \quad [e^x] \quad (\text{dove } [\cdot] \text{ indica la parte intera}).$$

### 3 IL CALCOLO DEI LIMITI

Esaminiamo ora più da vicino le proprietà dei limiti e i principali metodi per il loro calcolo effettivo.

#### 3.1 Proprietà fondamentali di limiti e continuità

Anzitutto possiamo enunciare i teoremi sui limiti di funzioni che discendono immediatamente dai corrispondenti teoremi sui limiti di successioni (che abbiamo dimostrato nel par. 1.3).

Utilizzeremo qui la nozione di “proprietà vera definitivamente per  $x \rightarrow c$ ”, introdotta alla fine del paragrafo precedente. Nei prossimi enunciati  $x_0$  e  $\ell$  saranno punti di  $\mathbb{R}^*$ , salvo avviso contrario.

**TEOREMA 3.13 (DEL CONFRONTO)** Se:

1. Per  $x \rightarrow c$ ,  $f(x) \rightarrow \ell$  e  $g(x) \rightarrow \ell$ ;
2.  $f(x) \leq h(x) \leq g(x)$  definitivamente per  $x \rightarrow c$

allora anche  $h(x) \rightarrow \ell$  per  $x \rightarrow c$ .

**COROLLARIO 3.14** Se:

1. Per  $x \rightarrow c$ ,  $g(x) \rightarrow 0$ ;
  2.  $|h(x)| \leq g(x)$  definitivamente per  $x \rightarrow c$
- allora anche  $h(x) \rightarrow 0$  per  $x \rightarrow c$ .

**COROLLARIO 3.15** Se  $f(x) \rightarrow 0$  per  $x \rightarrow 0$  e  $g(x)$  è limitata definitivamente per  $x \rightarrow c$ , allora  $f(x)g(x) \rightarrow 0$  per  $x \rightarrow c$ .

**TEOREMA 3.16 (DI PERMANENZA DEL SEGNO, 1<sup>a</sup> FORMA)**

Se per  $x \rightarrow c$ ,  $f(x) \rightarrow \ell > 0$ , allora  $f(x) > 0$  definitivamente per  $x \rightarrow c$ .

**TEOREMA 3.17 (DI PERMANENZA DEL SEGNO, 2<sup>a</sup> FORMA)**

Se per  $x \rightarrow c$ ,  $f(x) \rightarrow \ell$ , e  $f(x) \geq 0$  definitivamente per  $x \rightarrow c$ , allora  $\ell \geq 0$ .

**TEOREMA 3.18 (DI PERMANENZA DEL SEGNO PER FUNZIONI CONTINUE)**

Se  $f$  è continua in  $c$  e  $f(c) > 0$ , allora  $f(x) > 0$  definitivamente per  $x \rightarrow c$ .

**TEOREMA 3.19 (ALGEBRA DEI LIMITI, CASO DEI LIMITI FINITI)**

Se per  $x \rightarrow c$ ,  $f(x) \rightarrow \ell_1$  e  $g(x) \rightarrow \ell_2$  ( $\ell_1, \ell_2 \in \mathbb{R}$ ), allora per  $x \rightarrow c$  si ha:

1.  $f(x) \pm g(x) \rightarrow \ell_1 \pm \ell_2$ ;
2.  $f(x)g(x) \rightarrow \ell_1\ell_2$ ;
3.  $f(x)/g(x) \rightarrow \ell_1/\ell_2$  (purché  $\ell_2 \neq 0$ ,  $g(x) \neq 0$  definitivamente per  $x \rightarrow c$ ).

**TEOREMA 3.20 (ARITMETIZZAZIONE PARZIALE DI  $\infty$ )** Valgono per i limiti di funzioni gli stessi risultati di “aritmetizzazione parziale di  $\infty$ ” che valgono per i limiti di successioni (v. par. 1.3).

**DIMOSTRAZIONE DEI TEOREMI PRECEDENTI.** Mostriamo in che modo i teoremi sui limiti di successioni dimostrati nel par. 1.3 e la definizione successionale di limite di funzione implicano questi teoremi sui limiti di funzioni. Come si vedrà, l’idea è semplice e applicata in modo ripetitivo.

**TEOREMA DEL CONFRONTO** Sia  $x_n$  una qualsiasi successione tale che  $x_n \neq c \forall n$  e  $x_n \rightarrow c$  per  $n \rightarrow \infty$ . Vogliamo provare che  $h(x_n) \rightarrow \ell$  per  $n \rightarrow \infty$ . D’altro canto sappiamo che

$$\begin{aligned} f(x_n) &\leq h(x_n) \leq g(x_n) \text{ definitivamente, e} \\ f(x_n) &\rightarrow \ell \text{ e } g(x_n) \rightarrow \ell \end{aligned}$$

(perché per ipotesi  $f(x) \rightarrow \ell$  e  $g(x) \rightarrow \ell$  e questo significa che per  $x_n \rightarrow c$  si ha  $f(x_n) \rightarrow \ell$  e  $g(x_n) \rightarrow \ell$ ). Ma allora, per il teorema del confronto applicato alle successioni  $f(x_n), h(x_n), g(x_n)$ , si conclude che  $h(x_n) \rightarrow \ell$ , e il teorema è provato.

Dal teorema del confronto seguono i due corollari, esattamente come nel caso delle successioni.

**TEOREMA DI PERMANENZA DEL SEGNO** Sia  $x_n$  una qualsiasi successione tale che  $x_n \neq c$  per ogni  $n$  e  $x_n \rightarrow c$ ; allora  $f(x_n) \rightarrow \ell > 0$  (per l’ipotesi), quindi per il teorema di permanenza del segno per successioni (applicato alla successione  $f(x_n)$ ) si conclude che  $f(x_n) > 0$  definitivamente. Poiché questo vale per ogni successione  $x_n \rightarrow c$ , si conclude che  $f(x) > 0$  definitivamente per  $x \rightarrow c$ . (Si veda anche l’Esercizio 7, nei Complementi alla fine di questo capitolo.)

Da questo teorema si deduce il teorema nella sua seconda forma (come nel caso delle successioni), e anche il teorema di permanenza del segno per funzioni continue. Infatti: se  $f$  è continua in  $c$ , allora  $f(c) = \lim_{x \rightarrow c} f(x)$ , quindi l’ipotesi  $f(c) > 0$  significa che  $\lim_{x \rightarrow c} f(x) > 0$ , e questo per il teorema di permanenza del segno (1<sup>a</sup> forma) implica che  $f(x) > 0$  definitivamente per  $x \rightarrow c$ .

**TEOREMA SULL’ALGEBRA DEI LIMITI** Sia  $x_n$  una qualsiasi successione tale che  $x_n \neq c$  per ogni  $n$  e  $x_n \rightarrow c$ ; allora per l’ipotesi si ha che  $f(x_n) \rightarrow \ell_1$  e  $g(x_n) \rightarrow \ell_2$ ; dal teorema sull’algebra dei limiti per successioni si conclude quindi che  $f(x_n) \pm g(x_n) \rightarrow \ell_1 \pm \ell_2$ , e quindi  $f(x) \pm g(x) \rightarrow \ell_1 \pm \ell_2$ . In modo perfettamente analogo si dimostrano i punti 2 e 3 del teorema, ed il teorema di aritmetizzazione parziale di  $\infty$ .

Dal teorema dell’algebra dei limiti segue un analogo risultato per quanto riguarda le funzioni continue:

**TEOREMA 3.21 (ALGEBRA DELLE FUNZIONI CONTINUE)** Siano  $f, g$  due funzioni definite almeno in un intorno di  $x_0 \in \mathbb{R}$ , e continue in  $x_0$ . Allora:

1.  $f(x) \pm g(x)$  è continua in  $x_0$ ;
2.  $f(x)g(x)$  è continua in  $x_0$ ;
3.  $f(x)/g(x)$  è continua in  $x_0$ , purché  $g(x_0) \neq 0$ .

**DIMOSTRAZIONE.** Proviamo ad esempio la 3, essendo le altre analoghe (e più semplici). Per ipotesi sappiamo che  $f$  e  $g$  sono continue in  $x_0$ , ossia:

$$f(x) \rightarrow f(x_0) \text{ e } g(x) \rightarrow g(x_0) \text{ per } x \rightarrow x_0.$$

Inoltre  $g(x_0) \neq 0$  e quindi, per il teorema di permanenza del segno per funzioni continue,  $g(x) \neq 0$  definitivamente per  $x \rightarrow x_0$ . Allora per il teorema sull’algebra dei limiti si conclude che

$$\frac{f(x)}{g(x)} \rightarrow \frac{f(x_0)}{g(x_0)} \text{ per } x \rightarrow x_0,$$

ossia  $f/g$  è continua in  $x_0$ .

Vale anche il prossimo importante:

**TEOREMA 3.22 (DI CONTINUITÀ DELLE FUNZIONI ELEMENTARI)** Le seguenti funzioni elementari sono continue in tutti i punti del proprio insieme di definizione:

1. Potenze a esponente intero, razionale o reale;
2. Funzioni esponenziali;
3. Funzioni logaritmiche;
4. Funzioni trigonometriche elementari ( $\sin x, \cos x$ ).

Vedremo in seguito come, combinando tra loro in molti modi questi “componenti elementari” continui si ottengano ancora funzioni continue. Avremo quindi modo, molto spesso, di sapere *a priori* che una funzione è continua nel suo insieme di definizione (ossia senza applicare caso per caso la definizione di continuità). In questi casi, potremo allora sapere *a priori* che il limite di  $f(x)$  per  $x \rightarrow x_0$  si può effettivamente calcolare valutando semplicemente  $f(x_0)$ .

**DIMOSTRAZIONE.** Non dimostriamo il teorema in tutte le sue affermazioni. Ci limitiamo, a titolo d’esempio, a provare la continuità in tutto  $\mathbb{R}$  delle funzioni  $\sin x, \cos x$ .

Sappiamo già (v. Esempio 2.6) che  $\sin x$  è continuo in  $x = 0$ . Da questo segue che anche  $\cos x$  è continuo in  $x = 0$ . Infatti, uno sguardo alla circonferenza trigonometrica mostra che, se  $x$  è un angolo nel primo quadrante,  $\sin x + \cos x \geq 1$  (in quanto 1 è l'ipotenusa di un triangolo rettangolo di cateti  $\sin x, \cos x$ ). Ne segue che

$$\begin{aligned} 0 \leq 1 - \cos x &\leq \sin x \text{ per } x \in \left(0, \frac{\pi}{2}\right), \text{ e quindi} \\ 0 \leq 1 - \cos x &\leq |\sin x| \text{ per } x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right). \end{aligned}$$

Allora per il teorema del confronto  $1 - \cos x \rightarrow 0$  per  $x \rightarrow 0$ , quindi  $\cos x \rightarrow 1$  per  $x \rightarrow 0$ , e perciò  $\cos x$  è continua in  $x = 0$ .

Per provare ora la continuità di  $\sin x$  in un generico punto  $x_0$  consideriamo la catena di relazioni:

$$\begin{aligned} |\sin(x_0 + h) - \sin x_0| &= |\sin x_0 \cosh + \cos x_0 \sinh - \sin x_0| \leq \\ &\leq |\sin x_0| |\cosh - 1| + |\cos x_0| |\sinh|. \end{aligned}$$

Ora  $|\sinh|$  e  $|\cosh - 1|$  tendono a zero per  $h \rightarrow 0$ , per quanto appena dimostrato, mentre  $|\sin x_0|$  e  $|\cos x_0|$  sono costanti, dunque  $\sin(x_0 + h) - \sin x_0 \rightarrow 0$ , ossia  $\sin(x_0 + h) \rightarrow \sin x_0$  per  $h \rightarrow 0$ , e  $\sin x$  è continua in  $x_0$ . Una analoga catena di diseguaglianze dimostra la continuità di  $\cos x$ .

### ESEMPIO

Per il teorema 3.21 possiamo affermare ad esempio che:

1. Le funzioni potenza a esponente intero sono continue (in tutto  $\mathbb{R}$ ), in quanto  $f(x) = x$  è ovviamente continua, e  $f(x) = x^n$  è il prodotto di  $n$  funzioni continue.
2. I polinomi sono funzioni continue, in quanto ottenuti sommando funzioni del tipo  $cx^n$ , continue per il punto precedente.
3. Le funzioni razionali (cioè i quozienti di polinomi) sono funzioni continue, tranne nei punti in cui si annulla il denominatore (che è un polinomio, quindi si annulla in un numero finito di punti).
4. Avendo dimostrato che  $\sin x, \cos x$  sono continue in tutto  $\mathbb{R}$ , deduciamo che le funzioni  $\operatorname{tg} x, \operatorname{cotg} x$  sono continue nel loro insieme di definizione.

Un'altra operazione che “conserva” i limiti e la continuità è la *composizione di funzioni*:

**TEOREMA 3.23 (DI CAMBIO DI VARIABILE NEL LIMITE)** *Siano  $f, g$  due funzioni per cui è ben definita la composizione  $f \circ g$ , almeno definitivamente per  $x \rightarrow x_0$  ( $x_0 \in \mathbb{R}^*$ ), e supponiamo che:*

- i)  $g(x) \rightarrow t_0$  per  $x \rightarrow x_0$ ;
- ii) esiste  $\lim_{t \rightarrow t_0} f(t) = \ell \in \mathbb{R}^*$
- iii)  $g(x) \neq t_0$  definitivamente per  $x \rightarrow x_0$ .

Allora esiste

$$\lim_{x \rightarrow x_0} f(g(x)) = \lim_{t \rightarrow t_0} f(t).$$

I punti  $x_0, t_0 \in \mathbb{R}^*$ . L'ipotesi iii) non è necessaria nel caso in cui  $f$  sia continua in  $t_0$ , oppure (ovviamente) nel caso in cui  $t_0 = \pm\infty$ .

**DIMOSTRAZIONE.** Basta osservare che se  $x_n$  è una qualsiasi successione tale che  $x_n \neq x_0 \forall n$  e  $x_n \rightarrow x_0$  per  $n \rightarrow \infty$ , allora  $g(x_n) \rightarrow t_0$  per  $n \rightarrow \infty$  (per l'ipotesi (i)) e  $g(x_n) \neq t_0$  definitivamente (per l'ipotesi (iii)). Perciò

$$f(g(x_n)) \rightarrow \ell$$

(per l'ipotesi (ii)), e questo prova la tesi. Se  $t_0 = \pm\infty$  la condizione  $g(x) \neq \pm\infty$  è ovviamente verificata, mentre se  $f$  è continua in  $t_0$ ,  $\ell = f(t_0)$ , perciò nel caso risultasse  $g(x_n) = t_0$  per qualche  $n$ , si avrebbe

$$f(g(x_n)) = f(t_0) = \ell$$

e quindi la convergenza  $f(g(x_n)) \rightarrow \ell$  sarebbe comunque garantita.  $\diamond$

Dalla proprietà precedente, segue subito il prossimo:

**TEOREMA 3.24 (DI CONTINUITÀ DELLA FUNZIONE COMPOSTA)** *Siano:  $g$  una funzione definita almeno in un intorno di  $x_0$  e continua in  $x_0$ ,  $f$  una funzione definita almeno in un intorno di  $t_0 = g(x_0)$  e continua in  $t_0$ , allora  $f \circ g$  è definita almeno in un intorno di  $x_0$  ed è continua in  $x_0$ .*

**DIMOSTRAZIONE.** Poiché  $g$  è continua in  $x_0$ ,

$$\lim_{x \rightarrow x_0} g(x) = g(x_0) = t_0,$$

allora per il teorema del cambio di variabile nel limite, si ha:

$$\begin{aligned} \lim_{x \rightarrow x_0} f(g(x)) &= \lim_{t \rightarrow t_0} f(t) = \text{poiché } f \text{ è continua in } t_0 \\ &= f(t_0) = f(g(x_0)), \end{aligned}$$

e la tesi è dimostrata.  $\diamond$

Vogliamo ora trarre qualche conclusione dai vari teoremi sulle funzioni continue visti finora. Sappiamo che

- le funzioni elementari dell'Analisi matematica sono continue nel loro insieme di definizione;
- somma, prodotto, quoziente di funzioni continue danno funzioni continue (dove il denominatore non si annulla);
- composizione di funzioni continue dà una funzione continua.

Ne segue che tutte le funzioni che si possono ottenere con somma, prodotto, quoziente e composizioni da funzioni elementari sono continue nel loro insieme di definizione. Risultano pertanto continue funzioni come

$$e^{-x^2} \quad \sqrt{\sin x} \quad \log_a(1 + (\operatorname{tg} x)^2), \dots$$

Quindi, per esempio:

$$\lim_{x \rightarrow \pi} \sqrt{\sin x} = \sqrt{\sin \pi} = 0 \quad \lim_{x \rightarrow 1} e^{-x^2} = e^{-1} = \frac{1}{e}$$

$$\lim_{x \rightarrow 0} \log_a(1 + (\operatorname{tg} x)^2) = \log_a(1 + (\operatorname{tg} 0)^2) = \log_a 1 = 0$$

poiché, per le funzioni continue, se  $x_0$  è un punto del dominio, il limite si calcola semplicemente sostituendo  $x_0$  nell'espressione analitica della funzione.

Mostriamo ora vari esempi di applicazioni dei teoremi di questo paragrafo.



**Limiti di polinomi** Per calcolare il limite di un polinomio per  $x \rightarrow \pm\infty$  basta calcolare il limite del termine di grado massimo. Infatti se il polinomio è  $P(x) = a_0x^n + a_1x^{n-1} + \dots + a_n$ ,  $a_0 \neq 0$ , si ha:

$$P(x) = a_0x^n \left\{ 1 + \frac{a_1}{a_0x} + \frac{a_2}{a_0x^2} + \dots + \frac{a_n}{a_0x^n} \right\}$$

Tutti i termini dopo 1, in parentesi graffa, tendono a 0 per  $x \rightarrow \pm\infty$  e quindi  $\lim_{x \rightarrow \pm\infty} P(x) = \lim_{x \rightarrow \pm\infty} a_0x^n$ .

Per esempio:

$$\lim_{x \rightarrow +\infty} (3x^3 - x^2 + 1) = \lim_{x \rightarrow +\infty} 3x^3 = +\infty$$

$$\lim_{x \rightarrow -\infty} (-2x^3 + x + 120) = \lim_{x \rightarrow -\infty} -2x^3 = +\infty$$

**Limiti di funzioni razionali** Sia ora  $f(x) = P_n(x)/Q_m(x)$  dove  $P_n$  e  $Q_m$  sono, rispettivamente, polinomi di grado  $n$  e  $m$  dati da:

$$P_n(x) = a_0x^n + a_1x^{n-1} + \dots + a_n$$

$$Q_m(x) = b_0x^m + b_1x^{m-1} + \dots + b_m$$

Per calcolare il limite di  $P_n(x)/Q_m(x)$  per  $x \rightarrow \pm\infty$  basta calcolare il limite del rapporto dei monomi di grado massimo, cioè:

$$\lim_{x \rightarrow \pm\infty} \frac{P_n(x)}{Q_m(x)} = \lim_{x \rightarrow \pm\infty} \frac{a_0x^n}{b_0x^m}.$$

Infatti,

$$\lim_{x \rightarrow \pm\infty} \frac{P_n(x)}{Q_m(x)} = \lim_{x \rightarrow \pm\infty} \frac{a_0x^n \{ 1 + \text{termini che tendono a } 0 \}}{b_0x^m \{ 1 + \text{termini che tendono a } 0 \}} = \lim_{x \rightarrow \pm\infty} \frac{a_0x^n}{b_0x^m}.$$

Per esempio:

$$\lim_{x \rightarrow -\infty} \frac{3x^3 + 1}{x^2 + 2} = -\infty \quad \lim_{x \rightarrow +\infty} \frac{3x^3 + 1}{x^4 - 10x + 1} = 0 \quad \lim_{x \rightarrow -\infty} \frac{6x^5 + 7x^2}{7x^5 - 4} = \frac{6}{7}$$

**ESEMPIO** Una situazione simile analoga si ha nel calcolo del limite per  $x \rightarrow 0$  di certi quozienti:

$$\lim_{x \rightarrow 0} \frac{x^{2/3}}{x^2 + 3\sqrt[3]{x} + x}$$

È una sorta di “quoziente di polinomi” (per la verità, sono potenze a esponente razionale, non intero): per  $x \rightarrow 0$ , tutte le potenze scritte tendono a 0, e si ha una forma di indeterminazione  $\frac{0}{0}$ . Questa volta, i termini preponderanti sono quelli di esponente minimo, perché  $x \rightarrow 0$ :

$$\frac{x^{2/3}}{x^2 + 3\sqrt[3]{x} + x} = \frac{x^{2/3}}{3x^{1/3}(1 + \frac{x^{2/3}}{3} + \frac{x^{5/3}}{3})} = \frac{x^{1/3}}{3(1 + \frac{x^{2/3}}{3} + \frac{x^{5/3}}{3})} \rightarrow 0$$

perché  $x^{1/3} \rightarrow 0$  e  $3(1 + \frac{x^{2/3}}{3} + \frac{x^{5/3}}{3}) \rightarrow 3$ .



$$\lim_{x \rightarrow 0} x \sin \frac{1}{x} = 0$$

Infatti, poiché  $|\sin \frac{1}{x}| \leq 1$ , si ha

$$\left| x \sin \frac{1}{x} \right| \leq |x|$$

da cui, per il corollario del teorema del confronto,  $x \sin \frac{1}{x} \rightarrow 0$ . La conclusione è interessante perché il fattore  $\sin \frac{1}{x}$ , da solo, non ammette limite, quindi non si sarebbe potuto applicare il teorema sul limite del prodotto.



$$\lim_{x \rightarrow 0^\pm} e^{\frac{1}{x}} = \begin{cases} +\infty & \\ 0 & \end{cases}$$

Infatti, per  $x \rightarrow 0^\pm$ ,  $\frac{1}{x} \rightarrow \pm\infty$ , e per  $y \rightarrow \pm\infty$ ,  $e^y \rightarrow \begin{cases} +\infty & \\ 0 & \end{cases}$  (rispettivamente). Abbiamo applicato, implicitamente, il teorema sul cambio di variabile nel limite.



$$\lim_{x \rightarrow 0^+} \frac{x-1}{\sin x} = -\infty$$

perché  $(x-1) \rightarrow -1$  e  $\sin x \rightarrow 0^+$ . Abbiamo applicato il teorema sull'aritmetizzazione parziale di  $\infty$ .



$$\lim_{x \rightarrow -\infty} \left( \frac{1}{x} - 2 \right) x^3 = +\infty$$

perché  $(\frac{1}{x} - 2) \rightarrow -2$  e  $x^3 \rightarrow -\infty$ . Abbiamo applicato il teorema sull'aritmetizzazione parziale di  $\infty$ .



$$\lim_{x \rightarrow +\infty} \left( \frac{x + \sin x}{2x + \cos x} \right) = \frac{1}{2}.$$

Infatti:

$$\frac{x + \sin x}{2x + \cos x} = \frac{x(1 + \frac{\sin x}{x})}{2x(1 + \frac{\cos x}{2x})} = \frac{(1 + \frac{\sin x}{x})}{2(1 + \frac{\cos x}{2x})}.$$

Ora per il corollario del teorema del confronto,

$$\frac{\sin x}{x} \rightarrow 0 \text{ e } \frac{\cos x}{2x} \rightarrow 0, \text{ per } x \rightarrow +\infty.$$

Allora per il teorema sull'algebra dei limiti,

$$\frac{(1 + \frac{\sin x}{x})}{2(1 + \frac{\cos x}{2x})} \rightarrow \frac{(1 + 0)}{2(1 + 0)} = \frac{1}{2}.$$

### 3.2 Limiti notevoli

Vediamo ora alcune comuni tecniche di calcolo dei limiti che, combinando i teoremi generali sui limiti con l'uso di alcuni *limiti notevoli* che riguardano certe funzioni elementari, risultano piuttosto potenti. I limiti notevoli, inoltre, saranno utili anche nel calcolo differenziale.

#### Limiti notevoli di seno e coseno

Nel calcolare  $\lim_{x \rightarrow 0} \frac{\sin x}{x}$  si presenta la forma di indecisione  $\frac{0}{0}$ . Vogliamo dimostrare che

$$(3.1) \quad \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$$

Osserviamo subito che, essendo  $\sin x$  e  $x$  funzioni dispari,  $\frac{\sin x}{x}$  è funzione pari e quindi è sufficiente calcolare  $\lim_{x \rightarrow 0^+} \frac{\sin x}{x}$ . A tale scopo, osservando la figura 3.7 si vede che l'area del triangolo  $OPA$  è minore di quella del settore circolare  $OPA$ , a sua volta minore di quella del triangolo  $OTA$ . Ne segue

$$\frac{1}{2} \cdot 1 \cdot \sin x \leq \frac{1}{2} \cdot 1 \cdot x \leq \frac{1}{2} \cdot 1 \cdot \tan x$$

ossia

$$\sin x < x < \tan x$$

Dividendo per  $\sin x$ , che è positivo perché  $x \in (0, \frac{\pi}{2})$ , si ha

$$1 < \frac{x}{\sin x} < \frac{1}{\cos x} \quad \forall x \in \left(0, \frac{\pi}{2}\right)$$

e infine, passando ai reciproci, si ottiene

$$\cos x < \frac{\sin x}{x} < 1$$

Dal teorema del confronto, essendo  $\lim_{x \rightarrow 0} \cos x = 1$ , si deduce la (3.1).

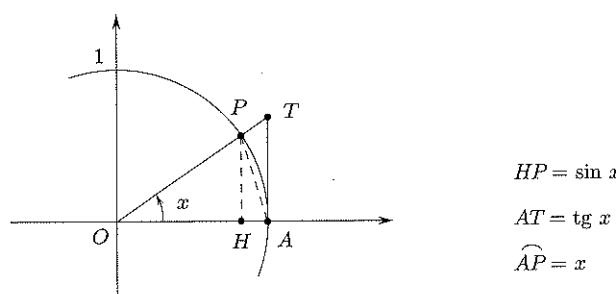


Figura 3.7.

Proviamo ora che:

(3.2)

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \frac{1}{2}$$

Infatti

$$\frac{1 - \cos x}{x^2} = \frac{1 - \cos^2 x}{x^2(1 + \cos x)} = \left(\frac{\sin x}{x}\right)^2 \frac{1}{1 + \cos x} \rightarrow \frac{1}{2}$$

perché  $\frac{\sin x}{x} \rightarrow 1$  e  $(1 + \cos x) \rightarrow 2$ .

#### Prolungamento per continuità di una funzione

In base al limite (3.1) dimostrato, la funzione  $f(x) = \frac{\sin x}{x}$ , inizialmente non definita per  $x = 0$ , può essere *prolungata per continuità* anche in  $x = 0$ , ponendo

$$f(x) = \begin{cases} \frac{\sin x}{x} & \text{se } x \neq 0 \\ 1 & \text{se } x = 0 \end{cases}$$

La funzione  $f$  così definita risulta continua anche in  $x = 0$ .

Più in generale: se una funzione  $f(x)$  non è definita in  $x_0$  ma esiste finito

$$\lim_{x \rightarrow x_0} f(x) = l$$

la funzione può essere prolungata per continuità anche in  $x_0$ , ponendo per definizione  $f(x_0) = l$ .

Lo studente si convinca del fatto che se invece la funzione  $f$  possiede in  $x_0$  una discontinuità a salto, un asintoto verticale, o comunque non ammette limite finito, non è possibile renderla continua in  $x_0$  alterandone la definizione *in un punto solo*.

Ogni volta che una funzione, inizialmente non definita in  $x_0$ , risulti essere prolungabile con continuità in  $x_0$ , si intende che la funzione sia stata effettivamente prolungata.

#### Limite di $(1 + \frac{1}{x})^x$ per $x \rightarrow \pm\infty$ e altri limiti notevoli collegati

Già sappiamo (v. Teorema 3.9) che per ogni successione  $\{a_n\}$  divergente a  $+\infty$  o  $-\infty$  si ha

$$\lim_{n \rightarrow +\infty} \left(1 + \frac{1}{a_n}\right)^{a_n} = e.$$

Questo fatto, per la definizione successionale di limite di funzione, implica immediatamente che:

(3.3)

$$\left(1 + \frac{1}{x}\right)^x \rightarrow e \quad \text{per } x \rightarrow \pm\infty$$

Dal limite (3.3) se ne possono dedurre alcuni altri.

Passando ai logaritmi nella (3.3), si ottiene

$$\log \left(1 + \frac{1}{x}\right)^x = x \log \left(1 + \frac{1}{x}\right) \rightarrow \log e = 1 \quad \text{per } x \rightarrow \pm\infty$$

perché la funzione  $\log x$  è continua.

Se si pone  $y = \frac{1}{x}$ ,  $x \rightarrow \pm\infty$  equivale a  $y \rightarrow 0^\pm$  e l'ultimo limite si può scrivere nella forma seguente

$$(3.4) \quad \boxed{\frac{\log(1+y)}{y} \rightarrow 1 \quad \text{se} \quad y \rightarrow 0}$$

Se poi nella (3.4) poniamo  $y = e^x - 1$ , allora  $y \rightarrow 0$  equivale a  $x \rightarrow 0$ , e sostituendo si ricava

$$\frac{\log(e^x)}{e^x - 1} = \frac{x}{e^x - 1} \rightarrow 1 \quad \text{se} \quad x \rightarrow 0$$

o, meglio:

$$(3.5) \quad \boxed{\frac{e^x - 1}{x} \rightarrow 1 \quad \text{se} \quad x \rightarrow 0}$$

Infine, se nella (3.4) poniamo  $y = (1+x)^\alpha - 1$ , con  $\alpha$  esponente reale qualsiasi, allora  $x \rightarrow 0$  equivale ad  $y \rightarrow 0$ , e si ha

$$\begin{aligned} \frac{\log(1+y)}{y} &= \frac{\log[1 + (1+x)^\alpha - 1]}{(1+x)^\alpha - 1} = \frac{\alpha \log(1+x)}{(1+x)^\alpha - 1} = \\ &= \frac{\alpha x}{(1+x)^\alpha - 1} \cdot \frac{\log(1+x)}{x} \rightarrow 1 \end{aligned}$$

e poiché anche  $\frac{\log(1+x)}{x} \rightarrow 1$ , se ne deduce:

$$(3.6) \quad \boxed{\lim_{x \rightarrow 0} \frac{(1+x)^\alpha - 1}{x} = \alpha}$$

### 3.3 Confronti e stime asintotiche

I limiti (3.1), (3.2), (3.4), (3.5) (3.6) sono espressi in modo forse più significativo usando il simbolo di equivalenza asintotica  $\sim$ , già introdotto per le successioni.

**DEFINIZIONE 3.16** Si dice che due funzioni  $f, g$  sono *asintotiche* per  $x \rightarrow c$  se

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = 1$$

e si scrive  $f \sim g$  per  $x \rightarrow c$ .

Il simbolo di asintotico gode di tutte le proprietà enunciate nel paragrafo 1.5. I limiti precedenti si possono allora riscrivere come segue:

$$(3.7) \quad \begin{aligned} \sin x &\sim x, \quad 1 - \cos x \sim \frac{1}{2}x^2, \quad e^x - 1 \sim x \\ \log(1+x) &\sim x, \quad (1+x)^\alpha - 1 \sim \alpha x \\ &\text{per } x \rightarrow 0 \end{aligned}$$

Le (3.7) si possono esprimere dicendo ad esempio che  $\sin x$ ,  $\log(1+x)$  e  $e^x - 1$ , in *prima approssimazione* o *al primo ordine*, si comportano come  $x$ , per  $x \rightarrow 0$ .

Le (3.7) possono essere generalizzate. Se  $\varepsilon(x)$  è una funzione che tende a zero<sup>3</sup> ed è definitivamente diversa da zero (cioè è un *infinitesimo*), si può scrivere

$$(3.8) \quad \begin{aligned} \sin \varepsilon(x) &\sim \varepsilon(x) \\ 1 - \cos \varepsilon(x) &\sim \frac{1}{2}\varepsilon(x)^2 \\ e^{\varepsilon(x)} - 1 &\sim \varepsilon(x) \quad \varepsilon(x) \rightarrow 0 \\ \log(1 + \varepsilon(x)) &\sim \varepsilon(x) \\ (1 + \varepsilon(x))^\alpha - 1 &\sim \alpha\varepsilon(x) \end{aligned}$$

Le (3.8) si deducono dalle (3.7) con il semplice cambio di variabile  $y = \varepsilon(x)$ .



$$\lim_{x \rightarrow 0} \frac{\log(1+2x)}{\sin 3x} = \frac{2}{3}$$

Il limite dà una forma di indeterminazione  $[\frac{0}{0}]$ . Tuttavia, per  $x \rightarrow 0$ ,  $\log(1+2x) \sim 2x$ ,  $\sin 3x \sim 3x$ , e quindi

$$\frac{\log(1+2x)}{\sin 3x} \sim \frac{2x}{3x} = \frac{2}{3}$$



$$\lim_{x \rightarrow 1} \frac{(x-1)^2}{e^{3(x-1)^2} - 1} = \frac{1}{3}$$

Il limite dà una forma di indeterminazione  $[\frac{0}{0}]$ . Tuttavia,  $e^{3(x-1)^2} - 1 \sim 3(x-1)^2$  per  $x \rightarrow 1$  (la funzione  $\varepsilon(x) = 3(x-1)^2$  è infinitesima per  $x \rightarrow 1$ ), perciò

$$\frac{(x-1)^2}{e^{3(x-1)^2} - 1} \sim \frac{(x-1)^2}{3(x-1)^2} = \frac{1}{3}$$



$$\lim_{x \rightarrow +\infty} (\sqrt[3]{x^3 + 2x^2 + 1} - x) = \frac{2}{3}$$

È una forma di indeterminazione del tipo  $[+\infty - \infty]$ . Tuttavia:

$$\begin{aligned} (\sqrt[3]{x^3 + 2x^2 + 1} - x) &= x \left( \sqrt[3]{1 + \left( \frac{2}{x} + \frac{1}{x^3} \right)} - 1 \right) \sim \\ &\sim x \left[ \frac{1}{3} \left( \frac{2}{x} + \frac{1}{x^3} \right) \right] \rightarrow \frac{2}{3}. \end{aligned}$$

Abbiamo usato la stima  $\sqrt[3]{1 + \varepsilon(x)} - 1 \sim \frac{1}{3}\varepsilon(x)$ , con  $\varepsilon(x) = (\frac{2}{x} + \frac{1}{x^3})$ .

Altri importanti limiti riguardano il comportamento all'infinito di potenze, esponenziali e logaritmi, analoghi a quelli validi nel caso delle successioni:

<sup>3</sup>Non ha importanza a che cosa tende  $x$ .

**TEOREMA 3.25 (GERARCHIA DEGLI INFINITI)** Consideriamo le tre famiglie di funzioni:

$$(\log_a x)^\alpha \quad x^\beta \quad b^x \quad \text{con } \alpha, \beta > 0, a, b > 1$$

Allora, per  $x \rightarrow +\infty$ , ognuna è un infinito di ordine inferiore rispetto a quella che le sta a destra. Esplcitamente:

$$(3.9) \quad \lim_{x \rightarrow +\infty} \frac{(\log_a x)^\alpha}{x^\beta} = 0 \quad \lim_{x \rightarrow +\infty} \frac{x^\beta}{b^x} = 0$$

il che si può esprimere a parole dicendo che qualunque potenza (positiva) di  $x$  prevale su qualunque potenza di  $\log x$ , e qualunque esponenziale (base > 1) di  $x$  prevale su qualunque potenza di  $x$ .

Inoltre ponendo  $y = \frac{1}{x}$  nel primo limite in (3.9) si trova<sup>4</sup>

$$(3.10) \quad \lim_{y \rightarrow 0^+} y^\beta (-\log y)^\alpha = 0 \quad (\beta > 0, \forall \alpha)$$



$$\lim_{x \rightarrow +\infty} \frac{x^3}{2^x} = 0$$

$$\lim_{x \rightarrow +\infty} \frac{(\log x)^2}{\sqrt{x}} = 0$$

$$\lim_{x \rightarrow 0^+} \sqrt{x} \log x = 0^-$$

$$\lim_{x \rightarrow 0^+} x^x = \lim_{x \rightarrow 0^+} e^{x \log x} = e^{0^-} = 1^-$$

$$\lim_{x \rightarrow 0^+} x e^{1/x} = \left[ x = \frac{1}{t} \right] = \lim_{t \rightarrow +\infty} \frac{e^t}{t} = +\infty.$$

Il teorema 3.25 potrà essere dimostrato più avanti con gli strumenti del calcolo differenziale (cap. 4, par. 4.4).

### 3.4 Stime asintotiche e grafici

Le stime asintotiche non servono solo per calcolare limiti, ma anche per tracciare il grafico qualitativo di una funzione nell'intorno di un certo punto, oppure nell'intorno di  $\pm\infty$ .



$$f(x) = \sqrt[3]{x} + x^2$$

<sup>4</sup>Perché ci vuole  $(-\log y)^\alpha$  e sarebbe errato scrivere  $(\log y)^\alpha$ ?

Si può ragionare così: la funzione è definita e continua su tutto  $\mathbb{R}$ ; per  $x \rightarrow \pm\infty$ ,  $f(x) \sim x^2$ . Dunque  $f(x) \rightarrow +\infty$ ; inoltre il suo grafico sarà simile, per  $x$  grande in valore assoluto, a quello di  $x^2$ . La funzione inoltre si annulla in  $x = 0$  e, per  $x \rightarrow 0$ ,  $f(x) \sim \sqrt[3]{x}$ ; dunque il suo grafico sarà simile, in un intorno di  $x = 0$ , a quello di  $\sqrt[3]{x}$ : in particolare, avrà tangente verticale nell'origine. Il più semplice grafico compatibile con queste informazioni è:

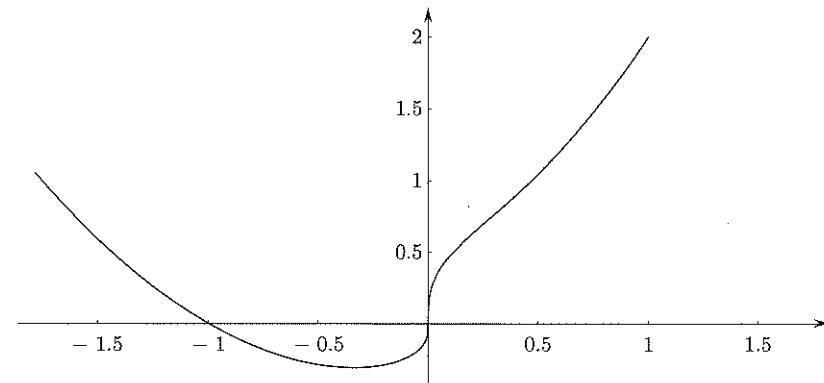


Figura 3.8.

Spesso l'andamento di una funzione nell'intorno di un punto (ad esempio il fatto che abbia tangente verticale o orizzontale) è prevedibile in base a un'opportuna stima asintotica, che consente di tracciare il grafico qualitativo di  $f$  (nell'intorno del punto) per confronto con quello di una funzione nota (ad esempio, una potenza a esponente razionale). Analoghe stime sono utili per  $x \rightarrow \pm\infty$ .

Uno dei problemi fisici all'origine della teoria quantistica è quello della radiazione del corpo nero. La legge proposta nel 1900 da Planck per la distribuzione spettrale della radiazione emessa da un corpo nero è:

$$f(\lambda) = \frac{8\pi hc\lambda^{-5}}{e^{hc/\lambda kT} - 1}$$

Qui  $f(\lambda)$  è la densità di energia raggiante corrispondente alla lunghezza d'onda  $\lambda$ , quando la temperatura del corpo è  $T$ ;  $h$  è la costante di Planck,  $k$  la costante di Boltzmann e  $c$  la velocità della luce. Questa legge è in accordo con i dati sperimentali, mentre la legge che si può dedurre usando la fisica classica (legge di Rayleigh-Jeans) è

$$f(\lambda) = 8\pi kT\lambda^{-4}$$

che è in accordo coi dati sperimentali solo per  $\lambda$  grande, mentre per  $\lambda$  piccolo è in profondo disaccordo: tende a  $+\infty$ , mentre la curva sperimentale tende a zero!

Studiamo la curva di Planck con opportune stime asintotiche.

Per  $\lambda \rightarrow +\infty$ ,

$$e^{hc/\lambda kT} - 1 \sim hc/\lambda kT$$

$$e^{-\frac{8\pi hc\lambda^{-5}}{e^{hc/\lambda kT} - 1}} \sim \frac{8\pi hc\lambda^{-5}}{hc/\lambda kT} = 8\pi kT\lambda^{-4}$$

ossia si ritrova la legge di Rayleigh-Jeans, in prima approssimazione; in particolare, la densità di energia raggiante tende a zero come  $\lambda^{-4}$ .

Per  $\lambda \rightarrow 0^+$ ,

$$\frac{8\pi hc\lambda^{-5}}{e^{hc/\lambda kT} - 1} \sim \frac{8\pi hc\lambda^{-5}}{e^{hc/\lambda kT}} \rightarrow 0^+$$

(l'esponenziale a denominatore tende a infinito più rapidamente di  $\lambda^{-5}$  a numeratore). Quindi la curva di Planck è in accordo coi dati sperimentali anche per  $\lambda$  piccolo. Poiché la funzione è sempre positiva, il suo grafico, in base a queste informazioni, sarà del tipo:

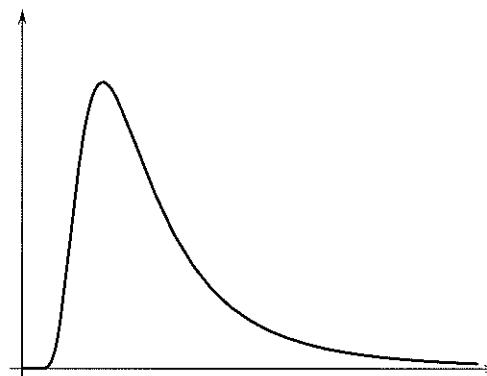


Figura 3.9.

### Crescita di una funzione all'infinito

Supponiamo di voler tracciare il grafico di una funzione  $f$  che, per  $x \rightarrow +\infty$  ( $o -\infty$ ) tende a  $+\infty$  ( $-\infty$ ). Per descrivere la velocità con cui la funzione tende all'infinito, sono utili le seguenti nozioni: diremo che, per  $x \rightarrow +\infty$ ,

$$f \text{ ha crescita } \begin{cases} \text{sopralineare} \\ \text{lineare} \\ \text{sottolineare} \end{cases} \quad \text{se} \quad \lim_{x \rightarrow +\infty} \frac{f(x)}{x} = \begin{cases} \pm\infty \\ m \text{ (finito e diverso da 0)} \\ 0 \end{cases}$$

(Analoghe definizioni si danno per  $x \rightarrow -\infty$ ). Solo nel caso in cui una funzione ha crescita lineare, è possibile che ammetta asintoto obliqua (v. par. 2 per il calcolo di tale asintoto).

Esempi di funzioni con crescita sopralineare per  $x \rightarrow +\infty$  sono gli esponenziali  $a^x$  e le potenze  $x^a$  con  $a > 1$ ; esempi di funzioni con crescita sottolineare per  $x \rightarrow +\infty$  sono i logaritmi  $\log_a x$  e le potenze  $x^a$  con  $0 < a < 1$ .

### Esercizi

**10** La funzione

$$f(x) = 2x + e^x + e^{1/x}$$

per  $x \rightarrow +\infty$  è asintotica a  $e^x$ ; pertanto tende a  $+\infty$  con crescita sopralineare; per  $x \rightarrow -\infty$  è asintotica a  $2x$ ; pertanto tende a  $-\infty$  linearmente; poiché

$$\lim_{x \rightarrow -\infty} [f(x) - 2x] = 1,$$

la funzione ha asintoto obliqua  $y = 2x + 1$  per  $x \rightarrow -\infty$ .

### Esercizi

Sui limiti notevoli.

$$\lim_{x \rightarrow +\infty} x \sin \frac{1}{x}$$

$$\lim_{x \rightarrow 0} \frac{\cos x - 1}{\sin^2 x}$$

$$\lim_{x \rightarrow +\infty} x \log \left( \frac{x+3}{x+1} \right)$$

Suggerimento: poiché  $\frac{x+3}{x+1} \rightarrow 1$ , si può porre  $\frac{x+3}{x+1} = 1 + \varepsilon(x)$  con  $\varepsilon(x)$  infinitesima; quindi...



$$\lim_{x \rightarrow +\infty} \left( \frac{x^2 + 3}{x^2 + 2} \right)^x$$

Suggerimento: dovendo calcolare il limite di una funzione del tipo  $f(x)^{g(x)}$  che dà una forma di indeterminazione del tipo  $[1^\infty]$ , riscriverla nella forma

$$f(x)^{g(x)} = e^{g(x) \log f(x)}$$

e cominciare a calcolare il limite dell'esponente (che è ora una forma di indeterminazione del tipo  $[\infty \cdot 0]$ ). Per far questo, seguire il suggerimento dell'esercizio precedente.

**11** Fornire una stima asintotica, per  $x \rightarrow 0$ , delle funzioni:  $\operatorname{tg} x$ ,  $\operatorname{cotg} x$ ,  $\operatorname{arctg} x$ ,  $\operatorname{arcsin} x$ .

$$\lim_{x \rightarrow \pm\infty} \frac{2^{\frac{x}{x}} + \log|x|}{x^2 + 1}$$

$$\lim_{x \rightarrow 0} \frac{e^{x^2} - 1}{\log(1 + 2x^2)}$$

$$\lim_{x \rightarrow 0} \frac{x(\sin 2x)^2}{\sin(x^3)}$$

$$\lim_{x \rightarrow \pm\infty} (\sqrt{x^2 + 3x + 2} - |x|)$$

$$\lim_{x \rightarrow +\infty} \frac{\log(\log x)}{1 + \log x}$$

$$\lim_{x \rightarrow 1} \frac{(\log x)^2}{(2x-2)^2}$$

$$\lim_{x \rightarrow 0} x e^{\sin \frac{1}{x}}$$

$$\lim_{x \rightarrow +\infty} x \log \left( \frac{x+5}{x-1} \right)$$

$$\lim_{x \rightarrow +\infty} x e^{\sin x}$$

Per ciascuna delle seguenti funzioni: dire dove è definita e dove è continua; calcolare i limiti alla frontiera dell'insieme di definizione (punti in cui non è definita ed eventualmente  $\pm\infty$ ); determinare tutti gli eventuali asintoti (verticali, orizzontali, obliqui); utilizzando opportune stime asintotiche nei punti in cui la funzione si annulla e all'infinito, prevedere l'andamento della funzione nell'intorno di tali punti.

$$\lim_{x \rightarrow 2} \frac{x^2 + 3x - 1}{x+2}$$

$$\lim_{x \rightarrow 1} \frac{x^3 + 2x - 1}{x+2}$$

$$\lim_{x \rightarrow +\infty} x e^{\frac{2x+1}{x+3}}$$

$$\lim_{x \rightarrow 1} \frac{\log x}{\sqrt[3]{x-1}}$$

$$\lim_{x \rightarrow 0} \frac{\cos x - 1}{x}$$

$$\lim_{x \rightarrow 0} \log \left( \frac{2x+1}{x^2+x-2} \right)$$

$$\lim_{x \rightarrow 1} \arctg \left( \frac{x+2}{x-1} \right)$$

$$\lim_{x \rightarrow 0} x \arctg \frac{1}{x}$$

$$\lim_{x \rightarrow +\infty} x \log \left( \frac{2x^2+3}{x^2+x+1} \right)$$

$$\lim_{x \rightarrow 0} 2x e^{-1/x}$$

$$\lim_{x \rightarrow 0} x^2 e^{-|x|}$$

$$\lim_{x \rightarrow +\infty} \frac{\sqrt[3]{x} \log|x|}{\log|x+1|}$$

$$\lim_{x \rightarrow 1} \frac{x^2 + 2x - 1}{x^2 - 1}$$

$$\lim_{x \rightarrow 0} x e^{-1/|x|}$$

**12** L'impedenza  $Z$  di un circuito elettrico  $RLC$ :  $R$  (resistenza),  $L$  (induttanza),  $C$  (capacità), è data da

$$Z(\omega) = \sqrt{R^2 + \left( L\omega - \frac{1}{C\omega} \right)^2} \quad (\omega > 0).$$

Studiare, con l'aiuto del computer, come varia  $Z$  al variare dei parametri  $R$ ,  $L$ ,  $C$ .

## 4 PROPRIETÀ GLOBALI DELLE FUNZIONI CONTINUE O MONOTONE SU UN INTERVALLO

### 4.1 Funzioni continue su un intervallo

Le proprietà delle funzioni continue che abbiamo incontrato finora si basano sostanzialmente sul concetto di continuità *in un punto*. Come vedremo ora, invece, è soprattutto il fatto che una funzione sia continua su un intervallo ad avere delle conseguenze interessanti. La proprietà di continuità su un intervallo ha una semplice interpretazione geometrica: il grafico della funzione, su quell'intervallo, si può tracciare “senza staccare la penna dal foglio”. Il concetto di funzione continua si presta quindi a tradurre in maniera rigorosa l'idea geometrica di *linea continua*, e quella fisica di *grandezza che varia con continuità*. Questo è vero, però, pur di utilizzare funzioni definite su un intervallo di  $\mathbb{R}$ . Come vedremo, infatti, le proprietà notevoli delle funzioni continue su un intervallo dipendono in maniera cruciale dall'assioma di continuità di  $\mathbb{R}$ , e non varrebbero, ad esempio, se il nostro ambiente di lavoro fosse il campo  $\mathbb{Q}$  dei razionali.

La prima proprietà di cui ci occupiamo, che chiamiamo *teorema degli zeri*, riguarda la risoluzione di un'equazione del tipo

$$(4.1) \quad f(x) = 0$$

Quando  $f$  è un polinomio di grado  $\leq 4$  esistono formule che forniscono le soluzioni della (4.1) mediante radicali. Se però  $f$  è un polinomio di grado  $> 4$  o una funzione più complicata, salvo casi particolarmente fortunati, non esistono formule per le soluzioni. Geometricamente, risolvere la (4.1) significa determinare le ascisse dei punti di intersezione tra il grafico di  $y = f(x)$  e l'asse delle ascisse. Naturalmente possono esserci infinite soluzioni, un numero finito di soluzioni, nessuna soluzione. Ogni soluzione si chiama *zero di  $f$* .

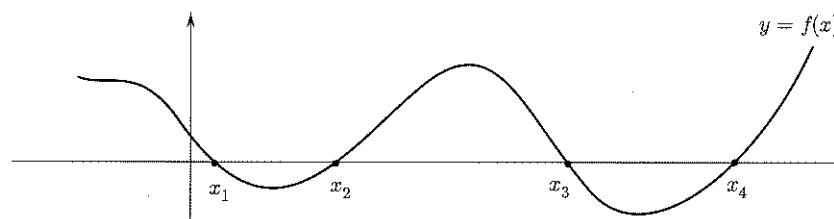


Figura 3.10. L'equazione  $f(x) = 0$  ha 4 soluzioni, ossia  $f$  ha 4 zeri.

Il teorema degli zeri dà alcune semplici condizioni sotto le quali esiste uno zero di  $f$  e anche un modo per calcolarlo.

**TEOREMA 3.26 (DEGLI ZERI)** Sia:

- i)  $f$  continua in  $[a, b]$
- ii)  $f(a) \cdot f(b) < 0$

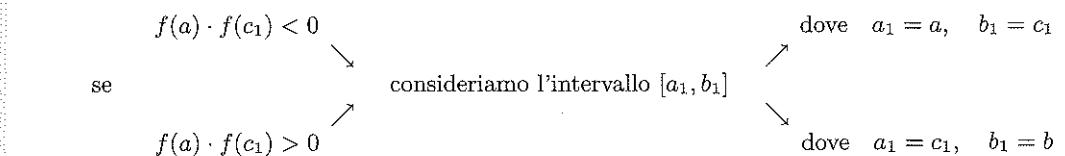
Allora esiste  $c \in (a, b)$  tale che  $f(c) = 0$ .

Se  $f$  è anche strettamente monotona, lo zero è unico.

**DIMOSTRAZIONE.** Costruiamo una successione che tende a uno zero di  $f$ . Poniamo  $c_1 = \frac{a+b}{2}$ , punto medio dell'intervallo  $[a, b]$ .

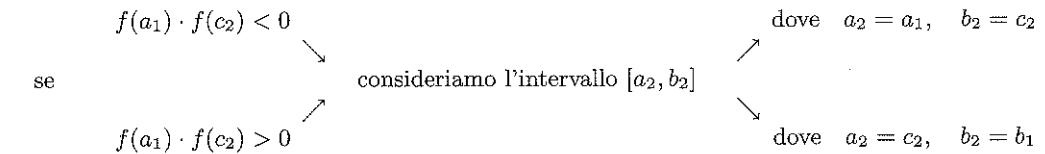
Se  $f(c_1) = 0$  siamo fortunati e il teorema è dimostrato.

Se  $f(c_1) \neq 0$  guardiamo il segno di  $f(a) \cdot f(c_1)$  e procediamo come segue:



Poniamo ora  $c_2 = \frac{b_1+a_1}{2}$ , punto medio dell'intervallo  $[a_1, b_1]$ , calcoliamo  $f(c_2)$  e procediamo come prima:

se  $f(c_2) = 0$  lo zero cercato è  $c_2$  e il teorema è dimostrato; se  $f(c_2) \neq 0$  guardiamo il segno di  $f(a_1) \cdot f(c_2)$ :



Continuando in questo modo troviamo una sequenza di intervalli  $[a_n, b_n]$  con le seguenti proprietà:

- 1)  $a_n \leq a_{n+1}$  e  $b_n \geq b_{n+1}$  ( $\{a_n\}$  cresce e  $\{b_n\}$  decresce)
- 2)  $b_n - a_n = \frac{b-a}{2^n}$  (ciascun intervallo è lungo la metà del precedente)
- 3)  $f(a_n) \cdot f(b_n) < 0$  (per come sono stati scelti  $a_n$  e  $b_n$  a ogni passo)

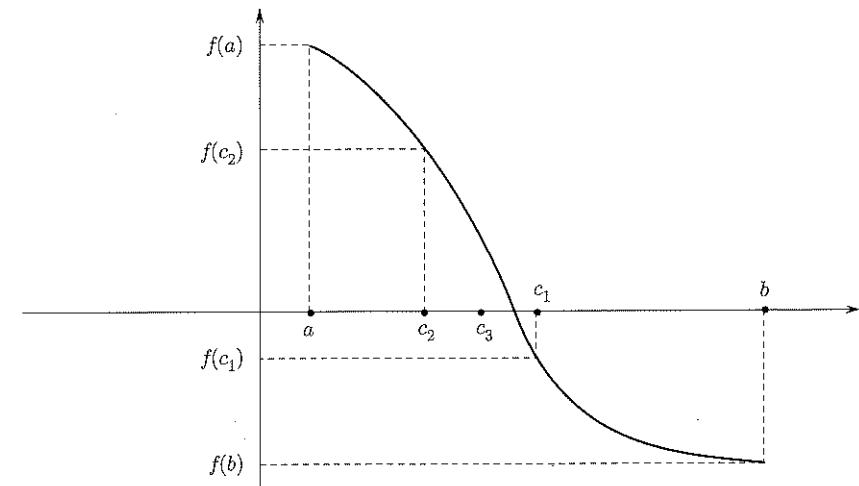


Figura 3.11.

Per la 1), possiamo allora dedurre che le successioni  $\{a_n\}$  e  $\{b_n\}$  hanno limite finito, in base al teorema di esistenza del limite delle successioni monotone; cioè:

$$a_n \rightarrow \ell_1 \quad \text{e} \quad b_n \rightarrow \ell_2, \quad n \rightarrow +\infty$$

(Si noti che le successioni sono anche limitate perché contenute in  $[a, b]$ .)

Dalla 2) deduciamo che  $b_n - a_n = \frac{b-a}{2^n} \rightarrow 0$  se  $n \rightarrow +\infty$  e perciò  $\ell_1 = \ell_2 = \ell$ .

Per la continuità di  $f$ , abbiamo allora che

$$f(a_n) \cdot f(b_n) \rightarrow f(\ell)^2$$

mentre dalla 3) e dal teorema della permanenza del segno deduciamo  $f(\ell)^2 \leq 0$ . Deve perciò essere  $f(\ell) = 0$  e così  $\ell$  è lo zero cercato.

Arrestando il procedimento indicato nella dimostrazione dopo  $n$  passi, si può assumere  $a_n$  o  $b_n$  come valore approssimato dello zero  $\ell$  di  $f$ , rispettivamente, per difetto o per eccesso. L'errore commesso è non superiore a  $\frac{b-a}{2^n}$ .

Si osservi che, se in  $[a, b]$  vi sono più zeri della funzione  $f$ , il procedimento illustrato non indica quale di essi venga determinato.

Una seconda importante proprietà delle funzioni continue è espressa dal seguente:

**TEOREMA 3.27 (DI WEIERSTRASS)** *Sia  $f : [a, b] \rightarrow \mathbb{R}$  una funzione continua. Allora  $f$  assume massimo e minimo in  $[a, b]$ , ossia: esistono  $x_m, x_M \in [a, b]$  tali che  $f(x_m) \leq f(x) \leq f(x_M)$  per ogni  $x \in [a, b]$ .*

Si dice che  $x_m$  è punto di minimo per  $f$ , e  $m = f(x_m)$  è il minimo di  $f$ ; analogamente,  $x_M$  è punto di massimo, e  $M = f(x_M)$  è il massimo di  $f$ .

**DIMOSTRAZIONE.** Proveremo che  $f$  ammette massimo in  $[a, b]$ . Analogamente si prova che ha minimo.

Cominciamo con la seguente osservazione sul concetto di estremo superiore: se  $E_1, E_2$  sono due sottoinsiemi non vuoti di  $\mathbb{R}$ ,

$$\sup(E_1 \cup E_2) = \max(\sup E_1, \sup E_2).$$

Questa proprietà è vera per insiemi sia limitati che illimitati (se  $E_i$  è superiormente illimitato si pone  $\sup E_i = +\infty$ ; si intende che  $\max(a, +\infty) = +\infty$ ). Sia ora

$$f : I \rightarrow \mathbb{R} \text{ con } I = I_1 \cup I_2$$

( $I, I_1, I_2$  intervalli). Applicando l'osservazione precedente agli insiemi  $E_i = \{f(x) : x \in I_i\}$  si ha:

$$\sup_I f = \max \left( \sup_{I_1} f, \sup_{I_2} f \right).$$

In particolare quindi, è vera almeno una delle due relazioni:

$$\sup_I f = \sup_{I_1} f; \quad \sup_I f = \sup_{I_2} f.$$

Veniamo alla dimostrazione del teorema. Useremo una costruzione iterativa simile a quella vista nella dimostrazione del teorema degli zeri. Consideriamo la funzione  $f : [a, b] \rightarrow \mathbb{R}$ , e poniamo

$$\Lambda = \sup_{[a, b]} f$$

(con  $\Lambda$  eventualmente uguale a  $+\infty$ ). Suddividiamo  $[a, b]$  in due intervalli uguali; per quanto osservato sopra, per uno di essi, che chiamiamo  $[a_1, b_1]$ , sarà vero che

$$\Lambda = \sup_{[a_1, b_1]} f.$$

Suddividiamo  $[a_1, b_1]$  in due intervalli uguali; per uno di questi, che chiamiamo  $[a_2, b_2]$ , sarà vero che

$$\Lambda = \sup_{[a_2, b_2]} f.$$

Procedendo per dicotomia, così facendo costruiamo una successione di intervalli  $[a_n, b_n]$ , ciascuno contenuto nei precedenti, con le proprietà:

1.  $a_n$  monotona crescente e limitata;  $b_n$  monotona decrescente e limitata;
2.  $b_n - a_n = \frac{b-a}{2^n} \rightarrow 0$ ;
3.  $\Lambda = \sup_{[a_n, b_n]} f$ .

Per lo stesso ragionamento fatto nella dimostrazione del teorema degli zeri (che utilizza il teorema di monotonia per le successioni), dai punti 1) e 2) segue che le successioni  $a_n$  e  $b_n$  convergono ad uno stesso limite  $x_0 \in [a, b]$ . Distinguiamo ora due casi:  $\Lambda < +\infty$ ,  $\Lambda = +\infty$  (proveremo che questo secondo caso in realtà non può verificarsi). Se  $\Lambda < +\infty$ , per ogni  $n$  esiste un punto  $t_n \in [a_n, b_n]$  tale che

$$(4.2) \quad \Lambda - \frac{1}{n} < f(t_n) \leq \Lambda.$$

Infatti: poiché  $\Lambda - \frac{1}{n}$  è minore di  $\Lambda$  che è il minimo dei maggioranti dei valori di  $f(x)$  in  $[a_n, b_n]$ ,  $\Lambda - \frac{1}{n}$  non è un maggiorante, quindi esiste  $t_n \in [a_n, b_n]$  con la proprietà (4.2).

Poiché  $t_n \in [a_n, b_n]$ ,  $a_n \rightarrow x_0$  e  $b_n \rightarrow x_0$ , per il teorema del confronto anche  $t_n \rightarrow x_0$ .

Ancora per il teorema del confronto, la (4.2) dà allora

$$\lim_{n \rightarrow \infty} f(t_n) = \Lambda.$$

D'altro canto, poiché  $f$  è continua e  $t_n \rightarrow x_0$ , si ha che  $\lim_{n \rightarrow \infty} f(t_n) = f(x_0)$ , perciò  $f(x_0) = \Lambda$ . Allora, poiché  $\Lambda$  è il sup dei valori di  $f$  in  $[a, b]$ ,  $\Lambda$  è il massimo di  $f$  in  $[a, b]$ , ed è assunto nel punto  $x_0$ . In questo caso perciò il teorema è dimostrato.

Supponiamo ora che sia  $\Lambda = +\infty$ . Allora per ogni  $n$  esiste un punto  $t_n \in [a_n, b_n]$  tale che

$$(4.3) \quad f(t_n) \geq n.$$

Ragionando come sopra si prova che  $t_n \rightarrow x_0$  per un certo  $x_0 \in [a, b]$ . Poiché  $f$  è continua in  $x_0$ ,

$$\lim_{n \rightarrow \infty} f(t_n) = f(x_0),$$

ma per la (4.3)  $\lim_{n \rightarrow \infty} f(t_n) = +\infty$ , assurdo perché in  $x_0$  la funzione deve avere un valore finito.

Mostriamo ora che le ipotesi del teorema di Weierstrass sono tutte necessarie, con i seguenti esempi:



**ESEMPIO**  $f(x) = x$  in  $(0, 1)$ .

La funzione è continua su un intervallo limitato, ma non chiuso. La funzione non ha né massimo né minimo (il suo estremo superiore, 1, e il suo estremo inferiore, 0, non sono assunti).

**ESSERE**  $f(x) = x$  in  $\mathbb{R}$ . La funzione è continua su un intervallo non limitato. La funzione non ha né massimo né minimo (non è nemmeno limitata).

$$f(x) = \begin{cases} x & \text{per } x \in (0, 1) \\ \frac{1}{2} & \text{per } x = 0, x = 1 \end{cases}$$

La funzione è definita su un intervallo chiuso e limitato  $[0, 1]$ , ma non è continua. La funzione non ha né massimo né minimo (il suo estremo superiore, 1, e il suo estremo inferiore, 0, non sono assunti).

**TEOREMA 3.28 (DEI VALORI INTERMEDI)** Se  $f$  è continua su  $[a, b]$ , allora per ogni valore  $\lambda$  compreso tra  $m$  e  $M$  (minimo e massimo di  $f$  in  $[a, b]$ ), esiste un ingresso  $x$  in  $[a, b]$  che ha il valore  $\lambda$  come uscita (proprietà dei valori intermedi).

**DIMOSTRAZIONE.** Quest'ultimo teorema è una semplice conseguenza del teorema di Weierstrass e del teorema degli zeri: sia  $m < \lambda < M$ ,  $f(x_2) = m$ ,  $f(x_1) = M$ . Allora la funzione  $g(x) = f(x) - \lambda$  è continua in  $[x_1, x_2]$  (supponendo  $x_1 < x_2$ ) e  $g(x_1) = f(x_1) - \lambda = M - \lambda > 0$ ,  $g(x_2) = f(x_2) - \lambda = m - \lambda < 0$ . Dal teorema degli zeri, esiste  $\ell$  tale che  $g(\ell) = 0$  e cioè  $f(\ell) = \lambda$ .

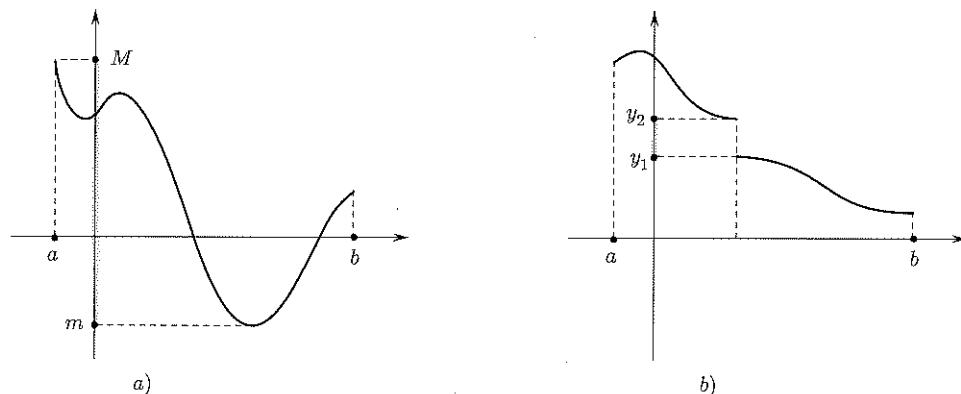


Figura 3.12. a) L'immagine di  $[a, b]$  è l'intervallo  $[m, M]$ . b) Una funzione che non ha la proprietà dei valori intermedi. Ogni valore  $\lambda$  tra  $y_1$  e  $y_2$  non è un'uscita di  $f$ .

Le proprietà dei due teoremi precedenti si possono sintetizzare nell'unico enunciato seguente:

**TEOREMA 3.29** Se  $f : [a, b] \rightarrow \mathbb{R}$  è continua, allora  $f([a, b]) = [m, M]$ ; cioè l'immagine di un intervallo  $[a, b]$  è l'intervallo di estremi  $m = \min_{[a,b]} f$  e  $M = \max_{[a,b]} f$  (vedi fig. 3.12a).

Riguardo al teorema dei valori intermedi, si consideri il prossimo semplice



**ESSERE** Sia  $f(x) = x^2$  e consideriamo  $f$  come funzione definita dall'insieme  $\mathbb{Q}$  dei razionali a  $\mathbb{Q}$  stesso (questo è lecito perché il quadrato di un razionale è un razionale). Allora  $f$  non ha la proprietà dei valori intermedi. Infatti, ad esempio,  $f(1) = 1$ ,  $f(2) = 4$ , ma  $f$  non assume tutti i valori razionali compresi tra 1 e 4: per esempio non assume mai il valore 2, o 3. In altre parole, la proprietà dei valori intermedi è vera per le funzioni continue grazie alle proprietà dell'insieme dei numeri reali. Questi, anzi, sono i motivi più interessanti per l'Analisi per cui è utile avere come ambiente di lavoro l'insieme dei reali e non l'insieme dei razionali.

Mostriamo come il teorema degli zeri permetta di dimostrare direttamente il teorema sull'esistenza della radice  $n$ -esima (enunciato nel cap.1, par. 5.1):

**TEOREMA 3.30** Per ogni  $y \in \mathbb{R}, y > 0$  e  $n \in \mathbb{N}, n \geq 1$ , esiste uno e un solo  $x \in \mathbb{R}, x > 0$ , tale che  $x^n = y$ . (Questo numero  $x$  si definisce radice  $n$ -esima di  $y$ ).

**DIMOSTRAZIONE.** Consideriamo la funzione  $f(x) = x^n$ . Questa funzione è continua su tutto  $\mathbb{R}$ , ad esempio perché è il prodotto di  $n$  funzioni continue  $g(x) = x$ . Mostriamo che esistono  $x_2 > x_1 > 0$  tali che  $x_1^n < y < x_2^n$ . Infatti:

se  $y \geq 1$  basta scegliere  $x_1 = 1, x_2 = y$  e si ha  $x_1^n \leq y < x_2^n$  ( $1 \leq y < y^n$ );  
se  $y < 1$ , basta scegliere  $x_1 = y, x_2 = 1$  e si ha  $x_1^n < y \leq x_2^n$  ( $y^n < y \leq 1$ ).

Allora possiamo applicare il teorema dei valori intermedi alla funzione continua  $f(x) = x^n$  sull'intervallo  $[x_1, x_2]$ : poiché  $f$  è strettamente crescente, il suo minimo è  $x_1^n$  e il suo massimo è  $x_2^n$ ; poiché  $x_1^n < y < x_2^n$ , esiste  $x_0 \in [x_1, x_2]$  tale che  $x_0^n = y$ .

## 4.2 Funzioni monotone su un intervallo

Ci occupiamo ora di funzioni monotone su un intervallo (e non necessariamente continue). Il prossimo teorema si può vedere come una estensione alle funzioni del teorema di monotonia per le successioni, visto nel par. 1.2. Questo teorema condivide con i teoremi sulle funzioni continue su un intervallo, che abbiamo discusso nel paragrafo precedente, due caratteristiche importanti:

la sua dimostrazione si basa sull'*assioma di continuità* di  $\mathbb{R}$ ;

il teorema assume come ipotesi una *proprietà globale* della funzione (ossia la sua monotonia su un intervallo).

**TEOREMA 3.31 (DI MONOTONIA)** Sia  $f : (a, b) \rightarrow \mathbb{R}$  una funzione monotona. Allora per ogni  $c \in (a, b)$  esistono finiti i limiti destro e sinistro, per  $x \rightarrow c$ ; ai due estremi  $a, b$  esistono i limiti destro (in  $a$ ) e sinistro (in  $b$ ), eventualmente infiniti.

**DIMOSTRAZIONE.** Questo teorema non sarà dimostrato utilizzando il teorema di monotonia per le successioni, ma piuttosto ripetendo una dimostrazione analoga.

Per fissare le idee, supponiamo  $f$  crescente in  $(a, b)$ . Sia  $c \in (a, b)$  e sia

$$\Lambda = \sup \{f(x) : x \in (a, c)\}.$$

Notiamo che  $\Lambda$  esiste finito per la proprietà dell'estremo superiore, in quanto  $f(c)$  è un maggiorante dell'insieme  $\{f(x) : x \in (a, c)\}$ . Proviamo che

$$\lim_{x \rightarrow c^-} f(x) = \Lambda.$$

Sia dunque  $x_n$  una qualsiasi successione in  $(a, c)$  tale che  $x_n \rightarrow c$ , e proviamo che  $f(x_n) \rightarrow \Lambda$ , ossia che per ogni  $\varepsilon > 0$  risulta definitivamente

$$\Lambda - \varepsilon < f(x_n) < \Lambda + \varepsilon.$$

La seconda diseguaglianza è ovvia in quanto  $f(x_n) \leq \Lambda$  per definizione di  $\Lambda$ , perché  $x_n \in (a, c)$ . Per provare la prima, osserviamo che essendo  $\Lambda - \varepsilon$  minore di  $\Lambda$ , cioè del minimo maggiorante di  $\{f(x) : x \in (a, c)\}$ , non è un maggiorante di tale insieme, perciò esiste un punto  $\bar{x} \in (a, c)$  tale che  $f(\bar{x}) \geq \Lambda - \varepsilon$ . Poiché  $f$  è crescente, ne segue che

$$f(x) \geq \Lambda - \varepsilon \quad \forall x \in (\bar{x}, c).$$

D'altra canto per ipotesi  $x_n < c \ \forall n$  e  $x_n \rightarrow c$  per  $n \rightarrow \infty$ , perciò  $x_n \in (\bar{x}, c)$  definitivamente. Ma allora

$$f(x_n) \geq \Lambda - \varepsilon \text{ definitivamente,}$$

che è quanto occorreva provare.

Analogamente si può dimostrare che esiste

$$\lim_{x \rightarrow c^+} f(x) = \inf \{f(x) : x \in (c, b)\}.$$

Per quanto riguarda i limiti ai due estremi dell'intervallo, se si vuole provare ad esempio che esiste

$$\lim_{x \rightarrow b^-} f(x),$$

si definirà ancora

$$\Lambda = \sup \{f(x) : x \in (a, b)\};$$

tuttavia in questo caso  $\Lambda$  potrebbe anche essere  $+\infty$  (non possiamo affermare che  $f(b)$  sia un maggiorante dell'insieme, perché in  $b$  la funzione non è definita). Nel caso in cui  $\Lambda < \infty$  si può ripetere la dimostrazione precedente; se invece  $\Lambda = \infty$ , il ragionamento si modifica come segue. Dall'ipotesi

$$\sup \{f(x) : x \in (a, b)\} = +\infty$$

segue che per ogni  $K > 0$  esiste  $\bar{x} \in (a, b)$  tale che  $f(\bar{x}) > K$ ; per la monotonia di  $f$ , allora,

$$f(x) > K \quad \forall x \in (\bar{x}, b),$$

allora presa una successione  $x_n \rightarrow b$ , si ha che  $x_n \in (\bar{x}, b)$  definitivamente, e quindi

$$f(x_n) > K \text{ definitivamente,}$$

perciò

$$\lim_{x \rightarrow b^-} f(x_n) = +\infty.$$



Una conseguenza del teorema di monotonia è che se una funzione è monotona in un intervallo  $(a, b)$ , i suoi eventuali punti di discontinuità in  $(a, b)$  sono necessariamente *discontinuità a salto*, ad eccezione degli estremi  $a, b$ , in cui può avversi anche un asintoto verticale.

### 4.3 Continuità e invertibilità

Sia  $f : I \rightarrow \mathbb{R}$ , con  $I$  intervallo. Occupiamoci ancora del problema dell'invertibilità di  $f$ , vedendo ora il problema dal punto di vista della continuità.

Sappiamo già che una funzione definita su un intervallo e strettamente monotona è invertibile (con inversa monotona; v. cap. 2, par. 4.2). Sappiamo anche che il viceversa non è vero in generale: esistono funzioni invertibili su un intervallo, e non monotone (si veda la figura seguente).

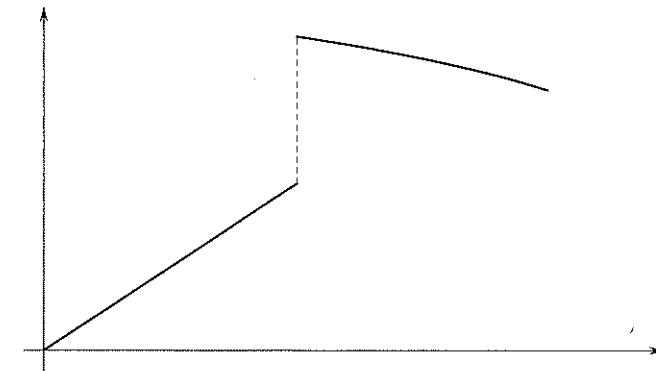


Figura 3.13.

Tuttavia, se aggiungiamo l'ipotesi della continuità il teorema si inverte:

**TEOREMA 3.32** *Sia  $f : I \rightarrow \mathbb{R}$ , con  $I$  intervallo, una funzione continua in  $I$ . Allora  $f$  è invertibile in  $I$  se e solo se è strettamente monotona. In tal caso la sua inversa è ancora strettamente monotona e continua.*

**DIMOSTRAZIONE.** Sappiamo già che se  $f$  è strettamente monotona è invertibile (indipendentemente dalle ipotesi che  $f$  sia continua, e che sia definita su un intervallo). Mostriamo che vale il viceversa, ossia che se è continua e invertibile, allora è strettamente monotona. Per assurdo, non sia strettamente monotona, allora esistono tre punti  $x_1 < x_2 < x_3$  in  $I$  tali che

$$f(x_1) < f(x_2) \text{ e } f(x_2) > f(x_3),$$

oppure tali che

$$f(x_1) > f(x_2) \text{ e } f(x_2) < f(x_3).$$

Per fissare le idee, supponiamo che valga la prima delle due alternative (nell'altro caso si ragionerà analogamente). Confrontiamo i valori di  $f(x_1)$  e  $f(x_3)$ ; non possono essere uguali perché  $f$  per ipotesi è invertibile, dunque  $f(x_1) < f(x_3)$  oppure  $f(x_1) > f(x_3)$ . Di nuovo, per fissare le idee supponiamo che valga la prima delle due alternative (nell'altro caso si ragionerà analogamente). Dunque sappiamo che:

$$x_1 < x_2 < x_3 \text{ e } f(x_1) < f(x_3) < f(x_2).$$

Poiché  $f$  è continua, per il teorema dei valori intermedi esiste  $x_0 \in (x_1, x_2)$  tale che  $f(x_0) = f(x_3)$ . Poiché  $x_0 \neq x_3$  (perché  $x_0 < x_2 < x_3$ ), ne segue che  $f$  non può essere invertibile, assurdo.

Questo dimostra la prima parte del teorema. Sia ora  $f$  continua, strettamente monotona e quindi invertibile in  $I$ , e sia  $g$  la sua funzione inversa, ancora strettamente monotona e

invertibile. Proviamo che  $g$  è continua. Per quanto osservato dopo il teorema di monotonia, la funzione  $g$ , strettamente monotona, o è continua, oppure ha dei punti di discontinuità a salto. In tal caso l'immagine di  $g$  non è un intervallo (ma è l'unione di almeno due intervalli disgiunti), il che è assurdo perché tale immagine è  $I$ . Dunque  $g$  è continua.

Si noti che nella dimostrazione del teorema precedente si sono utilizzati sia il teorema dei valori intermedi, sia il teorema di monotonia per le funzioni. Di nuovo, quindi, abbiamo fatto implicitamente uso, in modo essenziale, dell'assioma di continuità di  $\mathbb{R}$ . Il teorema appena dimostrato significa in particolare che:

**Una funzione continua e invertibile su un intervallo, ha inversa continua.**

Questo fatto completa la dimostrazione del teorema di continuità delle funzioni elementari; infatti:

la continuità di  $a^x$  implica la continuità di  $\log_a x$ ;  
la continuità di  $\sin x$ ,  $\cos x$ ,  $\operatorname{tg} x$  implica la continuità di  $\arcsin x$ ,  $\arccos x$ ,  $\operatorname{arctg} x$ .

## COMPLEMENTI

**7** Dimostrare che se per ogni successione  $x_n \rightarrow c$ , con  $x_n \neq c \forall n$ , la proprietà  $p(x_n)$  è vera definitivamente, allora la proprietà  $p(x)$  è vera definitivamente per  $x \rightarrow c$ .

Questo fatto è stato implicitamente utilizzato nella dimostrazione del teorema di permanenza del segno per le funzioni.

*Suggerimento:* ragionare per assurdo, scrivendo in maniera esplicita e quantitativa cosa significa la falsità della tesi, e costruire una successione che contraddice l'ipotesi.

**8** Dimostrare la seguente variante del Teorema di Weierstrass su intervalli illimitati:

“Sia  $f : \mathbb{R} \rightarrow \mathbb{R}$  continua,  $f(x) \geq 0$  per ogni  $x \in \mathbb{R}$  e supponiamo che  $f(x) \rightarrow 0$  per  $x \rightarrow \pm\infty$ . Allora  $f$  ha massimo in  $\mathbb{R}$ ”.

*Suggerimento:* sia  $x_0$  un punto in cui  $f(x_0) > 0$ ; poiché  $\lim_{x \rightarrow +\infty} f(x) = 0$ , esiste  $K > 0$  tale che  $f(x) < f(x_0) \forall |x| > K$ . Applicando il teorema di Weierstrass sull'intervallo  $[-K, K]$ ....

Mostrare poi con esempi che:

- il teorema è falso senza l'ipotesi  $f(x) \geq 0$ ;
- il teorema è falso senza l'ipotesi  $\lim_{x \rightarrow \pm\infty} f(x) = 0$ ;
- nelle ipotesi del teorema, il *minimo* di  $f$  in  $\mathbb{R}$  può non esistere.

**9** Dimostrare le seguenti varianti del Teorema degli zeri su intervalli illimitati e su intervalli aperti:

(a) “Sia  $f : \mathbb{R} \rightarrow \mathbb{R}$  continua e supponiamo che

$$\lim_{x \rightarrow +\infty} f(x) \cdot \lim_{x \rightarrow -\infty} f(x) < 0$$

(supponendo anche che i due limiti esistano, finiti o infiniti). Allora esiste  $c \in \mathbb{R}$  tale che  $f(c) = 0$ ”.

(b) “Sia  $f : (a, b) \rightarrow \mathbb{R}$  continua e supponiamo che

$$\lim_{x \rightarrow a^+} f(x) \cdot \lim_{x \rightarrow b^-} f(x) < 0$$

(supponendo anche che i due limiti esistano, finiti o infiniti). Allora esiste  $c \in (a, b)$  tale che  $f(c) = 0$ ”.

*Suggerimento:* usando l'ipotesi sui limiti, mostrare che in entrambi i casi esiste un intervallo  $[x_1, x_2]$  contenuto nel dominio della funzione, tale che  $f(x_1) \cdot f(x_2) < 0$ , e quindi...

**10** Utilizzando opportunamente i teoremi enunciati nell'esercizio 9, dimostrare che l'equazione

$$\operatorname{tg} x + e^x - 5 = 0$$

ha almeno una soluzione in  $(-\frac{\pi}{2}, \frac{\pi}{2})$ .

**11** Enunciando e dimostrando un'opportuna variante del teorema dimostrato nell'esercizio 8, provare che la funzione

$$f(x) = x^3 e^{-x^2}$$

ha massimo e minimo in  $\mathbb{R}$ .

## 4

# Calcolo differenziale per funzioni di una variabile

## ■ 1 INTRODUZIONE AL CALCOLO DIFFERENZIALE

In questa introduzione vogliamo far riflettere l'allievo su alcuni *problem*i da cui il calcolo differenziale è nato. Porremo questi problemi sotto forma di domande, a cui non risponderemo subito, ma nel seguito del capitolo. Come si vedrà, la risposta a questi problemi coinvolge la nozione di *limite*, discussa in precedenza. Anzi, si può dire che storicamente la nozione di limite sia stata introdotta principalmente per sviluppare le idee del calcolo differenziale, che introdurremo in questo capitolo, e del calcolo integrale, di cui ci occuperemo nel capitolo 6.

### Che cos'è la retta tangente in un punto a una curva?

Anche se geometricamente il concetto sembra intuitivo, non è così ovvio quale possa essere una definizione corretta. Consideriamo la più semplice figura curvilinea, ossia la circonferenza. Cos'è la tangente a una circonferenza in un suo punto  $P$ ? In base alla geometria elementare possiamo rispondere:

1. È la retta passante per  $P$  e perpendicolare al raggio passante per  $P$ .  
Oppure:
2. È l'unica retta passante per  $P$  che non interseca la circonferenza in altri punti.

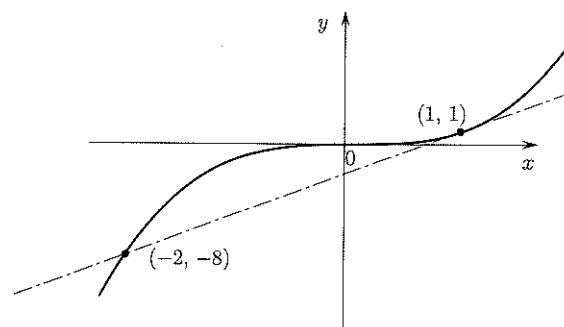


Figura 4.1. La retta tangente alla curva  $y = x^3$  nel punto  $(1, 1)$  è  $y = 3x - 2$ , che taglia la curva anche in  $(-2, -8)$ .

Queste definizioni si possono estendere a curve più generali? La definizione 1 certamente no: cos'è il raggio di una parabola, ad esempio? La definizione 2 sembra più incoraggiante perché "funziona" anche per le coniche (parabole, ellissi, iperbolici). Ma già per la curva  $y = x^3$  non funziona: la retta tangente a questa curva in un punto generico (escluso l'origine) taglia la curva anche in un altro punto (fig. 4.1); viceversa ci sono curve, come  $y = |x|$  che in un punto (l'origine) hanno infinite rette soddisfacenti questa definizione, ma intuitivamente nessuna di esse si può chiamare tangente (fig. 4.2).

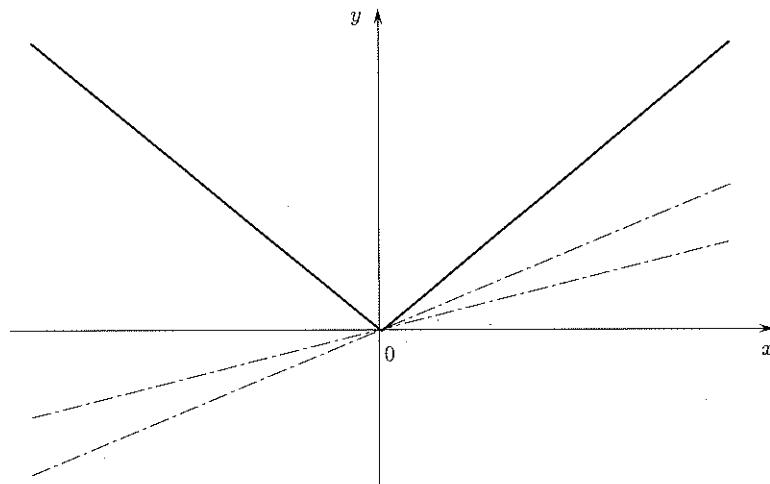


Figura 4.2. La funzione  $y = |x|$  ha infinite rette che tagliano il suo grafico solo nell'origine: nessuna di esse è "tangente".

Ci accorgiamo dunque di parlare comunemente di un oggetto geometrico ("retta tangente") di cui non sappiamo (ancora) dare una definizione precisa e generale.

In realtà non c'è speranza di definire la tangente come la retta che ha "meno intersezioni delle altre" con la curva. Ci vuole un'idea diversa. Geometricamente, la retta tangente può essere approssimata dalla retta che passa per due punti molto vicini posti sulla curva, il primo fissato (il punto di tangenza) e l'altro "mobile" (e scelto via via più vicino al punto di tangenza). Due punti individuano una retta; più i due punti si avvicinano, più questa retta si avvicinerà alla tangente; se però partissimo subito dal considerare "due punti coincidenti", avremmo un solo punto, per il quale passano infinite rette, e la retta che cerchiamo resterebbe indeterminata. Si tratta quindi di capire cosa accade della retta per i due punti quando il secondo punto, mobile, si avvicina sempre più al primo, senza però coincidere con esso. La "retta limite" – se esiste – si chiamerà retta tangente.

Notiamo che il problema della determinazione della tangente è importante non solo da un punto di vista geometrico, ma anche per un'importante applicazione: è facile convincersi (vedi fig. 4.3) che una curva il cui grafico sia "regolare" ha tangente orizzontale nei punti di massimo e di minimo (ed eventualmente anche in altri punti).

Perciò, per cercare i punti di massimo e minimo di una funzione, è utile saper scrivere analiticamente l'equazione della retta tangente alla curva in un punto gene-

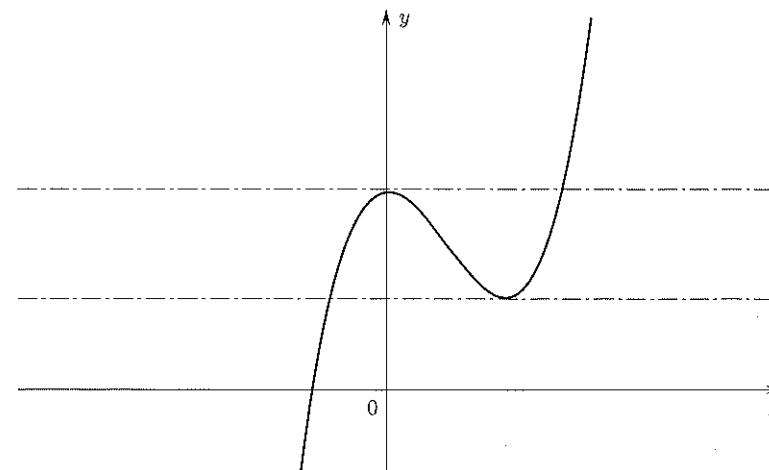


Figura 4.3.

rico, per vedere poi in quali punti essa è orizzontale. Questa idea si deve per primo a Fermat, che la elaborò intorno al 1630. Ad esempio, la meccanica insegna che un sasso lanciato verso l'alto (nel vuoto) con una velocità iniziale  $v_0$ , dopo un tempo  $t$  si trova a un'altezza

$$h(t) = v_0 t - \frac{1}{2} g t^2,$$

dove  $g$  è l'accelerazione di gravità, pari a  $9,8 \text{ m/s}^2$ . Problema: qual è il punto più alto raggiunto dal sasso? È il valore massimo assunto dalla funzione  $h(t)$ ; cerchiamo dunque il valore di  $t$  per cui la curva  $h(t)$  ha retta tangente orizzontale; in realtà questa curva è una parabola, e senza scomodare il calcolo infinitesimale sappiamo che ha valore massimo per  $t =$  ascissa del vertice, cioè  $t = v_0/g$ ; perciò la quota massima raggiunta è  $h(v_0/g) = \frac{1}{2} v_0^2/g$ . È chiaro però che in problemi più generali potremmo trovare una funzione  $h(t)$  "qualsiasi", e si pone il problema di avere un *metodo generale* per determinare il punto in cui la tangente è orizzontale, e prima ancora per calcolare il coefficiente angolare della retta tangente in un punto qualsiasi.

Un punto istruttivo di questa discussione è il seguente: senza il calcolo infinitesimale non si potrebbe non solo *risolvere* certi problemi, ma neppure *dar senso* ad essi (ad esempio, definire rigorosamente il concetto di retta tangente).

#### Che cos'è la velocità di un oggetto in moto?

Questo è un altro concetto che sembra ovvio ma non lo è. Consideriamo un punto materiale che si muove di moto vario, e proponiamoci di calcolare la sua velocità, nota la legge oraria del moto (ossia lo spazio percorso in funzione del tempo). Nella fisica galileiana l'unica velocità di cui si parla esplicitamente è la *velocità media*, ossia il rapporto tra lo spazio percorso in un certo intervallo di tempo e l'intervallo di tempo stesso; questo concetto è sufficiente per descrivere il moto uniforme. Ma in fenomeni come la caduta di un grave, l'oggetto cambia velocità a ogni istante. Che cos'è la velocità istantanea? Intuitivamente, è la velocità media in un intervallo di tempo brevissimo. Ma come si può definire rigorosamente? E come si calcola effettivamente?

Si noti che la parola “brevissimo” non ha alcun significato rigoroso, in matematica! Quello che si può pensare di fare, è calcolare la velocità media in un intervallo di tempo *variabile*  $\Delta t$ , ossia il rapporto tra lo spazio percorso nell’intervallo di tempo  $\Delta t$  e l’intervallo  $\Delta t$  stesso, e cercare di capire a quale valore si avvicina questa quantità quando  $\Delta t$  diventa sempre più piccolo (ma diverso da zero: se  $\Delta t = 0$  il procedimento perde significato, dandoci un rapporto 0/0). L’introduzione consapevole di questa nozione di velocità istantanea si deve a Newton (inizio 1700), ed è uno degli elementi di novità assoluta della fisica newtoniana rispetto a quella galileiana e pregalileiana. Per Newton, come per noi, la velocità è la *derivata rispetto al tempo* della funzione  $s(t)$  = “spazio percorso nel tempo  $t$ ”, ovvero il limite a cui tende il rapporto tra lo spazio  $\Delta s$  percorso in un intervallo di tempo  $\Delta t$  e l’intervallo  $\Delta t$ , quando  $\Delta t$  diventa sempre più piccolo. Procediamo ancora di un passo. Galileo aveva già osservato che il concetto di velocità è sempre *relativo* ad un sistema di riferimento, e che “quieta” o “moto rettilineo uniforme” possono essere due descrizioni dello stesso fenomeno osservato da due sistemi di riferimento diversi (entrambi ugualmente “lecliti”: i cosiddetti *sistemi di riferimento inerziali*). Ne segue che l’effetto di una *forza* (o se vogliamo di una “causa fisica”) non può essere la velocità. Qual è allora tale effetto? Newton è il primo ad affermarlo esplicitamente: tale effetto è l’*accelerazione*, ovvero la *velocità di variazione della velocità*. Questo concetto presuppone ovviamente quello di velocità istantanea, e introduce l’idea che il concetto matematico fisicamente interessante sia la *derivata della derivata della funzione  $s(t)$* , ossia la derivata seconda di  $s(t)$ .

Il *calcolo differenziale* (ovvero lo studio della nozione di derivata), che è parte del calcolo infinitesimale, è quindi inscindibilmente legato alla nascita della scienza moderna.

## 2 DERIVATA DI UNA FUNZIONE

### 2.1 Derivata e retta tangente

Riprendiamo ora da un punto di vista quantitativo la discussione fatta nel paragrafo precedente sul concetto di retta tangente; arriveremo così alla definizione di *derivata*.

Cartelli stradali del tipo



indicano la “pendenza media” del percorso. Nel caso indicato la pendenza media è del 10%. Che cosa vuol dire? Significa che ad ogni avanzamento di 1 km corrisponde un innalzamento (o un abbassamento) di circa 100 m = 0,1 km (fig. 4.4).

Il 10% che indica la pendenza è il rapporto

$$\frac{\text{variazione quota}}{\text{variazione percorso}}$$

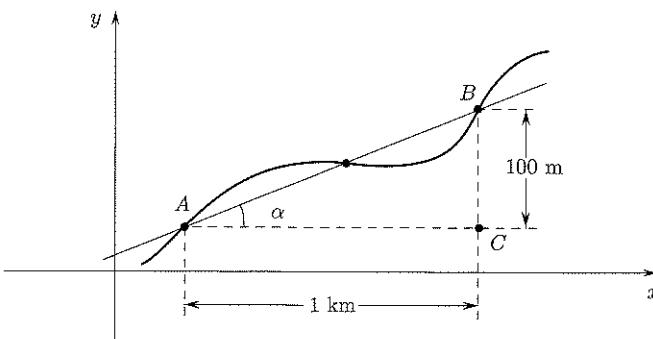


Figura 4.4.

che rappresenta il *tasso di variazione* della quota rispetto al percorso, relativamente al tratto di 1 km. Si noti che tasso di variazione positivo indica (in media!) innalzamento, mentre tasso negativo indica abbassamento.

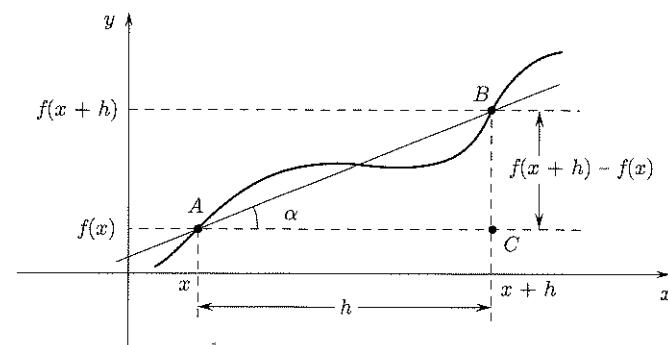


Figura 4.5.

Se immaginiamo il nostro percorso idealmente descritto dal grafico di una funzione  $y = f(x)$  possiamo pensare che le coordinate dei punti  $A$  e  $B$  (in fig. 4.5) siano rispettivamente  $(x, f(x))$  e  $(x + h, f(x + h))$ . Allora avremo

$$\overline{AC} = h \quad \text{e} \quad \overline{CB} = f(x + h) - f(x)$$

e quindi

$$\frac{\text{variazione quota}}{\text{variazione percorso}} = \frac{f(x + h) - f(x)}{h}$$

L’ultimo rapporto prende il nome di *rapporto incrementale di  $f$  relativo all’intervallo  $[x, x + h]$* . Geometricamente, essendo il triangolo  $ABC$  rettangolo, si ha che:

$$\frac{f(x + h) - f(x)}{h} = \tan \alpha = \text{coefficiente angolare della retta } AB$$

D’altra parte, osservando la figura 4.5, si vede che la pendenza media da  $A$  a  $B$  non è un dato preciso in quanto, durante il percorso, nonostante la pendenza media sia positiva, vi sono tratti in discesa, a pendenza perciò negativa.

Per ovviare all'inconveniente e raggiungere un'adeguata precisione occorre considerare tratti di percorso sempre più piccoli, ossia far tendere la lunghezza del percorso a zero. Naturalmente, nel caso della pendenza stradale, un dato medio relativo a lunghezze dell'ordine del km può essere un'informazione sufficiente all'automobilista, ma vedremo più avanti in altri esempi che il passaggio al limite citato risulterà molto significativo. Ritornando al nostro esempio, consideriamo il rapporto:

$$\frac{f(x+h) - f(x)}{h} = \operatorname{tg} \alpha$$

e passiamo al limite (supponendo che esista) per  $h \rightarrow 0$ . Che cosa succede geometricamente? Il punto  $A$ , di coordinate  $(x, f(x))$  rimane fisso, mentre il punto  $B$ , di coordinate  $(x+h, f(x+h))$  si muove verso  $A$ , mantenendosi sul grafico di  $f$  (fig. 4.6).

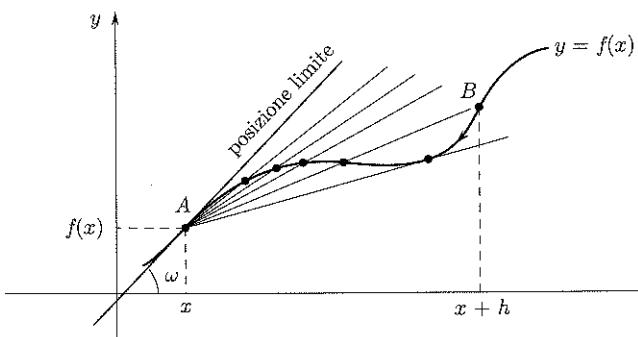


Figura 4.6.

Contemporaneamente la retta  $AB$  varia la sua pendenza assestandosi su una posizione limite. Ne scaturiscono i seguenti fondamentali concetti:

1. la retta limite prende il nome di *retta tangente* al grafico di  $f$  nel punto di ascissa  $x$ ;
2. la sua *pendenza* (o il suo *coefficiente angolare*) è data da  $\operatorname{tg} \omega$  (fig. 4.6) e prende il nome di *derivata prima* di  $f$  nel punto  $x$ .

In generale, abbiamo la seguente definizione.

**DEFINIZIONE 4.1** Sia  $f : (a, b) \rightarrow \mathbb{R}$ ;  $f$  si dice *derivabile* in  $x_0 \in (a, b)$  se esiste finito  $\lim_{h \rightarrow 0} \frac{f(x_0+h) - f(x_0)}{h}$ . Tale limite prende il nome di *derivata prima* (o semplicemente *derivata*) di  $f$  in  $x_0$  e si indica con uno dei simboli seguenti:

$$f'(x_0) \quad \left. \frac{df}{dx} \right|_{x=x_0} \quad Df(x_0) \quad \dot{f}(x_0)$$

$$\boxed{\lim_{h \rightarrow 0} \frac{f(x_0+h) - f(x_0)}{h} = f'(x_0)}$$

La retta di equazione:

$$(2.1) \quad \boxed{y = f(x_0) + f'(x_0)(x - x_0)}$$

si chiama *retta tangente* al grafico di  $f$  nel punto  $(x_0, f(x_0))$ .

Se  $f$  è derivabile in ogni punto di  $(a, b)$  è definita la funzione  $f' : (a, b) \rightarrow \mathbb{R}$ , *derivata di f*, data da

$$x \mapsto f'(x)$$

In questo caso possiamo anche chiederci se la funzione  $f'(x)$  sia a sua volta derivabile (in un punto o in tutto l'intervallo); in caso affermativo, chiameremo *derivata seconda di f* la derivata di  $f'$  e la indicheremo con uno dei simboli

$$f''(x_0) \quad \left. \frac{d^2 f}{dx^2} \right|_{x=x_0} \quad D^2 f(x_0) \quad \ddot{f}(x_0)$$

Vedremo più avanti qual è il significato geometrico e l'utilizzo della derivata seconda. In modo del tutto analogo si definirà la derivata di ordine  $n$ , o *derivata n-esima*, indicata con i simboli:

$$f^{(n)}(x_0) \quad \left. \frac{d^n f}{dx^n} \right|_{x=x_0} \quad D^n f(x_0)$$

## 2.2 Altre interpretazioni della derivata

La definizione di derivata come limite del rapporto incrementale indica che essa deve considerarsi come *tasso di variazione puntuale o istantaneo*, anziché medio.

Questo modo di vedere è particolarmente fruttuoso in molte situazioni concrete come indicato qui di seguito.

- Un oggetto è in moto lungo una traiettoria rettilinea con legge oraria  $s = s(t)$ ,  $t \geq 0$ . La funzione  $t \mapsto s(t)$  indica lo spazio percorso al tempo  $t$ . Allora

$$\frac{ds}{dt} = \text{velocità dell'oggetto}$$

- Analogamente, la velocità istantanea di variazione della velocità  $v(t)$  dell'oggetto, rappresenta (per definizione!) l'accelerazione istantanea dell'oggetto stesso:

$$\frac{dv}{dt} = \text{accelerazione dell'oggetto}$$

Poiché a sua volta  $v$  era la derivata della funzione  $s(t)$ , si ottiene che

$$\text{accelerazione} = \frac{d}{dt} \left( \frac{ds}{dt} \right) = \frac{d^2 s}{dt^2}$$

Ecco un primo esempio di interpretazione della derivata seconda di una funzione. Se sull'oggetto agisce una forza  $f = f(s, s')$ , la seconda legge della dinamica si può scrivere nella forma

$$m s'' = f(s, s')$$

dove  $m$  è la massa dell'oggetto.

- Un pendolo oscilla in un piano verticale attorno a un asse, formando con questo un angolo  $\theta = \theta(t)$ , come in figura. In questo caso

$$\frac{d\theta}{dt} = \text{velocità angolare del pendolo}$$

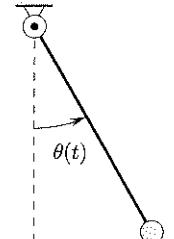


Figura 4.7.

- In un filo percorso da corrente sia  $q = q(t)$  la quantità di carica che attraversa una sezione del filo nell'istante  $t$ . Si ha:

$$\frac{dq}{dt} = \text{intensità di corrente}$$

- Se  $c = c(p)$  indica il costo di produzione di una quantità  $p$  (di un determinato bene), il rapporto tra  $c(p + \Delta p) - c(p)$  e  $\Delta p$  indica la variazione del costo proveniente da una variazione (media) unitaria della produzione e prende il nome di *costo marginale medio*. Passando al limite per  $\Delta p \rightarrow 0$  si ottiene il costo marginale; cioè, in questo caso

$$\frac{dc}{dp} = \text{costo marginale di produzione}$$

Invitiamo il lettore ad allungare la lista delle possibili interpretazioni della derivata con altri esempi (accelerazione angolare, densità di massa, ...).

### 2.3 Derivate di funzioni elementari

La tabella seguente contiene le derivate delle principali funzioni elementari.

$f$	$f'$	$f$	$f'$
1. $c$ (costante)	0	11. $e^x$	$e^x$
2. $x$	1	12. $\log x$ ( $x > 0$ )	$\frac{1}{x}$
3. $x^2$	$2x$	13. $a^x$ ( $a > 0$ )	$a^x \log a$
4. $\frac{1}{x}$	$-\frac{1}{x^2}$	14. $\log_a x$ ( $a > 0, a \neq 1$ )	$\frac{1}{x \log a}$
5. $\sqrt{x}$ ( $x \geq 0$ )	$\frac{1}{2\sqrt{x}}$ ( $x > 0$ )	15. $\operatorname{Sh}x$	$\operatorname{Ch}x$
6. $x^\alpha$ ( $\alpha \in \mathbb{R}, x > 0$ )	$\alpha x^{\alpha-1}$	16. $\operatorname{Ch}x$	$\operatorname{Sh}x$
7. $\sin x$	$\cos x$	17. $\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$
8. $\cos x$	$-\sin x$	18. $\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$
9. $\operatorname{tg} x$	$1 + \operatorname{tg}^2 x = \frac{1}{(\cos x)^2}$	19. $\operatorname{arctg} x$	$\frac{1}{1+x^2}$
10. $\operatorname{cotg} x$	$-(1 + \operatorname{cotg}^2 x) = -\frac{1}{\sin^2 x}$		

La 1 è immediata; 2, 3, 4, 5 sono casi particolari di 6, che abbiamo scritto esplicitamente perché si incontrano così frequentemente che meritano di essere ricordate. Proviamo ad esempio la:

$$3. \quad f(x) = x^2, \quad f'(x) = 2x$$

Si ha:

$$\frac{f(x+h) - f(x)}{h} = \frac{(x+h)^2 - x^2}{h} = \frac{x^2 + 2hx + h^2 - x^2}{h} = 2x + h$$

che tende a  $2x$  per  $h \rightarrow 0$ ; da qui, la formula.

6.

$$f(x) = x^\alpha, \quad f'(x) = \alpha x^{\alpha-1}$$

Sia  $x > 0$ . Scriviamo:

$$\frac{(x+h)^\alpha - x^\alpha}{h} = x^\alpha \frac{(1 + \frac{h}{x})^\alpha - 1}{h} \sim x^\alpha \frac{\alpha \frac{h}{x}}{h} = \alpha x^{\alpha-1},$$

dove abbiamo usato il limite notevole  $(1 + \varepsilon(h))^\alpha - 1 \sim \alpha \varepsilon(h)$  per  $h \rightarrow 0$ , con  $\varepsilon(h) = h/x$  (cfr. la (3.8) del cap. 3).

7, 8.

$$f(x) = \sin x, \quad f'(x) = \cos x$$

$$f(x) = \cos x, \quad f'(x) = -\sin x$$

Si ha, utilizzando le formule di addizione

$$\begin{aligned} \frac{\sin(x+h) - \sin x}{h} &= \frac{\sin x \cos h + \sin h \cos x - \sin x}{h} = \\ &= \sin x \frac{\cos h - 1}{h} + \frac{\sin h}{h} \cos x \xrightarrow[h \rightarrow 0]{} \cos x \end{aligned}$$

dove abbiamo usato i limiti notevoli  $\frac{\sin h}{h} \rightarrow 1$  e

$$\frac{\cos h - 1}{h} \sim -\frac{1}{2} \frac{h^2}{h} = -\frac{1}{2} h \rightarrow 0$$

Analogamente si mostra la seconda formula. Omettiamo i dettagli.

11.

$$f(x) = e^x, \quad f'(x) = e^x$$

Si ha:

$$\frac{f(x+h) - f(x)}{h} = \frac{e^{x+h} - e^x}{h} = e^x \frac{e^h - 1}{h} \xrightarrow[h \rightarrow 0]{} e^x$$

essendo  $\frac{e^h - 1}{h} \rightarrow 1$  per  $h \rightarrow 0$  (limite notevole).

Dalla formula si ricava che la funzione  $x \mapsto e^x$  soddisfa l'equazione  $y' = y$ .

12.

$$f(x) = \log x, \quad f'(x) = \frac{1}{x}$$

Abbiamo:

$$\frac{f(x+h) - f(x)}{h} = \frac{\log(x+h) - \log x}{h} = \frac{\log(1 + h/x)}{h} \sim \frac{h}{x} \cdot \frac{1}{h} = \frac{1}{x}$$

dove abbiamo usato la (3.8) del capitolo 3.

Le formule 9, 10 e 13-19 potranno essere dimostrate utilizzando le regole che vedremo nel prossimo paragrafo.

**Le equazioni differenziali soddisfatte dalle funzioni esponenziali e trigonometriche**

Osserviamo un fatto notevole che riguarda le funzioni esponenziali e trigonometriche. La 13 può essere riletta dicendo che le funzioni  $f(x) = a^x$  soddisfano l'*equazione differenziale*

$$f'(x) = kf(x)$$

con  $k$  costante opportuna. La legge esponenziale governa quindi i fenomeni in cui la *velocità di crescita (o diminuzione) di una grandezza è proporzionale alla grandezza stessa*. Questa semplice legge si ritrova in molte leggi fisiche e questo è certamente uno dei motivi a cui le funzioni esponenziali devono la loro importanza.

La 7 e la 8 ci dicono invece che, posto  $f(x) = \sin x$ ,

$$\frac{d^2f}{dx^2}(x) = \frac{d}{dx} \left( \frac{df}{dx} \right)(x) = \frac{d}{dx}(\cos x) = -\sin x = -f(x)$$

Perciò la funzione  $\sin x$  (e, come si vede con passaggi analoghi, anche la funzione  $\cos x$ ) soddisfa l'*equazione differenziale*

$$f''(x) = -f(x)$$

Pertanto le funzioni sinusoidali governano i fenomeni in cui l'*accelerazione con cui varia una grandezza è uguale alla grandezza stessa, cambiata di segno*. Tali sono ad esempio molti fenomeni di tipo vibratorio.

Non è esagerato dire che *questo* è il motivo fondamentale per cui le funzioni trigonometriche sono così importanti nel calcolo infinitesimale e nella matematica applicata: un motivo che apparentemente non ha nulla a che vedere con la trigonometria, da cui queste funzioni sono nate! Approfondiremo questi fatti nel vol.2, trattando le equazioni differenziali.



**ESEMPIO** Calcoliamo l'*equazione della retta tangente* al grafico delle seguenti funzioni

$$f_1(x) = e^x, \quad f_2(x) = x^3$$

nei punti di ascissa  $x = 2$ .

Abbiamo  $f_1(2) = e^2$  e

$$f'_1(x) = e^x \quad f'_1(2) = e^2$$

e quindi l'*equazione della retta* è

$$y = f_1(2) + f'_1(2)(x - 2) = e^2 + e^2(x - 2)$$

Per l'altra funzione:

$$f_2(2) = 8, \quad f'_2(x) = 3x^2 \quad (\text{formula 6 della tabella con } \alpha = 3)$$

e quindi  $f'_2(2) = 12$ ; pertanto si ottiene per la retta tangente, l'*equazione*

$$y = f_2(2) + f'_2(2)(x - 2) = 8 + 12(x - 2)$$

Riportiamo in figura 4.8 i grafici delle due funzioni, con le rispettive rette tangenti per  $x = 2$ :

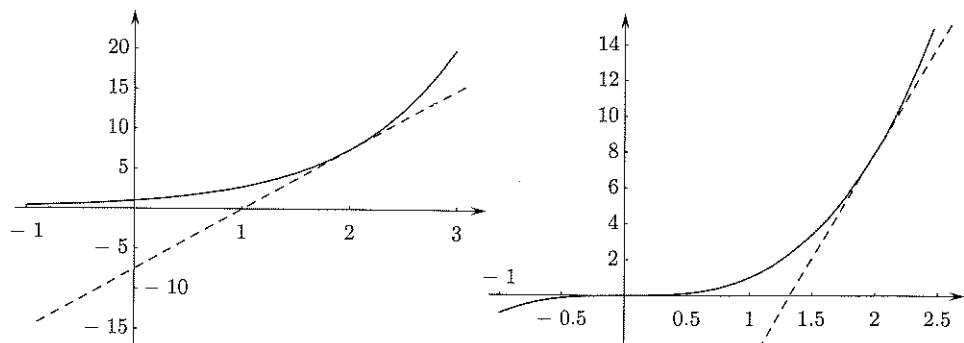


Figura 4.8.

**2.4 Punti angolosi, cuspidi, flessi a tangente verticale**

Se una funzione  $f$  è derivabile in un punto  $x_0$ , nel punto di coordinate  $(x_0, f(x_0))$  il grafico ha una retta tangente ben definita. Che cosa succede quando  $f$  non è derivabile in un punto? Vediamo alcuni esempi.

**Punti angolosi**

Sia  $f(x) = |x|$ . Essendo  $f(x) = x$  per  $x > 0$  e  $f(x) = -x$  per  $x < 0$ , si ha  $f'(x) = +1$  se  $x > 0$  e  $f'(x) = -1$  per  $x < 0$ , avendo  $f'$  il significato di coefficiente angolare. Nell'origine  $x = 0$ , occorre usare la definizione. Ora

$$\frac{f(h) - f(0)}{h} = \frac{|h|}{h}$$

e quindi, se  $h \rightarrow 0^+$ ,  $|h| = h$  e il limite del rapporto incrementale è 1, mentre se  $h \rightarrow 0^-$ ,  $|h| = -h$  e il limite è -1.

Si conclude che, non esistendo il limite del rapporto incrementale,  $f$  non è derivabile in  $x = 0$ . D'altra parte, ricordando il grafico di  $f(x) = |x|$  si vede che la tangente nell'origine non è ben definita.

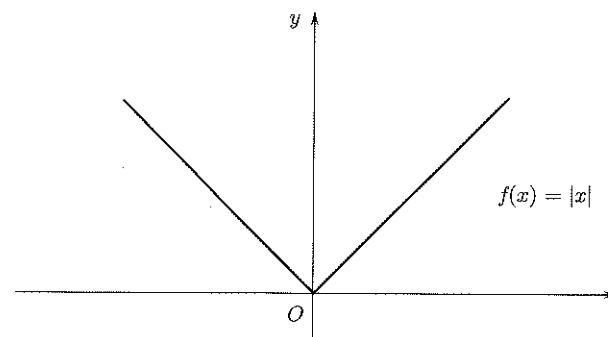


Figura 4.9. La funzione  $|x|$  non è derivabile in  $x = 0$ .

Tuttavia i limiti destro e sinistro del rapporto incrementale di  $|x|$  esistono finiti e in  $(0, 0)$  il grafico presenta "un angolo".

La circostanza merita una definizione.

**DEFINIZIONE 4.2** Siano  $f : (a, b) \rightarrow \mathbb{R}$ ,  $x_0 \in (a, b)$ . Se esiste finito  $\lim_{h \rightarrow 0^+} \frac{f(x_0+h)-f(x_0)}{h}$  (oppure  $\lim_{h \rightarrow 0^-} \frac{f(x_0+h)-f(x_0)}{h}$ ) allora  $f$  si dice *derivabile dalla destra* (oppure *dalla sinistra*); il limite si chiama *derivata destra* (oppure *sinistra*) e si indica con il simbolo  $f'_+(x_0)$  (oppure  $f'_-(x_0)$ ).

Nel caso in cui  $f$  sia continua e derivabile *da destra e da sinistra* (ma non derivabile) in  $x_0$  si dice che  $f$  ha un *punto angoloso* in  $x = x_0$ . Dunque,  $|x|$  ha un punto angoloso in  $x = 0$ .

Vale la pena ricordare la formula che esprime sinteticamente la derivata della funzione valore assoluto (fuori dall'origine):

$$|x'| = \operatorname{sgn}(x) = \begin{cases} 1 & \text{per } x > 0 \\ -1 & \text{per } x < 0 \end{cases}$$

#### Punti a tangente verticale. Cuspidi

Se  $f$  è continua in un punto  $x_0$  e

$$\lim_{h \rightarrow 0} \frac{f(x_0+h)-f(x_0)}{h} = +\infty \quad \text{oppure} \quad -\infty$$

$f$  non è derivabile in  $x_0$  ma, geometricamente, il grafico di  $f$  ha una retta tangente ben definita e parallela all'asse delle ordinate. Ammetteremo in tal caso la scrittura  $f'(x_0) = +\infty$ ,  $f'(x_0) = -\infty$  e parleremo di *flesso a tangente verticale* (fig. 4.10). Il concetto di flesso sarà definito più in generale in seguito (par. 5.2).

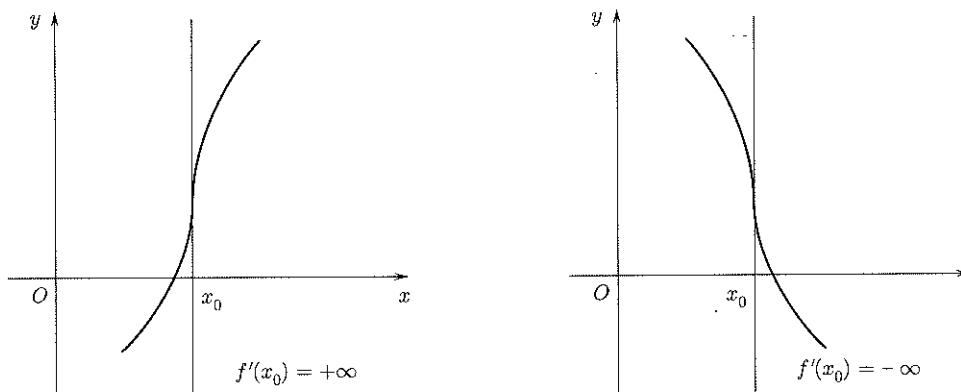


Figura 4.10.

Ad esempio, la funzione  $f(x) = \sqrt[3]{x}$  ha un punto a tangente verticale in  $x = 0$ .

Consideriamo ora la funzione  $f(x) = \sqrt[3]{|x|}$ , il cui grafico è riportato in figura 4.11. In questo caso si ha  $f'_+(0) = +\infty$ ,  $f'_-(0) = -\infty$  e si dice che in  $x = 0$ ,  $f$  ha una *cuspide*.

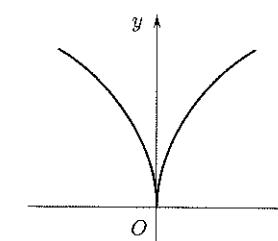


Figura 4.11.  $f(x) = \sqrt[3]{|x|}$ .

**DEFINIZIONE 4.3** Se  $f$  è continua in  $x_0$  e  $f'_+(x_0) = \pm\infty$ ,  $f'_-(x_0) = \mp\infty$  si dice che  $f$  ha in  $x_0$  una *cuspide*.

Nel caso misto in cui una delle due derivate è finita e l'altra infinita (con  $f$  continua) si parla ancora di punto angoloso.

Infine, se la funzione è definita solo per  $x \geq x_0$  e in tal punto ha derivata (destra) infinita, diremo semplicemente che in tal punto ha *tangente verticale*, senza parlare né di cuspide né di flesso. Ad esempio, la funzione  $\sqrt{x}$  ha un punto a tangente verticale in  $x = 0$ .

#### Continuità e derivabilità

Vale il seguente semplice ma importante

**TEOREMA 4.1** Se  $f$  è derivabile in un punto  $x_0$  allora  $f$  è continua in  $x_0$ .

DIMOSTRAZIONE. Scriviamo

$$f(x_0+h) - f(x_0) = \frac{f(x_0+h) - f(x_0)}{h} \cdot h \sim f'(x_0)h \quad \text{per } h \rightarrow 0.$$

Perciò  $\lim_{h \rightarrow 0} [f(x_0+h) - f(x_0)] = 0$  da cui  $\lim_{h \rightarrow 0} f(x_0+h) = f(x_0)$ , che è la continuità di  $f$  in  $x_0$ .

Come conseguenza, se una funzione è *discontinua* in  $x_0$ , non può essere derivabile in  $x_0$ .

Viceversa, se  $f$  è continua in  $x_0$ , non necessariamente  $f$  è derivabile in  $x_0$  come mostra  $f(x) = |x|$  che è continua in  $x = 0$  ma non ivi derivabile.

Consideriamo la funzione  $f$  periodica di periodo  $T$  che coincide con  $\frac{2A}{T}|x|$  nell'intervallo  $[-T/2, T/2]$  (fig. 4.12).

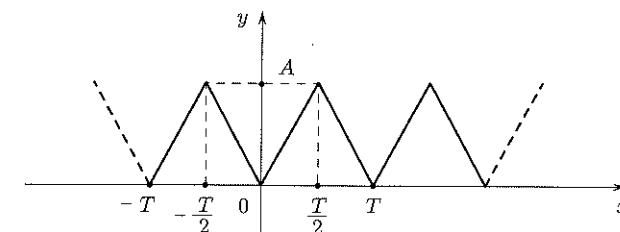


Figura 4.12.

Una funzione di questo genere modellizza un'onda o segnale triangolare di ampiezza  $A$ . Questa funzione presenta punti angolosi in  $x = \pm \frac{T}{2}, \pm T, \pm \frac{3}{2}T$  ecc.

La sua derivata non è definita in questi punti ed è costante a tratti; precisamente, vale  $\frac{2A}{T}$  per  $0 < x < \frac{T}{2}$ , vale  $-\frac{2A}{T}$  per  $\frac{T}{2} < x < T$  e si ripete con periodicità  $T$ , con un grafico illustrato in figura 4.13.

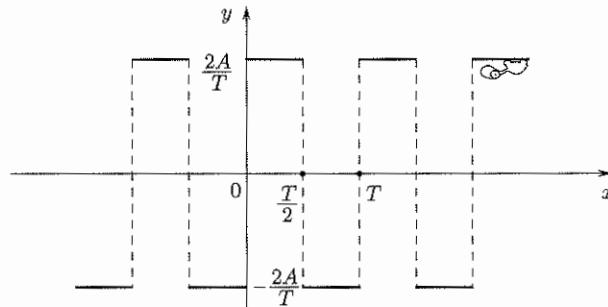


Figura 4.13.

La derivata del segnale triangolare di ampiezza  $\frac{2A}{T}$  rappresenta quella che si chiama onda quadrata.

### Esercizi

Utilizzando la definizione di derivata, determinare il comportamento nell'origine delle seguenti funzioni (tangente orizzontale, cuspide, flesso a tangente verticale...):

$$x^{1/3} \quad x^{4/3} \quad x^{2/3} \quad x^{5/3} \quad x^{1/2} \quad x^{3/2}$$

(Questo argomento permette di completare la giustificazione del grafico delle funzioni potenza a esponente razionale o reale, che abbiamo descritto nel cap. 2, par. 3.1).

2 Sia

$$f(x) = x \log x, \text{ per } x > 0$$

Dopo aver prolungato per continuità  $f$  anche in  $x = 0$ , calcolare la sua derivata destra in 0.

2 Sia

$$f(x) = e^{-1/x}, \text{ per } x > 0$$

Dopo aver prolungato per continuità (da destra)  $f$  anche in  $x = 0$ , calcolare la sua derivata destra in 0.

## 3 REGOLE DI CALCOLO DELLE DERIVATE

Vediamo ora la relazione tra l'operazione di derivata e le principali operazioni già note sulle funzioni; in particolare mostreremo la relazione tra:

$$\begin{aligned} \text{derivazione e} & \left\{ \begin{array}{l} \text{operazioni algebriche } (\pm, \cdot, /) \\ \text{composizione} \\ \text{inversione} \end{array} \right. \end{aligned}$$

### 3.1 Algebra delle derivate

**TEOREMA 4.2** Siano  $f, g : (a, b) \rightarrow \mathbb{R}$ , derivabili in  $(a, b)$ ; allora  $f \pm g$ ,  $f \cdot g$ ,  $f/g$  ( $g \neq 0$ ) sono derivabili in  $(a, b)$  e valgono le seguenti formule

$$(3.1) \quad (f \pm g)' = f' \pm g'$$

$$(3.2) \quad (f \cdot g)' = f' \cdot g + f \cdot g'$$

$$(3.3) \quad (f/g)' = (f' \cdot g - f \cdot g')/g^2$$

In particolare, dalla (3.2) si deduce

$$(3.4) \quad (k \cdot f)' = k \cdot f' \quad k \text{ costante}$$

essendo la derivata di una costante uguale a zero, e dalla (3.3) si deduce, per  $f = 1$ ,

$$(3.5) \quad \left( \frac{1}{g} \right)' = -\frac{g'}{g^2}$$

La (3.2) si dice *regola di Leibnitz* e si estende al prodotto di  $n$  fattori:

$$(f_1 f_2 \cdots f_n)' = f'_1 f_2 \cdots f_n + f_1 f'_2 \cdots f_n + \cdots + f_1 f_2 \cdots f'_n.$$

Come si vedrà dalla dimostrazione, il teorema ha in realtà un carattere *puntuale*, ossia: se  $f$  e  $g$  sono derivabili in un punto  $x_0 \in (a, b)$ , allora in quel punto sono derivabili anche  $f + g$ ,  $f \cdot g$ , ..., e valgono le formule scritte. Solitamente comunque il teorema si applica nella forma in cui l'abbiamo enunciato (cioè in situazioni in cui  $f$  e  $g$  sono derivabili in tutti i punti di un intervallo).

**DIMOSTRAZIONE.** Lasciamo per esercizio la dimostrazione di (3.1) e (3.4), molto semplici. Dimostriamo la (3.2): si ha, fissato  $x \in (a, b)$

$$f(x+h)g(x+h) - f(x)g(x) = f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)$$

e quindi

$$\begin{aligned} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} &= f(x+h) \cdot \frac{g(x+h) - g(x)}{h} + g(x) \frac{f(x+h) - f(x)}{h} \longrightarrow \\ &\longrightarrow f(x)g'(x) + f'(x)g(x) \end{aligned}$$

poiché  $f(x+h) \rightarrow f(x)$  quando  $h \rightarrow 0$ , essendo  $f$  continua in quanto derivabile.

Proviamo ora la (3.5):

$$\frac{1}{h} \left[ \frac{1}{g(x+h)} - \frac{1}{g(x)} \right] = \frac{g(x) - g(x+h)}{hg(x)g(x+h)} = -\frac{g(x+h) - g(x)}{h} \cdot \frac{1}{g(x)g(x+h)} \rightarrow -\frac{g'(x)}{g(x)^2}$$

dove si è sfruttato ancora il fatto che, essendo derivabile,  $g$  è continua, perciò  $g(x+h) \rightarrow g(x)$  per  $h \rightarrow 0$ .

Notiamo infine che dalle (3.2) e (3.5) si deduce anche la (3.3), infatti:

$$\begin{aligned}\left(\frac{f}{g}\right)' &= \left(f \cdot \frac{1}{g}\right)' = \text{per la (3.2) applicata a } f \text{ e } \frac{1}{g} \\ &= f' \cdot \frac{1}{g} + f \cdot \left(\frac{1}{g}\right)' = \text{per la (3.5)} \\ &= f' \cdot \frac{1}{g} + f \cdot \left(-\frac{g'}{g^2}\right) = \frac{f'g - fg'}{g^2}, \text{ cioè la (3.3).}\end{aligned}$$



**Calcoliamo la velocità** di un oggetto in moto rettilineo con legge oraria data da

$$s(t) = v_0 t + \frac{1}{2} g t^2$$

La derivata della funzione  $v_0 t$  è  $v_0$  mentre quella di  $\frac{1}{2} g t^2$  è  $\frac{1}{2} g \cdot 2t = gt$  (usando 2 volte la (3.4)). Usando ora la (3.1) si trova

$$v(t) = s'(t) = v_0 + gt$$

**Calcoliamo la velocità di variazione del volume di una sfera** rispetto al raggio.

Essendo  $V(r) = \frac{4}{3}\pi r^3$ , la velocità richiesta è (regola (3.4))

$$V'(r) = \frac{4}{3}\pi(3r^2) = 4\pi r^2 \quad (= \text{superficie della sfera di raggio } r!)$$

**Derivata della funzione tangente.** Poiché  $\operatorname{tg} x = \frac{\sin x}{\cos x}$ , usando la formula (3.3) si trova

$$\frac{d}{dx} \operatorname{tg} x = \frac{(\cos x)^2 + (\sin x)^2}{(\cos x)^2} = \frac{1}{(\cos x)^2} = 1 + (\operatorname{tg} x)^2$$

Analogamente si calcola la derivata della funzione cotangente.

**Se  $c = c(p)$**  è il costo di produzione di una quantità  $p$  di prodotto, il rapporto

$$\frac{c(p)}{p}$$

si chiama *costo medio* e rappresenta il *costo di una unità di prodotto*.

Per calcolare la derivata del costo medio, usiamo la formula (3.3); si trova

$$\frac{d}{dp} \frac{c(p)}{p} = \frac{c'(p) \cdot p - c(p)}{p^2}$$

### 3.2 Derivata di una funzione composta

**TEOREMA 4.3 (REGOLA DELLA CATENA)** Sia  $g \circ f$  la composta di due funzioni  $f$  e  $g$ . Se  $f$  è derivabile in un punto  $x$  e  $g$  è derivabile in  $y = f(x)$  allora  $g \circ f$  è derivabile in  $x$  e vale la formula:

(3.6)

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$$

**DIMOSTRAZIONE.** Si ha  $(g \circ f)(x+h) - (g \circ f)(x) = g(f(x+h)) - g(f(x))$ .

Se poniamo  $k = f(x+h) - f(x)$ ,  $y = f(x)$ , allora  $f(x+h) = y+k$ , e per la continuità di  $f$ ,  $h \rightarrow 0$  implica  $k \rightarrow 0$ . Con le nuove notazioni,

$$g(f(x+h)) - g(f(x)) = g(y+k) - g(y)$$

Osserviamo ora che la definizione di derivata

$$g'(y) = \lim_{k \rightarrow 0} \frac{g(y+k) - g(y)}{k}$$

si può riscrivere, per  $k \neq 0$ ,

$$\frac{g(y+k) - g(y)}{k} = g'(y) + \varepsilon(k)$$

dove  $\varepsilon(k)$  indica una quantità che tende a zero per  $k \rightarrow 0$ . Moltiplicando ambo i membri dell'equazione precedente per  $k$  si trova

$$g(y+k) - g(y) = g'(y) \cdot k + \varepsilon(k) \cdot k$$

relazione valida anche per  $k = 0$ . Dunque:

$$g(f(x+h)) - g(f(x)) = g'(y) \cdot k + \varepsilon(k) \cdot k$$

Dividendo per  $h$ , e osservando che  $k/h \rightarrow f'(x)$  si ottiene la (3.6).

La (3.6) si chiama *regola della catena*; usando le notazioni (di Leibniz)  $\frac{df}{dx}$  e  $\frac{dg}{dx}$  per le derivate di  $f$  e  $g$  e posto  $w = g(y)$ , la (3.6) acquista una forma più significativa:

$$(3.7) \quad \frac{dw}{dx} = \frac{dw}{dy} \cdot \frac{dy}{dx} \quad (\text{come se } dy \text{ si semplificasse})$$

La (3.7) esprime il fatto che il tasso di variazione di  $w$  rispetto a  $x$  è il prodotto dei tassi di variazione “intermedi”, di  $w$  rispetto a  $y$  e di  $y$  rispetto a  $x$ .

Come suggerisce il nome di “regola della catena”, la (3.7) può essere generalizzata alla composizione di un numero qualsiasi di funzioni, composte una con l'altra. Ad esempio per tre funzioni si ha

$$[f(g(h(x)))]' = f'(g(h(x))) \cdot g'(h(x)) \cdot h'(x)$$

Per usare questa regola, insieme alle altre dell'algebra delle derivate, occorre imparare a vedere una funzione complicata come composizione successiva di funzioni più semplici. Per individuare le componenti può essere utile immaginare come si calcola la funzione composta mediante una calcolatrice tascabile:



**Si voglia derivare**

$$w(x) = (\sin x)^3$$

Per calcolarla, occorre inserire il valore di  $x$ , calcolare  $\sin x$  e poi elevare tutto al cubo:

$$x^{\sin} \mapsto \sin x \stackrel{(\cdot)^3}{\mapsto} (\sin x)^3$$

Posto

$$f(x) = \sin x, g(y) = y^3$$

si ha allora  $w(x) = g(f(x))$ . Pertanto:

$$w'(x) = 3(\sin x)^2 \cdot \cos x$$

Usando la (3.7) si scriverebbe  $y = \sin x$ ,  $w = y^3$  e quindi

$$\frac{dw}{dx} = \frac{dw}{dy} \cdot \frac{dy}{dx} = 3y^2 \cdot \cos x = 3(\sin x)^2 \cdot \cos x$$

Derivata di  $f(t) = A \sin(\omega t + \varphi)$  e di  $g(t) = Ae^{-\alpha t} \cos(\omega t + \varphi)$

$$f'(t) = A \frac{d}{dt} \sin(\omega t + \varphi) = A \cos(\omega t + \varphi) \cdot \omega = A\omega \cos(\omega t + \varphi)$$

formula (3.4)                                  regola della catena

$$\begin{aligned} g'(t) &= A \frac{d}{dt} [e^{-\alpha t} \cos(\omega t + \varphi)] = A \left\{ \frac{d}{dt} e^{-\alpha t} \cdot \cos(\omega t + \varphi) + e^{-\alpha t} \frac{d}{dt} \cos(\omega t + \varphi) \right\} = \\ &\quad \text{formula (3.4)} \qquad \qquad \qquad \text{formula (3.2)} \\ &= A \left\{ e^{-\alpha t} (-\alpha) \cdot \cos(\omega t + \varphi) + e^{-\alpha t} (-\sin(\omega t + \varphi) \cdot \omega) \right\} = \\ &\quad \text{(catena)} \\ &= -Ae^{-\alpha t} \{-\alpha \cos(\omega t + \varphi) + \sin(\omega t + \varphi)\} \end{aligned}$$

Possiamo ora completare la dimostrazione delle formule contenute nella tabella delle derivate di funzioni elementari.

$$(a^x)' = (e^{x \log a})' =$$

(usando le formule per la derivata di  $e^t$  e per la derivata della funzione composta)

$$= e^{x \log a} \cdot \log a = a^x \log a$$

$$(\log_a x)' = \left( \frac{\log x}{\log a} \right)' =$$

(usando le formule per la derivata di  $\log x$  e per la derivata di  $kf(x)$ )

$$= \frac{1}{x \log a}$$

$$(\operatorname{Sh} x)' = \left( \frac{e^x - e^{-x}}{2} \right)' = \frac{1}{2}(e^x - (-e^{-x})) = \frac{e^x + e^{-x}}{2} = \operatorname{Ch} x$$

$$(\operatorname{Ch} x)' = \left( \frac{e^x + e^{-x}}{2} \right)' = \frac{1}{2}(e^x + (-e^{-x})) = \frac{e^x - e^{-x}}{2} = \operatorname{Sh} x$$

(Derivate di funzioni del tipo  $f(x)^{g(x)}$ ). I passaggi per il calcolo di  $(a^x)'$  si basano su un "trucco" di uso comune: riscrivere una funzione  $f(x)^{g(x)}$  con  $f(x) > 0$  nella forma seguente:

$$f(x)^{g(x)} = e^{g(x) \log f(x)}$$

Si è sfruttata la definizione di logaritmo  $e^{\log A} = A$  e la proprietà dei logaritmi  $\log[f(x)^{g(x)}] = g(x) \log f(x)$ . A questo modo si può calcolare la derivata di una funzione di questo tipo:

$$\begin{aligned} [f(x)^{g(x)}]' &= [e^{g(x) \log f(x)}]' = e^{g(x) \log f(x)} \cdot [g(x) \log f(x)]' = \\ &= f(x)^{g(x)} \left[ g'(x) \log f(x) + g(x) \frac{f'(x)}{f(x)} \right] \end{aligned}$$

Ad esempio:

$$(x^x)' = (e^{x \log x})' = x^x [x \log x]' = x^x [\log x + 1]$$

(Valore assoluto di una funzione). Consideriamo una funzione del tipo:

$$|f(x)|$$

Sappiamo che il valore assoluto non è derivabile là dove il suo argomento si annulla. Tuttavia, nei punti in cui  $f(x) \neq 0$ , la derivazione di funzione composta dà:

$$|f(x)|' = \operatorname{sgn}(f(x)) \cdot f'(x) = \begin{cases} f'(x) & \text{per } f(x) > 0 \\ -f'(x) & \text{per } f(x) < 0 \end{cases}$$

In generale, ci aspettiamo che la funzione  $|f(x)|$  presenti punti angolosi nei punti in cui  $f(x)$  si annulla. Ad esempio, la funzione

$$e^{|x+1|}$$

ha un punto angoloso in  $x = -1$ .

(Derivata di alcune funzioni logaritmiche). Le seguenti situazioni si presentano di frequente:

$$(\log|x|)' = \frac{1}{|x|} \cdot \operatorname{sgn}(x) = \frac{1}{x} \quad (\text{senza modulo!})$$

$$(\log|f(x)|)' = \frac{f'(x)}{f(x)}$$

(per la formula precedente e la derivazione di funzioni composte);

$$(\log(cx))' = (\log c + \log x)' = \frac{1}{x} \quad (\text{come la derivata di } \log x!);$$

$$\left( \log \left( \frac{ax+b}{cx+d} \right) \right)' = (\log(ax+b) - \log(cx+d))' = \frac{a}{ax+b} - \frac{c}{cx+d}$$

(ossia, talvolta conviene usare le proprietà dei logaritmi per trasformare una funzione logaritmica prima di calcolarne la derivata).

Il seguente esempio mette in luce le possibilità di calcolo connesse con la formula della catena.



Un contenitore cilindrico con raggio di base  $R = 1$  m e altezza 3 m è pieno d'acqua. Da un rubinetto posto in prossimità del fondo vengono prelevati 10 litri al minuto, (fig. 4.14).

Con quale velocità l'altezza dell'acqua decresce?

Sia  $h$  l'altezza (in decimetri) della colonna d'acqua e  $V$  il suo volume (in  $\text{dm}^3$ ).

Vogliamo trovare  $dh/dt$ , sapendo che  $dV/dt = -10 \text{ dm}^3/\text{min}$ .

$$\frac{dV}{dt} = \frac{dV}{dh} \cdot \frac{dh}{dt}$$

essendo  $V = 100\pi h \text{ dm}^3$  otteniamo  $dV/dh = 100\pi \text{ dm}^2$  e infine

$$\frac{dh}{dt} = -\frac{10}{100\pi} \text{ dm/min} = -\frac{1}{10\pi} \text{ dm/min}$$

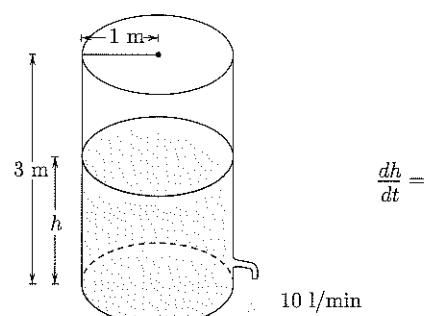


Figura 4.14.

### Derivata logaritmica. Elasticità

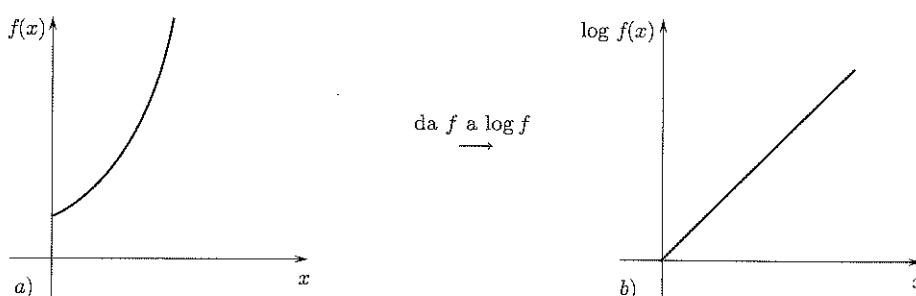
Data  $f > 0$ , si chiama *derivata logaritmica* di  $f$  la derivata di  $\log f$ . In formule:

$$(3.8) \quad \text{derivata logaritmica di } f = \frac{d}{dx} \log f(x) = \frac{f'(x)}{f(x)}$$

La derivata logaritmica ha il significato di *tasso di incremento relativo* di  $f$  rispetto a  $x$ .

In casi concreti tale tasso è spesso più significativo del tasso assoluto  $f'$ . Ad esempio, un incremento annuo di capitale di 1 miliardo su un totale di 10 ha un effetto ben diverso di un incremento di 1 miliardo su 1000 miliardi. Nel primo caso il tasso di variazione relativo è  $1/10$  nel secondo  $1/1000$ , mentre uguali sono gli incrementi assoluti.

Per i motivi sopraesposti, quando interessa visualizzare gli incrementi relativi si ricorre a grafici in scala *semilogaritmica*: sull'asse delle ascisse si collocano i valori di  $x$ , mentre su quello delle ordinate quelli di  $\log f(x)$ . A questo tipo di rappresentazione si ricorre anche quando  $f(x)$  cresce così rapidamente da richiedere una compressione troppo elevata della scala (unità di misura) sull'asse delle ordinate o in quella delle ascisse. Basti pensare al grafico di  $f(x) = e^x$  che in scala semilogaritmica coincide con... la bisettrice del 1° quadrante:

Figura 4.15. a) Grafico di  $f(x) = e^x$  in scala normale; b) grafico di  $e^x$  in una scala semilogaritmica.

Altra variante possibile è quella dei grafici in scala logaritmica, in cui invece di  $x$  sull'asse delle ascisse si collocano i valori di  $\log x$ ; lo stesso si fa con  $f$  sull'asse delle ordinate.

La pendenza della retta tangente a un grafico in scala logaritmica (cioè la derivata di  $\log f$  rispetto a  $\log x$ ) rappresenta il tasso di variazione relativa di  $f$  rispetto a variazioni relative di  $x$ ; questa quantità prende il nome di *elasticità* di  $f$  e si indica con  $E(x)$ .

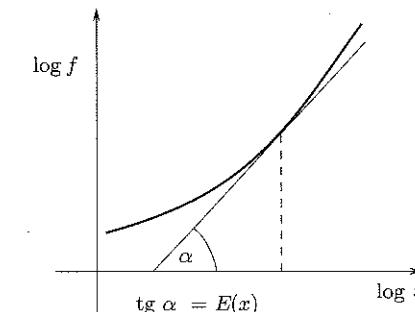


Figura 4.16.

Per trovare l'espressione analitica di  $E(x)$ , osserviamo che, per il teorema di derivazione delle funzioni composte, si ha, posto  $u = \log x$ :

$$\frac{d \log f}{dx} = \frac{d \log f}{du} \cdot \frac{du}{dx}$$

ovvero

$$\frac{f'(x)}{f(x)} = E(x) \frac{1}{x}$$

da cui si ricava

$$E(x) = x \frac{f'(x)}{f(x)}$$



Se  $f(x) = x^\alpha$ ,  $x > 0$ ,

$$E(x) = x \frac{\alpha x^{\alpha-1}}{x^\alpha} = \alpha$$

ossia: le potenze hanno elasticità costante. In realtà, si può mostrare che sono le uniche funzioni ad avere questa proprietà.

### 3.3 Derivata di funzione inversa

**TEOREMA 4.4**  $f : (a, b) \rightarrow \mathbb{R}$  continua e invertibile in  $(a, b)$  e  $g = f^{-1}$  la sua inversa, definita in  $f((a, b))$ . Supponiamo inoltre che esista  $f'(x_0) \neq 0$  per un certo  $x_0 \in (a, b)$ . Allora  $g$  è derivabile in  $y_0 = f(x_0)$  e

$$(3.9) \quad g'(y_0) = \frac{1}{f'(x_0)}.$$

DIMOSTRAZIONE. Sia:

$$\begin{aligned} f(x_0) &= y_0; & g(y_0) &= x_0; \\ f(x_0 + h) &= y_0 + k; & g(y_0 + k) &= x_0 + h \end{aligned}$$

Consideriamo il rapporto incrementale di  $g$  in  $y_0$ :

$$\frac{g(y_0 + k) - g(y_0)}{k} = \frac{h}{f(x_0 + h) - f(x_0)}.$$

Se  $k \neq 0$ ,  $f(x_0 + h) - f(x_0) \neq 0$  e quindi anche  $h \neq 0$ ; dunque l'ultimo quoziente si può riscrivere anche nella forma

$$\frac{1}{\frac{f(x_0 + h) - f(x_0)}{h}}.$$

Inoltre, per  $k \rightarrow 0$  si ha  $g(y_0 + k) \rightarrow g(y_0)$  perché  $g$  è continua, essendo l'inversa di una funzione continua su un intervallo (si veda il Cap.3, par. 4.3); d'altro canto  $h = g(y_0 + k) - g(y_0)$ , quindi per  $k \rightarrow 0$  anche  $h \rightarrow 0$ , e per ipotesi

$$\frac{1}{\frac{f(x_0 + h) - f(x_0)}{h}} \rightarrow \frac{1}{f'(x_0)}$$

quindi esiste

$$\lim_{k \rightarrow 0} \frac{g(y_0 + k) - g(y_0)}{k} = \frac{1}{f'(x_0)}.$$

Osserviamo che, assumendo la derivabilità di  $f^{-1}$ , la (3.9) segue subito dall'identità  $g(f(x)) = x$  e dalla regola della catena:

$$g'(f(x)) \cdot f'(x) = 1$$

da cui, se  $f'(x) \neq 0$ , la (3.9). La dimostrazione che abbiamo dato, tuttavia, è necessaria per dedurre la derivabilità di  $g$  dalle nostre ipotesi su  $f$ .

La (3.9) ha un semplice significato geometrico, ricordando che i grafici di  $f$  e  $g = f^{-1}$  sono simmetrici rispetto alla bisettrice  $y = x$  (fig. 4.17).

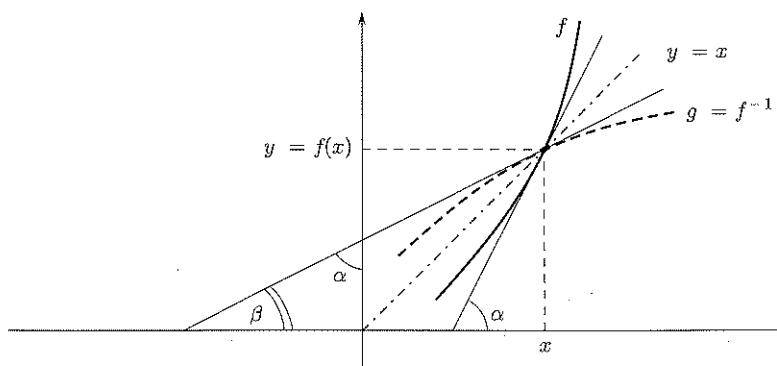


Figura 4.17. Gli angoli  $\alpha$  e  $\beta$  sono complementari ( $\alpha + \beta = \pi/2$ ) e quindi  $f'(x) = \operatorname{tg} \alpha = \operatorname{tg}(\frac{\pi}{2} - \beta) = \frac{1}{\operatorname{tg} \beta} = \frac{1}{g'(y)}$ .

Con la notazione di Leibniz, posto  $y = f(x)$ ,  $x = g(y)$ , (3.9) si scrive nella forma

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} \quad (!)$$

Si faccia attenzione al fatto che nella formula di derivazione della funzione inversa, le derivate  $f'$  e  $g'$  sono calcolate in *due punti diversi*: è questa la principale attenzione da avere nell'applicazione di questo teorema, come mostreranno i prossimi esempi.



**Esempio** Derivata dell'arcotangente, dell'arcoseno e dell'arcocoseno.

Poniamo  $y = \operatorname{tg} x$ ,  $x = \operatorname{arctg} y$  con  $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$ ,  $y \in \mathbb{R}$ .

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{1 + (\operatorname{tg} x)^2} = \frac{1}{1 + y^2}$$

**Esempio** Poniamo ora  $y = \sin x$ ,  $x = \arcsin y$ , con  $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ ,  $y \in [-1, 1]$ .

Poiché per quei valori di  $x$  si ha  $\cos x = \sqrt{1 - (\sin x)^2} = \sqrt{1 - y^2}$ :

$$\frac{dx}{dy} = \frac{1}{\cos x} = \frac{1}{\sqrt{1 - y^2}}$$

**Esempio** Analogamente, se  $y = \cos x$ ,  $x = \arccos y$  con  $x \in [0, \pi]$ ,  $y \in [-1, 1]$ ,

$$\frac{dx}{dy} = \frac{1}{-\sin x} = \frac{1}{-\sqrt{1 - y^2}}$$

Si osservi che le funzioni  $\arcsin x$ ,  $\arccos x$ , pur essendo definite e continue in  $[-1, 1]$ , non sono derivabili agli estremi dell'intervallo: precisamente, presentano in questi punti tangente verticale.

L'utilità del teorema di derivazione della funzione inversa consiste nel fatto che permette di calcolare la derivata di  $g$  anche in situazioni in cui  $g$  non si sa scrivere esplicitamente:



**Esempio** Sia  $f(x) = x + e^x$ . La funzione è strettamente crescente in tutto  $\mathbb{R}$ , dunque invertibile; sia  $g$  la sua inversa. Calcoliamo, ad esempio,  $g'(y_0)$  per  $y_0 = f(0) = 1$ . Si ha:

$$f'(x) = 1 + e^x; f'(0) = 2 \neq 0; g'(1) = \frac{1}{f'(0)} = \frac{1}{2}.$$



In base alle regole di calcolo delle derivate e alla tabella delle derivate delle funzioni elementari, calcolare la derivata delle seguenti funzioni:

1)  $3x^4 + 5x + x^{3/2} - 2x^{-3}$

2)  $e^{-3x}(x^2 + 2x - 1)$

3)  $\log|x|, \log 3x, \log \left| \frac{x+2}{3-x} \right|$

4)  $\frac{1}{a} \operatorname{arctg} \frac{x}{a}, a > 0$

- $x \log x$
- $e^{2x} (2 \sin 3x - 4 \cos 3x)$
- $\cot g x, \operatorname{Th} x, \operatorname{Coth} x$
- SettSh  $x$ , SettCh  $x$  (Suggerimento: usare il teorema sulla derivata della funzione inversa)
- $\frac{x^2+3x-2}{2x+1}$
- $\frac{ax+b}{cx+d}$

Scrivere l'equazione della retta tangente al grafico di  $y = f(x)$  nel punto  $(x_0, f(x_0))$ :

- $f(x) = \sin x, x_0 = \frac{\pi}{3}$
- $f(x) = 3x^2 + 2x + 1, x_0 = 2$
- $f(x) = \log x, x_0 = 1$
- $f(x) = a^x, x_0 = 2$

Qual è il tasso di variazione del volume di una sfera rispetto al suo raggio? E rispetto all'area della sua superficie?

In un triangolo isoscele  $ABC$  (fig. 4.18) il vertice  $C$  si muove perpendicolarmente alla base  $AB$  in modo che l'area del triangolo cresca ad una velocità di  $4 \text{ cm}^2/\text{s}$ . La base  $AB$  è lunga  $3 \text{ cm}$ . A quale velocità cresce l'altezza  $CH$ ? E il lato  $CB$ ?

Un parallelepipedo di base  $1 \text{ m} \times 2 \text{ m}$  e altezza  $5 \text{ m}$  è pieno d'acqua. Da un rubinetto posto in prossimità del fondo vengono prelevati  $20 \text{ litri al minuto}$ . Con quale velocità l'altezza dell'acqua decresce?

Una funzione è nota dal suo grafico (fig. 4.19); tracciare con buona approssimazione il grafico della sua derivata.

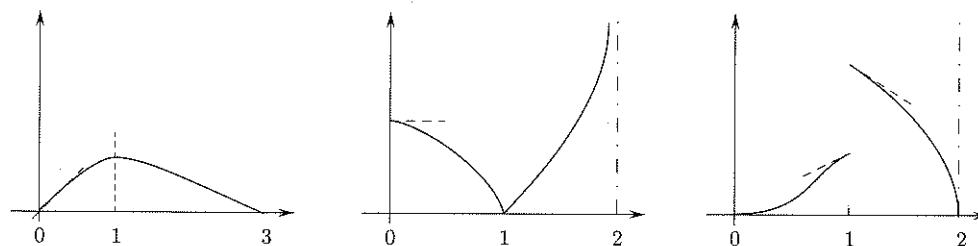


Figura 4.19.

Calcolare la derivata delle seguenti funzioni, dove esiste. Studiare i punti di non derivabilità, stabilendo se si tratta di punti angolosi (in questo caso, calcolare la derivata destra o sinistra), punti di cuspide, flessi a tangente verticale, punti a tangente verticale, tracciando un grafico locale della funzione in quei punti. Prima di eseguire il calcolo della derivata, cercare di prevedere quali saranno i punti di non derivabilità, in base alla forma della funzione.

- $\arctg \frac{1+x}{1-x}$
- $e^{\frac{x+2}{3}}$
- $2^{x^2+3x}$
- $\log_2 |3x|$
- $x^x \log x$
- $\log |\log x|$
- $|x^2 + 3x - 4|$
- $\sqrt[3]{2x^2 + x - 1}$
- $e^{\sqrt{x}}$
- $\log^2(1 + \sqrt[3]{x})$
- $|x^3|$
- $e^{-|x|}$
- $e^x \sqrt[3]{\frac{x+2}{x-3}}$

Calcolare  $f'(0)$  in base alla definizione. Calcolare poi  $f'(x)$  per  $x \neq 0$ , e stabilire se la derivata prima è continua in  $x = 0$ :

$$\begin{aligned} \text{f}(x) &= \begin{cases} x^2 \sin \frac{1}{x} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases} \\ \text{f}(x) &= \begin{cases} e^{-1/x^2} & \text{per } x > 0 \\ x^2 & \text{per } x \leq 0 \end{cases} \end{aligned}$$

## ■ 4 IL TEOREMA DEL VALOR MEDIO E LE SUE CONSEGUENZE

### 4.1 Punti stazionari. Massimi e minimi locali

Uno degli usi più proficui del calcolo differenziale consiste nella ricerca dei massimi e minimi (in breve, nell'*ottimizzazione*) di una funzione. Ecco alcuni problemi di ottimizzazione in ambiti diversi.

- Ottica geometrica.** Le leggi della riflessione e della rifrazione si possono trovare utilizzando il “principio di Fermat”: “il percorso scelto” da un raggio luminoso per collegare due punti  $A$  e  $B$  è quello che richiede il *minor* tempo di percorrenza.
- Geometria elementare.** Inscrivere in un cono circolare retto di altezza  $h$  e raggio di base  $R$ , un cilindro di volume massimo.
- Econometria elementare.** Sia  $c(p)$  il costo di produzione di una quantità  $p$  di un certo prodotto. L’efficienza della produzione si può misurare col rapporto  $\frac{c(p)}{p} =$  costo medio per unità di prodotto. Per raggiungere il massimo di efficienza occorre dunque minimizzare la funzione  $\frac{c(p)}{p}$ .
- Dalla vita comune!** Un tubo di lunghezza  $4 \text{ m}$  (sezione piccola) deve essere trasportato attraverso il cunicolo raffigurato in figura 4.20. È possibile il trasporto?

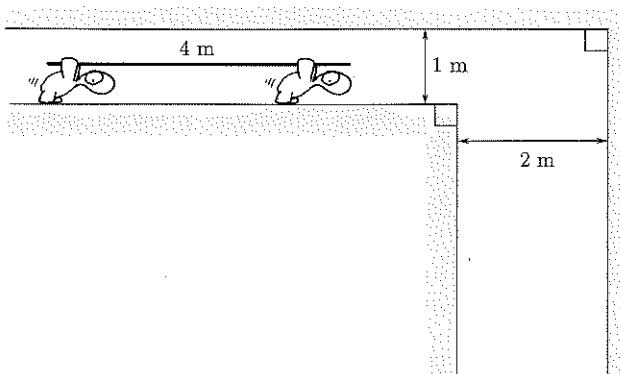


Figura 4.20. Passerà il tubo lungo 4 m?

Torniamo alla teoria. Richiamiamo le nozioni di massimo e minimo assoluti (o globali), per una funzione  $f : [a, b] \rightarrow \mathbb{R}$ :

**DEFINIZIONE 4.4** Si dice che  $M$  è *massimo* di  $f$  in  $[a, b]$  e  $x_0 \in [a, b]$  è *punto di massimo* se

$$f(x_0) = M \geq f(x), \quad \text{per ogni } x \in [a, b].$$

Analoga definizione per il *minimo*. Vi sono altri tipi di estremo (massimo o minimo) e cioè gli *estremi locali*.

**DEFINIZIONE 4.5** Si dice che  $M$  è *massimo locale* (o relativo) per  $f$  e che  $x_0$  è *punto di massimo locale* se:

esiste un intervallo  $(x_0 - \delta, x_0 + \delta)$  tale che  $M = f(x_0) \geq f(x)$   
per ogni  $x \in (x_0 - \delta, x_0 + \delta) \cap [a, b]$

Analogamente per un *minimo locale*.

Notiamo espressamente che:

- il minimo e il massimo globale di  $f$  (se esistono) sono *unici* (naturalmente i punti di max e min possono essere più di uno);
- massimi e minimi locali possono essere più di uno. Evidentemente ogni estremo globale è anche locale.

Le seguenti figure illustrano varie situazioni.

- La funzione in figura 4.21 a) presenta:

- massimo globale  $M = f(x_2)$ ;  $x_2$  unico punto di massimo globale.
- minimo globale  $m = f(a)$ ;  $a$  unico punto di minimo globale.
- un massimo locale (non globale) in  $x = x_0$ .
- due minimi locali (non globali) in  $x = x_1$  (che è anche un punto angoloso) e  $x = b$  (estremo destro di  $[a, b]$ ).

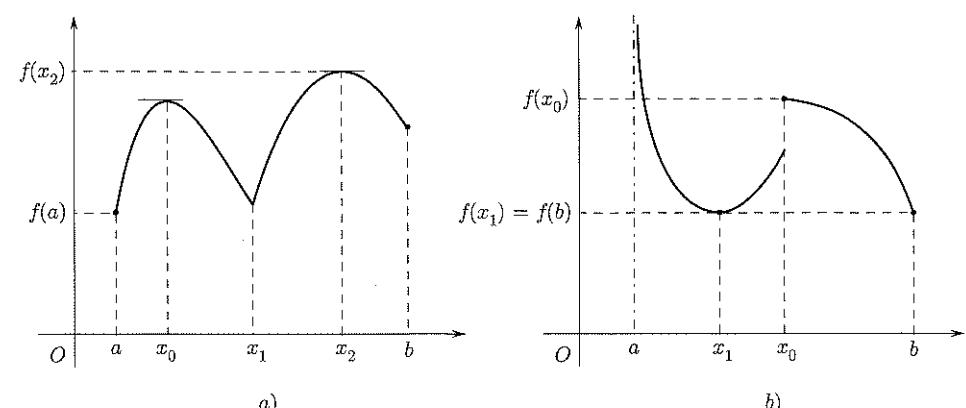


Figura 4.21.

- La funzione in figura 4.21 b) presenta:

- un massimo locale in  $x = x_0$  (punto di discontinuità a salto).
- minimo globale  $m = f(b) = f(x_1)$ ;  $b$  e  $x_1$  sono punti di minimo globale.

Il massimo globale di  $f$  non esiste (si ha  $\lim_{x \rightarrow a^+} f(x) = +\infty$ ).

Le figure mostrano che in un punto di estremo (locale o globale)  $f$  può non essere derivabile ed essere perfino discontinua. Se però  $f : [a, b] \rightarrow \mathbb{R}$  è derivabile in un punto  $x_0$  che sia di massimo o minimo locale e che sia diverso da  $a$  e da  $b$ , allora in  $x_0$  la derivata si annulla, ossia la tangente al grafico in  $(x_0, f(x_0))$  è orizzontale (fig. 4.21). Precisamente:

**TEOREMA 4.5 (DI FERMAT)** Sia  $f : [a, b] \rightarrow \mathbb{R}$ , derivabile in  $x \in (a, b)$ . Se  $x$  è punto di estremo locale allora

$$f'(x) = 0$$

**DIMOSTRAZIONE.** Sia, ad esempio,  $x$  punto di max locale. Allora, per  $z$  abbastanza vicino a  $x$ , si ha  $f(z) \leq f(x)$ . Perciò:

$$z < x \implies \frac{f(z) - f(x)}{z - x} \geq 0 \quad \text{e quindi} \quad f'_-(x) = \lim_{z \rightarrow x^-} \frac{f(z) - f(x)}{z - x} \geq 0$$

(abbiamo applicato il teorema della permanenza del segno, cap. 3, par. 3.1). D'altra parte

$$z > x \implies \frac{f(z) - f(x)}{z - x} \leq 0 \quad \text{e quindi} \quad f'_+(x) = \lim_{z \rightarrow x^+} \frac{f(z) - f(x)}{z - x} \leq 0$$

Essendo  $f$  derivabile in  $x$ , si ha  $f'_-(x) = f'_+(x) = 0$ .

**Punti stazionari**

I punti in cui  $f'$  si annulla, si dicono *punti stazionari* per  $f$ . Abbiamo appena visto che, se  $x$  non si trova agli estremi dell'intervallo nel quale  $f$  è definita, allora

$$x \text{ di estremo locale} \implies x \text{ stazionario}$$

Vi possono però essere punti stazionari che *non* sono di estremo. Ad esempio la funzione  $f(x) = x^3$  ha  $f'(x) = 3x^2$  che si annulla nell'origine, ma  $x = 0$  non è punto di estremo (vedi fig. 4.22).

Si tratta di un punto di *flesso* (o di inflessione) a tangente orizzontale come vedremo più avanti.

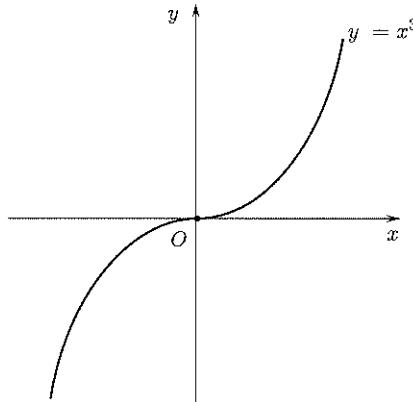


Figura 4.22.  $y = x^3$  ha un punto di flesso in  $x = 0$ , stazionario.

**4.2 Teorema del valor medio. Test di monotonia**

Supponiamo che per raggiungere una località  $B$ , partendo da una località  $A$  distante 160 km da  $B$ , si impieghino 2 h. La velocità media durante il percorso è dunque di 80 km/h. Se il viaggio non ha subito interruzioni o altre irregolarità, certamente in qualche istante la velocità è stata esattamente di 80 km/h.

Cioè, in qualche istante:

$$\text{velocità media} = \text{velocità istantanea}$$

È concettualmente questo il contenuto del teorema del valor medio, che come vedremo, è denso di conseguenze.

**TEOREMA 4.6 (DEL VALOR MEDIO O DI LAGRANGE)** *Sia  $f$  derivabile in  $(a, b)$  e continua in  $[a, b]$  (cioè continua fino agli estremi dell'intervallo). Allora*

$$(4.1) \quad \text{esiste } c \in (a, b) \text{ tale che } \frac{f(b) - f(a)}{b - a} = f'(c)$$

Il significato geometrico del teorema è illustrato nella figura seguente.

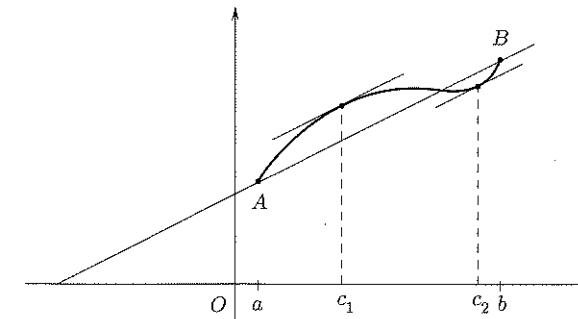


Figura 4.23.

Si ha:

$$\frac{f(b) - f(a)}{b - a} = \text{pendenza della retta } AB$$

$$f'(c) = \text{pendenza della retta tangente al grafico di } f \text{ nel punto } (c, f(c)).$$

La (4.1) esprime dunque il fatto che nel punto  $(c, f(c))$  la tangente al grafico di  $f$  è parallela alla retta  $AB$ . In figura 4.23 esistono due di tali punti, di ascissa  $c_1, c_2$ .

**DIMOSTRAZIONE DEL TEOREMA DEL VALOR MEDIO.** Osserviamo che la retta  $AB$  ha equazione

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

e consideriamo la funzione

$$w(x) = f(x) - \left[ f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right]$$

È facile verificare che:  $w(a) = w(b) = 0$ ,  $w$  è continua in  $[a, b]$  e  $w$  è derivabile in  $(a, b)$ .

Poiché

$$w'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

la (4.1) equivale a dimostrare che esiste  $c \in (a, b)$  tale che  $w'(c) = 0$ .

Essendo  $w$  continua in  $[a, b]$ , per il teorema di Weierstrass esistono due punti  $x_1$  e  $x_2$  in  $[a, b]$  tali che

$$\begin{aligned} f(x_1) &= \text{massimo di } f \text{ in } [a, b] = M \\ f(x_2) &= \text{minimo di } f \text{ in } [a, b] = m \end{aligned}$$

Se  $M = m$ , allora  $w(x)$  è costante in  $[a, b]$ , e quindi  $w'(x) = 0 \forall x \in [a, b]$ .

Se  $M > m$ , almeno uno dei due punti  $x_1, x_2$  non si trova agli estremi dell'intervallo, essendo  $w(a) = w(b) = 0$ .

Il teorema di Fermat implica allora che nel punto di massimo o minimo che risulta interno (eventualmente entrambi) la derivata di  $w$  si annulla e il teorema è così dimostrato.



Sia  $f(x) = x^2$ . Allora  $f'(x) = 2x$  e il teorema afferma che in ogni intervallo  $[a, b]$  esiste un numero  $c$  tale che

$$\frac{b^2 - a^2}{b - a} = 2c \quad \text{da cui}$$

$$c = \frac{a+b}{2} = \text{media aritmetica di } a \text{ e } b$$

Cioè: ogni corda  $AB$  della parabola  $y = x^2$  è parallela alla tangente nel punto di ascissa uguale alla media aritmetica delle ascisse di  $A$  e  $B$ .

Sia  $f(x) = \frac{1}{x}$ . Allora  $f'(x) = -\frac{1}{x^2}$  e dal teorema si deduce l'esistenza di  $c \in [a, b]$  tale che

$$\frac{\frac{1}{b} - \frac{1}{a}}{b - a} = -\frac{1}{c^2}, \quad \text{ossia} \quad \frac{1}{ab} = \frac{1}{c^2}$$

da cui

$$c = \sqrt{ab} = \text{media geometrica di } a \text{ e } b$$

Cioè: ogni corda  $AB$  dell'iperbole  $y = \frac{1}{x}$  è parallela alla tangente nel punto di ascissa uguale alla *media geometrica* delle ascisse di  $A$  e  $B$ .

Si noti che  $\sqrt{ab} \leq \frac{a+b}{2}$  e cioè che

$$\text{media geometrica} \leq \text{media aritmetica}$$

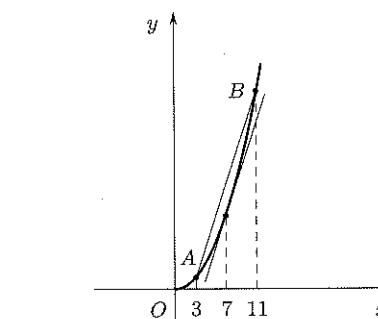


Figura 4.24.

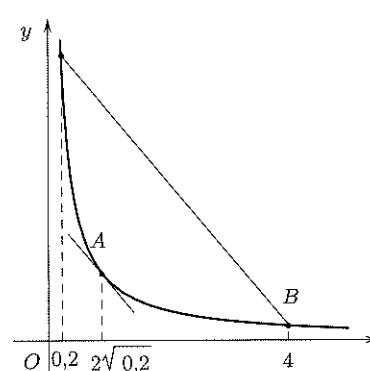


Figura 4.25.

Ricordando il significato geometrico di derivata, si deduce subito che se una funzione derivabile è crescente oppure decrescente in un intervallo  $(a, b)$  la sua derivata è  $\geq 0$  oppure  $\leq 0$  rispettivamente. Il teorema di Lagrange permette di dimostrare anche il viceversa. Le due implicazioni sono contenute nel prossimo:

**TEOREMA 4.7 (TEST DI MONOTONIA)** Sia  $f : (a, b) \rightarrow \mathbb{R}$ , derivabile. Allora

$$\begin{aligned} f &\text{ crescente} & \Leftrightarrow & f'(x) \geq 0 \\ &&& \forall x \in (a, b) \\ f &\text{ decrescente} & \Leftrightarrow & f'(x) \leq 0 \end{aligned}$$

**DIMOSTRAZIONE.** Infatti, considerata  $f : (a, b) \rightarrow \mathbb{R}$  e due punti qualunque  $x, z \in (a, b)$

$$f \begin{cases} \text{crescente} \\ \text{decrescente} \end{cases} \Leftrightarrow \frac{f(z) - f(x)}{z - x} \begin{cases} \geq 0 \\ \leq 0 \end{cases}$$

Passando al limite per  $z \rightarrow x$ , per il teorema della permanenza del segno, dalle due precedenti relazioni si ottiene

$$\begin{aligned} f &\text{ crescente} & \Rightarrow & f'(x) \geq 0 \\ f &\text{ decrescente} & \Rightarrow & f'(x) \leq 0 \end{aligned} \quad \forall x \in (a, b)$$

Viceversa, sia, ad esempio,  $f'(x) \geq 0$  per ogni  $x \in (a, b)$ , e proviamo che allora  $f$  è crescente in  $(a, b)$ . Prendiamo dunque due punti qualsiasi  $x_1, x_2 \in (a, b)$ ,  $x_1 < x_2$ , e mostriamo che  $f(x_1) \leq f(x_2)$ . Infatti, applicando il teorema di Lagrange ad  $f$  sull'intervallo  $[x_1, x_2]$  abbiamo che esiste  $c \in (x_1, x_2)$  tale che

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c)$$

Poiché  $f'(c) \geq 0$  e  $x_2 - x_1 > 0$ , ne segue  $f(x_2) - f(x_1) \geq 0$ , cioè la tesi. ◆

Con la stessa dimostrazione si vede anche che: se  $f'(x) = 0$  per ogni  $x \in (a, b)$ , allora  $f$  è costante in  $(a, b)$ . Poiché l'implicazione inversa (se  $f$  è costante in  $(a, b)$  allora ha derivata nulla) è ovvia, risulta dimostrata la seguente

**PROPOSIZIONE 4.1 (CARATTERIZZAZIONE DELLE FUNZIONI A DERIVATA NULLA)**  
Sia  $f : (a, b) \rightarrow \mathbb{R}$ . Allora

$$f' = 0 \text{ in } (a, b) \Leftrightarrow f \text{ è costante in } (a, b)$$

Un errore da evitare è usare la proposizione 4.1 su insiemi più generali degli intervalli.



Consideriamo la funzione

$$f(x) = \arctg x + \arctg \frac{1}{x}, \quad \text{per } x \neq 0$$

e calcoliamo

$$f'(x) = \frac{1}{1+x^2} + \frac{1}{1+\frac{1}{x^2}} \left( -\frac{1}{x^2} \right) = 0 \quad \text{per ogni } x \neq 0$$

Si può applicare la proposizione 4.1 a  $f$  e concludere che  $f$  è costante? La risposta è no: si può solo affermare che è costante su ciascuno dei due intervalli  $(0, +\infty)$  e  $(-\infty, 0)$ . Per sapere quanto vale, è sufficiente calcolare  $f$  in un punto "comodo" di ciascun intervallo, per esempio:

$$f(1) = \arctg 1 + \arctg 1 = 2 \cdot \frac{\pi}{4} = \frac{\pi}{2}; \quad \text{dunque } f(x) = \frac{\pi}{2} \text{ per ogni } x > 0$$

$$f(-1) = \arctg(-1) + \arctg(-1) = -\frac{\pi}{2}; \quad \text{dunque } f(x) = -\frac{\pi}{2} \text{ per ogni } x < 0$$

Abbiamo dunque dimostrato l'identità:

$$\arctg x + \arctg \frac{1}{x} = \begin{cases} \frac{\pi}{2} & \text{se } x > 0 \\ -\frac{\pi}{2} & \text{se } x < 0 \end{cases}$$

**Ricerca di massimi e minimi**

Supponiamo di avere la solita funzione  $f : [a, b] \rightarrow \mathbb{R}$  e di volerne cercare i massimi e i minimi.

Se  $f$  è derivabile si può procedere nel modo seguente.

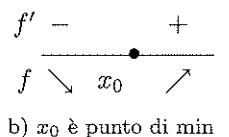
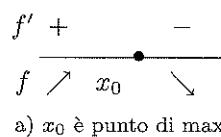
*Passo 1.* Si calcolano  $f(a)$  e  $f(b)$ .

*Passo 2.* Si calcola  $f'(x)$  e si risolve l'equazione

$$f'(x) = 0$$

In tal modo si trovano i punti stazionari, tra i quali vi sono gli eventuali punti di estremo locale, interni a  $(a, b)$ .

*Passo 3.* Se non vi sono punti stazionari,  $f(a)$  o  $f(b)$  sono punti di estremo globale. Se viceversa  $x = x_0$  è punto stazionario, occorre stabilirne la natura. A tale scopo si può studiare il segno di  $f'$  in un intorno di  $x_0$ , ricordando che  $f' \geq 0$  implica  $f$  crescente,  $f' \leq 0$  implica  $f$  decrescente. I casi che si presentano più comunemente sono illustrati qui di seguito:



$$f'(x_0) = 0$$

*Passo 4.* Trovati gli eventuali punti di estremo locale, si calcola il valore di  $f$  in questi punti e lo si confronta con  $f(a)$  e  $f(b)$ .

Vediamo degli esempi numerici

**Esempio**

Sia  $f(x) = xe^{-x^2}$  con  $x \in [0, 2]$ .

$$1. f(0) = 0, f(2) = 2e^{-4}$$

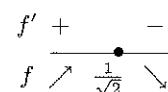
$$2. f'(x) = e^{-x^2} + xe^{-x^2}(-2x) = e^{-x^2}(1 - 2x^2)$$

$$f'(x) = 0 \iff 1 - 2x^2 = 0 \implies x = \pm \frac{1}{\sqrt{2}}$$

Solo  $x_0 = \frac{1}{\sqrt{2}} \in [0, 2]$  e perciò è questo l'unico punto stazionario.

3. Studiamo ora il segno di  $f'$ , vicino a  $x_0 = \frac{1}{\sqrt{2}}$ :

$$f'(x) \geq 0 \text{ per } 2x^2 \leq 1 \quad -\frac{1}{\sqrt{2}} \leq x \leq \frac{1}{\sqrt{2}}$$



Si deduce quindi che  $x_0 = \frac{1}{\sqrt{2}}$  è punto di massimo locale.

4.  $f\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}e^{-1/2} = \frac{1}{\sqrt{2}e}$  è maggiore sia di  $f(0) = 0$  che di  $f(2) = \frac{2}{e^4}$ . Si conclude quindi che:

$$f(0) = 0 \text{ è minimo globale}$$

$$f\left(\frac{1}{\sqrt{2}}\right) \text{ è massimo globale}$$

Si noti che in questo caso, il fatto che esista un solo punto stazionario con ordinata maggiore di quella agli estremi dell'intervallo, permette di trarre le stesse conclusioni anche senza lo studio del segno della derivata prima.

Un grafico qualitativo di  $f$  è il seguente

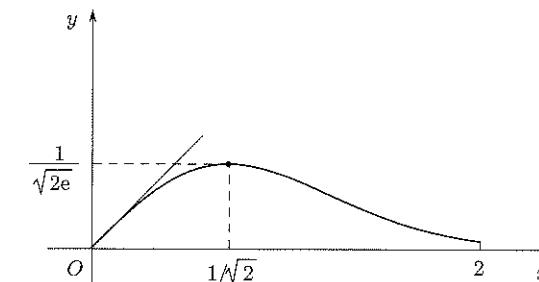


Figura 4.26. Grafico in  $[0, 2]$  di  $f(x) = xe^{-x^2}$ ; si noti che  $f(x) \sim x$  per  $x \rightarrow 0$ .

Se volessimo studiare i massimi e minimi locali di  $f(x)$  su tutto  $\mathbb{R}$ , il segno della derivata prima ci direbbe che

$$x = \frac{1}{\sqrt{2}} \text{ è punto di massimo locale; } x = -\frac{1}{\sqrt{2}} \text{ è punto di minimo locale.}$$

**Figura di diffrazione della luce attraverso una fenditura.** Un fascio di luce che attraversa una piccola fenditura la cui larghezza è dello stesso ordine di grandezza della lunghezza d'onda della luce, produce, su uno schermo su cui incide, una figura di interferenza. L'intensità luminosa in un punto dello schermo è data dalla funzione:

$$I = I_0 \left( \frac{\sin \frac{\varphi}{2}}{\frac{\varphi}{2}} \right)^2$$

dove  $I_0$  è l'intensità massima (che si ottiene nel punto centrale) e l'angolo  $\varphi$  è collegato alla differenza di fase tra le due onde luminose che passano ai due estremi della fenditura. Cerchiamo i massimi e minimi di  $I$  in funzione di  $\varphi$ . Anzitutto, conviene porre  $t = \varphi/2$  e cercare massimi e minimi della funzione

$$I = I_0 \left( \frac{\sin t}{t} \right)^2$$

(questo significa semplicemente cambiare scala sull'asse  $\varphi$ ; basterà ricordare poi che  $\varphi = 2t$ ). Il grafico qualitativo di questa funzione può essere tracciato facilmente, senza calcoli: la funzione  $\frac{\sin t}{t}$  è pari, vale 1 in 0, ha infinite oscillazioni smorzate per  $t \rightarrow \infty$ ; di conseguenza la funzione  $I = I_0 \left( \frac{\sin t}{t} \right)^2$  è pari, non negativa, vale  $I_0$  in 0, ed ha infinite oscillazioni smorzate:

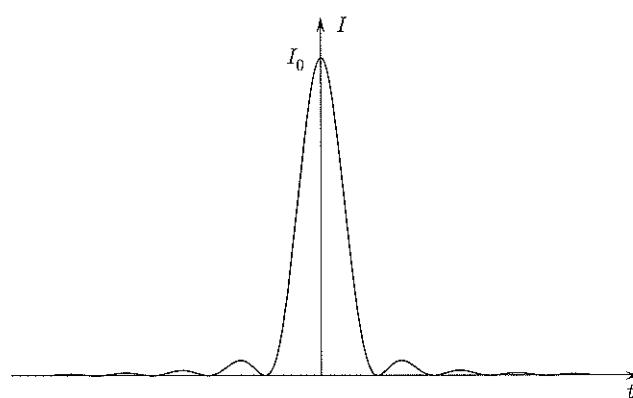


Figura 4.27.

Vediamo quindi che  $I$  ha infiniti punti di minimo, nei punti in cui si annulla,  $t = k\pi$  (cioè  $\varphi = 2k\pi$ ), e infiniti punti di massimo, di cui il massimo assoluto è per  $t = 0$  ( $\varphi = 0$ ), e vale  $I_0$ . Il problema è determinare gli altri punti di massimo relativo. Questi si trovano risolvendo l'equazione

$$\frac{dI}{dt} = 0 \quad \text{ossia} \quad I_0 2 \frac{\sin t}{t} \left( \frac{t \cos t - \sin t}{t^2} \right) = 0$$

Il fattore  $\sin t$  si annulla nei punti di minimo che già conosciamo; i punti di massimo sono dunque le soluzioni di

$$t \cos t - \sin t = 0$$

equazione che non può essere risolta in modo esatto, ma che si studia facilmente con un *confronto grafico*: riscritta nella forma

$$t = \operatorname{tg} t$$

possiamo osservare dal grafico delle funzioni  $t, \operatorname{tg} t$ , dove sono collocate le soluzioni (v. fig. 4.28).

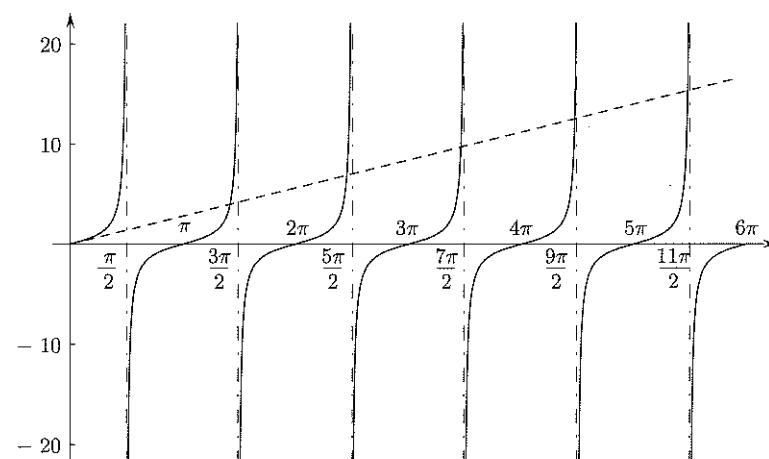


Figura 4.28.

Il primo punto positivo di massimo si trova poco prima di  $\frac{3}{2}\pi$ ; il secondo poco prima di  $\frac{5}{2}\pi$ , e così via; più ci si allontana dall'origine, più questi punti tendono a coincidere esattamente con i valori  $t = \frac{3}{2}\pi + k\pi$  (quindi  $\varphi = 3\pi + 2k\pi$ ). Dunque ogni punto di massimo si trova approssimativamente a metà strada tra due punti di minimo successivi. I valori dei massimi relativi saranno all'incirca

$$I \left( \frac{3}{2}\pi + k\pi \right) = \frac{I_0}{(\frac{3}{2}\pi + k\pi)^2}$$

Una localizzazione più precisa delle soluzioni dell'equazione  $t = \operatorname{tg} t$  si potrà dare con i metodi esposti nel paragrafo 7.5.

### Successioni monotone

Come abbiamo visto nel capitolo 3, le successioni monotone hanno importanti proprietà. D'altro canto, non sempre è facile dimostrare la monotonia di una successione per via algebrica. Il calcolo differenziale ci offre un metodo utile.



Sia

$$a_n = \frac{\log n}{n}$$

Sappiamo che  $a_n \geq 0$  e  $a_n \rightarrow 0$  (per confronto tra infiniti). Chiediamoci se la successione è monotona decrescente. Per definizione, ciò significa che

$$a_{n+1} \leq a_n \quad \text{ossia} \quad \frac{\log(n+1)}{n+1} \leq \frac{\log n}{n}$$

Poiché al crescere di  $n$  sia il numeratore che il denominatore crescono, non è facile provare questa diseguaglianza. D'altro canto, sia

$$f(x) = \frac{\log x}{x} \quad \text{per } x > 0$$

(Ora la variabile  $x$  non è un intero, ma un numero reale!). Calcoliamo:

$$f'(x) = \frac{1 - \log x}{x^2} \leq 0 \quad \text{per } x \geq e$$

Ne segue che  $f$  è decrescente per  $x \geq e$ ; di conseguenza la successione  $a_n = f(n)$  è decrescente almeno per  $n \geq 3$  (il primo intero  $> e$ ). Nel capitolo 5, parlando di *serie a segni alterni*, vedremo un'applicazione di questa osservazione.

Il passaggio “dal discreto al continuo” (cioè dagli interi ai reali) è un modo per avere a disposizione gli strumenti del calcolo differenziale. Attenzione a non usare indiscriminatamente questo expediente. Si ragioni sul fatto che in questo modo *non si potrebbe* studiare la monotonia di successioni come

$$\frac{n!}{2^n} \quad \frac{n^2 + (-1)^n n}{n^3 + 1} \quad \frac{n + (-2)^n}{3^n} \dots$$

### 4.3 Soluzione di alcuni problemi di massimo e minimo

Ritorniamo in questo paragrafo ai problemi posti all'inizio del paragrafo 4.1, presentandone la soluzione.

**Legge della riflessione**

Vogliamo ricavare la legge sugli angoli di riflessione dal principio di Fermat. A tale scopo consideriamo, in un mezzo omogeneo, un raggio luminoso riflesso da uno specchio curvo. Immaginiamo che la sezione dello specchio coincida con il grafico di una funzione  $y = f(x)$ , derivabile. Possiamo sempre scegliere il sistema di riferimento in modo che il punto di riflessione sia l'origine  $(0, 0)$  e che l'asse  $x$  sia diretto come la tangente allo specchio nell'origine. Di conseguenza abbiamo

$$f(0) = 0 \quad f'(0) = 0$$

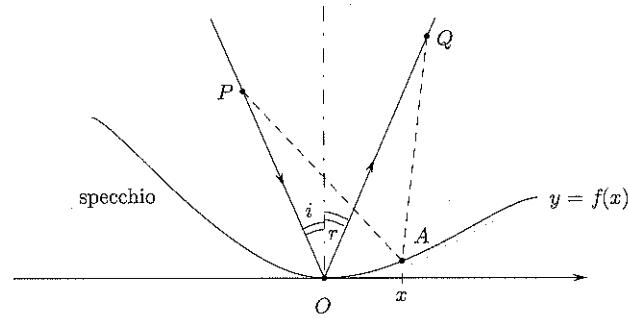


Figura 4.29.

Con riferimento alla figura 4.29,  $P$  è un punto "sul raggio incidente" e  $Q$  è un punto "sul raggio riflesso". Il punto  $A$  è un altro punto sullo specchio. Quello che dobbiamo mostrare è che, poiché il tempo impiegato dal raggio per andare da  $P$  a  $Q$  è minimo (principio di Fermat), deve risultare  $i = r$ .

Ora, la velocità della luce è costante in un mezzo omogeneo e non varia dopo la riflessione e, pertanto, minimizzare il tempo di percorrenza da  $P$  a  $Q$  equivale a minimizzare il percorso  $S = \overline{PA} + \overline{AQ}$ .

Siano  $(x_0, y_0)$  le coordinate di  $P$ ,  $(x_1, y_1)$  quelle di  $Q$  e  $(x, f(x))$  quelle di  $A$ . Si ha:

$$S(x) = \sqrt{(x - x_0)^2 + (f(x) - y_0)^2} + \sqrt{(x - x_1)^2 + (f(x) - y_1)^2}$$

La funzione  $S = S(x)$  deve avere un minimo in  $x = 0$ , ascissa del punto di riflessione e perciò deve essere  $S'(0) = 0$ .

Essendo

$$S'(x) = \frac{(x - x_0) + [f(x) - y_0] \cdot f'(x)}{\sqrt{(x - x_0)^2 + (f(x) - y_0)^2}} + \frac{(x - x_1) + [f(x) - y_1] \cdot f'(x)}{\sqrt{(x - x_1)^2 + (f(x) - y_1)^2}}$$

$S'(0) = 0$  equivale a (ricordando che  $f(0) = f'(0) = 0$ )

$$(4.2) \quad \frac{-x_0}{\sqrt{x_0^2 + y_0^2}} + \frac{-x_1}{\sqrt{x_1^2 + y_1^2}} = 0$$

Ma

$$\sqrt{x_0^2 + y_0^2} = \overline{PO} \quad \text{e} \quad \frac{-x_0}{\sqrt{x_0^2 + y_0^2}} = \sin i$$

mentre

$$\sqrt{x_1^2 + y_1^2} = \overline{QO} \quad \text{e} \quad \frac{x_1}{\sqrt{x_1^2 + y_1^2}} = \sin r.$$

Da (4.2) si deduce

$$\sin i = \sin r$$

che implica (essendo  $0 < i, r < \frac{\pi}{2}$ ),  $i = r$ .

**Legge della rifrazione**

Quando un raggio luminoso passa da un mezzo omogeneo a un altro (aria  $\rightarrow$  acqua, ad esempio) la sua velocità cambia. Supponiamo che i due mezzi siano separati da una superficie piana e che la velocità (scalare) nei due mezzi sia  $v_1$  nel primo e  $v_2$  nel secondo.

Nel passare da un mezzo all'altro, un raggio luminoso cambia direzione di propagazione (viene cioè *rifratto*) ma raggio incidente e raggio rifratto giacciono in un piano verticale alla superficie di separazione.

Fissiamo il sistema di riferimento in modo che l'asse  $x$  sia sulla superficie di separazione e che il punto di rifrazione sia l'origine (fig. 4.30).

Sia  $P = (x_0, y_0)$  un punto sul raggio incidente,  $Q = (x_1, y_1)$  un punto sul raggio rifratto e  $A = (x, 0)$  un altro punto sull'asse  $x$ .

In base al principio di Fermat, il percorso  $POQ$ , scelto dal raggio luminoso, è quello che minimizza il tempo di percorrenza da  $P$  a  $Q$ . Ora, per un qualunque percorso  $PAQ$  il tempo  $T = T(x)$  di percorrenza è dato da (tempo = spazio/velocità)

$$T(x) = \frac{\overline{PA}}{v_1} + \frac{\overline{AQ}}{v_2} = \frac{\sqrt{(x - x_0)^2 + y_0^2}}{v_1} + \frac{\sqrt{(x - x_1)^2 + y_1^2}}{v_2}$$

e quindi il principio di Fermat implica  $T'(0) = 0$ .

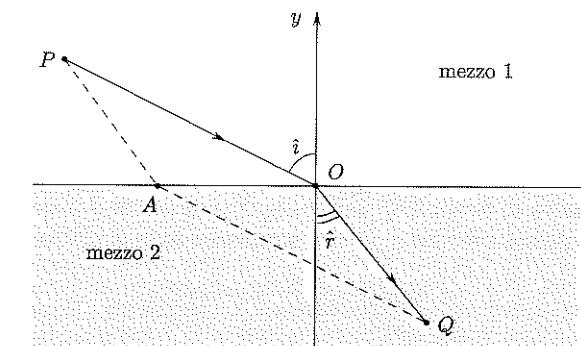


Figura 4.30.

Poiché

$$T'(x) = \frac{x - x_0}{v_1 \sqrt{(x - x_0)^2 + y_0^2}} + \frac{x - x_1}{v_2 \sqrt{(x - x_1)^2 + y_1^2}}$$

l'equazione  $T'(0) = 0$  equivale a

$$(4.3) \quad \frac{1}{v_1} \frac{-x_0}{\sqrt{x_0^2 + y_0^2}} - \frac{1}{v_2} \frac{x_1}{\sqrt{x_1^2 + y_1^2}} = 0$$

Osservando che  $\frac{-x_0}{\sqrt{x_0^2 + y_0^2}} = \sin \hat{i}$  e  $\frac{x_1}{\sqrt{x_1^2 + y_1^2}} = \sin \hat{r}$ , dalla (4.3) si deduce

$$\boxed{\frac{\sin \hat{i}}{\sin \hat{r}} = \frac{v_1}{v_2}}$$

che si chiama *legge di Snell* della rifrazione.

Inscrivere in un cono circolare retto di altezza  $h$  e raggio di base  $R$ , un cilindro di volume massimo.

Con riferimento alla figura 4.31, occorre scegliere  $\overline{CH} = x$  in modo che il volume  $\pi \overline{CH} \cdot \overline{CD}^2$  sia massimo. Poiché i triangoli  $CAB$  e  $DAK$  sono simili, si ha  $\overline{DA} = x \cdot \frac{R}{h}$  e quindi  $\overline{CD} = R - \overline{DA} = R - \frac{R}{h}x$ .

Il volume da massimizzare è quindi dato da

$$\begin{aligned} V(x) &= \pi x \cdot \left(R - \frac{R}{h}x\right)^2 = \\ &= \pi \left(\frac{R}{h}\right)^2 x \cdot (h-x)^2 \quad 0 < x < h \end{aligned}$$

Si ha  $V'(x) = \pi \left(\frac{R}{h}\right)^2 [(h-x)^2 - 2x(h-x)]$ . Annullando  $V'(x)$  si trova

$$\begin{aligned} 0 &= (h-x)^2 - 2x(h-x) = (h-x)[h-x-2x] = \\ &= (h-x)(h-3x) \end{aligned}$$

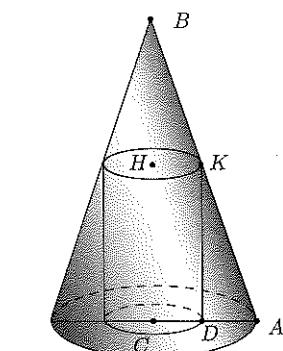


Figura 4.31.  $AC = R$ ,  $CB = h$ .

da cui  $x = \frac{h}{3}$ , essendo  $x = h$  corrispondente a un cono degenere di volume 0. Poiché  $V'(x) > 0$  per  $x < \frac{h}{3}$ ,  $x = \frac{h}{3}$  è effettivamente punto di massimo, corrispondente al volume

$$V\left(\frac{h}{3}\right) = \frac{4}{27} \pi h R^2$$

Consideriamo ora il problema dell'efficienza della produzione di un dato bene, il cui costo medio (per unità di prodotto) sia  $c(p)/p$ . Per ottenere il massimo di efficienza occorre minimizzare il costo medio e perciò, supponendo la funzione  $c = c(p)$  derivabile, dovrà essere

$$\frac{d}{dp} \frac{c(p)}{p} = \frac{c'(p) \cdot p - c(p)}{p^2} = 0$$

Se  $p_0$  è punto di minimo dovrà quindi verificarsi che

$$(4.4) \quad c'(p_0) = \frac{c(p_0)}{p_0}$$

ossia: costo marginale = costo medio

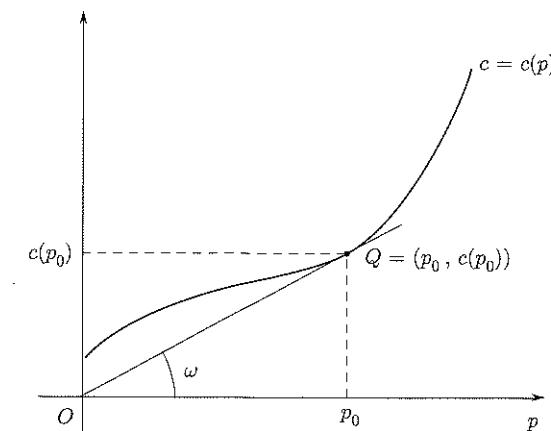


Figura 4.32.  $\operatorname{tg} \omega = c'(p_0) = c(p_0)/p_0$ .

La (4.4) si può riscrivere come  $p_0 \frac{c'(p_0)}{c(p_0)} = 1$  e significa che l'*elasticità di c nel punto  $p_0$*  è uguale a 1.

La (4.4) ha un'interessante interpretazione geometrica, illustrata nella figura 4.32: Se  $p_0$  verifica (4.4), la retta  $OQ$ , di coefficiente angolare  $\frac{c(p_0)}{p_0}$ , coincide con la retta tangente al grafico di  $c = c(p)$  nel punto  $Q$ , di coefficiente angolare  $c'(p_0)$ .

Queste considerazioni indicano che non sempre esiste un punto  $p_0$  che soddisfa (4.4) e che, se esiste, può non essere unico (vedere figure 4.33 a), b)).

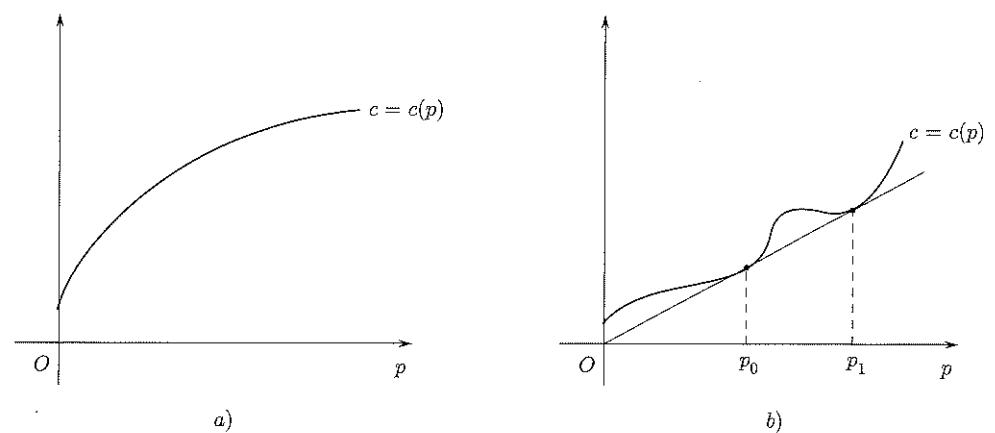


Figura 4.33. a) Non esistono punti ad elasticità 1; b)  $p_0$  e  $p_1$  sono punti ad elasticità 1.

Consideriamo il problema del passaggio di un tubo di lunghezza 4 m, con sezione trascurabile, attraverso il cunicolo (bidimensionale) in figura 4.34 e scegliamo il riferimento cartesiano come in figura 4.34.

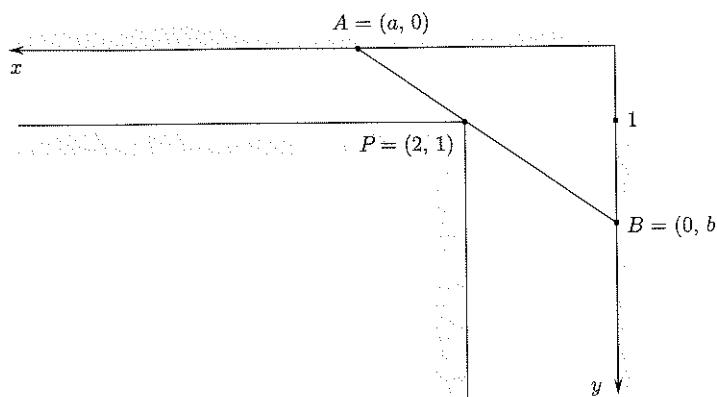


Figura 4.34.

Un minimo di riflessione indica che il passaggio è possibile se la lunghezza del tubo è minore della minima lunghezza dei possibili segmenti  $APB$ .

L'equazione della retta  $AP$  è

$$y = \frac{1}{2-a}(x - a)$$

Se  $a > 2$ , la retta interseca il semiasse positivo delle  $y$  nel punto  $B = (0, \frac{a}{a-2})$ . La lunghezza del segmento  $AB$ , da minimizzare, è data allora da:

$$L(a) = \sqrt{a^2 + \left(\frac{a}{a-2}\right)^2}$$

Poiché la radice quadrata è una funzione crescente basterà minimizzare il suo argomento

$$L_0(a) = a^2 + \left(\frac{a}{a-2}\right)^2 \quad a > 2$$

Si ha

$$L'(a) = 2a\left(1 - \frac{2}{(a-2)^3}\right)$$

$L'(a) = 0$  per  $(a-2)^3 = 2$  cioè  $a = \sqrt[3]{2} + 2$ , che è punto di minimo come facilmente si verifica. La lunghezza minima è quindi

$$L(\sqrt[3]{2} + 2) = \sqrt{5 + 3\sqrt[3]{4} + 3\sqrt[3]{16}} \simeq 4.16$$

Essendo  $4.16 > 4$ , il tubo può passare.

Trovare il massimo o il minimo di una funzione è anche un modo per *dimostrare disuguaglianze*, altra tipica applicazione del calcolo differenziale.

Trovare la più piccola costante  $c_n > 0$  per cui è vera la seguente diseguaglianza:

$$(a+b)^n \leq c_n(a^n + b^n) \text{ per ogni } a, b > 0$$

(dove  $n$  è un intero positivo fissato).

Anzitutto osserviamo che, per  $a, b > 0$ , è sempre

$$(a^n + b^n) < (a+b)^n$$

(basta sviluppare la potenza di  $(a+b)$ ); si chiede di dimostrare che vale anche una diseguaglianza di segno opposto, per una opportuna costante  $c_n$ . Si osservi che la costante *più piccola* per cui questo è vero è la costante *migliore*, cioè quella per cui la diseguaglianza dà l'informazione più precisa. Anzitutto riduciamo il problema ad un problema in una sola variabile anziché due: basta raccogliere  $b^n$  ad ambo i membri e semplificare:

$$\left(\frac{a}{b} + 1\right)^n \leq c_n \left(\left(\frac{a}{b}\right)^n + 1\right)$$

e quindi porre  $t = a/b$ :

$$(t+1)^n \leq c_n(t^n + 1).$$

Sia ora

$$f(t) = \frac{(t+1)^n}{t^n + 1}.$$

La diseguaglianza da dimostrare è  $f(t) \leq c_n$  per ogni  $t > 0$ , e la minima costante per cui questo è vero è il *massimo di  $f$  in  $(0, \infty)$* . Ci siamo così ricondotti ad un consueto problema di massimizzazione di una funzione di una variabile. Calcoliamo

$$f'(t) = \frac{n(t+1)^{n-1}(t^n + 1) - nt^{n-1}(t+1)^n}{(t^n + 1)^2} = \frac{n(t+1)^{n-1}}{(t^n + 1)^2}(1 - t^{n-1}) \geq 0 \text{ per } t \leq 1.$$

Quindi  $t = 1$  è il punto di massimo, e il massimo è

$$c_n = f(1) = 2^{n-1}.$$

#### 4.4 Il teorema di de l'Hospital

Una notevole applicazione del calcolo differenziale si ha nel calcolo dei limiti che si presentano nelle forme di indecisione  $\left[\frac{0}{0}\right]$  e  $\left[\frac{\infty}{\infty}\right]$ . Precisamente si ha:

**TEOREMA 4.8 (DI DE L'HOSPITAL)** Siano  $f, g$  funzioni derivabili in un intervallo  $(a, b)$  con  $g', g' \neq 0$  in  $(a, b)$ . Se

$$\text{i)} \lim_{x \rightarrow a^+} f(x) = \lim_{x \rightarrow a^+} g(x) = 0 \text{ oppure } \pm\infty$$

$$\text{ii)} \lim_{x \rightarrow a^+} f'(x)/g'(x) = L \in \mathbb{R}^*$$

Allora

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = L$$

Il teorema continua a valere se  $a = -\infty$  oppure se si considera il limite per  $x \rightarrow b^-$  (anziché per  $x \rightarrow a^+$ ), con  $b \leq +\infty$ .

**DIMOSTRAZIONE.** Nel caso  $f(x), g(x) \rightarrow 0$ . Daremo prima un'idea intuitiva (ma non conclucente) della dimostrazione, e poi mostreremo come la si possa rendere rigorosa. Sia  $x_n$  una successione tendente ad  $a^+$ ; prolunghiamo per continuità  $f$  e  $g$  in  $a$  ponendo  $f(a) = g(a) = 0$ . Allora

$$(4.5) \quad \frac{f(x_n)}{g(x_n)} = \frac{f(x_n) - f(a)}{g(x_n) - g(a)}.$$

Se applichiamo a  $f, g$  separatamente il teorema di Lagrange sull'intervallo  $[a, x_n]$ , otteniamo che l'ultimo quoziente scritto è uguale a:

$$\frac{f'(t_n)(x_n - a)}{g'(t_n^*)(x_n - a)} = \frac{f'(t_n)}{g'(t_n^*)}$$

dove  $t_n, t_n^*$  sono due punti opportuni che cadono nell'intervallo  $(a, x_n)$ . Poiché quando  $x_n \rightarrow 0$  anche  $t_n$  e  $t_n^* \rightarrow 0$ , sembra "ragionevole" che il limite del quoziente di  $f'/g'$  sia uguale al limite del quoziente  $f/g$ . Tuttavia questo non si può affermare rigorosamente, perché le successioni  $t_n, t_n^*$  sono a priori diverse tra loro. Per aggirare il problema occorre modificare leggermente l'argomentazione seguita. Riprendiamo dunque la dimostrazione dalla (4.5), e definiamo

$$h(x) = f(x_n)g(x) - g(x_n)f(x)$$

Notiamo che  $h(a) = h(x_n) = 0$ . La funzione  $h$  soddisfa le ipotesi del teorema di Lagrange sull'intervallo  $[a, x_n]$ , dunque esiste  $t_n \in (a, x_n)$  tale che

$$h'(t_n) = \frac{h(x_n) - h(a)}{x_n - a} = 0$$

ovvero, calcolando

$$h'(x) = f(x_n)g'(x) - g(x_n)f'(x), \quad f(x_n)g'(t_n) - g(x_n)f'(t_n) = 0$$

Dunque per ogni  $x_n$  esiste un punto  $t_n \in (a, x_n)$  tale che

$$\frac{f(x_n)}{g(x_n)} = \frac{f'(t_n)}{g'(t_n)}$$

Per  $n \rightarrow \infty$ ,  $t_n \rightarrow a^+$ , perciò  $\frac{f'(t_n)}{g'(t_n)} \rightarrow L$ , e di conseguenza anche  $\frac{f(x_n)}{g(x_n)} \rightarrow L$ , che è quanto volevamo dimostrare.

Come prima applicazione del teorema di de L'Hospital, dimostriamo il teorema 3.25 del capitolo 3, relativo alla "gerarchia degli infiniti" nel calcolo dei limiti.

Sia da calcolare

$$\lim_{x \rightarrow +\infty} \frac{x^\alpha}{e^{\beta x}} \quad \alpha > 0, \beta > 0$$

Posto  $f(x) = x$ ,  $g(x) = e^{\gamma x}$  ( $\gamma > 0$ ) si ha  $f'(x) = 1$ ,  $g'(x) = \gamma e^{\gamma x}$  e poiché

$$\lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)} = \lim_{x \rightarrow +\infty} \frac{1}{\gamma e^{\gamma x}} = 0$$

si ha che  $\lim_{x \rightarrow +\infty} \frac{x}{e^{\gamma x}} = 0$ . Osservando ora che

$$\frac{x^\alpha}{e^{\beta x}} = \left( \frac{x}{e^{(\beta/\alpha)x}} \right)^\alpha = \left( \frac{x}{e^{\gamma x}} \right)^\alpha \text{ con } \gamma = \frac{\beta}{\alpha}$$

si ha  $\lim_{x \rightarrow +\infty} \frac{x^\alpha}{e^{\beta x}} = \left( \lim_{x \rightarrow +\infty} \frac{x}{e^{\gamma x}} \right)^\alpha = 0$ .

Per provare ora che

$$\lim_{x \rightarrow +\infty} \frac{(\log_a x)^\alpha}{x^\beta} = 0$$

è sufficiente eseguire il cambio di variabile  $t = \log_a x$ ,  $x = a^t$ , e ricondursi al limite appena dimostrato.

Un'altra applicazione tipica del teorema di de L'Hospital è il calcolo di limiti che coinvolgono forme di indeterminazione non risolubili mediante i limiti notevoli (ossia, al prim'ordine).



$$\lim_{x \rightarrow 0} \frac{x - \sin x}{x^3} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

La stima al prim'ordine  $\sin x \sim x$  non è sufficiente a risolvere tale forma di indeterminazione, in quanto porta a

$$\frac{x - \sin x}{x^3} = \frac{1}{x^2} \left( 1 - \frac{\sin x}{x} \right) \rightarrow ? \quad (\infty \cdot 0)$$

(infatti la quantità tra parentesi tende a zero, per la stima asintotica).

Il teorema di de L'Hospital dà:

$$\lim_{x \rightarrow 0} \frac{(x - \sin x)'}{(x^3)'} = \lim_{x \rightarrow 0} \frac{1 - \cos x}{3x^2} = \frac{1}{6}$$

(per il limite notevole del coseno). Dunque anche il limite di partenza vale  $1/6$ .

Per usare efficacemente il teorema, è utile talvolta fare qualche passaggio preliminare (come una stima asintotica, oppure un cambio di variabile), in modo che la successiva applicazione del teorema semplifichi l'espressione, anziché complicarla. Talvolta, inoltre, il teorema va applicato più volte consecutivamente, per sciogliere la forma di indeterminazione: in questo caso, comunque, già dopo la prima applicazione ci si dovrebbe accorgere che l'ordine di infinitesimo (o di infinito) a numeratore o a denominatore si è abbassato. Se così non è, significa che non si sta seguendo la strada giusta per calcolare il limite!



$$\lim_{x \rightarrow 0} \frac{e^{-1/x^2}}{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Se applichiamo direttamente il teorema di de L'Hospital, l'espressione si complica (provare!). Invece, col cambio di variabile  $x^2 = 1/t$  ci si riconduce a

$$\lim_{t \rightarrow +\infty} \frac{\sqrt{t}}{e^t}$$

che per confronto di infiniti (cioè per il teorema di de L'Hospital!) tende a zero.

**Teorema di de L'Hospital: precauzioni per l'uso**

- Il teorema si usa per quozienti, *non per prodotti!*
- Il teorema si usa per quozienti che siano effettive *forme di indeterminazione!*
- Il teorema prescrive di calcolare il quoziente delle derivate, *non la derivata del quoziente!*
- Se il limite di  $f'/g'$  non esiste, nulla si può affermare sul limite di  $f/g$ . In particolare, *non è lecito concludere che non esiste nemmeno il limite di  $f/g$ .*



Si voglia calcolare:

$$\lim_{x \rightarrow +\infty} \frac{x - \sin x}{x + \sin x} = \begin{bmatrix} +\infty \\ +\infty \end{bmatrix}$$

Il valore del limite si può stabilire immediatamente:

$$\frac{x - \sin x}{x + \sin x} \sim \frac{x}{x} = 1$$

Supponiamo di volerlo calcolare col teorema di de L'Hospital:

$$\lim_{x \rightarrow +\infty} \frac{(x - \sin x)'}{(x + \sin x)'} = \lim_{x \rightarrow +\infty} \frac{1 - \cos x}{1 + \cos x} \text{ non esiste}$$

Se concludessimo che il limite di partenza non esiste, diremmo il falso (quel limite vale 1). Il punto è che se  $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$  non esiste (né finito né infinito), cade una delle ipotesi del teorema, e in base ad esso non si può semplicemente concludere nulla.

**4.5 Limite della derivata e derivabilità**

Supponiamo di voler studiare la derivabilità o meno di una funzione del tipo:

$$f(x) = x |\log x|.$$

La funzione è definita e continua per  $x > 0$ , ed è certamente derivabile per  $x \neq 1$ , mentre in  $x = 1$ , annullandosi l'argomento del modulo, sospettiamo la presenza di un punto angoloso. Poiché

$$f(x) = \begin{cases} x \log x & \text{per } x \geq 1 \\ -x \log x & \text{per } 0 < x \leq 1 \end{cases}$$

possiamo calcolare

$$f'(x) = \begin{cases} \log x + 1 & \text{per } x > 1 \\ -(\log x + 1) & \text{per } 0 < x < 1. \end{cases}$$

Rimane da studiare la derivabilità di  $f$  in  $x = 1$ . È facile osservare che

$$\lim_{x \rightarrow 1^+} f'(x) = \lim_{x \rightarrow 1^+} (\log x + 1) = 1 \text{ mentre}$$

$$\lim_{x \rightarrow 1^-} f'(x) = \lim_{x \rightarrow 1^-} (-\log x - 1) = -1.$$

Possiamo usare queste informazioni per affermare che la derivata destra e sinistra di  $f$ , in  $x = 1$ , valgono 1,  $-1$ , rispettivamente? (Se è così,  $x = 1$  è un punto angoloso). In altre parole, è vero ad esempio che

$$\lim_{x \rightarrow 1^+} f'(x) = f'_+(1) \text{ ossia:}$$

il limite destro della derivata per  $x \rightarrow 1$  è uguale alla derivata destra in  $x = 1$ ?

Si osservi che la cosa non è ovvia perché, per definizione, la derivata destra è il limite destro del rapporto incrementale:

$$f'_+(1) = \lim_{h \rightarrow 0^+} \frac{f(1+h) - f(1)}{h}$$

che però è molto più scomodo da calcolare, in pratica, rispetto al limite della derivata. Il teorema di Lagrange permette di rispondere affermativamente a questa domanda:

**TEOREMA 4.9** *Sia  $f : [a, b] \rightarrow \mathbb{R}$ , continua in  $a$ , derivabile in  $(a, b)$ , ed esista (finito o infinito)*

$$\lim_{x \rightarrow a^+} f'(x) = m \in \mathbb{R}^*.$$

Allora esiste

$$f'_+(a) = m.$$

*Detto in parole: se la funzione è continua in  $a$  ed esiste il limite destro della derivata, allora esiste la derivata destra, e coincide con quel limite. Un analogo enunciato vale per la derivata sinistra, e quindi per la derivata.*

**DIMOSTRAZIONE.** Sia  $h > 0$ . Applicando il teorema di Lagrange a  $f$  sull'intervallo  $[a, a+h]$ , otteniamo che esiste  $t_h \in [a, a+h]$  tale che

$$\frac{f(a+h) - f(a)}{h} = f'(t_h).$$

Per  $h \rightarrow 0^+$  si ha che  $t_h \rightarrow a^+$ , quindi per ipotesi  $f'(t_h) \rightarrow m$ . Di conseguenza esiste

$$f'_+(a) = \lim_{h \rightarrow 0^+} \frac{f(a+h) - f(a)}{h} = m.$$

Nell'esempio precedente,  $f$  è continua in  $x = 1$ , quindi il teorema è applicabile e permette di concludere che

$$f'_+(1) = 1, f'_-(1) = -1, \text{ per cui } x = 1 \text{ è un punto angoloso.}$$

Si considerino anche i prossimi esempi.



Sia

$$f(x) = x^2 \log |x|,$$

definita per  $x \neq 0$ . Sappiamo però che esiste

$$\lim_{x \rightarrow 0} x^2 \log |x| = 0,$$

perciò possiamo prolungare  $f$  per continuità definendo  $f(0) = 0$ , ed  $f$  così risulta continua anche in 0. Chiediamoci se  $f$  è anche derivabile in 0. Cominciamo a calcolare, per  $x \neq 0$ ,

$$f'(x) = 2x \log|x| + x.$$

Ora calcoliamo

$$\lim_{x \rightarrow 0} f'(x) = \lim_{x \rightarrow 0} (2x \log|x| + x) = 0.$$

Perciò per il teorema precedente esiste  $f'(0) = 0$ , in particolare  $f$  è derivabile in  $x = 0$ .

Sia

$$f(x) = \operatorname{arctg} \frac{1}{x}.$$

Calcoliamo

$$f'(x) = \frac{1}{1 + \frac{1}{x^2}} \cdot \left( -\frac{1}{x^2} \right) = -\frac{1}{1 + x^2}.$$

Osserviamo che

$$\lim_{x \rightarrow 0} f'(x) = \lim_{x \rightarrow 0} \left( -\frac{1}{1 + x^2} \right) = -1.$$

Si può concludere allora che esiste  $f'(x) = -1$ ? No! In questo caso viene meno l'ipotesi di continuità di  $f$  in 0: si osservi che  $f$  non è definita per  $x = 0$ , e

$$\lim_{x \rightarrow 0^\pm} f(x) = \pm \frac{\pi}{2}$$

dunque  $x = 0$  è un punto di discontinuità a salto, ed  $f$  non è derivabile in  $x = 0$ . Si osservi dunque che l'ipotesi di continuità, nel teorema precedente, è essenziale.

Sia

$$f(x) = x^2 \sin \left( \frac{1}{x} \right).$$

La funzione non è definita in  $x = 0$ , ma si può prolungarla per continuità ponendo  $f(0) = 0$  (infatti,  $f(x) \rightarrow 0$  per  $x \rightarrow 0$ ). Dunque ora  $f$  è continua in 0; per  $x \neq 0$ , inoltre,  $f$  è certamente derivabile e si ha:

$$f'(x) = 2x \sin \left( \frac{1}{x} \right) + x^2 \cos \left( \frac{1}{x} \right) \cdot \left( -\frac{1}{x^2} \right) = 2x \sin \left( \frac{1}{x} \right) - \cos \left( \frac{1}{x} \right).$$

Osserviamo ora che

$$\lim_{x \rightarrow 0} f'(x) \text{ non esiste}$$

(né finito né infinito). Possiamo concludere che non esiste  $f'(0)$ ? No! Si noti che il teorema non permette questa conclusione, ed in effetti, in questo caso, calcolando  $f'(0)$  mediante la definizione si ha:

$$f'(0) = \lim_{h \rightarrow 0} \frac{f(h) - 0}{h} = \lim_{h \rightarrow 0} h \sin \left( \frac{1}{h} \right) = 0,$$

perciò in realtà  $f$  è derivabile e  $f'(0) = 0$ . Quando il limite della derivata non esiste (né finito né infinito), il teorema non è applicabile; in particolare non è detto che non esista la derivata nel punto considerato: occorre calcolarla dalla definizione.

### Esercizi

Dopo aver stabilito l'insieme di definizione delle seguenti funzioni, determinarne i punti di massimo e minimo e tracciarne sommariamente il grafico.

$x^2 e^x$

$x^3 + x^2 - x + 1$

$\sqrt[3]{x} e^x$

$x^4 - 8x^3 + 22x^2 - 24x + 12$

$x^2 \log x$

$x^5 - 5x^4 + 5x^3 + 1$

$\sqrt{x} \log x$

$\frac{x+2}{x^2+1}$

### Problemi di massimo e minimo.

Impostare e risolvere i seguenti problemi di massimo o minimo, cercando di imitare le tecniche illustrate negli esempi del paragrafo 4.3. In particolare, impostare il problema in modo da ridurlo ad un problema di massimizzazione o minimizzazione di una funzione di una variabile.

Una ditta produttrice di birra desidera minimizzare il costo della lattina. Essendo di materiale omogeneo e volume fissato (33 cl) occorre minimizzare la superficie totale del cilindro di volume pari a 33 cl. Quali sono le dimensioni (altezza e diametro) della lattina?

Un uomo deve raggiungere un punto che si trova sull'altra sponda di un fiume, 100 metri più a valle; il fiume è rettilineo e largo 10 metri; l'uomo può correre sulla sponda del fiume con velocità  $v$ , quindi tuffarsi e attraversare a nuoto il fiume, con velocità inferiore, pari a  $\delta v$  ( $0 < \delta < 1$ ).

Determinare dopo quanti metri di corsa l'uomo si deve tuffare, affinché sia minimo il tempo impiegato a raggiungere la meta.

Se l'uomo è un nuotatore provetto,  $\delta$  sarà quasi uguale a 1: determinare il valore esatto di  $\delta$  per il quale all'uomo conviene tuffarsi immediatamente, senza percorrere neanche un metro sulla terraferma.

Si vuole costruire una scatola, senza coperchio, col vincolo che la base sia quadrata e la superficie totale della scatola misuri 108 cm. Di quali dimensioni (lato della base e altezza) dev'essere affinché il volume sia massimo possibile? E quanto sarà il volume?

Si determini l'altezza e il raggio di base del cono di volume minimo circoscritto ad una sfera di raggio  $r$ . Si dimostri poi che il suddetto cono è anche quello di superficie minima totale.

Sia  $n$  un intero  $\geq 2$ . Determinare (in dipendenza da  $n$ ) le migliori costanti  $c_1, c_2$  per cui si abbia:

$$c_1(a+b) \leq \left( \sqrt[n]{a} + \sqrt[n]{b} \right)^n \leq c_2(a+b) \quad \forall a, b \in \mathbb{R}.$$

(Dire che le costanti sono le migliori significa che  $c_1$  è la massima e  $c_2$  la minima costante che rendono vere le diseguaglianze precedenti). Suggerimento: imitare il procedimento seguito nell'ultimo esempio del paragrafo 4.3.

Monotonia di successioni.

Si stabilisca se le seguenti successioni  $a_n$  sono monotone (crescenti o decrescenti), almeno per  $n \geq n_0$  opportuno:

$$\begin{aligned} & \frac{n+3}{n^2-2n+4}; \quad \frac{n^2}{\log n}; \quad \frac{n!}{2^n}; \\ & n - \sqrt{n^2 + 1}; \quad ne^{-n}; \quad n \sin n. \end{aligned}$$

Calcolare i seguenti limiti utilizzando opportunamente, se occorre, il teorema di de L'Hospital.

1)  $\lim_{x \rightarrow 1} \frac{\cos(\pi/2x)}{\sin(\pi x)}$

2)  $\lim_{x \rightarrow +\infty} \frac{x}{2^{x^2}}$

3)  $\lim_{x \rightarrow 0} \frac{e^x - x - 1}{x^2}$

4)  $\lim_{x \rightarrow 2} \frac{\sqrt{x^2 - 4}}{\sin(x-2)}$

5)  $\lim_{x \rightarrow +\infty} x^{1/x}$

6)  $\lim_{x \rightarrow 2} \frac{x^2 - 4}{\sqrt{\sin(x-2)^2}}$

7)  $\lim_{x \rightarrow 0} \frac{x^2 \cos x + 1 - e^{x^2}}{x^2 \sin^2 x}$

8)  $\lim_{x \rightarrow 0} \frac{x - \sin x}{x^2 \sin x}$

9)  $\lim_{x \rightarrow 0} \frac{x \sin x + \log(1-x^2)}{x^2 (2x+x^2)^2}$

10)  $\lim_{x \rightarrow 1} \frac{x \log x}{x^2 - 1}$

11)  $\lim_{x \rightarrow +\infty} x \left( \arctg x - \frac{\pi x + 2}{2x+1} \right)$

12)  $\lim_{x \rightarrow 1} \frac{x^2 - 3x + 3}{\cos \frac{\pi x}{2}}$

13) Sia  $f(x) = \frac{\sin x}{x}$ . Dopo aver prolungato per continuità  $f$  in  $x = 0$ , calcolare  $f'(0)$  e  $f''(0)$ . (Suggerimento: Nei due limiti che si chiede di calcolare, è utile il teorema di de L'Hospital).

14) Come l'esercizio precedente, per la funzione  $\left(\frac{\sin x}{x}\right)^2$ .

15) Il teorema di de L'Hospital a volte può produrre... il moto perpetuo. Si provi a calcolare il seguente limite (elementare!) applicando più volte il teorema:

$$\lim_{x \rightarrow +\infty} \frac{\sqrt{x^2 + 1}}{x}$$

16) Stabilire se le seguenti funzioni sono continue e se sono derivabili, nel punto  $x = 0$ :

$$f_1(x) = \begin{cases} x^2 \log x & \text{per } x > 0 \\ e^x - 1 & \text{per } x \leq 0 \end{cases}$$

$$f_2(x) = x|x|$$

$$f_3(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{per } x > 0 \\ 0 & \text{per } x = 0 \\ e^{1/x} & \text{per } x < 0 \end{cases}$$

Suggerimento: utilizzare la definizione di derivata oppure il criterio visto nel par. 4.5, scegliendo caso per caso il metodo opportuno.

## 5 DERIVATA SECONDA

La derivata seconda di una funzione ha vari significati geometrici, che ci permetteranno di meglio studiare le proprietà del grafico di una funzione.

### 5.1 Significato geometrico della derivata seconda

Se la derivata prima ha come significato geometrico quello di pendenza del grafico, la derivata seconda rappresenta la velocità di variazione di tale pendenza e pertanto costituisce una misura del grado di scostamento del grafico dall'andamento rettilineo.

Per precisare meglio il concetto, consideriamo la situazione illustrata in figura 4.35.

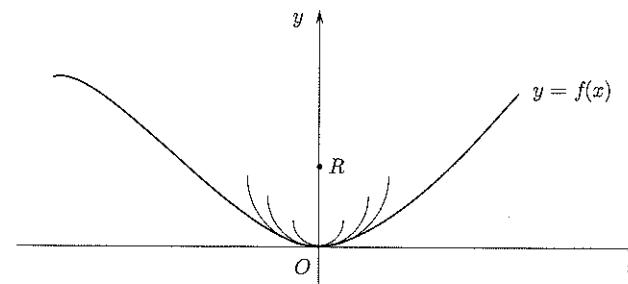


Figura 4.35.

La funzione  $f$  in figura, soddisfa le condizioni  $f(0) = f'(0) = 0$ ,  $f''(0) \geq 0$ .

Consideriamo ora la famiglia di semicirconferenze con centro sull'asse  $y$  tangente al grafico di  $f$  nell'origine.

L'equazione della famiglia è

$$y = R - \sqrt{R^2 - x^2}$$

come facilmente si verifica e, per ogni  $R$ , si ha  $y'(0) = y(0) = 0$ .

Tra queste semicirconferenze selezioniamo quella che non soltanto ha la stessa tangente in  $x = 0$  ma anche la stessa velocità di variazione della pendenza in  $x = 0$ .

In altri termini, scegliamo  $R$  in modo che

$$(5.1) \quad y''(0) = f''(0).$$

Si ha,

$$y'(x) = \frac{x}{\sqrt{R^2 - x^2}}, \quad y''(x) = \frac{R^2}{(R^2 - x^2)^{3/2}}$$

e quindi  $y''(0) = \frac{1}{R}$ . Se dunque vogliamo che sia soddisfatta la (5.1) occorre scegliere  $R$  in modo che

$$(5.2) \quad f''(0) = \frac{1}{R}.$$

La (5.2) esprime il significato geometrico della derivata seconda: nella situazione descritta,  $f''(0)$  rappresenta il reciproco del raggio della circonferenza che "meglio approssima"  $f$  in  $x = 0$ ;  $\frac{1}{R}$  prende il nome di *curvatura* (del grafico) di  $f$  in  $x = 0$  e

$R$  è il *raggio di curvatura*. Ripetendo un'analogia costruzione in un punto generico  $x$  del grafico di  $f$  (senza supporre a priori  $f(x) = f'(x) = 0$ ) si trova che il cerchio che meglio approssima il grafico nel punto di ascissa  $x$  ha raggio  $R(x)$  dato da:

$$\frac{1}{R(x)} = \frac{|f''(x)|}{(1 + f'(x)^2)^{3/2}}$$



Per una parabola del tipo  $y = ax^2$  si ha  $y'(x) = 2ax$ ,  $y''(x) = 2a$  e quindi

$$\frac{1}{R(x)} = \frac{2|a|}{(1 + 4a^2x^2)^{3/2}}.$$

Si noti che la curvatura è massima per  $x = 0$ , cioè nel vertice.

## 5.2 Derivata seconda, concavità e convessità

Abbiamo visto che il significato geometrico della derivata seconda è legato al concetto di *curvatura* della curva grafico della funzione. Vedremo ora come rendere più preciso, qualitativamente e quantitativamente, questa idea di curvatura del grafico attraverso il concetto di *convessità*. Si tratta di una nozione che in matematica ha una portata molto vasta, di cui qui potremo solo toccare i primi aspetti.

### Convessità e corde

Cominciamo col ricordare che, in geometria, si dice che un *insieme*  $F$  (pensiamo ad esempio a un sottoinsieme del piano) è *convesso* se, per ogni coppia di punti  $P_1, P_2 \in F$  avviene che il segmento che congiunge  $P_1$  a  $P_2$  è interamente contenuto in  $F$ :

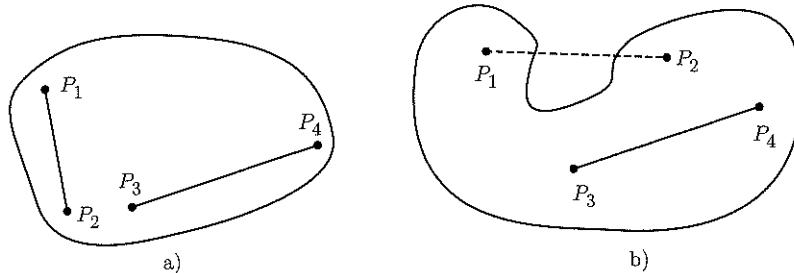


Figura 4.36. a) un insieme convesso; b) un insieme non convesso.

Coerentemente a questa idea, diamo allora la seguente

**DEFINIZIONE 4.6** Consideriamo una funzione  $f : I \rightarrow \mathbb{R}$  (dove  $I$  è un intervallo). Si chiama *epigrafico* (o sopagrafico) di  $f$  l'insieme

$$\text{epif} = \{(x, y) \in \mathbb{R}^2 : x \in I \text{ e } y \geq f(x)\}.$$

Si dice allora che  $f$  è convessa in  $I$  se il suo epigrafico è un insieme convesso. Si dice che  $f$  è concava in  $I$  se  $-f$  è convessa in  $I$ .

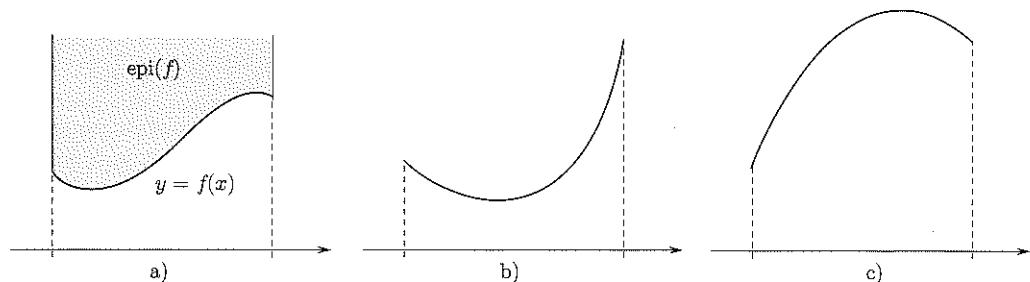


Figura 4.37. a) epigrafico; b) funzione convessa; c) funzione concava.

Segnaliamo subito che esiste una seconda terminologia, alternativa, che si usa talvolta per esprimere questi concetti: anziché  $f$  convessa in  $(a, b)$  si può dire  $f$  *concava verso l'alto* in  $(a, b)$ , e invece di concava in  $(a, b)$  si può dire *concava verso il basso* in  $(a, b)$ .

Per non creare confusione, in questo paragrafo non useremo mai questa seconda terminologia.

Si dimostra facilmente che la definizione precedente è equivalente alla prossima, che fa riferimento al grafico anziché all'epigrafico della funzione:

**DEFINIZIONE 4.7** Sia  $f : I \rightarrow \mathbb{R}$  dove  $I$  è un intervallo. Si dice che  $f$  è convessa (concava) in  $I$  se per ogni coppia di punti  $x_1, x_2 \in I$  si ha che il segmento ("corda") di estremi  $(x_1, f(x_1)), (x_2, f(x_2))$  non ha punti sotto (sopra) il grafico di  $f$ .

Questa definizione può essere riformulata anche mediante una diseguaglianza analitica. Con riferimento alla figura 4.38, esprimiamo il generico punto  $(z_t, f_t)$  del segmento congiungente i punti  $(x_1, f(x_1)), (x_2, f(x_2))$  come *combinazione lineare convessa* dei suoi estremi:

$$z_t = (1 - t)x_1 + tx_2; \quad f_t = (1 - t)f(x_1) + tf(x_2), \quad t \in [0, 1].$$

Si osservi che al variare di  $t \in [0, 1]$  il punto  $z_t$  percorre il segmento  $[x_1, x_2]$  sull'asse  $x$ , e il punto del piano  $(z_t, f_t)$  percorre il segmento di estremi  $(x_1, f(x_1)), (x_2, f(x_2))$ . La seconda definizione di convessità che abbiamo dato si esprime dunque così:

$$f(z_t) \leq f_t \quad \forall x_1, x_2 \in I, \forall t \in [0, 1]$$

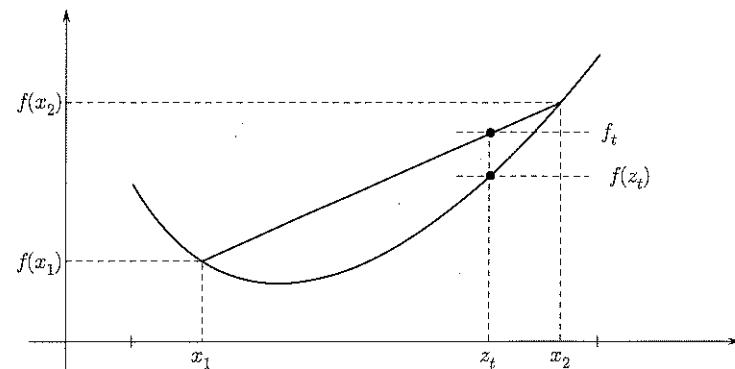


Figura 4.38.

o, più esplicitamente

$$f((1-t)x_1 + tx_2) \leq (1-t)f(x_1) + tf(x_2) \quad \forall x_1, x_2 \in I, \forall t \in [0, 1].$$

Questa diseguaglianza esprime il modo in cui viene spesso sfruttata l'informazione che una funzione è convessa. Se nella precedente diseguaglianza vale sempre il  $<$  la funzione si dice *strettamente convessa*. Per le funzioni concave (o strettamente concave) varrà una analoga diseguaglianza col segno  $\geq$  ( $\geq$  o  $>$ ).

Si noti che la definizione di funzione convessa non richiede a priori che  $f$  sia una funzione continua o derivabile, quindi è molto generale:

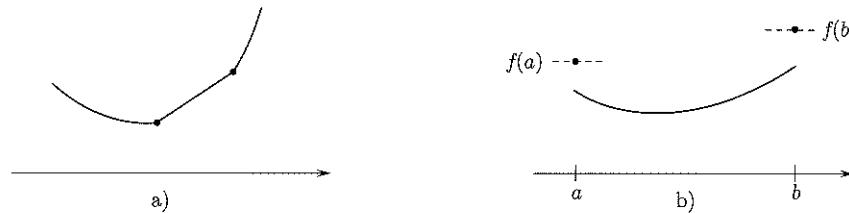


Figura 4.39. a) funzione convessa con due punti angolosi; b) funzione convessa con due discontinuità agli estremi di un intervallo.

Si può dimostrare che le situazioni esemplificate dalla figura precedente sono effettivamente “le peggiori possibili”, nel senso che:

**TEOREMA 4.10** Una funzione convessa (o concava) su un intervallo  $I$  è continua in  $I$ , salvo al più negli estremi dell’intervallo. Inoltre possiede derivata destra e sinistra in ogni punto interno dell’intervallo.

Dunque, punti angolosi all’interno e punti di discontinuità agli estremi sono i soli comportamenti irregolari permessi ad una funzione convessa o concava.

#### Convessità e derivate

Abbiamo visto che la definizione di convessità ha senso anche per funzioni non derivabili. Ad ogni modo, se sappiamo a priori che la funzione è derivabile una volta o due volte nell’intervallo considerato, allora la convessità risulta avere relazioni interessanti con la derivata prima e seconda della funzione:

**TEOREMA 4.11** Sia  $f : (a, b) \rightarrow \mathbb{R}$ .

- a) Se  $f$  è derivabile in  $(a, b)$ , allora  $f$  è convessa (concava) in  $(a, b)$  se e solo se  $f'$  è crescente (decrescente) in  $(a, b)$ .
- b) Se  $f$  è derivabile due volte in  $(a, b)$ , allora  $f$  è convessa (concava) in  $(a, b)$  se e solo se  $f''(x) \geq 0$  ( $\leq 0$ ) per ogni  $x \in (a, b)$ .

Non dimostriamo il punto (a); il punto (b) segue da (a) per il test di monotonia, applicato ad  $f'$ .

Come conseguenza di questo teorema, lo studio del segno della derivata seconda ci permette di decidere della convessità o concavità di una funzione, aggiungendo quindi un elemento utile allo studio del grafico della funzione.

Per le funzioni strettamente convesse o concave, il punto (a) del teorema si può enunciare in modo analogo così:

a') Se  $f$  è derivabile in  $(a, b)$ , allora  $f$  è strettamente convessa (concava) in  $(a, b)$  se e solo se  $f'$  è strettamente crescente (decrescente) in  $(a, b)$ .

Invece, per quanto riguarda il punto (b) si può dimostrare solo la seguente implicazione:

b') Se  $f$  è derivabile due volte in  $(a, b)$  e  $f''(x) > 0$  ( $< 0$ ) per ogni  $x \in (a, b)$ , allora  $f$  è strettamente convessa (concava).

Il viceversa però non vale: ad esempio,  $f(x) = x^4$  è strettamente convessa in  $\mathbb{R}$  ma  $f''(0) = 0$ .



Le funzioni esponenziali  $f(x) = a^x$  sono strettamente convesse in  $\mathbb{R}$ , per qualunque base  $a \leq 1$ . Infatti:

$$f'(x) = a^x \log a; \quad f''(x) = a^x (\log a)^2 > 0 \quad \forall x \in \mathbb{R}, \quad \forall a > 0, a \neq 1.$$



Le funzioni logaritmiche  $f(x) = \log_a x$  sono strettamente concave in  $(0, \infty)$  se  $a > 1$ , strettamente convesse in  $(0, \infty)$  se  $0 < a < 1$ . Infatti:

$$f'(x) = \frac{1}{x \log a}; \quad f''(x) = -\frac{1}{x^2 \log a} \begin{cases} < 0 & \forall x > 0, & \text{se } a > 1 \\ > 0 & \forall x > 0, & \text{se } 0 < a < 1. \end{cases}$$



Le funzioni potenza a esponente reale  $\alpha$ ,  $f(x) = x^\alpha$  sono strettamente convesse in  $(0, \infty)$  se  $\alpha > 1$  oppure  $\alpha < 0$ , sono strettamente concave in  $(0, \infty)$  se  $0 < \alpha < 1$ . Infatti:

$$f'(x) = \alpha x^{\alpha-1}; \quad f''(x) = \alpha(\alpha-1)x^{\alpha-2}$$

che ha, per ogni  $x > 0$ , il segno di  $\alpha(\alpha-1)$ .

Lo studente si renda conto del significato geometrico di queste affermazioni osservando i grafici delle funzioni ora citate.

#### Convessità e rette tangenti

Un’utile caratterizzazione geometrica della convessità, su cui ritorneremo in seguito (par. 7.4) coinvolge le *rette tangenti* al grafico della funzione:

**TEOREMA 4.12** Una funzione  $f : (a, b) \rightarrow \mathbb{R}$ , derivabile in  $(a, b)$ , è strettamente convessa (concava) in  $(a, b)$  se e solo se comunque si scelga un punto  $x_0 \in (a, b)$  si ha che il grafico di  $f$  si mantiene in tutto  $(a, b)$  sopra (sotto) il grafico della sua retta tangente in  $(x_0, f(x_0))$ , tranne che in quel punto.



Sappiamo che  $f(x) = e^x$  è convessa in tutto  $\mathbb{R}$ . Ne segue che se, ad esempio, costruiamo la retta tangente al grafico di  $f$  in  $x = 0$ , ossia  $y = x + 1$ , risulta  $e^x \geq 1 + x \forall x \in \mathbb{R}$ . La retta tangente al grafico di  $f$  in  $x = 1$ , invece, è  $y = ex$ . Di conseguenza risulta anche  $e^x \geq ex \forall x \in \mathbb{R}$ , e così via.

È utile talvolta sapere che una funzione (derivabile) convessa sta *sopra* le proprie rette tangenti e *contemporaneamente* sta *sotto* le proprie corde (definizione di convessità). Questo permette di concludere che, presi due punti qualsiasi sul grafico di una funzione convessa, il grafico tra quei due punti cade tutto nel triangolo che ha per lati la corda che li unisce e le rette tangenti al grafico nei due punti (fig. 4.40).

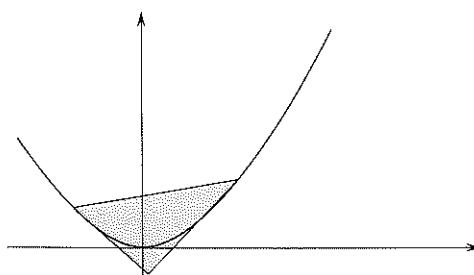


Figura 4.40.

Questa limitazione può essere utile, ad esempio, quando si cerca di localizzare l'intersezione tra due curve (si veda l'esercizio 106).

#### Punti di flesso

Ovviamente, il *verso della concavità* di una funzione (ossia il fatto che sia convessa o concava) può cambiare, nel suo insieme di definizione; questo ci conduce al concetto di *punto di flesso*:

**DEFINIZIONE 4.8** Sia  $f : (a, b) \rightarrow \mathbb{R}$  una funzione e  $x_0 \in (a, b)$  sia un punto di derivabilità per  $f$ , oppure sia  $f'(x_0) = \pm\infty$ . Il punto  $x_0$  si dice di *flesso* per  $f$  se esiste un intorno destro  $(x_0, x_0 + h)$ ,  $h > 0$ , in cui  $f$  è strettamente convessa (concava), e un intorno sinistro  $(x_0 - h, x_0)$  in cui è strettamente concava (convessa).

Attraversando un punto di flesso, la derivata seconda di  $f$  (se esiste) cambia segno. Ci aspettiamo allora che in questo punto  $f''$  si annulli. Si può infatti dimostrare che:

**TEOREMA 4.13** Sia  $x_0$  un punto di flesso per  $f$ ; se esiste  $f''(x_0)$ , allora  $f''(x_0) = 0$ .

Notiamo che, se sapessimo che  $f''$  esiste in un intorno di  $x_0$  ed è continua in  $x_0$ , allora la tesi del teorema seguirebbe dal teorema dei valori intermedi per le funzioni continue (applicato ad  $f''$ ). Ad ogni modo, si può dimostrare il teorema appena enunciato anche senza queste ipotesi ulteriori.

L'implicazione opposta a quella enunciata dal teorema non è vera, come mostra l'esempio 5.8: un punto in cui la derivata seconda si annulla può non essere di flesso.

Il significato geometrico dei punti di flesso è ulteriormente chiarito dal prossimo:

**TEOREMA 4.14** Se  $f : (a, b) \rightarrow \mathbb{R}$  è derivabile in  $(a, b)$  e  $x_0 \in (a, b)$  è un punto di flesso, allora il grafico di  $f(x)$  attraversa la propria retta tangente in  $(x_0, f(x_0))$  (si veda la fig. 4.41):

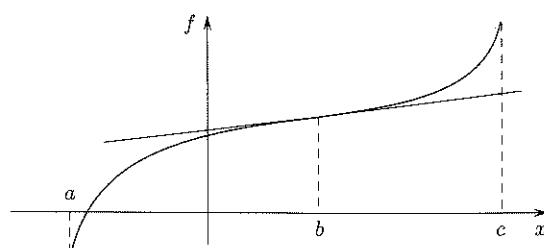


Figura 4.41.

**DIMOSTRAZIONE.** Tracciamo la retta tangente al grafico di  $f(x)$  nel punto di ascissa  $x_0$ . Se  $f$  è (ad esempio) concava in  $(a, x_0)$ , poiché  $f$  è derivabile, il grafico di  $f$  sta sotto la retta in  $(a, x_0)$  (si veda il teorema su convessità e rette tangenti); d'altra parte  $f$  è convessa in  $(x_0, b)$ , perciò il suo grafico sta sopra la retta in  $(x_0, b)$ . Di conseguenza in  $x_0$  il grafico attraversa la retta tangente.

#### ESEMPIO

Sia  $f(x) = x^3$ . Poiché  $f'(x) = 3x^2 \geq 0$ , la funzione è crescente su tutto  $\mathbb{R}$ . Ha un punto stazionario per  $x = 0$ , che non sarà però punto di massimo o minimo, perché la funzione è crescente.  $f''(x) = 6x \geq 0$  per  $x \geq 0$ . Dunque la funzione è convessa per  $x > 0$ , concava per  $x < 0$  e ha un punto di flesso in  $x = 0$ , a tangente orizzontale.

Sia  $f(x) = e^{-x^2}$ . Poiché  $f'(x) = -2xe^{-x^2} \geq 0$  per  $x \leq 0$ , la funzione cresce per  $x < 0$ , decresce per  $x > 0$  perciò ha un punto di massimo relativo in  $x = 0$ .

$f''(x) = e^{-x^2}(4x^2 - 2) \geq 0$  per  $x^2 \geq \frac{1}{2}$ ;  $x \geq \frac{1}{\sqrt{2}}$ ;  $x \leq -\frac{1}{\sqrt{2}}$ . La funzione è convessa per questi valori, concava per  $-\frac{1}{\sqrt{2}} \leq x \leq \frac{1}{\sqrt{2}}$ , ha punti di flesso in  $x = \pm \frac{1}{\sqrt{2}}$ , con retta tangente di pendenza  $f'(\pm \frac{1}{\sqrt{2}}) = \mp \sqrt{\frac{2}{e}}$  (v. fig. 4.42).

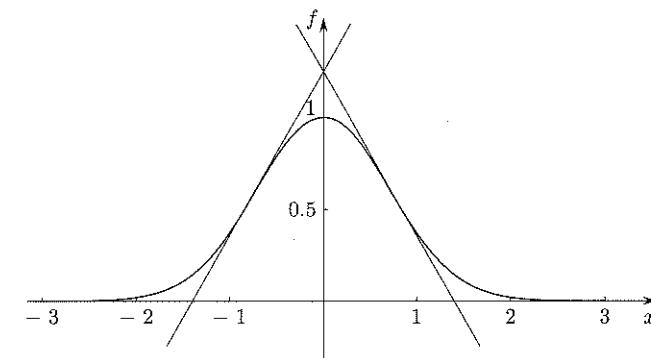


Figura 4.42.

Sia  $f(x) = x^4$ . Poiché  $f'(x) = 4x^3 \geq 0$  per  $x > 0$ , la funzione è crescente per  $x > 0$ , decrescente per  $x < 0$  e ha un punto di minimo in  $x = 0$ .  $f''(x) = 12x^2 \geq 0$  per ogni  $x$ , quindi la funzione è convessa in tutto  $\mathbb{R}$ . Perciò il punto  $x = 0$ , in cui  $f''$  si annulla, non è un punto di flesso.

Sia  $f(x) = \sqrt[3]{x}$ . Per  $x > 0$  la funzione è concava, per  $x < 0$  la funzione è convessa (come si deduce dal segno della derivata seconda); in  $x = 0$  è  $f'(0) = +\infty$ . In base alla nostra definizione,  $x = 0$  è un punto di flesso (a tangente verticale). Questo è anche coerente alla terminologia che abbiamo introdotto discutendo i punti di non derivabilità, nel par. 2.4. Si vede che anche in questo caso il grafico attraversa la tangente nel punto di flesso; si noti che  $f''(0)$  non esiste (v. fig. 4.43).

Sia  $f(x) = x|x|$ . Per  $x > 0$  è  $f(x) = x^2$ , convessa; per  $x < 0$  è  $f(x) = -x^2$ , concava; si noti che esiste  $f'(0) = 0$ ; dunque  $x = 0$  è un punto di flesso. Tuttavia,  $f''(0)$  non esiste.

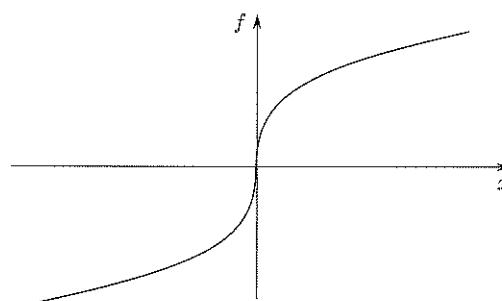


Figura 4.43.

## 6 STUDIO DEL GRAFICO DI UNA FUNZIONE

Il calcolo differenziale sviluppato fin qui ci permette ora di affrontare in modo completo un tipico problema, quello di tracciare il grafico di una funzione  $f$  di una variabile. Anche se l'avvento dei mezzi elettronici di calcolo ha reso possibile la costruzione automatica dei grafici di funzioni, riteniamo ugualmente utile che lo studente sappia destreggiarsi manualmente in questo compito, armato unicamente di carta e matita. A questo proposito, daremo ora qualche regola generale su come procedere, e illustreremo le più comuni situazioni che si possono verificare, attraverso esempi ed esercizi.

*Alcuni passi obbligati:*

- Determinare l'insieme su cui  $f$  è definita, ossia il dominio di  $f$ ,  $\text{Dom}(f)$ .
- Quindi, calcolare i limiti (o limiti destri, sinistri) alla frontiera di tale insieme (compresi, eventualmente,  $\pm\infty$ ). In questa fase si determinano gli eventuali asintoti orizzontali e verticali e alcuni eventuali punti di discontinuità.
- Se per  $x \rightarrow \pm\infty$  la funzione tende a  $\pm\infty$ , occorre chiedersi se la funzione presenta un asintoto obliquo. Più in generale, può essere utile una *stima asintotica* che ci dica se la funzione tende a  $\infty$  in modo sopra o sottolineare (e in tal caso certamente *non ha* asintoto obliquo) o in modo lineare (e in tal caso può avere o non avere asintoto obliquo). La stima asintotica all'infinito dà normalmente anche un'informazione sulla *concavità* di  $f$  all'infinito.
- Calcolare  $f'(x)$ , nei punti in cui esiste. Studiare accuratamente i punti in cui  $f$  è continua ma non derivabile e stabilirne la natura (punti angolosi, di flesso a tangente verticale, di cuspide). In punti angolosi o agli estremi del dominio è utile il calcolo di derivate destre o sinistre che determinano la pendenza del grafico in quei punti.
- Studiare quindi il segno di  $f'$  (sul suo insieme di definizione), per ottenere le informazioni sulla monotonia di  $f$  e sui suoi punti di massimo e minimo relativi.

Vi sono altre indagini che *in certi casi* possono essere utili:

- Osservare (quando c'è!) un'eventuale simmetria di  $f$  (pari o dispari) e restringere quindi lo studio a  $x \geq 0$ .

- Osservare (quando c'è!) l'eventuale periodicità di  $f$  e restringere quindi lo studio a un periodo.
- Determinare gli zeri della funzione e studiarne il segno. Si tenga presente che *solo nei casi particolarmente semplici* è possibile fare questo a priori; in generale, queste informazioni si possono dedurre solo al termine dello studio, come conseguenza delle altre informazioni.
- Calcolare la derivata seconda e studiarne il segno, per dedurne informazioni sulla concavità e i flessi di  $f$ . Si tenga presente che in molti casi l'espressione di  $f''$  è troppo complessa perché se ne riescano a determinare esattamente gli zeri. In questo caso, il calcolo di  $f''$  non aggiunge niente a ciò che un utilizzo attento di limiti, stime asintotiche e derivata prima permette già di affermare.

Un'ultima avvertenza: è utile tracciare il grafico gradualmente, inserendo le informazioni via via raccolte (eventualmente correggendosi), anziché raccogliere tutte le informazioni e poi fare il grafico: i progressi graduali aiutano a controllare la coerenza del procedimento e a capire quali informazioni è ancora utile raccogliere.



$$f(x) = e^{-|x|} \sqrt{x^2 - 5x + 6}$$

Ci aspettiamo: un punto angoloso in  $x = 0$ , per la presenza di  $|x|$ ; punti a tangente verticale, dove si annulla il radicando.

Insieme di definizione:  $x^2 - 5x + 6 \geq 0; (x-3)(x-2) \geq 0; x \leq 2; x \geq 3$ .

$$\lim_{x \rightarrow \pm\infty} f(x) = 0$$

Perciò  $y = 0$  è asintoto orizzontale per  $x \rightarrow \pm\infty$ .

$$f(x) \geq 0 \text{ per ogni } x. f(x) = 0 \text{ per } x = 2, 3.$$

Per  $x \rightarrow 3^+$ ,  $f(x) \sim e^{-3} \sqrt{x-3} \rightarrow 0$  e ha tangente verticale in  $x = 3$ .

Per  $x \rightarrow 2^-$ ,  $f(x) \sim e^{-2} \sqrt{(2-x)} \rightarrow 0$  e ha tangente verticale in  $x = 2$ .

$$\begin{aligned} f'(x) &= e^{-|x|} \left( -\text{sgn}(x) \sqrt{x^2 - 5x + 6} + \frac{1}{2\sqrt{x^2 - 5x + 6}} (2x - 5) \right) = \\ &= \frac{e^{-|x|}}{2\sqrt{x^2 - 5x + 6}} (-2\text{sgn}(x)(x^2 - 5x + 6) + 2x - 5) = \\ &= \begin{cases} \frac{e^{-x}}{2\sqrt{x^2 - 5x + 6}} (-2x^2 + 12x - 17) & \text{se } x > 0 \\ \frac{e^x}{2\sqrt{x^2 - 5x + 6}} (2x^2 - 3x + 1) & \text{se } x < 0 \end{cases} \end{aligned}$$

Per derivare la funzione contenente il modulo, abbiamo usato la scrittura compatta  $|x'| = \text{sgn}(x)$  per  $x \neq 0$ ; solo al termine del calcolo, per rendere più leggibile l'espressione trovata, l'abbiamo riscritta esplicitamente per  $x > 0$  e  $x < 0$ .

Calcoliamo:

$$\lim_{x \rightarrow 0^\pm} f'(x) = \begin{cases} -\frac{17}{2\sqrt{6}} \\ \frac{1}{2\sqrt{6}} \end{cases}$$

$x = 0$  è punto angoloso per  $f$  ( $f$  non è derivabile e le tangenti, da destra e sinistra, hanno pendenze diverse).

$$\text{Per } x > 0 \quad f'(x) \geq 0 \text{ se } 2x^2 - 12x + 17 \leq 0 \quad \frac{6-\sqrt{2}}{2} \leq x \leq \frac{6+\sqrt{2}}{2}$$

$$\frac{6-\sqrt{2}}{2} \simeq 2,29; \quad \frac{6+\sqrt{2}}{2} \simeq 3,7, \text{ quindi}$$

per  $0 \leq x \leq 2$ ,  $f(x)$  è decrescente;

per  $2 \leq x \leq \frac{6+\sqrt{2}}{2}$ ,  $f(x)$  è sempre crescente;

per  $x \geq \frac{6+\sqrt{2}}{2}$ ,  $f(x)$  è sempre decrescente.

$x = \frac{6+\sqrt{2}}{2}$  punto di massimo relativo. Il massimo è circa  $M = f(3,7) = 0,027$ .

Per  $x < 0$ ,  $f'(x) \geq 0$  se  $2x^2 - 3x + 1 \geq 0$ ,  $x \leq \frac{1}{2}$ ,  $x \geq 1$ , quindi:

per  $x < 0$ ,  $f(x)$  è sempre crescente.

Si vede anche che  $x = 0$  (in cui  $f$  non è derivabile) è punto di massimo; il massimo (assoluto) è  $M = \sqrt{6}$ . Il grafico qualitativo è in figura 4.44. Ci sono anche due punti di flesso, uno in  $(0, 2)$  e uno in  $(3, +\infty)$ . Si omette il calcolo di  $f''$ .

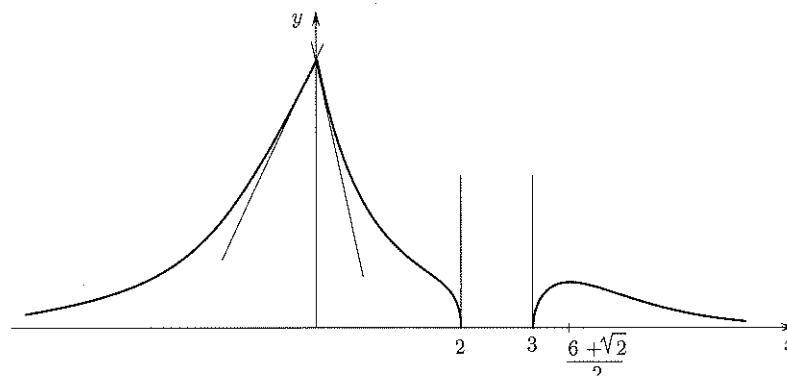


Figura 4.44.



$$f(x) = xe^{\frac{x+2}{x-1}}$$

Insieme di definizione:  $x \neq 1$ .  $f(x) \geq 0$  per  $x \geq 0$

$$\text{Per } x \rightarrow 1^\pm, \frac{x+2}{x-1} \rightarrow \pm\infty, f(x) \rightarrow \begin{cases} +\infty \\ 0^+ \end{cases}$$

$x = 1$  asintoto verticale per  $x \rightarrow 1^+$ . Per  $x \rightarrow \pm\infty$ ,  $\frac{x+2}{x-1} \rightarrow 1$ ,  $f(x) \sim xe$ , perciò  $f(x) \rightarrow \pm\infty$  linearmente. Vediamo se c'è asintoto obliquo:

$$\lim_{x \rightarrow \pm\infty} [f(x) - xe] = \lim_{x \rightarrow \pm\infty} xe \left[ e^{\frac{x+2}{x-1}-1} - 1 \right]$$

Ora,  $\frac{x+2}{x-1} - 1 = \frac{3}{x-1} \rightarrow 0$ , perciò

$$e^{\frac{x+2}{x-1}-1} - 1 \sim \frac{3}{x-1} \quad xe \left[ e^{\frac{x+2}{x-1}-1} - 1 \right] \sim \frac{3xe}{x-1} \rightarrow 3e$$

Quindi la funzione ha asintoto obliquo  $y = xe + 3e$  per  $x \rightarrow \pm\infty$ .

(Si presti attenzione al calcolo del limite precedente: se  $\frac{f(x)}{x} \rightarrow m$ , il limite di  $[f(x) - mx]$  è sempre una forma di indeterminazione  $[\infty - \infty]$ ; raccogliendo  $mx$ , si ha  $mx \left[ \frac{f(x)}{mx} - 1 \right]$ , che è ora una forma di indeterminazione del tipo  $[\infty \cdot 0]$ ; questa va risolta generalmente applicando qualche limite notevole all'espressione tra  $[\cdot]$ , per darne una stima asintotica).

$$\begin{aligned} f'(x) &= e^{\frac{x+2}{x-1}} \left( 1 + x \frac{(x-1)-(x+2)}{(x-1)^2} \right) = \frac{e^{\frac{x+2}{x-1}}}{(x-1)^2} [(x-1)^2 - 3x] = \\ &= \frac{e^{\frac{x+2}{x-1}}}{(x-1)^2} (x^2 - 5x + 1) \end{aligned}$$

Per ogni  $x \neq 1$   $f'(x)$  è definita; per  $x \rightarrow 1^-$ ,  $f'(x) \rightarrow 0^-$  (l'esponenziale va a zero più rapidamente di  $(x-1)^2$ ). Il grafico quindi arriva in  $x = 1$  con tangente orizzontale, da sinistra.

$$f'(x) \geq 0 \text{ per } x^2 - 5x + 1 \geq 0$$

$$x \geq \frac{5+\sqrt{21}}{2} \simeq 4,8 \quad x \leq \frac{5-\sqrt{21}}{2} \simeq 0,2$$

$$x = \frac{5+\sqrt{21}}{2} \text{ punto di min rel.} \quad x = \frac{5-\sqrt{21}}{2} \text{ punto di max rel.}$$

Grafico:

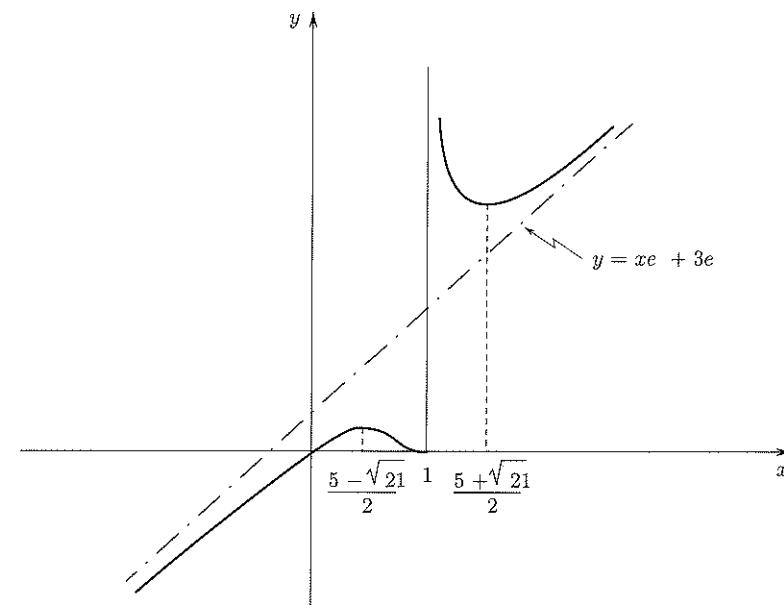


Figura 4.45.



Studiare le seguenti funzioni, utilizzando i suggerimenti forniti, e tracciarne il grafico.



$$e^x \sqrt[3]{\frac{x+3}{x-2}}$$

*Suggerimento.* Ci aspettiamo: flesso a tangente verticale dove si annulla il radicando, asintoto verticale dove si annulla il denominatore.



$$\left| \frac{x^2 + x - 2}{2x + 3} \right|$$

*Suggerimento.* Ci aspettiamo: punti angolosi dove si annulla il numeratore; asintoto verticale dove si annulla il denominatore. Conviene studiare la funzione senza modulo e poi...



$$e^{3/\log x}$$

Nei punti in cui la funzione non è definita ma ha limite (destro o sinistro) finito, studiare anche il limite della derivata prima, per sapere con quale pendenza la curva arriva in tali punti.



$$\arctg \left( \frac{x+1}{x-3} \right) + \frac{x}{4}$$

*Suggerimento.* Questo è un tipico esempio in cui l'equazione  $f(x) = 0$  non è risolubile con passaggi algebrici: inutile quindi perder tempo cercando di studiare il segno di  $f(x)$  e le sue intersezioni con l'asse  $x$ : sarà lo studio complessivo a darci informazioni al riguardo. Studiare in particolare il comportamento di  $f(x)$  all'infinito e il punto di discontinuità che si trova: calcolare anche i limiti di  $f'(x)$  in tale punto, per tracciare un grafico accurato.

Studiare le seguenti funzioni e tracciarne il grafico.

$$\frac{x^2 + 3x + 2}{x + 3}$$

$$x e^{1/(1-|x|)}$$

$$\frac{x^3 + 1}{x + 2}$$

$$\frac{x}{\log(1+x)}$$

$$\frac{x^2 + 2x - 3}{5x^2 + x}$$

$$\sqrt{\frac{x^2 - 1}{x + 3}}$$

$$x \log^{1/3} x$$

$$\log \left( \frac{x^2 + 3x - 4}{x + 1} \right)$$

$$\sqrt{x} |\log x|$$

$$e^{-|\frac{2x+1}{x-2}|}$$

$$\sqrt{x} \log^2 x$$

$$e^{\frac{x+2}{x^2-1}}$$

$$x^3 \log |x|$$

$$e^{\frac{1}{x-1}} \log x$$

$$1 - |x| e^{-x}$$

$$\frac{\log x}{\sqrt[3]{x-1}}$$

Le seguenti funzioni, tratte da modelli reali, contengono qualche costante o parametro. Tracciarne il grafico qualitativo.

(*Curva logistica*). Si tracci, al variare del parametro  $k$ , il grafico della curva

$$N(t) = \frac{k N_0 e^t}{k - N_0 + N_0 e^t}$$

dove  $N_0$  è una costante positiva. (Questa funzione rappresenta, sotto opportune ipotesi e in opportune unità di misura, il numero di individui di una data specie all'istante  $t$ , se all'istante 0 tale numero è  $N_0$ ; la costante  $k$  è detta *capacità dell'ambiente*; questo esempio sarà ripreso nel vol. 2°, trattando le equazioni differenziali).

(*Funzione di Langevin*). Si tracci per  $y > 0$  il grafico della funzione:

$$L(y) = m_0 \left( \frac{\operatorname{Ch} y}{\operatorname{Sh} y} - \frac{1}{y} \right)$$

dove  $m_0$  è una costante positiva. Questa funzione entra nella descrizione dei fenomeni di *polarizzazione magnetica per orientamento*.

Siano  $\gamma, \omega_0$  costanti positive. Tracciare il grafico delle due funzioni:

$$n_1(\omega) = 1 + \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2} \quad n_2(\omega) = \frac{\gamma \omega}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2}$$

(queste funzioni entrano nella descrizione degli indici di rifrazione delle onde elettromagnetiche nei dielettrici).

(*Effetto Doppler relativistico*). Se la sorgente e il ricevitore di un'onda si allontanano con velocità relativa  $v$  e l'onda è emessa con frequenza  $\nu_0$ , la frequenza ricevuta sarà

$$\nu = \nu_0 \frac{\sqrt{1 - v^2/c^2}}{1 + v/c}$$

dove  $c$  è la velocità della luce (costante positiva). Si tracci il grafico della funzione  $\nu = \nu(v)$  nel dominio fisicamente ammissibile,  $|v| < c$ .

(*Angolo di rifrazione*). Dalla legge di Snell della rifrazione (cfr. par. 4.3) si ricava che nel passaggio di un raggio di luce da un mezzo a un altro, l'angolo  $\theta_r$  di rifrazione dipende dall'angolo  $\theta_i$  di incidenza in base alla legge:

$$\theta_r = \arcsin \left( \frac{\sin \theta_i}{n} \right)$$

dove  $n$  = indice di rifrazione relativo del mezzo 2 rispetto al mezzo 1, può essere maggiore o minore di 1 (ma è positivo). Studiare questa funzione di  $\theta_i$  e tracciarne il grafico, distinguendo i casi  $n > 1$ ,  $n < 1$ . L'angolo  $\theta_i$  è sempre compreso in  $(0, \frac{\pi}{2})$ . (In uno dei due casi è necessaria una limitazione più stretta e, per angoli di incidenza non ammissibili, non si ha rifrazione).

(*Distribuzione spettrale della radiazione emessa da un corpo nero*). Nell'esempio 3.40 del capitolo 3 si è studiata mediante stime asintotiche la funzione

$$f(\lambda) = \frac{8\pi hc \lambda^{-5}}{e^{hc/\lambda kT} - 1}$$

( $h, c, k$  costanti positive;  $T$  = temperatura assoluta, parametro positivo;  $\lambda$  = lunghezza d'onda). Se ne completa ora lo studio calcolando la derivata prima e dimostrando che il punto di massimo di  $f$  cade (ad esempio) nell'intervallo  $(\frac{hc}{5kT}, \frac{hc}{4kT})$ . Ciò significa che la lunghezza d'onda per cui è massima la densità di energia raggiante è inversamente proporzionale alla temperatura assoluta  $T$  (legge dello spostamento di Wien).

Infine, si pensi ora  $\lambda$  fissato e si tracci il grafico di  $f$  come funzione di  $T$ .

## 7 CALCOLO DIFFERENZIALE E APPROXIMAZIONI

### 7.1 Differenziale e approssimazione lineare. Il simbolo di “o piccolo”

Un’operazione molto frequente sia in matematica che nelle sue applicazioni è quella di *linearizzazione*.

Sostanzialmente si tratta di approssimare una data quantità, che dipende in modo non lineare da una o più variabili, con una che dipende da queste linearmente.

Naturalmente, come in tutti i tipi di approssimazione, occorre fornire informazioni (qualitative o quantitative) sull’errore commesso.

Un esempio elementare di linearizzazione consiste nell’approssimare l’incremento subito da una data funzione  $f$ , in conseguenza di una variazione del suo argomento da  $x_0$  a  $x_0 + dx$ , sostituendo alla funzione stessa la retta tangente nel punto  $x_0$ .

Precisamente, sia  $f : (a, b) \rightarrow \mathbb{R}$  derivabile in un punto  $x_0$  e diamo a  $x_0$  un incremento  $dx$  (che pensiamo molto piccolo in valore assoluto, cioè  $|dx| \ll 1$ ). In conseguenza,  $f$  subisce un incremento

$$\Delta f = f(x_0 + dx) - f(x_0)$$

che, in generale, pensando  $x_0$  fissato, non è proporzionale a  $dx$  ossia non è lineare rispetto a  $dx$ .

Come indicato in figura 4.46, sostituiamo al grafico di  $f$ , vicino a  $x$ , la retta tangente nel punto  $P = (x_0, f(x_0))$ .

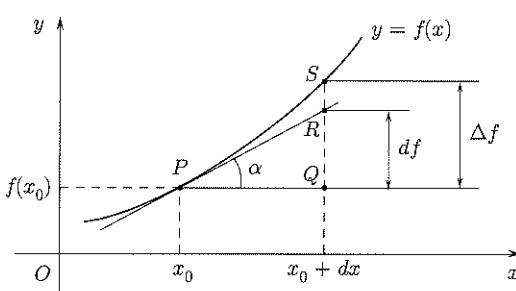


Figura 4.46.  $df = \tan \alpha \cdot dx$ .

L’incremento valutato lungo la retta tangente è pari alla lunghezza (con segno!) del segmento  $QR$  e cioè uguale a  $\tan \alpha \cdot dx$ , ovvero a  $f'(x_0)dx$ , ricordando che  $f'(x_0) = \tan \alpha$ . Tale incremento, proporzionale a  $dx$ , prende il nome di *differenziale* di  $f$  nel punto  $x_0$  e si indica col simbolo  $df(x_0)$ :

$$df(x_0) = f'(x_0)dx$$

Qual è l’errore commesso nell’approssimazione? In altri termini, che cosa si può dire sulla differenza tra  $\Delta f$  e  $df(x_0)$ ?

Per rispondere, è sufficiente ricorrere alla definizione di derivata; la formula

$$\lim_{dx \rightarrow 0} \frac{f(x_0 + dx) - f(x_0)}{dx} = f'(x_0)$$

si può scrivere

$$(7.1) \quad \frac{f(x_0 + dx) - f(x_0)}{dx} = f'(x_0) + \varepsilon(dx)$$

dove  $\varepsilon(dx)$  indica una quantità che tende a zero se  $dx \rightarrow 0$ ; è cioè un *infinitesimo* per  $dx \rightarrow 0$ .

Moltiplicando per  $dx$ , dalla (7.1) si ottiene infine

$$(7.2) \quad f(x_0 + dx) - f(x_0) = f'(x_0)dx + dx \cdot \varepsilon(dx).$$

Dalla (7.2), si ricava subito

$$\Delta f - df(x_0) = dx \cdot \varepsilon(dx)$$

La quantità  $dx \cdot \varepsilon(dx)$  è una quantità che, divisa per  $dx$ , tende a zero: questo significa che  $dx \cdot \varepsilon(dx)$  è *infinitesimo di ordine superiore rispetto a  $dx$* . Introduciamo un simbolo utile a indicare sinteticamente questa circostanza:

**DEFINIZIONE 4.9** Date due funzioni  $f(x), g(x)$ , definite in un intorno di  $x_0$ , si dice che

$$f(x) = o(g(x)) \text{ per } x \rightarrow x_0$$

(si legge “ $f(x)$  è o piccolo di  $g(x)$ ”) se

$$\frac{f(x)}{g(x)} \rightarrow 0 \text{ per } x \rightarrow x_0$$

Se  $g(x)$  è un infinitesimo, dire che  $f(x) = o(g(x))$  significa che  $f(x)$  è infinitesimo di ordine superiore rispetto a  $g(x)$ . Ad esempio, per  $x \rightarrow 0$ ,  $x^2 = o(x)$ ,  $e^{-1/x^2} = o(x^4)$ , ecc. In particolare, il simbolo  $o(1)$  denota semplicemente una quantità infinitesima.

Con questa simbologia, si scrive:

$$(7.3) \quad f(x_0 + dx) - f(x_0) = f'(x_0)dx + o(dx) \text{ per } dx \rightarrow 0$$

o, più sinteticamente:

$$\Delta f(x_0) = df(x_0) + o(dx) \text{ per } dx \rightarrow 0$$

La (7.3) esprime quella che si chiama un’approssimazione di  $\Delta f$  al prim’ordine. Si scrive anche, talvolta,

$$\Delta f \approx df \text{ al prim’ordine, vicino a } x_0$$

Le regole di calcolo dei differenziali sono analoghe a quelle di derivazione:

$d(f \pm g) = df \pm dg$
$d(f \cdot g) = gdf + f dg$
$d\left(\frac{f}{g}\right) = \frac{gdf - f dg}{g^2}$

Si noti poi che  $\frac{df}{f} = d\log|f|$  = differenziale logaritmico di  $f$ .



Consideriamo

$$f(x) = \sqrt{1+x}$$

e approssimiamo al primo ordine  $f(x) - f(0) = \sqrt{1+x} - 1$ .

Si noti che qui  $x_0 = 0$  e quindi  $x$  coincide con l'incremento  $dx$ . Essendo

$$f'(x) = \frac{1}{2\sqrt{1+x}} \quad f'(0) = \frac{1}{2}$$

si ha

$$(7.4) \quad \boxed{\sqrt{1+x} - 1 \approx \frac{1}{2}x} \quad \text{al primo ordine}$$

o anche

$$\sqrt{1+x} = 1 + \frac{1}{2}x + o(x), \text{ per } x \rightarrow 0$$

Più in generale, considerando  $f(x) = (1+x)^\alpha$ , si ha

$$f'(x) = \alpha(1+x)^{\alpha-1} \quad f'(0) = \alpha$$

e quindi ritroviamo (v. cap. 3, par. 3.3, formula (3.8))

$$(7.5) \quad \boxed{(1+x)^\alpha - 1 \approx \alpha x} \quad \text{al primo ordine}$$

**Potenziale di un dipolo elettrico.** Un'interessante applicazione della (7.4) si trova nel calcolo dell'approssimazione al primo ordine del potenziale creato da un dipolo elettrico. Con riferimento alla figura 4.47, consideriamo un dipolo costituito da due cariche  $+q$  e  $-q$  concentrate in due punti  $R$  e  $S$  rispettivamente, a distanza  $2a$  l'una dall'altra.

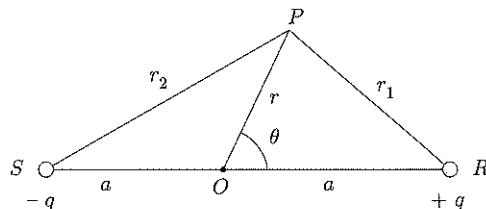


Figura 4.47.

Se  $P$  è un punto a distanza  $r$  dal punto medio  $O$  e a distanza  $r_1$  da  $R$ ,  $r_2$  da  $S$ , il potenziale in  $P$  è assegnato dalla formula (in opportune unità di misura)

$$V(P) = \frac{q}{4\pi\epsilon_0} \left( \frac{1}{r_1} - \frac{1}{r_2} \right)$$

Vogliamo valutare il potenziale quando  $a \ll r$  ( $a$  molto minore di  $r$ ), mantenendo costante il momento di dipolo  $\mu = a \cdot q$ .

Dal teorema di Carnot in trigonometria, si ha:

$$r_1^2 = r^2 + a^2 - 2ar \cos \theta = r^2 \left[ 1 + \left( -2\frac{a}{r} \cos \theta + \frac{a^2}{r^2} \right) \right]$$

$$r_2^2 = r^2 + a^2 + 2ar \cos \theta = r^2 \left[ 1 + \left( 2\frac{a}{r} \cos \theta + \frac{a^2}{r^2} \right) \right]$$

Ricordiamo ora che, se  $x$  è vicino a zero,  $\sqrt{1+x} \approx 1 + \frac{1}{2}x$  al primo ordine e quindi, essendo  $\frac{a}{r}$  vicino a zero e  $\frac{a}{r} \ll \frac{a^2}{r^2}$ , si può scrivere, al primo ordine:

$$r_1 \approx r \left( 1 - \frac{a}{r} \cos \theta \right) \quad r_2 \approx r \left( 1 + \frac{a}{r} \cos \theta \right)$$

Ne segue che, al primo ordine,  $r_1 - r_2 \approx 2a \cos \theta$ ,  $r_1 r_2 \approx r^2$  e infine

$$V(P) = \frac{q}{4\pi\epsilon_0} \left( \frac{1}{r_1} - \frac{1}{r_2} \right) = \frac{q}{4\pi\epsilon_0} \frac{r_2 - r_1}{r_1 r_2} \approx \frac{2a \cos \theta}{r^2} \frac{q}{4\pi\epsilon_0} = \frac{2\mu \cos \theta}{4\pi\epsilon_0 r^2}$$

che è l'approssimazione cercata. Si noti la dipendenza da  $r$  e dall'angolo  $\theta$ .

**Oscillazioni del pendolo.** Una massa puntiforme  $m$  oscilla, appesa a un filo di lunghezza  $l$ . Detto  $\theta(t)$  l'angolo formato dal filo con la verticale all'istante  $t$ , la forza agente sul punto nella direzione del moto è  $-mg \sin \theta$  (v. fig. 4.48)

D'altro canto, questa forza deve uguagliare, per la seconda legge della dinamica, il prodotto della massa per l'accelerazione, pari a  $ml\theta''$ . Si ottiene quindi l'equazione

$$ml\theta'' = -mg \sin \theta$$

ossia

$$\theta'' = -\frac{g}{l} \sin \theta$$

Tale equazione è difficile da studiare direttamente. Se però vogliamo studiare solo le piccole oscillazioni del punto attorno all'equilibrio, potremo assumere che  $\theta$  sia piccolo; si può allora studiare in prima approssimazione l'equazione linearizzata

$$\theta'' = -\frac{g}{l} \theta$$

(Abbiamo sfruttato  $\sin \theta \approx \theta$ , al prim'ordine, per  $\theta$  piccolo). Questa può essere risolta completamente. Si trova (ponendo  $\omega^2 = g/l$ )

$$\theta(t) = A \cos(\omega t + \theta_0)$$

con  $\theta_0$ ,  $A$  costanti da determinarsi mediante le condizioni iniziali del sistema. (Lo studente verifichi che le funzioni  $\theta(t)$  trovate risolvono effettivamente l'equazione. Nel secondo volume, trattando le equazioni differenziali si proverà che effettivamente non vi sono altre soluzioni).

**Energia cinetica classica e relativistica.** Nella meccanica classica, l'energia cinetica di un punto materiale di massa  $m$  in movimento, con velocità scalare  $v$ , è data da

$$E_c = \frac{1}{2}mv^2$$

La teoria della relatività modifica varie definizioni e leggi della meccanica classica. Anzitutto, la massa  $m$  non è più una costante, ma dipende dalla velocità, secondo la legge:

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

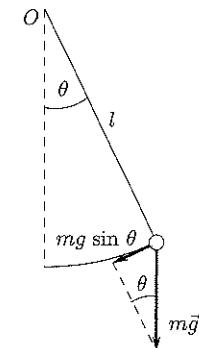


Figura 4.48.

dove  $c$  è la velocità della luce nel vuoto e  $m_0$  è la “massa a riposo” della particella (questa è costante); in secondo luogo, vale la famosa equazione dell’energia:

$$E = mc^2$$

L’energia della particella a riposo è invece

$$E_0 = m_0c^2$$

e l’energia cinetica relativistica è per definizione pari a  $(E - E_0)$ . Sostituendo ad  $m$  la sua espressione e troviamo  $E$  come funzione della velocità della particella:

$$E = \frac{m_0c^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Nell’ipotesi che la velocità della particella sia piccola rispetto a quella, elevatissima, della luce,  $v/c$  sarà molto piccolo. Linearizziamo allora l’espressione trovata come funzione di  $(v/c)^2$  (usando la (7.5) con  $\alpha = -1/2$ ):

$$\begin{aligned} E &= m_0c^2 \cdot \left(1 - \left(\frac{v}{c}\right)^2\right)^{-1/2} \approx m_0c^2 \left(1 + \frac{1}{2} \frac{v^2}{c^2}\right) = \\ &= m_0c^2 + \frac{1}{2}m_0v^2 = E_0 + E_c \end{aligned}$$

dove  $E_0$  è l’energia a riposo della particella, ed  $E_c$  l’energia cinetica classica. L’energia cinetica relativistica, che per definizione è pari a  $E - E_0$ , nell’approssimazione valida per velocità piccole rispetto a quella della luce coincide dunque con l’energia cinetica classica.

## 7.2 Limiti notevoli e sviluppi

Abbiamo visto che il processo di linearizzazione permette di approssimare una funzione derivabile, localmente, mediante la sua retta tangente, ossia una funzione lineare. D’altro canto, anche i “limiti notevoli” a suo tempo incontrati permettevano di scrivere simili risultati di approssimazione, per certe funzioni particolari. Ad esempio, per la funzione  $y = \sin x$ , il processo di linearizzazione, in  $x = 0$ , porta a scrivere

$$\sin x = x + o(x)$$

mentre a suo tempo abbiamo visto che

$$\sin x \sim x \text{ per } x \rightarrow 0$$

Che relazione c’è tra queste due affermazioni? Sono in effetti equivalenti, come mostra il prossimo fatto generale:

**TEOREMA 4.15 (RELAZIONE TRA “O PICCOLO” E “ASINTOTICO”)** Vale la seguente equivalenza:

$$\text{Per } x \rightarrow x_0, f(x) \sim g(x) \text{ se e solo se } f(x) = g(x) + o(g(x))$$

Il lettore per esercizio dimostri le due implicazioni contenute nell’affermazione precedente: basta usare le definizioni.

Questo permette di rileggere i limiti notevoli che abbiamo finora espresso con stime asintotiche, mediante uguaglianze che coinvolgono “o piccolo”. Ad esempio,

per  $x \rightarrow 0$ :

$$\sin x \sim x, \text{ quindi } \sin x = x + o(x);$$

$$1 - \cos x \sim \frac{1}{2}x^2, \text{ quindi } 1 - \cos x = \frac{1}{2}x^2 + o(x^2).$$

Riflettiamo sull’ultima: l’uguaglianza  $1 - \cos x = \frac{1}{2}x^2 + o(x^2)$  si può riscrivere nella forma:

$$(7.6) \quad \cos x = 1 - \frac{1}{2}x^2 + o(x^2)$$

(questo è ovvio: abbiamo spostato dei termini da un membro a un altro di una uguaglianza). Notiamo, invece, che la relazione asintotica  $1 - \cos x \sim \frac{1}{2}x^2$  non dice la stessa cosa della relazione  $\cos x \sim 1 - \frac{1}{2}x^2$ . Infatti, per  $x \rightarrow 0$ ,  $1 - \frac{1}{2}x^2 \sim 1$ , perciò la stima  $\cos x \sim 1 - \frac{1}{2}x^2$  contiene la stessa informazione che dire semplicemente “ $\cos x \rightarrow 1$ ”, ed è altrettanto vera che la stima  $\cos x \sim 1 + 40x^{13}$ ! Questa osservazione mostra un vantaggio del simbolo di “o piccolo” rispetto a quello di asintotico: un’uguaglianza si può riscrivere in vari modi, è più facile da usare senza errori, rispetto ad una stima asintotica.

Per esercizio, il lettore riscriva mediante il simbolo di “o piccolo” i limiti notevoli che riguardano le funzioni  $\sqrt{1+x}$ ,  $e^x$ ,  $\log(1+x)$ ,  $x^\alpha$ , per  $x \rightarrow 0$ .

## 7.3 Formula di Taylor-MacLaurin con resto secondo Peano

Vogliamo ora generalizzare il procedimento di “approssimazione per linearizzazione” a quello di “approssimazione polinomiale”. In altre parole, ci chiediamo: data una funzione, derivabile tutte le volte che sarà necessario, esiste un polinomio che, nell’intorno di un punto fissato, approssima la funzione meglio della sua retta tangente?

L’esempio (7.6) della funzione coseno si può rileggere in tal senso: la funzione  $\cos x$  è approssimata dalla parabola  $y = 1 - \frac{1}{2}x^2$  meglio che dalla retta tangente  $y = 1$ ,

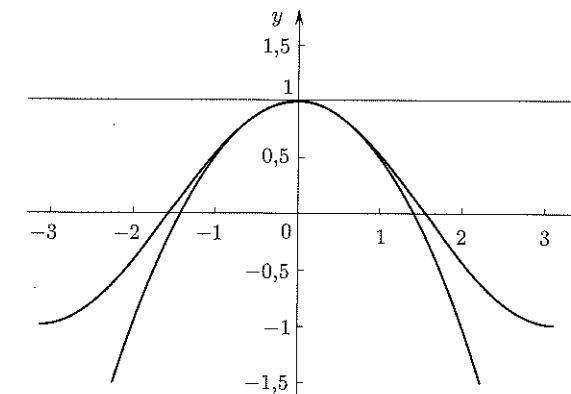


Figura 4.49. La funzione  $\cos x$  è approssimata dal polinomio  $y = 1 - \frac{1}{2}x^2$  meglio che dalla retta  $y = 1$ , vicino ad  $x = 0$ .

per  $x \rightarrow 0$ : infatti, lo scarto tra la funzione e questo polinomio di secondo grado è  $o(x^2)$ , cioè tende a zero più rapidamente di  $x^2$  (e non solo più rapidamente di  $x$ ). Per semplicità, cominciamo a ragionare nell'intorno del punto  $x_0 = 0$ .

Procediamo in 2 passi:

- Individuiamo un polinomio “candidato” ad approssimare bene la funzione, cercando un polinomio che abbia tutte le derivate fino all’ordine  $n$  uguali a quelle di  $f(x)$ , nel punto  $x = 0$ . Affinché questo sia sempre possibile, il grado del polinomio dev’essere almeno  $n$ . (Infatti, la derivata  $n$ -esima di un polinomio di grado minore di  $n$  è identicamente nulla, quindi non potrebbe essere uguale a  $f^{(n)}(0)$ , in generale). Facendo i calcoli, si trova che:

**TEOREMA 4.16 (POLINOMIO DI MACLAURIN)** *Data una funzione  $f$  derivabile  $n$  volte in  $x = 0$ , esiste uno e un sol polinomio di grado  $\leq n$ , chiamiamolo  $T_n$ , con la proprietà che:*

$$T_n(0) = f(0), T'_n(0) = f'(0), \dots, T_n^{(n)}(0) = f^{(n)}(0)$$

e questo polinomio, detto polinomio di MacLaurin di  $f(x)$  di grado  $n$ , è:

$$\begin{aligned} T_n(x) &= f(0) + xf'(0) + \frac{1}{2}x^2f''(0) + \frac{1}{3!}x^3f'''(0) + \dots + \frac{1}{n!}x^n f^{(n)}(0) = \\ &= \sum_{k=0}^n \frac{f^{(k)}(0)}{k!} x^k \end{aligned}$$

(avendo posto  $f^{(0)} = f$ ).

Notiamo esplicitamente che la richiesta che  $f$  sia derivabile  $n$  volte in  $x = 0$  implica in particolare che  $f$  sia derivabile  $n - 1$  volte in un intorno di 0. Notiamo anche che il polinomio  $T_n$  assegnato dal teorema precedente, solitamente è proprio di grado  $n$ , ma può avere grado minore se  $f^{(n)}(0) = 0$ .

È interessante osservare la coerenza dimensionale della formula che assegna il polinomio di MacLaurin: supponiamo che  $f$  abbia le dimensioni di una lunghezza  $[L]$  e  $x$  abbia le dimensioni di un tempo  $[T]$ . Allora  $[f'] = [L] \cdot [T]^{-1}$ ,  $[f''] = [L] \cdot [T]^{-2}$ , e in generale  $[f^{(k)}] = [L] \cdot [T]^{-k}$ . Questo significa che ogni addendo  $\frac{f^{(k)}(0)}{k!} x^k$  del polinomio  $T_n(x)$  ha la struttura

(costante adimensionale) · (fattore di dimens.  $[L] \cdot [T]^{-k}$ ) · (fattore di dimens.  $[T]^k$ )

Perciò il polinomio di MacLaurin ha le dimensioni di una lunghezza, esattamente come la funzione  $f$  che si vuole approssimare.

- Proviamo ora che il polinomio trovato approssima bene  $f(x)$ , in un intorno di  $x = 0$ . Precisamente, vale il:

**TEOREMA 4.17 (FORMULA DI MACLAURIN ALL’ORDINE  $n$ , CON RESTO SECONDO PEANO)** *Sia  $f : (a, b) \rightarrow \mathbb{R}$  derivabile  $n$  volte in  $0 \in (a, b)$ . Allora*

$$f(x) = T_n(x) + o(x^n) \text{ per } x \rightarrow 0$$

*La formula precedente si dice “formula di MacLaurin di ordine  $n$ , con resto secondo Peano”.*

La formula ha la struttura:

funzione da approssimare = polinomio approssimante + errore di approssimazione

dove l’errore di approssimazione è il termine  $o(x^n)$ , detto *resto secondo Peano*.

Per  $x \rightarrow 0$ , il resto secondo Peano è tanto più piccolo quanto maggiore è  $n$ . Lo spirito della formula è dunque il seguente: conoscendo un numero abbastanza alto di derivate di  $f$  nel punto  $x = 0$ , si può approssimare sempre meglio  $f$ , in un intorno di  $x = 0$ .

**DIMOSTRAZIONE.** Proviamo per semplicità il teorema nel caso  $n = 2$ , ossia:

$$f(x) = f(0) + xf'(0) + \frac{1}{2}x^2f''(0) + o(x^2) \text{ per } x \rightarrow 0$$

Occorre provare che

$$f(x) - \left[ f(0) + xf'(0) + \frac{1}{2}x^2f''(0) \right] = o(x^2) \text{ per } x \rightarrow 0$$

ossia (per definizione di “o piccolo”) che:

$$\lim_{x \rightarrow 0} \frac{f(x) - [f(0) + xf'(0) + \frac{1}{2}x^2f''(0)]}{x^2} = 0.$$

Questo limite dà una forma di indeterminazione  $[0/0]$ , che calcoliamo con De L’Hospital:

$$\lim_{x \rightarrow 0} \frac{f'(x) - [f'(0) + xf''(0)]}{2x}$$

dà ancora una forma  $[0/0]$ . Osserviamo ora che l’ultimo limite trovato è zero, per il seguente motivo. Per ipotesi,  $f$  è derivabile due volte, quindi  $f'$  è derivabile; se applichiamo il teorema di linearizzazione alla funzione  $f'(x)$  otteniamo

$$f'(x) = f'(0) + xf''(0) + o(x)$$

da cui segue che

$$\frac{f'(x) - [f'(0) + xf''(0)]}{x} \rightarrow 0.$$

Il caso generale ( $n$  qualsiasi) si può dimostrare per *induzione su  $n$*  (v. capitolo 1, par. 7). Abbiamo già provato che la tesi vale per  $n = 2$ . Supponiamo ora di sapere che per qualsiasi funzione derivabile  $n - 1$  volte in  $x = 0$  vale la formula di Mac Laurin all’ordine  $n$ , e proponiamoci di dimostrarlo che

$$f(x) = T_n(x) + o(x^n) \text{ per } x \rightarrow 0, \text{ ossia che}$$

$$(7.7) \quad \lim_{x \rightarrow 0} \frac{f(x) - T_n(x)}{x^n} = 0.$$

Scriviamo, provvisoriamente,  $T_{n,f}$  al posto di  $T_n$  per mettere in evidenza la dipendenza dalla funzione  $f$ . Applicando il teorema di De L’Hospital al limite (7.7) si ha:

$$(7.8) \quad \lim_{x \rightarrow 0} \frac{f'(x) - T'_{n,f}(x)}{nx^{n-1}}.$$

Osserviamo ora che la derivata del polinomio  $T_n(x)$  di  $f$  non è altro che il polinomio  $T_{n-1}(x)$  della funzione  $f'$ , cioè:

$$T'_{n,f}(x) = T_{n-1,f'}(x)$$

come si verifica direttamente dalla definizione di  $T_n$ . D'altro canto, per l'ipotesi induttiva, applicata alla funzione  $f'$ , sappiamo che

$$f'(x) = T_{n-1,f'}(x) + o(x^{n-1}), \text{ ossia}$$

$$\lim_{x \rightarrow 0} \frac{f'(x) - T_{n-1,f'}(x)}{x^{n-1}} = 0,$$

quindi il limite (7.8) è zero e, per il teorema di De L'Hospital, anche il limite (7.7) è zero, come volevamo dimostrare.

#### Formula di Taylor all'ordine $n$

Tutto questo discorso si può generalizzare ad un punto  $x_0 \neq 0$ : data una funzione derivabile  $n$  volte in  $x_0$ , si può costruire il suo polinomio di Taylor

$$T_{n,x_0}(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

e si può dimostrare che vale un analogo risultato di approssimazione locale:

**TEOREMA 4.18 (FORMULA DI TAYLOR ALL'ORDINE  $n$ , CON RESTO SECONDO PEANO)** Sia  $f : (a, b) \rightarrow \mathbb{R}$  derivabile  $n$  volte in  $x_0 \in (a, b)$ . Allora

$$f(x) = T_{n,x_0}(x) + o((x - x_0)^n) \text{ per } x \rightarrow x_0.$$



#### Approssimazione all'ordine $n$

Sia  $f(x) = e^x$ . Essendo  $f^{(n)}(x) = e^x$ ,  $f^{(n)}(0) = 1$  per ogni  $n$ , si ha:

$$e^x = T_{n,0}(x) + o(x^n) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + o(x^n), \text{ per } x \rightarrow 0$$

Sia  $f(x) = \sin x$ . Si ha:

$$f'(x) = \cos x \quad f''(x) = -\sin x \quad f'''(x) = -\cos x$$

e poi

$$f^{(iv)}(x) = \sin x = f(x) \quad f^{(v)}(x) = \cos x = f'(x) \quad \text{ecc.}$$

Ne segue che per  $n = 2k$ , pari, le derivate di ordine  $n$  sono nulle per  $x = 0$ , mentre quelle di ordine  $n = 2k + 1$ , dispari, sono alternativamente  $+1$  e  $-1$ . In sintesi:

$$f^{(2k)}(0) = 0 \quad f^{(2k+1)}(0) = (-1)^k \quad (k \geq 0)$$

I polinomi di MacLaurin della funzione seno contengono perciò solo termini di grado dispari:

$$\sin x = T_1(x) + o(x) = x + o(x)$$

$$\sin x = T_3(x) + o(x^3) = x - \frac{x^3}{6} + o(x^3)$$

$$\sin x = T_5(x) + o(x^5) = x - \frac{x^3}{6} + \frac{x^5}{120} + o(x^5)$$

⋮

$$\sin x = T_{2n+1}(x) + o(x^{2n+1}) = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + o(x^{2n+1})$$

La figura 4.50 illustra il miglioramento dell'approssimazione di  $\sin x$  mediante i suoi polinomi di MacLaurin, all'aumentare del grado del polinomio.

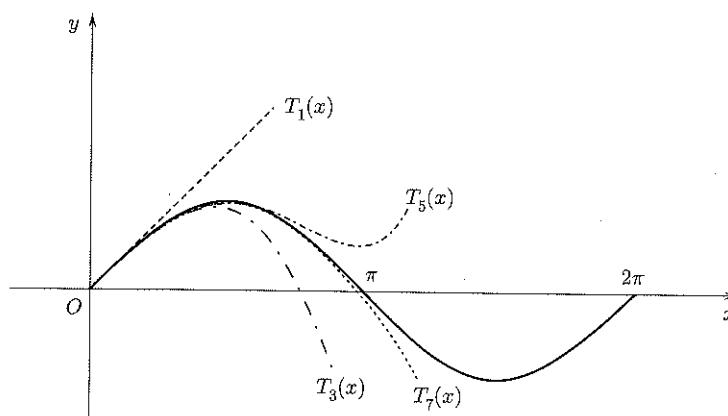


Figura 4.50.

Nel quadro che segue abbiamo raggruppato gli sviluppi di MacLaurin di alcune funzioni elementari, con il resto di Peano.

$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + o(x^n)$
$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + o(x^{2n+1})$
$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \dots + (-1)^n \frac{x^{2n}}{(2n)!} + o(x^{2n})$
$\operatorname{Sh} x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + \frac{x^{2n+1}}{(2n+1)!} + o(x^{2n+1})$
$\operatorname{Ch} x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots + \frac{x^{2n}}{(2n)!} + o(x^{2n})$
$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \dots + (-1)^{n-1} \frac{x^n}{n} + o(x^n)$
$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2} x^2 + \dots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} x^n + o(x^n)$

#### Proprietà del simbolo di “o piccolo”

Quando si opera su sviluppi di Taylor-MacLaurin con resto secondo Peano, ad esempio per calcolare limiti, ci si trova ad eseguire calcoli che coinvolgono  $o(x^n)$ . Per procedere correttamente è utile comprendere alcune proprietà del simbolo di “o piccolo”. Ricordiamo che

$$f(x) = o(g(x)) \text{ per } x \rightarrow x_0 \text{ se } \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0.$$

Ad esempio, per  $x \rightarrow 0$ ,  $x^2 = o(x)$ ;  $x^3 = o(x)$ ;  $x^3 = o(x^2)$ .

Il simbolo di  $o(\dots)$  non denota quindi una particolare funzione, ma qualsiasi funzione abbia la proprietà espressa dalla definizione. Da questo fatto seguono alcune proprietà della relazione di “ $o$  piccolo” che possono apparire “strane”. Ne illustriamo alcune su esempi che si incontrano frequentemente:

$$o(x) - o(x) = o(x) \quad (\text{non fa zero!})$$

$$o(-3x^2) = o(x^2), \quad -o(x) = o(x) \quad (\text{coefficienti numerici moltiplicativi sono irrilevanti})$$

Per  $x \rightarrow 0$ ,  $o(x) + o(x^2) = o(x)$  (l’errore più grossolano, cioè  $o(x)$ , ingloba quello più fine).

Analogamente si avrebbe:

$$\text{per } x \rightarrow +\infty, o(x) + o(x^2) = o(x^2)$$

Il simbolo di “ $o$  piccolo” si comporta al seguente modo con somme e prodotti (verificarlo in base alla definizione):

$$f \cdot o(g) = o(f \cdot g), \text{ ad esempio:}$$

$$xo(x^2) = o(x^3) \quad \frac{o(x^3)}{x} = o(x^2) \quad \frac{o(x^2)}{x^2} = o(1)$$

(ricordiamo che  $o(1)$  indica semplicemente una funzione che tende a zero per  $x \rightarrow x_0$ , il punto considerato). Ed anche:

$$o(f) \cdot o(g) = o(f \cdot g), \text{ ad esempio:}$$

$$o(x)o(x^2) = o(x^3)$$

#### 7.4 Formula di Taylor-MacLaurin con resto secondo Lagrange

Nella formula di Taylor con resto secondo Peano, l’informazione che abbiamo sull’errore commesso nell’approssimare  $f$  col suo polinomio di Taylor-MacLaurin è di tipo dinamico: al tendere a zero dell’incremento  $(x - x_0)$ , sappiamo che il resto tende a zero più rapidamente di  $(x - x_0)^n$ . Questa informazione, come abbiamo visto, è utile ad esempio nel calcolo dei limiti. Per un valore fissato dell’incremento  $(x - x_0)$ , però, la formula di Taylor con resto secondo Peano non dice nulla sull’entità dell’errore commesso. In varie questioni di calcolo approssimato è essenziale invece stimare l’errore che si commette approssimando una funzione col suo polinomio di Taylor-MacLaurin, quando l’incremento  $(x - x_0)$  ha un valore fissato, o un valore che non supera una soglia fissata.

Ad esempio, sappiamo che il polinomio di MacLaurin al 3° ordine di  $e^x$  è

$$T_3(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$$

Se volessimo utilizzare questo fatto per calcolare un valore approssimato di  $\sqrt{e}$ , ossia pensassimo di approssimare

$$e^{1/2} \text{ con } T_3\left(\frac{1}{2}\right) = \frac{79}{48}$$

come potremmo stimare a priori (ossia senza conoscere già il valore vero di  $\sqrt{e}$ ) di quanto stiamo sbagliando?

A questo tipo di problema risponde il prossimo risultato, che dà un modo alternativo di quantificare l’errore di approssimazione commesso:

**TEOREMA 4.19 (FORMULA DI TAYLOR ALL’ORDINE  $n$ , CON RESTO SECONDO LAGRANGE)** *Sia  $f : [a, b] \rightarrow \mathbb{R}$  derivabile  $n+1$  volte in  $[a, b]$ , e sia  $x_0 \in [a, b]$ . Allora esiste un punto  $c$  compreso tra  $x_0$  e  $x$  tale che*

$$(7.9) \quad f(x) = T_{n,x_0}(x) + \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}.$$

*La formula precedente si dice “formula di Taylor (o di MacLaurin se  $x_0 = 0$ ) di ordine  $n$ , con resto secondo Lagrange”.*

La formula ha la struttura:

funzione da approssimare = polinomio approssimante + errore di approssimazione

dove l’errore di approssimazione è il termine  $\frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}$ , detto *resto secondo Lagrange*.

Per  $n = 0$  la formula (7.9) coincide col Teorema di Lagrange.

Si noti che il punto  $c$  dipende da  $x_0$ ,  $x$  e  $n$ , ed è compreso tra  $x_0$  e  $x$ . Se si riesce a dimostrare che  $|f^{(n+1)}(t)| \leq M$  per ogni  $t$  compreso tra  $x_0$  e  $x$ , allora la formula di Taylor con resto secondo Lagrange dice che

$$|f(x) - T_{n,x_0}(x)| \leq \frac{M}{(n+1)!} |x - x_0|^{n+1}$$

che è appunto una stima dell’errore di approssimazione commesso.



Torniamo al problema di approssimare  $\sqrt{e}$  mediante la formula di Taylor. Ponendo, nella (7.9),  $f(x) = e^x$ ,  $x_0 = 0$ ,  $x = \frac{1}{2}$ ,  $n = 3$ , otteniamo:

$$\sqrt{e} = T_3\left(\frac{1}{2}\right) + \frac{e^c}{4!} \left(\frac{1}{2}\right)^4.$$

Poiché per  $0 \leq c \leq \frac{1}{2}$  è  $e^c \leq \sqrt{e} < \sqrt{3}$ <sup>1</sup>, si ottiene

$$\left| \sqrt{e} - T_3\left(\frac{1}{2}\right) \right| \leq \frac{\sqrt{3}}{2^4 \cdot 4!} \simeq 0.0045$$

Questo significa che approssimando  $\sqrt{e}$  con il valore  $\frac{79}{48}$  si commette un errore non superiore a 0.0045.

<sup>1</sup>Abbiamo usato la diseguaglianza grossolana  $e < 3$  per non cadere in un circolo vizioso: per stimare il valore di  $\sqrt{e}$  sembrerebbe necessario sapere già quanto vale  $\sqrt{e}!$  Invece, basta conoscere un valore che maggiora  $\sqrt{e}$ , e questo valore è, ad esempio,  $\sqrt{3}$ .

**DIMOSTRAZIONE DEL TEOREMA 4.19** Proviamo la (7.9) nel caso  $n = 1$ . Ponendo per comodità  $x_0 = a$ ,  $x = b$  l'enunciato diviene: se  $f : [a, b] \rightarrow \mathbb{R}$  è derivabile 2 volte in  $[a, b]$ , allora esiste un punto  $c \in [a, b]$  tale che

$$(7.10) \quad f(b) = f(a) + f'(a)(b-a) + \frac{1}{2}f''(c)(b-a)^2.$$

Poniamo  $f(b) - \{f(a) + f'(a)(b-a)\} = k(b-a)^2$  e cerchiamo di determinare la forma di  $k$ . Consideriamo la funzione

$$g(x) = f(b) - f(x) - f'(x)(b-x) - k(b-x)^2$$

e applichiamo ad essa il teorema del valor medio. Poiché  $g(b) = g(a) = 0$  (la seconda uguaglianza segue dalla definizione di  $k$ ) si trova che esiste  $c \in (a, b)$  tale che

$$0 = \frac{g(b) - g(a)}{b-a} = g'(c)$$

Ma  $g'(x) = -f'(x) - f''(x)(b-x) + f'(x) + 2k(b-x) = (b-x)[2k - f''(x)]$  e quindi  $g'(c) = 0$  implica, essendo  $c \neq b$ :

$$k = \frac{1}{2}f''(c)$$

Con lo stesso metodo si può dimostrare la (7.9) per  $n$  qualsiasi. Definiamo:

$$g(x) = f(b) - \sum_{j=0}^n \frac{f^{(j)}(x)(b-x)^j}{j!} - k(b-x)^{n+1}$$

con  $k$  definito implicitamente dall'identità

$$(7.11) \quad f(b) - T_{n,a}(b) = k(b-a)^{n+1}.$$

Ora si procede così:

a) Si verifica che  $g(b) = g(a) = 0$ :

$$g(b) = f(b) - f(b) = 0$$

$$g(a) = f(b) - T_{n,a}(b) - k(b-a)^{n+1} = 0 \text{ per definizione di } k.$$

b) Si applica il teorema di Lagrange a  $g$  in  $[a, b]$ , e si mostra che l'affermazione “esiste  $c \in (a, b)$  tale che  $g'(c) = 0$ ” è esattamente la (7.9). Cominciamo a calcolare  $g'$ :

$$\begin{aligned} g'(x) &= -\sum_{j=0}^n \frac{f^{(j+1)}(x)(b-x)^j}{j!} + \sum_{j=1}^n \frac{f^{(j)}(x)(b-x)^{j-1}}{(j-1)!} + k(n+1)(b-x)^n \\ &= -\sum_{j=0}^n \frac{f^{(j+1)}(x)(b-x)^j}{j!} + \sum_{j=0}^{n-1} \frac{f^{(j+1)}(x)(b-x)^j}{j!} + k(n+1)(b-x)^n \\ &= -\frac{f^{(n+1)}(x)(b-x)^n}{n!} + k(n+1)(b-x)^n \end{aligned}$$

(dove nel secondo passaggio abbiamo eseguito una traslazione di indici nella sommatoria).

Allora  $g'(c) = 0$  significa:

$$-\frac{f^{(n+1)}(c)(b-c)^n}{n!} + k(n+1)(b-c)^n = 0$$

ossia

$$k = \frac{f^{(n+1)}(c)}{(n+1)!}$$

che, inserita nella (7.11), dà la tesi.

### Formula di Taylor e convessità

Consideriamo la formula (7.9) con  $n = 1$ :

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(c)(x-x_0)^2$$

e supponiamo che in ogni punto di  $(a, b)$  si abbia  $f''(x) \geq 0$ , ovvero  $f$  sia convessa in  $(a, b)$ . Allora si ha  $\frac{1}{2}f''(c)(x-x_0)^2 \geq 0$  e perciò si può scrivere, per ogni coppia di punti  $x_0, x \in (a, b)$ :

$$(7.12) \quad f(x) \geq f(x_0) + f'(x_0)(x-x_0)$$

La (7.12) significa geometricamente che il grafico di  $f(x)$  si mantiene in tutto  $(a, b)$  sopra il grafico della sua retta tangente in  $x_0$  (e questo vale per ogni scelta del punto  $x_0 \in (a, b)$ ). Abbiamo dunque dimostrato che se una funzione (due volte derivabile) è convessa, è anche “convessa per tangentì” (vedi Teorema 4.12 par. 5.2).

Se invece  $f$  è concava, ossia  $f''(x) \leq 0$  in tutto  $(a, b)$ , la diseguaglianza (7.12) vale con “ $\leq$ ” anziché  $\geq$ : il grafico di  $f(x)$  si mantiene in tutto  $(a, b)$  sotto il grafico della sua retta tangente in  $x_0$ .

### 7.5 Risoluzione approssimata di equazioni: il metodo di Newton

Supponiamo di voler risolvere l'equazione

$$f(x) = 0$$

che equivale a cercare le intersezioni del grafico di  $f$  con l'asse  $x$ , e supponiamo di aver accertato (mediante confronto grafico, ad esempio) l'esistenza di un'unica soluzione  $x = \alpha$ , all'interno di un intervallo  $[a, b]$ . Si vuole dare una valutazione approssimata di  $\alpha$ . Il metodo di Newton, che ora illustriamo, serve proprio a questo scopo. L'idea è quella di costruire una successione convergente ad  $\alpha$ ; questa successione, come si vedrà, è definita in modo ricorsivo, ossia:

- si assegna esplicitamente il primo termine  $x_0$ ;
- si assegna esplicitamente la legge con cui calcolare  $x_{n+1}$  a partire da  $x_n$ , qualunque sia  $n$ .

In questo modo, il termine  $x_n$  può essere costruito a partire da  $x_0$  con  $n$  iterazioni del medesimo algoritmo. Il metodo, perciò, si presta molto bene al calcolo automatico.

Illustriamo il metodo con riferimento alla figura 4.51. Partiamo da  $x_0 = a$  e “linearizziamo” l'equazione  $f(x) = 0$  sostituendo a  $f$  la retta tangente al suo grafico nel punto  $(a, f(a))$ . Tale retta ha equazione

$$y = f(x_0) + f'(x_0)(x-x_0)$$

Invece di risolvere  $f(x) = 0$  risolviamo

$$f(x_0) + f'(x_0)(x-x_0) = 0$$

ottenendo

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \rightarrow \text{prima approssimazione di } \alpha$$

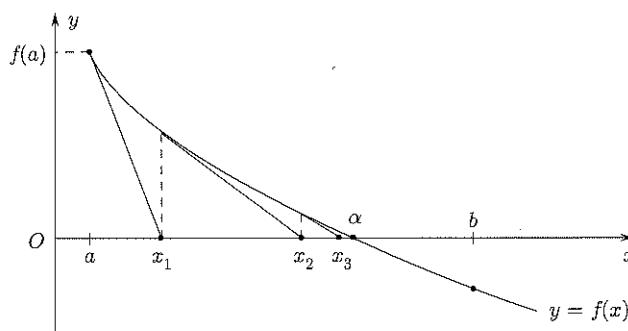


Figura 4.51.

Procediamo ora con  $x_1$  al posto di  $x_0 = a$ , sostituendo  $f$  con la retta tangente in  $(x_1, f(x_1))$ . Anziché  $f(x) = 0$  si risolve l'equazione

$$f(x_1) + f'(x_1)(x - x_1) = 0$$

ottenendo

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \rightarrow \text{seconda approssimazione di } \alpha$$

Proseguendo con  $x_2$  come per  $x_1$  e continuando si perviene alla legge di ricorrenza

$$(7.13) \quad \begin{cases} x_0 = a \\ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \end{cases}$$

Nella situazione descritta in figura 4.51, si vede che  $x_n \uparrow \alpha$ .

Si tratta ora di dimostrare rigorosamente che, sotto opportune ipotesi, la successione  $x_n$  effettivamente converge alla soluzione cercata. Vale in proposito il seguente

**TEOREMA 4.20** *Sia  $f : [a, b] \rightarrow \mathbb{R}$  due volte derivabile in  $[a, b]$ , e supponiamo che:*

1.  $f(a) \cdot f(b) < 0$ ;
2.  $f'(x), f''(x)$  hanno segno costante in  $[a, b]$ .
3.  $f(a) \cdot f''(a) > 0$ .

*Definiamo per ricorrenza la successione*

$$\begin{cases} x_0 = a \\ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \end{cases}$$

*Allora esiste uno e un sol punto  $c \in (a, b)$  tale che  $f(c) = 0$ , e la successione  $x_n$  converge a  $c$  per difetto.*

Se invece di valere la 3 vale la

3'.  $f(b) \cdot f''(b) > 0$ , allora la successione

$$\begin{cases} x_0 = b \\ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \end{cases}$$

converge a  $c$  per eccesso.

**OSSERVAZIONE** Notiamo che, sotto le ipotesi 1 e 2, è senz'altro verificata la 3 o la 3'. Infatti  $f''$  ha lo stesso segno in  $a$  e  $b$  per la 2, mentre  $f$  ha segni opposti in  $a$  e  $b$ , per la 1, perciò o vale la 3 o vale la 3'. Questo significa che il metodo è applicabile ogni volta che le ipotesi 1 e 2 sono verificate: si tratta solo, a quel punto, di scegliere opportunamente se porre  $x_0 = a$  oppure  $b$ . A sua volta, la condizione 2, apparentemente restrittiva, sarà in generale soddisfatta pur di scegliere un intervallo  $[a, b]$  abbastanza piccolo.

**DIMOSTRAZIONE.** Poiché  $f$  è continua in  $[a, b]$  e vale la 1, per il teorema degli zeri esiste almeno un  $c \in (a, b)$  tale che  $f(c) = 0$ . Inoltre, poiché  $f'(x)$  ha segno costante in  $(a, b)$ , la  $f$  è strettamente monotona, quindi tale punto  $c$  è unico (una funzione strettamente monotona non può annullarsi in due punti distinti). Questo prova esistenza e unicità del punto  $c$  in cui  $f$  si annulla. Proveremo che la successione  $x_n$  è *monotona*. Da questo seguirà (teorema di monotonia) che la successione è convergente. Mostriamo per intanto che, se  $x_n$  converge, certamente converge a  $c$ .

Infatti, se  $x_n \rightarrow \beta \in [a, b]$ , passando al limite nell'uguaglianza

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

si ha

$$\beta = \beta - \frac{f(\beta)}{f'(\beta)}$$

(abbiamo sfruttato il fatto che  $f$  è continua, e anche  $f'$  è continua, in quanto  $f'$  è derivabile, perché esiste per ipotesi  $f''$ ). Dall'ultima uguaglianza segue  $f(\beta) = 0$ , per cui  $\beta = c$  (unico punto in cui la funzione si annulla).

Proviamo dunque che  $x_n$  è monotona, sotto le ipotesi (1), (2), (3). Per fissare le idee, supponiamo che sia  $f(a) < 0$ . Allora per (1) è  $f(b) > 0$ . Poiché per la (2)  $f'(x)$  ha segno costante in  $[a, b]$ , dovrà essere  $f'(x) > 0$  in tutto  $[a, b]$  (se valesse l'altra diseguaglianza,  $f$  sarebbe decrescente, e non potrebbe essere  $f(a) < 0 < f(b)$ ). Poiché  $f(a) < 0$  e vale la (3),  $f''(a) < 0$ , dunque, poiché per la (2)  $f''(x)$  ha segno costante,  $f''(x) < 0$  in tutto  $[a, b]$ . Siamo quindi nella situazione della seguente figura:

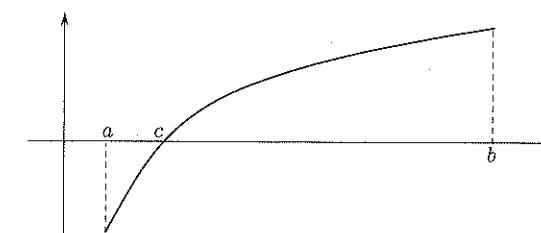


Figura 4.52.

Consideriamo ora la funzione

$$g(x) = x - \frac{f(x)}{f'(x)}.$$

Notiamo che è:

$$\begin{aligned} x_{n+1} &= g(x_n); \\ g(c) &= c \end{aligned}$$

(perché  $f(c) = 0$ ); inoltre

$$g'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2} > 0 \iff f(x) < 0$$

(perché  $f''(x) < 0$  in tutto  $[a, b]$ ), quindi  $g'(x) > 0$  per  $x \in [a, c]$  (vedi grafico di  $f$ ), ossia: in  $[a, c]$  la funzione  $g$  è strettamente crescente. Tenendo presenti queste cose, proviamo ora che:

$$(7.14) \quad \forall n \text{ è: } x_n < x_{n+1} < c.$$

Infatti: per  $n = 0$  si ha:

$$x_1 = a - \frac{f(a)}{f'(a)} > a$$

(perché  $f(a) < 0$  e  $f'(a) > 0$ ), ossia  $x_1 > x_0$ . Inoltre  $a < c$ , e poiché  $g$  è crescente, questo implica

$$\begin{aligned} g(a) &< g(c), \text{ ossia} \\ x_1 &< c. \end{aligned}$$

Quindi abbiamo provato che

$$x_0 < x_1 < c.$$

Applicando ora la  $g$  alle disuguaglianze precedenti ( $g$  è strettamente crescente in  $[a, c]$ , e i tre punti stanno in questo intervallo, quindi  $g$  conserva le disuguaglianze), si ha

$$\begin{aligned} g(x_0) &< g(x_1) < g(c), \text{ ossia} \\ x_1 &< x_2 < c. \end{aligned}$$

Applicando ancora la  $g$ , troveremo  $x_2 < x_3 < c$ , e così via. Perciò la (7.14) è vera, e in particolare  $x_n$  è monotona. Questo conclude la dimostrazione.  $\star$



Consideriamo ancora l'equazione

$$f(x) = e^{-x} - x = 0$$

Abbiamo:

$$f'(x) = -e^{-x} - 1 < 0 \quad \text{in } [0, 1]$$

$$f''(x) = e^{-x} > 0 \quad \text{in } [0, 1]$$

Inoltre  $f(0) = 1 > 0$  e  $f(1) < 0$ .

La successione (7.13), è in questo caso

$$x_0 = 0, \quad x_{n+1} = x_n + \frac{e^{-x_n} - x_n}{e^{-x_n} + 1} = \frac{x_n + 1}{1 + e^{x_n}}.$$

Usando una calcolatrice si trova

$$x_1 = 0,500$$

$$x_2 = 0,5663\dots$$

$$x_3 = 0,5671\dots$$

$$x_4 = 0,5671\dots$$

Si vede che già dopo 4 iterazioni si può dire  $c = 0,5671\dots$  con  $e^{-x_4} - x_4 \approx 4,55 \cdot 10^{-7}$ .



#### Approssimazioni al primo e al secondo ordine

La velocità  $v(t)$  di un corpo di massa  $m$  che cade nell'aria è regolata dalla legge

$$v(t) = \frac{mg}{h} \left( 1 - e^{-\frac{h}{m}t} \right)$$

dove  $g$  è l'accelerazione di gravità e  $h$  un coefficiente che esprime l'attrito dell'aria. (Questo è vero se si assume la forza di attrito proporzionale alla velocità). Linearizzando l'espressione, si trovi come varia la velocità nei primi istanti di caduta. (Cioè si approssimi  $v(t)$  per  $t \rightarrow 0$ ). L'approssimazione così ottenuta è per eccesso o per difetto? Per rispondere, si scriva ora l'approssimazione al second'ordine (mediante la formula di Taylor), e si confrontino le due espressioni.

(Cfr. l'esercizio precedente). Se si assume invece la forza di attrito proporzionale al quadrato della velocità, la velocità del corpo in caduta risulta:

$$v(t) = \sqrt{\frac{mg}{h}} \cdot \frac{e^{2\sqrt{\frac{gh}{m}}t} - 1}{e^{2\sqrt{\frac{gh}{m}}t} + 1}$$

con lo stesso significato dei simboli (il coefficiente  $h$  ha ora dimensioni diverse). Linearizzando l'espressione, si trovi come varia la velocità nei primi istanti di caduta.

L'oscillazione di una pallina fissata all'estremità di una molla e libera di muoversi su una retta è descritta dalla funzione:

$$x = Ae^{-kt} \cos(\omega t)$$

con  $A, k, \omega$  costanti positive. Linearizzando l'espressione, si approssimi  $x(t)$  per  $t \rightarrow 0$ . L'approssimazione così ottenuta è per eccesso o per difetto? Per rispondere, si scriva ora l'approssimazione al second'ordine, e si confrontino le due espressioni.

Usando la formula di Taylor al second'ordine col resto secondo Lagrange, si dia una maggiorazione del massimo errore che si commette approssimando  $\sin x$  con  $x$ , quando  $0 \leq x \leq \frac{\pi}{6}$ , e si precisi se l'errore commesso è per eccesso o per difetto.

Il massimo errore commesso è in realtà per  $x = \frac{\pi}{6}$ : si calcoli questo errore. Il risultato è coerente con la stima effettuata?

Usando la formula di Taylor al second'ordine col resto secondo Lagrange, si dia una maggiorazione del massimo errore che si commette approssimando  $\sqrt{1+x}$  con  $1 + \frac{1}{2}x$ , quando  $0 \leq x \leq \frac{1}{2}$ , e si precisi se l'errore commesso è per eccesso o per difetto.

Il massimo errore commesso è in realtà per  $x = \frac{1}{2}$ : si calcoli questo errore. Il risultato è coerente con la stima effettuata?

Dimostrare, sfruttando la definizione dei simboli di “o piccolo” e di “asintotico”, tutte le proprietà enunciate nel paragrafo “Proprietà del simbolo di o piccolo”.

#### Metodo di Newton

Nell'esempio 4.5 (diffrazione della luce attraverso una fenditura) abbiamo visto che la ricerca dei massimi dell'intensità luminosa nella figura di interferenza per la luce che attraversa una fenditura porta a risolvere l'equazione  $\operatorname{tg} t = t$ . Calcolare col metodo di Newton un valore approssimato della prima soluzione positiva di tale equazione, che si trova (come mostra il grafico) poco prima di  $\frac{3}{2}\pi$ .

(Suggerimento: applicare il metodo alla funzione  $f(t) = \operatorname{tg} t - t$  sull'intervallo  $[\frac{5}{4}\pi, \frac{3}{2}\pi - 0.1]$ , usando come valore iniziale  $a = \frac{3}{2}\pi - 0.1$ . Calcolare le iterate con l'aiuto di un computer).

Calcolare con approssimazione l'unica soluzione reale dell'equazione  $x^2 + 1 = \frac{1}{x}$ . A tal fine:

- Mostrare che il metodo di Newton è applicabile sull'intervallo  $[\frac{1}{2}, 1]$ ;
- Scrivere esplicitamente l'algoritmo iterativo.
- Calcolare esplicitamente le prime iterazioni, finché le prime due cifre decimali si stabilizzano.

Calcolare con approssimazione l'unica soluzione reale dell'equazione  $\log x = \frac{1}{x}$ . A tal fine:

- Mostrare che il metodo di Newton è applicabile sull'intervallo  $[1, 2]$ ;
- Scrivere esplicitamente l'algoritmo iterativo.
- Calcolare esplicitamente le prime iterazioni, finché le prime due cifre decimali si stabilizzano.

#### Convessità

Siano  $f, g$  due funzioni definite su tutto  $\mathbb{R}$ , derivabili due volte. Supponiamo che  $f$  sia convessa e  $g$  concava, su tutto  $\mathbb{R}$ . Utilizzando le proprietà geometriche enunciate nel paragrafo 5.2 per le funzioni concave e convesse, dimostrare che  $f$  e  $g$  non possono avere più di due intersezioni.

Calcolare i seguenti limiti utilizzando opportunamente, se necessario, gli sviluppi di Taylor.

$$\lim_{x \rightarrow +\infty} \frac{\log \sqrt{1+x^2}}{\sqrt[4]{x}}$$

$$\lim_{x \rightarrow +\infty} x^\alpha (\sqrt{x^2 + 2x + 3} - x - 1)$$

$$\lim_{x \rightarrow +\infty} (\sqrt{x^2 + x + 1} - x)$$

Sviluppare in serie di Taylor in vicinanza di  $x = 0$  all'ordine 5 le seguenti funzioni:

$$\frac{\sin x}{x}, \quad \log(1 + \cos x), \quad \frac{1}{\cos x}, \quad \frac{1}{\sqrt{3+x}}$$

Se per  $x \rightarrow 0$  (o  $x \rightarrow 0^\pm$  se il caso lo richiede) si verifica che  $f(x) \sim kx^\alpha$  ( $k \neq 0$ ,  $\alpha > 0$ ), si dice che  $\alpha$  è l'ordine di infinitesimo di  $f$  rispetto all'infinitesimo campione  $x$ .

Utilizzando opportunamente gli sviluppi di Taylor, determinare l'ordine di infinitesimo per  $x \rightarrow 0$  delle seguenti funzioni

$$i) \quad \sin x - x \quad ii) \quad (\sin x)^2 - x^2 \quad iii) \quad \cos(2\sqrt{x}) - e^{-2x}$$

Gli sviluppi di Taylor e il teorema di L'Hospital sono due diversi strumenti per il calcolo dei limiti; spesso entrambi i metodi possono essere utilizzati. Provare a rifare gli esercizi 57, 58, 60, 61, 62, 63, 64 utilizzando opportunamente gli sviluppi di Taylor.

## COMPLEMENTI

**1** Si considerino le funzioni iperboliche inverse  $\operatorname{Sh}^{-1}x$ ,  $\operatorname{Ch}^{-1}x$  (v. cap. 2, par. 4.4). Calcolare la derivata di queste funzioni inverse:

- applicando il teorema della derivata della funzione inversa, sfruttando perciò le formule per la derivata di  $\operatorname{Sh}x$  e  $\operatorname{Ch}x$ , e la relazione tra le due funzioni;
- derivando direttamente l'espressione esplicita della funzione inversa, calcolata nel cap. 2, par. 4.4. (Naturalmente, si deve trovare lo stesso risultato).

**2** Fare un esempio di funzione derivabile  $n$  volte ma non  $n-1$  volte, per  $n$  intero positivo qualsiasi.

**3** Si consideri la funzione

$$f_\alpha(x) = \begin{cases} x^\alpha \sin \frac{1}{x} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

Stabilire per quali valori del parametro reale  $\alpha$ , la funzione  $f_\alpha$ : è limitata; è continua; è derivabile in 0; ha derivata prima continua in 0; è derivabile due volte.

**4** Dimostrare il Teorema enunciato nel par. 7.2 sulla relazione tra “asintotico” e “o piccolo”.

**5** Dimostrare la seguente relazione tra i polinomi di MacLaurin di  $f$  e  $f'$  (relazione che abbiamo utilizzato nella dimostrazione della formula di Taylor con resto secondo Peano):

$$T'_{n,f}(x) = T_{n-1,f'}(x).$$

6 Dimostrare che dato un qualsiasi polinomio di grado  $n$ ,

$$p_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$$

e assegnato un punto  $x_0 \in \mathbb{R}$ , si può esprimere  $p_n$  anche nella forma di polinomio di grado  $n$  centrato in  $x_0$ :

$$p_n(x) = b_n (x - x_0)^n + b_{n-1} (x - x_0)^{n-1} + \dots + b_0$$

per opportuni coefficienti  $b_k$  determinati dai coefficienti  $a_k$ . (Questo fatto è implicitamente utilizzato nel provare il Teorema che caratterizza il polinomio di Taylor).

7 (a) Sia  $f$  una funzione invertibile in  $(a, b)$ , e  $g$  la sua funzione inversa. Nell'ipotesi che  $f$  sia derivabile 2 volte in  $x_0 \in (a, b)$  con  $f'(x_0) \neq 0$ , dimostrare la formula:

$$g''(y_0) = -\frac{f''(x_0)}{[f'(x_0)]^3} \quad (\text{con } y_0 = f(x_0)).$$

(b) Sapendo che

$$f(x) = a + b(x - x_0) + c(x - x_0)^2 + o(x - x_0)^2,$$

scrivere lo sviluppo di Taylor al second'ordine per  $g$  centrato in  $y_0$ , in funzione dei coefficienti  $a, b, c$ .

(c) Supponendo che  $f$  sia 3 volte derivabile in  $x_0$ , calcolare anche  $g'''(y_0)$ .

8 La formula di Taylor fornisce un ulteriore criterio per lo studio della natura dei punti stazionari. Sia  $f$  una funzione derivabile tutte le volte che ci occorrerà in  $x_0$ , e sia  $f'(x_0) = 0$ . Dimostrare che:

se il minimo intero  $n$  per cui risulta  $f^{(n)}(x_0) \neq 0$  è pari, allora  $x_0$  è un punto di massimo o minimo relativo, e precisamente: è di massimo (minimo) se  $f^{(n)}(x_0) < 0 (> 0)$ ; se invece il minimo intero  $n$  per cui risulta  $f^{(n)}(x_0) \neq 0$  è dispari, allora  $x_0$  non è un estremante.

# 5

## Serie

### 1 SERIE NUMERICHE

#### 1.1 Definizione e primi esempi

Nel capitolo 3 abbiamo introdotto il concetto di *limite di successione*, come punto di partenza per poter definire quello di *limite di funzione*, che è il fondamento concettuale del calcolo infinitesimale nel *continuo*, ossia il calcolo differenziale, trattato nel cap.4, e integrale, che tratteremo nel cap.6. In questo capitolo vedremo come il concetto di limite di successione permetta un altro importante sviluppo del calcolo infinitesimale, questa volta nell'ambito del *discreto*. Si tratta del concetto di *serie*, che come vedremo consente di estendere l'elementare operazione algebrica di somma a un numero infinito di addendi.

La possibilità di sommare infiniti numeri, magari tutti positivi, e ottenere un risultato finito, era qualcosa di paradossale per gli antichi filosofi greci: Zenone di Elea (intorno al 400 a.C.) formulò il celebre paradosso di Achille e la tartaruga.

In una gara di corsa Achille ( $A$ ) e la tartaruga ( $T$ ) partono inizialmente (tempo  $t_0$ ) dalle posizioni  $x_0$  e  $x_1 > x_0$  rispettivamente (fig. 5.1). In un tempo  $t_1$ ,  $A$  raggiungerà la posizione  $x_1$  ma, nello stesso tempo,  $T$  avrà raggiunto una posizione  $x_2 > x_1$ . Per

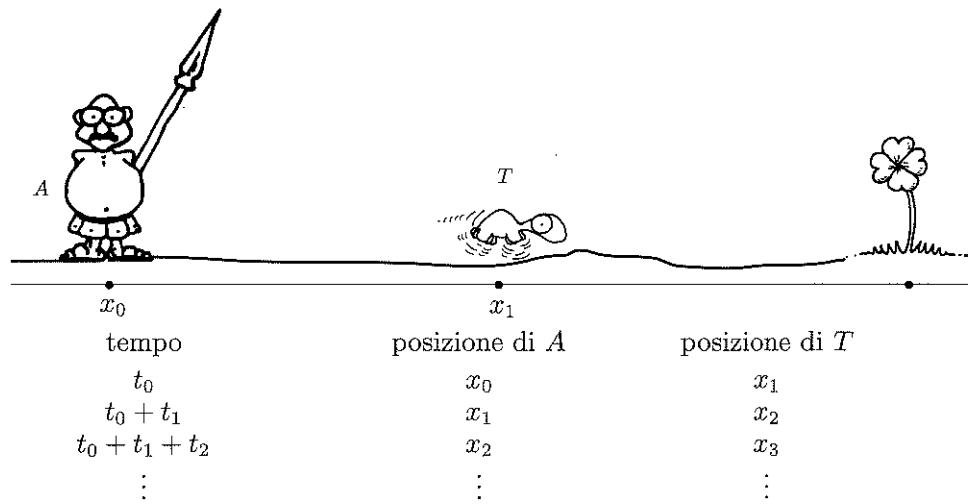


Figura 5.1.

arrivare a  $x_2$ ,  $A$  impiegherà un tempo  $t_2$ , ma in questo tempo  $T$  si sarà spostata in  $x_3$  e così via...

O si ammette che la somma di infinite quantità positive:  $t_0 + t_1 + t_2 + \dots$  possa dare un risultato finito, o si deve concludere che Achille non riuscirà mai a raggiungere la tartaruga!

D'altra parte non è così paradossale che una somma di infiniti addendi positivi, una volta che sia stata ben definita, possa dare un risultato finito; convinciamoci con questo semplice esempio. Immaginiamo di misurare un'asta lunga 2 m usando questo bizzarro procedimento: dividiamo l'asta a metà e misuriamo il primo pezzo: otterremo 1; poi dividiamo a metà il secondo pezzo e misuriamone la prima parte: otterremo  $\frac{1}{2}$ ; il rimanente pezzo sia diviso ancora a metà... e così via indefinitamente.

Otteniamo una somma infinita

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^n} + \dots$$

la quale, secondo le aspettative, dovrebbe avere come risultato 2.

Vediamo ora di precisare il discorso fatto con opportune definizioni.

Data una successione di numeri reali  $\{a_n\}$ , chiamiamo *serie dei termini*  $a_n$  la *scrittura formale*

$$(1.1) \quad \sum_{n=0}^{\infty} a_n$$

che si legge “serie (ma anche “somma”) per  $n$  da 0 a  $\infty$  di  $a_n$ ”. Per dare significato a questo simbolo, che intuitivamente rappresenta l'operazione di somma degli infiniti addendi  $a_n$ , costruiamo anzitutto un'altra successione,  $\{s_n\}$ , i cui termini sono così definiti:

$$(1.2) \quad \begin{aligned} s_0 &= a_0 \\ s_1 &= a_0 + a_1 \\ s_2 &= a_0 + a_1 + a_2 \\ &\dots \\ s_n &= a_0 + a_1 + \dots + a_n \\ &\dots \end{aligned}$$

Il numero  $s_n$  viene detto *somma parziale* (o ridotta)  $n$ -esima della serie (1.1), e la successione  $\{s_n\}$  si dice *successione delle somme parziali* della serie (1.1). Notare che le (1.2) si possono condensare nell'unica scrittura sintetica<sup>1</sup>:

$$(1.3) \quad s_n = \sum_{k=0}^n a_k \quad \text{per } n = 0, 1, 2, \dots$$

<sup>1</sup>Si presta attenzione, nelle formule precedenti, al diverso ruolo giocato dai due indici ( $k$  e  $n$ ): l'indice rispetto a cui si somma ( $n$  nella (1.1);  $k$  nella (1.3)) è sempre un indice muto: ad esempio la (1.3) si potrebbe scrivere anche così:

$$s_n = \sum_{j=0}^n a_j \quad \text{per } n = 0, 1, 2, \dots$$

**DEFINIZIONE 5.1** Diremo che la serie (1.1) è *convergente*, *divergente*, *irregolare*, se la successione  $\{s_n\}$  delle sue somme parziali è *convergente*, *divergente* o *irregolare*, rispettivamente. In particolare, se  $\{s_n\}$  è convergente,  $s_n \rightarrow s$ , diremo che  $s$  è la *somma della serie*, e scriveremo:

$$\sum_{n=0}^{\infty} a_n = s$$

In questo caso dunque vale la relazione:

$$\sum_{n=0}^{\infty} a_n = \lim_{n \rightarrow +\infty} \sum_{k=0}^n a_k = \lim_{n \rightarrow +\infty} s_n$$

L'ultima formula scritta è interessante perché spiega in che modo il concetto di serie traduca con precisione l'idea di “somma di infiniti addendi”: si calcola il *limite*, per  $n \rightarrow \infty$ , della *somma finita* dei primi  $n$  addendi.

L'espressione “studiare il carattere della serie” significa stabilire se la serie è *convergente* o *divergente*, o *irregolare*.

Talvolta, invece di sommare a partire da 0, si parte da un indice  $N > 0$ ; scriveremo allora  $\sum_{n=N}^{\infty} a_n$ . Per indicare una serie useremo talvolta anche il simbolo sintetico  $\sum a_n$ . In questo caso sarà chiaro dal contesto a partire da quale indice  $N$  occorra sommare.

**OSSERVAZIONE** Parlare di una serie numerica coinvolge sempre *due* diverse successioni: la successione  $\{a_n\}$  dei *termini della serie* ( $a_n$  si dice anche “termine generale della serie”) e la successione  $\{s_n\}$  delle sue *somme parziali*. Lo studente presti bene attenzione, volta per volta, a quale delle due successioni si riferiscono le affermazioni fatte. Per esempio, abbiamo visto che la serie  $\sum a_n$  si dice convergente se la successione  $\{s_n\}$  è convergente (non se la successione  $\{a_n\}$  è convergente!).



**(Serie geometrica)** Sia  $a_n = q^n$ ,  $q \in \mathbb{R}$ . Se  $q \neq 1$ , utilizzando la formula (1.1) del capitolo 1, abbiamo

$$s_n = 1 + q + q^2 + \dots + q^n = \frac{1 - q^{n+1}}{1 - q}.$$

Se invece  $q = 1$  abbiamo  $s_n = n + 1$ . Prendendo il limite, per  $n \rightarrow +\infty$ , otteniamo:

$$\lim_{n \rightarrow \infty} s_n = \begin{cases} \frac{1}{1 - q} & \text{se } |q| < 1 \\ +\infty & \text{se } q \geq 1 \\ \text{non esiste} & \text{se } q \leq -1 \end{cases}$$

e pertanto la serie  $\sum_{n=0}^{\infty} q^n$  è

$\begin{cases} \text{convergente (con somma } \frac{1}{1 - q} \text{)} & \text{se }  q  < 1 \\ \text{divergente a } +\infty & \text{se } q \geq 1 \\ \text{irregolare} & \text{se } q \leq -1 \end{cases}$
--

 (Serie armonica) È la serie

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n} + \dots$$

Dimostreremo (nel cap. 6 par. 8.3) che la successione  $s_n$  delle somme parziali è  $\geq \log(n+1)$  (e perciò, in particolare, tende a  $+\infty$ ).

 (Serie di Mengoli) È la serie:

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$$

Osservando che  $\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$ , si riesce a dare un'espressione semplice alla successione  $s_n$ :

$$s_n = \sum_{k=1}^n \left[ \frac{1}{k} - \frac{1}{k+1} \right] = \left[ 1 - \frac{1}{2} \right] + \left[ \frac{1}{2} - \frac{1}{3} \right] + \dots + \left[ \frac{1}{n} - \frac{1}{n+1} \right] =$$

(grazie al fatto che i termini si semplificano 2 a 2)

$$= 1 - \frac{1}{n+1}$$

Dunque  $s_n \rightarrow 1$ , ossia la serie converge e ha somma 1.

La serie di Mengoli è il più semplice esempio di *serie telescopica*, che significa quanto segue. Il termine generale  $a_k$  ha la forma  $(b_k - b_{k+1})$  (dove  $b_k$  è un'altra opportuna successione) e di conseguenza, grazie alle cancellazioni, si ha

$$s_n = b_1 - b_{n+1}$$

Se il termine  $b_n \rightarrow 0$ , la serie è convergente e ha somma  $b_1$ .

Supponiamo che la serie  $\sum_{n=1}^{\infty} a_n$  sia convergente e che  $s$  sia la sua somma. Ciò significa che  $s_n \rightarrow s$  e quindi

$$a_n = (s_n - s_{n-1}) \rightarrow (s - s) = 0$$

Ne segue che:

**TEOREMA 5.1** Condizione necessaria affinché una serie  $\sum_{n=0}^{\infty} a_n$  converga è che il termine generale  $a_n$  tenda a zero.

Come mostra l'esempio della *serie armonica*, la condizione non è sufficiente. Comunque, se il termine generale non tende a zero, certamente la serie non converge. Ad esempio  $\sum_{n=1}^{\infty} \cos \frac{1}{n}$  non converge, perché  $\cos \frac{1}{n} \rightarrow 1$ .

Talvolta è utile anche il seguente fatto:

**TEOREMA 5.2** Supponiamo che

$$\sum_{k=0}^{\infty} a_k$$

sia convergente. Allora per ogni  $n$  risulta convergente anche la serie (che si dice resto della serie di partenza)

$$R_n = \sum_{k=n}^{\infty} a_k; \text{ inoltre } R_n \rightarrow 0 \text{ per } n \rightarrow \infty.$$

Detto in parole: il resto di una serie convergente è infinitesimo.

**DIMOSTRAZIONE.** Sia  $s_n = \sum_{k=0}^n a_k$  e sia

$$S_h^{R_n} = \sum_{k=n}^h a_k \text{ (per } h \geq n\text{)}$$

la somma parziale  $h$ -esima della serie  $R_n$ . Dall'identità, valida per ogni  $h > n$ ,

$$S_h^{R_n} + s_{n-1} = s_h$$

si ha, passando al limite per  $h \rightarrow \infty$ ,

$$R_n + s_{n-1} = s.$$

Per  $n \rightarrow \infty$ ,  $s_{n-1} \rightarrow s$  e quindi  $R_n = s - s_{n-1} \rightarrow 0$ .

Nei prossimi paragrafi studieremo vari criteri di convergenza, che forniscono alcune condizioni sufficienti affinché una serie converga; il problema di calcolarne esplicitamente la somma, invece, esula per il momento dai nostri obiettivi.

## 1.2 Serie a termini non negativi

Per una serie a termini non negativi si hanno alcuni semplici criteri sufficienti di convergenza. Cominciamo con l'osservare che la successione delle somme parziali di una tale serie sarà crescente, poiché

$$s_{n+1} = s_n + a_{n+1} \geq s_n$$

Perciò, per il teorema sull'esistenza del limite di una *successione monotona* (v. cap. 3, par. 1.2), esiste  $\lim_{n \rightarrow +\infty} s_n$ : tale limite è finito oppure  $+\infty$  a seconda che la successione  $\{s_n\}$  sia limitata oppure no. In ogni caso risulta:  $\lim_{n \rightarrow +\infty} s_n = \sup_{n \in \mathbb{N}} \{s_n\}$ .

Possiamo perciò affermare che:

una serie  $\sum_{n=0}^{\infty} a_n$  a termini non negativi è convergente o è divergente a  $+\infty$ .  
Essa converge se e solo se la successione delle somme parziali  $n$ -esime è limitata.

### Criterio del confronto

Siano  $\sum a_n$  e  $\sum b_n$  due serie a termini non negativi e tali che

$$a_n \leq b_n \quad \text{definitivamente}$$

Allora valgono le seguenti implicazioni

$$\text{i)} \sum b_n \text{ convergente} \implies \sum a_n \text{ convergente}$$

$$\text{ii)} \sum a_n \text{ divergente} \implies \sum b_n \text{ divergente.}$$

La serie  $\sum b_n$  viene detta *maggiorante*, la  $\sum a_n$  *minorante*.

Questo e i prossimi criteri saranno dimostrati alla fine del paragrafo.



La serie  $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ , con  $\alpha \leq 1$  è divergente.

Infatti l'affermazione è vera per  $\alpha = 1$  (serie armonica); per  $\alpha < 1$  la serie data è maggiorante della serie armonica, perciò diverge.

#### Criterio del confronto asintotico

Se le due successioni (a termini positivi)  $\{a_n\}$  e  $\{b_n\}$  sono asintotiche,

$$a_n \sim b_n$$

allora le corrispondenti serie  $\sum a_n$  e  $\sum b_n$  hanno lo stesso carattere, cioè o sono entrambe convergenti o sono entrambe divergenti.



La serie  $\sum_{n=1}^{\infty} \frac{1}{n^2}$  converge perché  $\frac{1}{n^2} \sim \frac{1}{n(n+1)}$  e la serie  $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$  (di Mengoli) converge.

La serie  $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$  con  $\alpha \geq 2$  converge per il criterio del confronto:  $\frac{1}{n^\alpha} \leq \frac{1}{n^2}$  e  $\sum_{n=1}^{\infty} \frac{1}{n^2}$  converge.

Abbiamo quindi stabilito il carattere della *serie armonica generalizzata*  $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$  per  $\alpha \geq 2$  (convergente) e per  $\alpha \leq 1$  (divergente). Si può dimostrare (una prova di questo fatto sarà fornita nel cap. 6 par. 8.3) che nei casi intermedi ( $1 < \alpha < 2$ ) c'è ancora convergenza. Ricapitolando:

$$\boxed{\sum_{n=1}^{\infty} \frac{1}{n^\alpha} \text{ converge se } \alpha > 1; \text{ diverge se } \alpha \leq 1}$$

La serie  $\sum_{n=1}^{\infty} \frac{5n+\cos n}{3+2n^3}$  converge, perché  $\frac{5n+\cos n}{3+2n^3} \sim \frac{5}{2} \cdot \frac{1}{n^2}$ . (Ovviamente la costante moltiplicativa è ininfluente sul carattere della serie).

Un'osservazione importante riguarda la differenza tra *stabilire il carattere della serie* (ad esempio: converge oppure no) e *calcolare la somma della serie* (nel caso converga). In generale, quando affermiamo in base al criterio del confronto asintotico che la serie  $\sum a_n$  converge perché la serie  $\sum b_n$  converge, ciò non significa affatto che le due serie abbiano la stessa somma. Intuitivamente, ciò è dovuto al fatto che l'affermazione  $a_n \sim b_n$  ci dice che gli addendi delle due somme sono "simili" quando  $n$  è grande;

ma il valore della somma della serie dipende da *tutti* gli addendi, anche i primi. Ad esempio, abbiamo visto che

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1, \quad \text{e} \quad \frac{1}{n(n+1)} \sim \frac{1}{n^2}$$

Tuttavia si può dimostrare (lo vedremo nel vol.2°, trattando le *serie di Fourier*) che

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

Ciò vale a maggior ragione quando si applica il criterio del confronto per affermare la convergenza di una serie.

L'ultimo esempio fatto è interessante anche per un altro motivo. La serie  $\sum_{n=1}^{\infty} \frac{1}{n^2}$ , convergente, ha per somma un numero irrazionale, nonostante il fatto che ogni termine della successione delle somme parziali sia razionale. Ciò significa che se il nostro ambiente di lavoro fosse  $\mathbb{Q}$ , questa serie non sarebbe convergente.

I criteri di convergenza per le serie a termini positivi si basano tutti sulla proprietà dell'estremo superiore, di cui gode l'insieme dei numeri reali.

Facciamo ora un'osservazione importante sull'applicazione pratica del criterio del confronto asintotico. Tutti gli strumenti che abbiamo incontrato nei capitoli 3 e 4 per stabilire stime asintotiche per *funzioni* (limiti notevoli, sviluppi di MacLaurin...) si possono applicare anche per ottenere stime asintotiche per successioni, e quindi forniscono strumenti potenti per lo studio del carattere di una serie a termini positivi:



Si studi il carattere delle seguenti serie

$$\text{a)} \sum_{n=1}^{\infty} \left( \frac{e^{1/n} - 1}{n} \right); \quad \text{b)} \sum_{n=1}^{\infty} \left( \frac{1}{n} - \sin \frac{1}{n} \right); \quad \text{c)} \sum_{n=1}^{\infty} \log \left( \frac{n+1}{n+2} \right)$$

a)  $\frac{e^{1/n} - 1}{n} \sim \frac{1}{n^2}$  (sfruttando il limite notevole  $e^{\varepsilon(x)} - 1 \sim \varepsilon(x)$  per  $\varepsilon(x) \rightarrow 0$ ), perciò la serie, a termini positivi, per confronto asintotico con la serie di  $1/n^2$ , converge.

b) Dallo sviluppo di MacLaurin della funzione  $\sin x$ , si ha (ponendo  $x = \frac{1}{n}$ ):

$$\frac{1}{n} - \sin \frac{1}{n} = \frac{1}{n} - \left( \frac{1}{n} - \frac{1}{6n^3} + o\left(\frac{1}{n^3}\right) \right) \sim \frac{1}{6n^3}.$$

Dunque si tratta di una serie a termini positivi, e per confronto asintotico con la serie di  $1/n^3$ , questa serie converge.

c)  $\log \left( \frac{n+1}{n+2} \right) \sim \frac{n+1}{n+2} - 1 = \frac{-1}{n+2} \sim -\frac{1}{n}$ . (Si è sfruttato il limite notevole  $\log(1 + \varepsilon(x)) \sim \varepsilon(x)$  per  $\varepsilon(x) \rightarrow 0$ ). Dunque si tratta di una serie a termini negativi, il cui termine generale è asintotico a  $-1/n$ . Per confronto con la serie armonica, questa serie diverge ( $\rightarrow -\infty$ ).

**Criterio della radice**

Sia  $\sum a_n$  una serie a termini non negativi. Se esiste il limite

$$\lim_{n \rightarrow +\infty} \sqrt[n]{a_n} = l$$

e  $l > 1$  la serie diverge, se  $l < 1$  la serie converge; se  $l = 1$  nulla si può concludere.



**ESEMPIO** Sia data  $\sum_{n=1}^{\infty} \frac{a^n}{n^n}$  con  $a \geq 0$ , risulta, per  $n \rightarrow +\infty$

$$\sqrt[n]{\frac{a^n}{n^n}} = \frac{a}{n} \rightarrow 0$$

e perciò la serie data converge.

**ESEMPIO** Sia data  $\sum_{n=1}^{\infty} n^{\alpha} a^n$ , con  $\alpha \in \mathbb{R}$ ,  $a > 0$ ; risulta

$$\sqrt[n]{n^{\alpha} a^n} = a n^{\alpha/n}$$

Calcoliamo  $\lim_{n \rightarrow +\infty} n^{\alpha/n}$ ; basterà calcolare (cfr. il confronto di infiniti, cap.3, par.1.5)

$$\lim_{n \rightarrow +\infty} \log(n^{\alpha/n}) = \lim_{n \rightarrow +\infty} \frac{\alpha}{n} \log n = 0 \quad \forall \alpha \in \mathbb{R}$$

e perciò  $\lim_{n \rightarrow +\infty} n^{\alpha/n} = 1$ . Dunque, se  $a < 1$  la serie converge. Se  $a > 1$  la serie diverge. Se

$a = 1$  la serie diventa  $\sum_{n=1}^{\infty} n^{\alpha}$  e già sappiamo che per  $\alpha < -1$  questa serie è convergente, per  $\alpha \geq -1$  divergente.

**Criterio del rapporto**

Sia  $\sum a_n$  una serie a termini positivi. Se esiste il limite

$$\lim_{n \rightarrow +\infty} \frac{a_{n+1}}{a_n} = l$$

e  $l > 1$  la serie diverge, se  $l < 1$  la serie converge; se  $l = 1$  nulla si può concludere.



**ESEMPIO** La serie  $\sum_{n=0}^{\infty} \frac{1}{n!}$  è convergente. Infatti:

$$\lim_{n \rightarrow +\infty} \frac{1/(n+1)!}{1/n!} = \lim_{n \rightarrow +\infty} \frac{1}{n+1} = 0$$

Più avanti mostreremo che la somma di questa serie è  $e$ .

**Dimostrazione dei criteri di convergenza per le serie a termini positivi**

Proviamo ora i criteri di convergenza che abbiamo enunciato. Come si vedrà, questi sono una conseguenza della regolarità delle serie a termini positivi, che a sua volta dipende dal teorema di esistenza del limite per successioni monotone (v. cap. 3, par. 1.2).

**DIMOSTRAZIONE DEL CRITERIO DEL CONFRONTO**

Con le stesse notazioni impiegate nell'enunciato, sia:

$$s_n = \sum_{k=1}^n a_k; \quad s_n^* = \sum_{k=1}^n b_k$$

Poiché  $0 \leq a_k \leq b_k$  per ogni  $k$ , sommando membro a membro le diseguaglianze per  $k$  da 1 a  $n$ , si ottiene  $0 \leq s_n \leq s_n^*$ . Sappiamo già che una serie a termini positivi è regolare, ossia o converge o diverge (non può oscillare). Dunque le affermazioni *i*) e *ii*) sono logicamente equivalenti, perciò basta dimostrare la seconda, che è immediata perché: dire che  $\sum_{n=1}^{\infty} a_n$  diverge significa, per definizione di serie divergente, che  $s_n \rightarrow +\infty$ .

D'altro canto è  $s_n \leq s_n^*$ , perciò, per il criterio del confronto per le successioni, anche  $s_n^* \rightarrow +\infty$ , ossia  $\sum_{n=1}^{\infty} b_n$  diverge.

**DIMOSTRAZIONE DEL CRITERIO DEL CONFRONTO ASINTOTICO**

Dire che  $a_n \sim b_n$  per  $n \rightarrow \infty$  significa che  $\frac{a_n}{b_n} \rightarrow 1$  per  $n \rightarrow \infty$ ; questo implica che per ogni  $\varepsilon > 0$  sia definitivamente  $1 - \varepsilon < \frac{a_n}{b_n} < 1 + \varepsilon$ , ad esempio (scegliendo  $\varepsilon = \frac{1}{2}$ ) che sia

$$\frac{1}{2} < \frac{a_n}{b_n} < \frac{3}{2}$$

ossia (poiché per ipotesi  $b_n > 0$ )

$$\frac{1}{2} b_n < a_n < \frac{3}{2} b_n \text{ definitivamente.}$$

Per il teorema del confronto, la prima delle due diseguaglianze implica che se  $\sum a_n$  converge, anche  $\sum b_n$  converge, mentre la seconda implica che se  $\sum a_n$  diverge, anche  $\sum b_n$  diverge. Questo significa appunto che le serie hanno lo stesso carattere.

**DIMOSTRAZIONE DEL CRITERIO DELLA RADICE**

Supponiamo prima che sia:

$$\lim_{n \rightarrow +\infty} \sqrt[n]{a_n} = l < 1$$

Poiché  $\sqrt[n]{a_n} \rightarrow l$ , allora fissato comunque un  $\varepsilon > 0$ , definitivamente  $\sqrt[n]{a_n} \leq l + \frac{\varepsilon}{2}$ .

D'altro canto  $l < 1$ , perciò è anche  $l < 1 - \varepsilon$  per un  $\varepsilon > 0$  opportuno.

Per questo  $\varepsilon$  si ha dunque che, definitivamente

$$\sqrt[n]{a_n} \leq l + \frac{\varepsilon}{2} < (1 - \varepsilon) + \frac{\varepsilon}{2} = 1 - \frac{\varepsilon}{2}$$

e quindi

$$a_n < \left(1 - \frac{\varepsilon}{2}\right)^n.$$

Per confronto con la serie geometrica convergente  $\sum_{n=1}^{\infty} \left(1 - \frac{\varepsilon}{2}\right)^n$ , la serie di partenza converge.

Se ora, invece, è

$$\lim_{n \rightarrow +\infty} \sqrt[n]{a_n} = l > 1$$

con un ragionamento simile si deduce che

$$a_n > \left(1 + \frac{\varepsilon}{2}\right)^n$$

definitivamente, per un certo  $\varepsilon > 0$ . Dunque  $a_n \rightarrow +\infty$ , e la serie diverge.  $\diamond$

#### DIMOZRAZIONE DEL CRITERIO DEL RAPPORTO

È simile alla precedente. Supponiamo prima che sia:

$$\lim_{n \rightarrow +\infty} \frac{a_{n+1}}{a_n} = l < 1$$

Ragionando come nella dimostrazione precedente, si ha che, per qualche  $\varepsilon > 0$  e intero  $n_0$ ,

$$\frac{a_{n+1}}{a_n} < \left(1 - \frac{\varepsilon}{2}\right)$$

per ogni  $n \geq n_0$ . Ciò implica, ragionando iterativamente, che

$$a_{n+1} < \left(1 - \frac{\varepsilon}{2}\right) a_n < \left(1 - \frac{\varepsilon}{2}\right) \cdot \left(1 - \frac{\varepsilon}{2}\right) a_{n-1} < \dots < \left(1 - \frac{\varepsilon}{2}\right)^{n-n_0+1} a_{n_0}.$$

Per confronto con la serie geometrica convergente  $\sum_{n=1}^{\infty} \left(1 - \frac{\varepsilon}{2}\right)^n a_1$ , la serie di partenza converge.

Se ora, invece, è

$$\lim_{n \rightarrow +\infty} \frac{a_{n+1}}{a_n} = l > 1$$

con un ragionamento simile si deduce che

$$a_{n+1} > \left(1 + \frac{\varepsilon}{2}\right)^n a_1$$

definitivamente, per un certo  $\varepsilon > 0$ . Dunque  $a_n \rightarrow +\infty$ , e la serie diverge.  $\diamond$

Vale la pena osservare il seguente fatto, che talvolta ci sarà utile: come si vede dalle dimostrazioni appena svolte, sia nel criterio della radice che in quello del rapporto, se  $l > 1$  non solo la serie non converge ma, addirittura, il termine generale della serie tende a  $+\infty$ .

**OSSERVAZIONE** Abbiamo già osservato che il carattere di una serie non cambia se noi ne alteriamo un numero finito di termini. Di conseguenza, i criteri visti per le serie a termini non negativi, si applicano anche alle serie con termini *definitivamente* non negativi. Raccogliendo un segno meno dall'intera serie, poi, si vede che questi si

possono applicare anche alle serie a termini non positivi, e quindi alle serie a termini *definitivamente non positivi*. In sintesi, quindi, questi criteri si applicano alle serie che hanno tutti i termini (tranne un numero finito) dello stesso segno. Nel prossimo paragrafo esamineremo invece le serie che possiedono infiniti termini positivi e infiniti termini negativi.

### 1.3 Serie a termini di segno variabile

Veniamo ora allo studio delle serie a termini di segno qualsiasi. Diamo anzitutto una definizione:

**DEFINIZIONE 5.2** Una serie  $\sum a_n$  si dirà assolutamente convergente se converge la serie (a termini non negativi)  $\sum |a_n|$ .  $\diamond$

Vale il seguente teorema:

**TEOREMA 5.3** Se la serie  $\sum a_k$  converge assolutamente, allora converge.

**DIMOSTRAZIONE.** Lo proveremo come conseguenza del teorema di monotonia, decomponendo il termine generale della serie nella sua parte positiva e negativa. Per ogni  $k$ , definiamo:

$$\begin{aligned} s_n^+ &= \text{somma dei soli termini } a_k \geq 0, \text{ per } k = 0, 1, \dots, n \\ s_n^- &= -(\text{somma dei soli termini } a_k < 0, \text{ per } k = 0, 1, \dots, n) \\ s_n &= \sum_{k=0}^n a_k. \end{aligned}$$

Poiché  $s_n = s_n^+ - s_n^-$ , sarà sufficiente mostrare che le successioni  $s_n^+$  e  $s_n^-$  sono convergenti per concludere che  $s_n$  converge, e quindi che la serie  $\sum_{k=0}^{\infty} a_k$  converge.

Osserviamo che  $s_n^+$  e  $s_n^-$  sono successioni monotone non decrescenti (aumentando  $n$  di un'unità si somma ad  $s_n^+, s_n^-$  una quantità positiva, oppure non si somma nulla). Inoltre, per la diseguaglianza triangolare:

$$s_n^+ \leq \sum_{k=0}^n |a_k|; \quad s_n^- \leq \sum_{k=0}^n |a_k|.$$

D'altro canto, per ipotesi la serie  $\sum a_k$  converge assolutamente, ossia  $\sum_{k=0}^{\infty} |a_k|$  converge, e quindi la quantità  $\sum_{k=0}^n |a_k|$  è limitata; perciò le successioni  $s_n^+, s_n^-$  sono superiormente limitate e non decrescenti, e pertanto convergono, per il teorema di monotonia. Il teorema è così dimostrato.  $\diamond$

Dunque la convergenza assoluta implica la convergenza (ordinaria), detta anche *convergenza semplice*; il viceversa non è vero.

#### Esempio

La serie  $\sum_{n=1}^{\infty} (-1)^n / n^{\alpha}$ , con  $\alpha > 1$  è convergente, anzi assolutamente convergente. Infatti si ha

$$\left| \frac{(-1)^n}{n^{\alpha}} \right| = \frac{1}{n^{\alpha}}$$

e la serie  $\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}$  è convergente per  $\alpha > 1$ . Se  $\alpha = 1$  nulla, per il momento, possiamo concludere circa la convergenza della serie data, poiché la serie dei valori assoluti corrispondente è la serie armonica, che è divergente.

Mostreremo ora che  $\sum_{n=1}^{\infty} (-1)^n/n$  è convergente, fornendo così un esempio di serie convergente, ma non assolutamente convergente.

#### Serie a termini di segno alternato. Criterio di Leibniz

Tra le serie a termini di segno variabile, un caso particolarmente semplice è costituito dalle serie a segni alternati, per le quali vale il seguente criterio di convergenza:

**TEOREMA 5.4 (CRITERIO DI LEIBNIZ)** *Sia data la serie*

$$\sum_{n=0}^{\infty} (-1)^n a_n \text{ con } a_n \geq 0 \forall n.$$

Se:

- (i) la successione  $\{a_n\}$  è decrescente;
- (ii)  $a_n \rightarrow 0$  per  $n \rightarrow \infty$

allora la serie è convergente. Inoltre, le somme parziali di indice pari approssimano la somma per eccesso, quelle di indice dispari per difetto; il resto della serie è maggiorato, in valore assoluto, dal primo termine trascurato. In formule, detta  $s$  la somma della serie, si ha:

$$(i) \quad s_{2n} = \sum_{k=0}^{2n} (-1)^k a_k \downarrow s;$$

$$(ii) \quad s_{2n+1} = \sum_{k=0}^{2n+1} (-1)^k a_k \uparrow s;$$

$$(iii) \quad |R_n| = \left| \sum_{k=n}^{\infty} (-1)^k a_k \right| \leq a_n.$$

Osserviamo che, analogamente a quanto osservato riguardo ai criteri di convergenza per le serie a termini positivi, il criterio di Leibniz può essere applicato anche se i termini sono definitivamente di segno alterno e la successione  $a_n$  è definitivamente decrescente.

Si noti che la successione delle somme parziali di una serie a segni alternati ha la forma:

$$\begin{aligned} s_0 &= a_0 \\ s_1 &= a_0 - a_1 \\ s_2 &= a_0 - a_1 + a_2 \\ &\dots \\ s_n &= a_0 - a_1 + a_2 - a_3 + \dots + (-1)^n a_n \end{aligned}$$

dove gli  $a_i$  sono tutti  $\geq 0$ . Se allora la successione  $a_n$  tende a zero monotonamente, l'andamento di  $a_n$  e di  $s_n$  è del tipo descritto dalla figura seguente, che sostanzialmente contiene la dimostrazione del criterio di Leibniz:

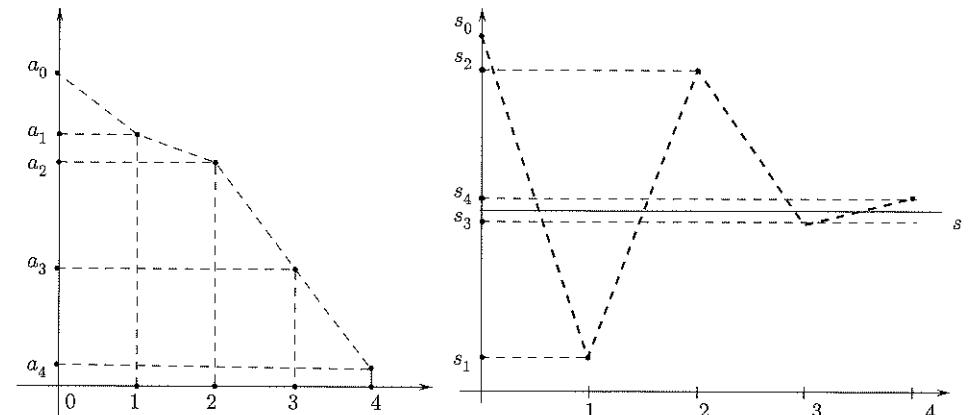


Figura 5.2.

**DIMOOSTRAZIONE.** Dimostriamo formalmente il teorema. Consideriamo la successione delle somme parziali,  $s_n$ , e le due successioni, estratte da questa,  $s_{2n}$  e  $s_{2n+1}$ . Consideriamo dapprima la successione  $s_{2n}$ , così formata:

$$\begin{aligned} s_0 &= a_0 \\ s_2 &= s_0 - a_1 + a_2 \leq s_0 \text{ perché } a_2 \leq a_1 \\ s_4 &= s_2 - a_3 + a_4 \leq s_2 \text{ perché } a_4 \leq a_3 \\ &\dots \end{aligned}$$

Come si vede, la successione  $\{s_{2n}\}$  è monotona decrescente. In modo analogo si dimostra che la successione  $\{s_{2n+1}\}$  è monotona crescente. Inoltre,

$$s_1 \leq s_{2n+1} = s_{2n} - a_{2n+1} \leq s_{2n} \leq s_0$$

perciò  $\{s_{2n+1}\}$  è superiormente limitata e  $\{s_{2n}\}$  è inferiormente limitata. Le due successioni sono quindi convergenti, per il teorema di monotonia; possiamo anche dire che convergono allo stesso limite, perché

$$0 \leq s_{2n} - s_{2n+1} \leq a_{2n+1} \rightarrow 0.$$

Detto  $s$  questo limite comune, abbiamo allora provato le (i). Inoltre, dal fatto che  $s_{2n+1} \rightarrow s$  e  $s_{2n} \rightarrow s$  si deduce che  $s_n \rightarrow s$ .

Proviamo ora (ii). Poiché  $s_{2n+1} \uparrow s$  e  $s_{2n} \downarrow s$ , possiamo scrivere

$$\begin{aligned} s_{2n-1} &\leq s \leq s_{2n} \\ s_{2n+1} &\leq s \leq s_{2n} \end{aligned}$$

da cui si deduce

$$\begin{aligned} 0 &\leq s - s_{2n-1} \leq s_{2n} - s_{2n-1} = a_{2n} \\ 0 &\leq s_{2n} - s \leq s_{2n} - s_{2n+1} = a_{2n+1}. \end{aligned}$$

Perciò (per  $n$  sia pari che dispari) si ha

$$|R_n| = |s - s_{n-1}| \leq a_n$$

che è la (ii).



Applicando il criterio di Leibniz osserviamo che le serie

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \quad \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$$

sono entrambe convergenti (la prima è assolutamente convergente, la seconda no). La prima serie è rapidamente convergente. Sommando infatti i primi 6 termini della serie

$$1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} = \frac{11}{30} = 0,3\bar{6}$$

si ottiene un valore per difetto della somma con un errore che non supera  $1/6! = 1/720$ . Mostreremo nel par. 2.1 che tale somma è  $\frac{1}{e}$ .

La seconda serie è lentamente convergente; per avere un valore della somma approssimato (per eccesso) a meno di  $1/100$  bisogna sommare 99 termini! Come vedremo (par. 2.1), la somma di questa serie è  $\log 2$ ; ma la serie non è adatta al calcolo approssimato di questo numero.



$$\sum_{n=1}^{\infty} (-1)^n \frac{n-1}{n^2+n}$$

È una serie a segni alterni,  $a_n = \frac{n-1}{n(n+1)}$  è positivo e tende a zero. Poiché  $a_n \sim \frac{1}{n}$ , la serie non converge assolutamente; vediamo dunque se è applicabile il criterio di Leibniz. Dobbiamo verificare se è decrescente, ovvero:

$$\frac{a_{n+1}}{a_n} \leq 1 \quad \frac{n}{(n+1)(n+2)} \leq \frac{n-1}{n(n+1)}$$

ossia

$$n^2 \leq n^2 + n - 2$$

vero per  $n \geq 2$ . Dunque si può applicare il criterio di Leibniz e concludere che la serie converge.



$$\sum_{n=2}^{\infty} (-1)^n \frac{\log n}{n}$$

Si tratta di una serie a segni alterni, poiché  $a_n = \frac{\log n}{n}$  è positiva; inoltre  $a_n \rightarrow 0$  (gerarchia degli infiniti). La serie non converge assolutamente, perché  $\frac{\log n}{n} > \frac{1}{n}$  (e  $\sum \frac{1}{n}$  diverge); vediamo se è applicabile il criterio di Leibniz. In questo caso non è facile provare la monotonia della successione per via puramente algebrica, cioè studiando la disequazione

$$a_{n+1} \leq a_n \quad \frac{\log(n+1)}{n+1} \leq \frac{\log n}{n}.$$

E' più comodo invece studiare la monotonia della successione  $a_n$  mediante il calcolo differenziale (si veda quanto osservato a riguardo nel cap. 4, par. 4.2): posto

$$f(x) = \frac{\log x}{x} \quad \text{per } x > 0$$

(con  $x$  variabile reale), calcoliamo:

$$f'(x) = \frac{1 - \log x}{x^2} \leq 0 \quad \text{per } x \geq e$$

Ne segue che  $f$  è decrescente per  $x \geq e$ ; di conseguenza la successione  $a_n = f(n)$  è decrescente almeno per  $n \geq 3$  (il primo intero  $> e$ ). Pertanto, per il criterio di Leibniz la serie converge.



$$\sum_{n=1}^{\infty} \frac{n+1+(-1)^n n^2}{n^3}$$

Scriviamo:

$$\sum_{n=1}^{\infty} \frac{n+1+(-1)^n n^2}{n^3} = \sum_{n=1}^{\infty} \frac{n+1}{n^3} + \sum_{n=1}^{\infty} \frac{(-1)^n}{n}$$

La prima serie converge perché è a termini positivi e  $a_n \sim \frac{1}{n^2}$ ; la seconda converge per il criterio di Leibniz; quindi la serie di partenza converge. Abbiamo qui usato un fatto generale:

$$\text{se } \sum_{k=1}^{\infty} a_k \text{ converge e } \sum_{k=1}^{\infty} b_k \text{ converge, allora } \sum_{k=1}^{\infty} (a_k + b_k) \text{ converge,}$$

come si verifica vedendo la serie come limite della successione delle somme parziali, e applicando il teorema sul limite della somma.



$$\sum_{n=1}^{\infty} (-1)^n \left[ \frac{\sqrt{n} + (-1)^n}{n} \right]$$

Scriviamo:

$$\sum_{n=1}^{\infty} (-1)^n \left[ \frac{\sqrt{n} + (-1)^n}{n} \right] = \sum_{n=1}^{\infty} \frac{(-1)^n}{\sqrt{n}} + \sum_{n=1}^{\infty} \frac{1}{n}$$

La prima serie converge per il criterio di Leibniz; la seconda diverge (serie armonica); quindi la serie di partenza diverge. Più in generale:

$$\text{se } \sum_{k=1}^{\infty} a_k \text{ converge e } \sum_{k=1}^{\infty} b_k \text{ diverge, allora } \sum_{k=1}^{\infty} (a_k + b_k) \text{ diverge,}$$

per lo stesso argomento usato nell'esempio precedente.

Questo esempio è interessante anche per il seguente motivo. Si noti che:

$$a_n = \frac{\sqrt{n} + (-1)^n}{n} \geq 0 \text{ e } a_n \rightarrow 0$$

Dunque la serie di partenza è una serie a segni alterni, e il suo termine generale tende a zero. Se affermassimo che essa converge per il criterio di Leibniz, diremmo il falso (abbiamo visto che la serie diverge). Il punto è che non abbiamo verificato l'altra ipotesi del criterio di Leibniz, ossia che  $\frac{\sqrt{n}+(-1)^n}{n}$  sia monotona decrescente. Evidentemente, quest'ipotesi non è verificata. (Per esercizio lo studente tabuli i primi 10 valori della successione, e osservi cosa accade).



Dare una stima asintotica del termine generale delle seguenti serie a termini positivi. Dedurne il carattere della serie (converge o diverge?).

$$\textcircled{1} \quad \sum_{n=1}^{\infty} \frac{n+3}{2n^3 + 2n + 2}$$

$$\textcircled{2} \quad \sum_{n=0}^{\infty} \frac{\sqrt[3]{n}}{\sqrt{n^2 + n + 1}}$$

Stabilire se le seguenti serie a termini positivi convergono o divergono, precisando il criterio utilizzato.

$$\textcircled{3} \quad \sum_{n=1}^{\infty} \frac{n!}{n^n}$$

$$\textcircled{4} \quad \sum_{n=1}^{\infty} \frac{n^{1/n}}{n!}$$

$$\textcircled{5} \quad \sum_{n=0}^{\infty} \frac{2^n}{e^{2n}}$$

$$\textcircled{6} \quad \sum_{n=1}^{\infty} \log\left(\frac{n+2}{n+4}\right)$$

$$\textcircled{7} \quad \sum_{n=0}^{\infty} \cos\left(\frac{n+2}{n^2+4}\right)$$

$$\textcircled{8} \quad \sum_{n=2}^{\infty} \sin\left(\frac{n+2}{n^2+4}\right)$$

Stabilire se le seguenti serie a termini di segno variabile convergono assolutamente e/o semplicemente, precisando il criterio utilizzato.

$$\textcircled{9} \quad \sum_{n=1}^{\infty} \frac{(-1)^n}{\sqrt{n}}$$

$$\textcircled{10} \quad \sum_{n=1}^{\infty} \frac{\cos\left(\frac{\pi}{2}n\right)}{n}$$

$$\textcircled{11} \quad \sum_{n=2}^{\infty} \frac{\sin(\log n)}{n^2 \log n}$$

Studiare il carattere delle seguenti serie, precisando il criterio utilizzato e giustificandone l'applicazione.

$$\textcircled{12} \quad \sum_{n=1}^{\infty} \frac{\sin n + (-1)^n n}{n^2}$$

$$\textcircled{13} \quad \sum_{n=1}^{\infty} \frac{n \cdot 2^n}{e^{n/2}}$$

$$\textcircled{14} \quad \sum_{n=1}^{\infty} \frac{1}{\sqrt{n^3 - n}}$$

$$\textcircled{15} \quad \sum_{n=1}^{\infty} \frac{n + \log n}{(n + \cos n)^3}$$

$$\textcircled{16} \quad \sum_{n=1}^{\infty} \frac{\log n}{n^2}$$

$$\textcircled{17} \quad \sum_{n=1}^{\infty} \frac{2^n + 1}{3^n + n}$$

\textcircled{18} Sapendo che  $\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$  se  $|q| < 1$ , scrivere un algoritmo per calcolare  $\frac{1}{0,815}$  con un errore inferiore a  $10^{-3}$ .

\textcircled{19} Calcolare la lunghezza della spezzata poligonale inscritta nel triangolo  $ABC$  ottenuta mandando successivamente, a partire da  $A$ , le perpendicolari ai lati  $CB$  e  $AB$ . Sia  $a = \overline{AC}$  e  $\gamma$  l'angolo  $\widehat{ACB}$  (v. fig. 5.3).

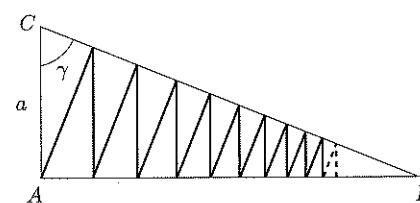


Figura 5.3.

\textcircled{20} Data una circonferenza di raggio  $R$ , si consideri un triangolo equilatero inscritto, quindi una seconda circonferenza inscritta in questo triangolo, poi ancora un triangolo equilatero inscritto nella seconda circonferenza, e così via. Calcolare la somma delle aree degli infiniti triangoli così ottenuti (v. fig. 5.4).

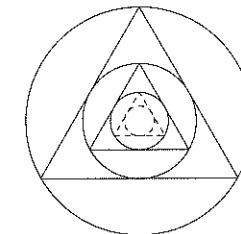


Figura 5.4.

## 2 SERIE DI TAYLOR. ESPONENZIALE COMPLESSO

Il concetto di serie numerica, discusso nel paragrafo precedente, è il primo passo per introdurre serie di tipo più generale, ossia *serie di funzioni*, che si possono vedere come serie numeriche dipendenti da un parametro  $x$ : quando queste serie convergono per ogni valore  $x$  appartenente a un certo intervallo, rappresentano delle funzioni di nuovo tipo. Nel 2° volume di questo corso incontreremo due importanti classi di serie di funzioni, ovvero le serie trigonometriche e le serie di potenze. Qui (par. 2.1) daremo solo un primo cenno alle serie di potenze, dette anche *serie di Taylor*, che si possono vedere come una estensione naturale degli sviluppi di Taylor studiati nel cap.4. Un'altra estensione naturale dei concetti visti finora sulle serie è quello che consiste nello studiare serie *nel campo complesso*: come vedremo, questo permetterà di gettare nuova luce sulle funzioni esponenziali e trigonometriche, scoprendo la loro sorprendente parentela. Di questo ci occuperemo nel par. 2.2.

Il lettore noterà come, mentre nel par.1 ci siamo perlopiù concentrati sul problema di discutere il carattere di una serie, e solo in pochissimi casi ne abbiamo saputo calcolare la somma, in questo paragrafo incontreremo diverse situazioni interessanti in cui si riesce a calcolare la somma: in generale, infatti, il modo naturale per saper calcolare la somma di una serie numerica è vederla (quando questo è possibile) come serie di funzioni, calcolata per un particolare valore della variabile. Questa osservazione potrà essere meglio compresa nello studio delle serie di funzioni che si svilupperà nel 2° vol.

### 2.1 Serie di Taylor delle trascendenti elementari

Riprendiamo ora, per darne una nuova generalizzazione, la formula di Taylor con resto secondo Lagrange (v. cap.4, par. 7.4), che riscriviamo nella forma:

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + E_n(x)$$

dove  $E_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}$ , e  $c$  è un opportuno punto tra  $x_0$  e  $x$ .

Se la funzione  $f$  ha derivate di ogni ordine, si può scrivere la serie

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

detta *serie di Taylor per la funzione f* (centrata in  $x_0$ ). Non abbiamo, per il momento, alcuna garanzia che la serie sia convergente e che la sua somma sia  $f(x)$ . Ciò sarà vero, in un certo punto  $x \in (a, b)$ , se l'errore  $E_n(x) \rightarrow 0$  per  $n \rightarrow +\infty$ ; infatti ciò equivale a dire che la somma parziale  $n$ -esima della serie,  $T_{n,x_0}(x)$ , ammette limite finito e tale limite è precisamente  $f(x)$ .

Se accade che, per ogni  $x$  in un intervallo  $I$ ,  $E_n(x) \rightarrow 0$ , diremo che  $f(x)$  è *sviluppabile in serie di Taylor* nell'intervallo  $I$ .

Ci sono funzioni (infinitamente differenziabili) che sono sviluppabili in serie di Taylor su tutta la retta, altre che lo sono solo su un intervallo limitato, altre per le quali l'intervallo si riduce a un punto solo (il punto  $x_0$ ).

Ci occupiamo ora principalmente delle tre funzioni elementari

$$e^x \quad \cos x \quad \sin x$$

che mostreremo essere sviluppabili in serie di Taylor su tutto  $\mathbb{R}$ .

#### La serie esponenziale

Scriviamo lo sviluppo di MacLaurin all'ordine  $n$  per  $e^x$ , con resto secondo Lagrange: per ogni intero  $n$  ed  $x \in \mathbb{R}$  esiste un punto  $c$ , compreso tra 0 e  $x$ , tale che

$$e^x = \sum_{k=0}^n \frac{x^k}{k!} + \frac{x^{n+1}}{(n+1)!} e^c$$

Fissiamo ora  $x$  e facciamo tendere  $n$  a  $+\infty$ . Il punto  $c$  può variare con  $n$  ma, essendo sempre compreso tra 0 e  $x$ , si può comunque affermare che

$$e^c \leq \begin{cases} e^x & \text{se } x > 0 \\ 1 & \text{se } x < 0 \end{cases}$$

In ogni caso, per  $x$  fissato e  $n \rightarrow +\infty$ ,  $e^c$  si mantiene limitato, mentre

$$\frac{x^{n+1}}{(n+1)!} \rightarrow 0$$

(confronto di infiniti che si dimostra, ad esempio, col criterio del rapporto, v. cap. 3, Teorema 3.11).

Ne segue quindi:

$$e^x = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{x^k}{k!} = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad \text{per ogni } x \in \mathbb{R}$$

Abbiamo quindi dimostrato che la funzione  $e^x$  si può scrivere come somma di una serie di potenze, la sua serie di Taylor, convergente per ogni  $x \in \mathbb{R}$ . In particolare, per  $x = 1$  otteniamo

$$(2.1) \quad e = \sum_{k=0}^{\infty} \frac{1}{k!}.$$

A questo punto conosciamo *due diversi algoritmi* per il calcolo del numero  $e$ : il primo (la definizione stessa di  $e$ ) consiste nel calcolare

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

ovvero, dal punto di vista del calcolo numerico approssimato, calcolare il valore di  $\left(1 + \frac{1}{n}\right)^n$  per un  $n$  molto grande; il secondo consiste nel calcolare la somma della serie (2.1) ovvero, dal punto di vista del calcolo numerico approssimato, calcolare la somma

dei primi  $n$  termini della serie per un  $n$  molto grande. Si verifica che il secondo metodo è molto più efficiente del primo. A titolo di esempio, si confrontino i seguenti valori:

$n$	$\left(1 + \frac{1}{n}\right)^n$	$\sum_{k=0}^n \frac{1}{k!}$	$e$
1	2,	2	2,71828...
2	2,25	2,5	2,71828...
3	2,37037	2,66667	2,71828...
4	2,44141	2,70833	2,71828...
5	2,48832	2,71667	2,71828...
6	2,52163	2,71806	2,71828...
7	2,5465	2,71825	2,71828...
8	2,56578	2,71828	2,71828...
9	2,58117	2,71828	2,71828...
10	2,59374	2,71828	2,71828...

#### Le serie delle funzioni trigonometriche elementari

Un discorso completamente analogo si può ripetere per le funzioni seno e coseno. Il termine  $f^{(n+1)}(c)$  che compare nelle rispettive formule di MacLaurin, ha la forma  $\pm \cos c$  o  $\pm \sin c$ , pertanto ha valore assoluto  $\leq 1$ . Questa limitazione è quel che basta a ripetere la dimostrazione precedente e concludere che

$$\boxed{\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}}$$

$$\boxed{\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}}$$

per ogni  $x \in \mathbb{R}$ .

Le formule trovate sono molto interessanti, per almeno due motivi. Anzitutto, forniscono un algoritmo efficace per il calcolo numerico approssimato delle funzioni trigonometriche elementari, per valori arbitrari delle variabili: il fatto che le serie siano a segni alterni, unito al carattere rapidamente decrescente (in valore assoluto) del termine generale, rendono queste serie particolarmente adatte al calcolo approssimato.



Si voglia calcolare  $\sin(0,8)$ . Posto

$$s_n = \sum_{k=0}^n (-1)^k \frac{(0,8)^{2k+1}}{(2k+1)!}$$

si ha

$$s_{2n-1} \leq \sin(0,8) \leq s_{2n} \text{ per ogni } n.$$

Ad esempio, per  $n = 2$  valgono le disuguaglianze:

$$\begin{aligned} s_3 &\leq \sin(0,8) \leq s_4 \\ \text{ma } s_3 &\simeq s_4 \simeq 0,717356 \end{aligned}$$

il che significa che anche  $\sin(0,8) \simeq 0,717356$ .

In questo caso la convergenza è molto veloce e basta un  $n$  abbastanza piccolo; concettualmente, comunque, il fatto di avere a disposizione una *serie* che converge a  $\sin x$  garantisce di poter eseguire il calcolo con precisione arbitraria.

Approfondendo questa osservazione, notiamo che, ad eccezione dei (pochi) casi in cui  $x$  è un angolo notevole per cui i valori di  $\sin x$  e  $\cos x$  sono noti dalla geometria elementare (es.  $\sin \frac{\pi}{6}$ ), oppure sono ricavabili da questi mediante opportune identità trigonometriche (es.  $\sin \frac{\pi}{12} = \sin(\frac{1}{2} \cdot \frac{\pi}{6}) = \dots$ ), la “definizione” trigonometrica di  $\sin x, \cos x$  non dà alcun criterio *operativo* per il calcolo effettivo del valore della funzione; in un certo senso, quindi, le serie che abbiamo appena scritto si possono considerare come la prima definizione completamente rigorosa di queste funzioni “elementari”!

#### Serie di potenze

In generale, si dice *serie di potenze* una serie del tipo

$$\sum_{k=0}^{\infty} a_k x^k$$

con  $a_k$  costanti reali (o complessi, come vedremo nel prossimo paragrafo), e  $x$  variabile reale (o complessa). La più semplice serie di potenze è la serie geometrica, che converge per  $|x| < 1$ :

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

Altri sviluppi in serie di potenze per funzioni notevoli sono i seguenti:

$$\operatorname{Sh} x = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!} \quad \text{per ogni } x \in \mathbb{R}$$

$$\operatorname{Ch} x = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} \quad \text{per ogni } x \in \mathbb{R}$$

$$\log(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k} \quad \text{per } |x| < 1$$

$$\text{per } \alpha \in \mathbb{R} \quad (1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \quad \text{per } |x| < 1$$

dove si è indicato con

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\dots(\alpha-k+1)}{k!}$$

il coefficiente binomiale  $k$ -esimo.

Nel secondo volume riprenderemo lo studio delle serie di potenze, inquadrandolo nel punto di vista più generale dello studio delle *serie di funzioni*.

#### 2.2 Serie nel campo complesso. Esponenziale complesso

Cominciamo a precisare il concetto di serie a termini complessi. Nel paragrafo 5.1, parlando di serie numeriche, ci siamo concentrati sulle serie a termini reali; tuttavia, le nozioni viste si estendono in modo naturale al caso di una serie

$$\sum_{n=0}^{\infty} a_n \quad \text{con } a_n \in \mathbb{C}$$

Infatti, una serie è definita mediante la successione delle somme parziali  $s_n = \sum_{k=0}^n a_k$ ; per dar senso al concetto di serie a termini complessi è sufficiente dunque dire cosa si intenda per limite di una successione  $s_n$  a valori complessi. Diremo che

$$s_n \rightarrow l \quad (l \in \mathbb{C}) \quad \text{se } |s_n - l| \rightarrow 0.$$

Poiché il modulo di un numero complesso è un numero reale, la nozione di limite nel campo complesso è ricondotta in questo modo a quella di limite nel campo reale.



**TEOREMA** Serie geometrica nel campo complesso. Consideriamo la serie geometrica

$$\sum_{n=0}^{\infty} z^n \quad \text{con } z \in \mathbb{C}.$$

Si può ripetere il ragionamento visto nel par.1.1 a proposito della serie geometrica nel campo reale:

$$s_n = \sum_{k=0}^n z^k = \frac{1 - z^{n+1}}{1 - z}.$$

Se ora  $z$  è un qualsiasi numero complesso di modulo  $|z| < 1$ , si ha

$$|z^{n+1}| = |z|^{n+1} \rightarrow 0, \quad \text{e } s_n \rightarrow \frac{1}{1-z}.$$

Pertanto

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z} \quad \text{per ogni } z \in \mathbb{C}, |z| < 1.$$

Per le serie a termini complessi si dimostra che continua a valere il *criterio della convergenza assoluta*:

**TEOREMA 5.5** Sia  $\sum_{n=0}^{\infty} a_n$  una serie a termini  $a_n$  complessi. Se la serie (a termini reali non negativi)  $\sum_{n=0}^{\infty} |a_n|$  converge, allora converge anche  $\sum_{n=0}^{\infty} a_n$ .

Una dimostrazione di questo teorema è suggerita nei Complementi. Questa semplice condizione permette di provare la convergenza di una serie a termini complessi applicando alla serie dei moduli i criteri che conosciamo per le serie a termini positivi. Applicheremo ora questi fatti alla serie esponenziale.

### Esponenziale complesso

La serie  $\sum_{n=0}^{\infty} \frac{z^n}{n!}$ , la cui somma è  $e^z$  per ogni  $z \in \mathbb{R}$ , si chiama *serie esponenziale*; per mezzo di essa è possibile definire l'espressione  $e^z$  quando  $z = x + iy$  è un numero complesso. Si osserva infatti che la serie

$$\sum_{n=0}^{\infty} \frac{z^n}{n!}$$

è assolutamente convergente per ogni  $z \in \mathbb{C}$  (infatti la serie a termini positivi  $\sum_{n=0}^{\infty} |z|^n/n!$  è convergente, come si vede ad esempio applicando il criterio del rapporto) e perciò la sua somma è un numero complesso che indicheremo con  $e^z$ . Si dimostra che questa definizione è coerente con le proprietà formali dell'esponenziale:

**TEOREMA 5.6** *Definendo*

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

si ha:

$$e^{z_1+z_2} = e^{z_1} \cdot e^{z_2} \quad \forall z_1, z_2 \in \mathbb{C}.$$

Una traccia della dimostrazione di questo teorema sarà fornita nei Complementi. Risulta allora, posto  $z = x + iy$ ,

$$e^z = e^x \cdot e^{iy}$$

dove

$$\begin{aligned} e^{iy} &= \sum_{n=0}^{\infty} \frac{(iy)^n}{n!} = 1 + iy - \frac{y^2}{2} - i\frac{y^3}{3!} + \frac{y^4}{4!} + i\frac{y^5}{5!} - \frac{y^6}{6!} - i\frac{y^7}{7!} + \dots = \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{y^{2k}}{(2k)!} + i \sum_{k=0}^{\infty} (-1)^k \frac{y^{2k+1}}{(2k+1)!} = \cos y + i \sin y \end{aligned}$$

Analogamente si mostra che

$$e^{-iy} = \cos y - i \sin y.$$

Abbiamo così trovato un legame (nel campo complesso) tra le funzioni trigonometriche ( $\sin, \cos$ ) e la funzione esponenziale. Le formule seguenti, dette *formule di Eulero*, sono valide per ogni  $x \in \mathbb{R}$ :

$e^{ix} = \cos x + i \sin x$	$\cos x = \frac{e^{ix} + e^{-ix}}{2}$
$e^{-ix} = \cos x - i \sin x$	$\sin x = \frac{e^{ix} - e^{-ix}}{2i}$

La formula di Eulero  $e^{iy} = \cos y + i \sin y$  è estremamente utile nel calcolo coi numeri complessi. Essa permette di trasformare la forma trigonometrica dei numeri complessi nella forma esponenziale, al modo seguente:

$$z = \rho(\cos \theta + i \sin \theta) = \rho e^{i\theta}$$

con  $\rho \geq 0$  modulo e  $\theta \in \mathbb{R}$  argomento del numero complesso  $z$ . La forma esponenziale dei numeri complessi serve agli stessi scopi della forma trigonometrica, ma è più compatta da scrivere e porta a formule più facili da ricordare.

Ad esempio il teorema di de Moivre sul prodotto di numeri complessi in forma trigonometrica:

$$\rho_1(\cos \theta_1 + i \sin \theta_1) \cdot \rho_2(\cos \theta_2 + i \sin \theta_2) = \rho_1 \rho_2 (\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2))$$

è estremamente naturale in forma esponenziale, perché formalmente è un'applicazione ovvia delle proprietà delle potenze:

$$\rho_1 e^{i\theta_1} \cdot \rho_2 e^{i\theta_2} = \rho_1 \rho_2 e^{i(\theta_1 + \theta_2)}$$

### Il logaritmo nel campo complesso

La funzione logaritmo può essere definita nel campo complesso usando, formalmente, la "solita" definizione: diciamo che

$$w = \log z \Leftrightarrow z = e^w$$

Dato  $z = \rho e^{i\theta}$ , cerchiamo il suo logaritmo  $w = x + iy$ . Dovrà essere:

$$\rho e^{i\theta} = e^{x+iy} = e^x e^{iy}$$

da cui deduciamo

$$\begin{cases} \rho = e^x, & \text{cioè } x = \ln \rho \\ e^{i\theta} = e^{iy}, & \text{cioè } y = \theta + 2k\pi \quad k \in \mathbb{Z} \end{cases}$$

(si ricordi la periodicità di  $e^{i\theta}$ !) In definitiva, si ha:

$$\log(\rho e^{i\theta}) = \ln \rho + i(\theta + 2k\pi) \quad k \in \mathbb{Z}$$

Ciò significa che, nel campo complesso, il logaritmo di un numero non è univocamente determinato, ma ogni numero (diverso da zero) ha infiniti logaritmi (una successione). Si chiama *valore principale del logaritmo* quello che si ottiene per  $k = 0$  e  $\theta \in [0, 2\pi)$ .

Ad esempio, calcoliamo

$$\log(1+i)$$

Poiché  $|1+i| = \sqrt{2}$  e  $\arg(1+i) = \frac{\pi}{4}$ , si ha:

$$\log(1+i) = \frac{1}{2} \ln 2 + i \left( \frac{\pi}{4} + 2k\pi \right)$$

Il logaritmo principale di  $(1+i)$  è  $\frac{1}{2} \ln 2 + i \frac{\pi}{4}$ .

**L'elevamento a potenza complessa**

Si può ora dare senso, nel campo complesso, anche alle potenze a esponente qualsiasi,  $z^\alpha$ , con  $\alpha \in \mathbb{C}$ . Si noti che finora sapevamo calcolare  $z^\alpha$  solo con  $\alpha$  intero (e allora  $z^\alpha$  era univocamente determinato) o  $\alpha = m/n$  razionale (mediante estrazione di radice  $n$ -esima e allora  $z^\alpha$  indicava  $n$  numeri). Vedremo ora che in generale se  $\alpha$  è complesso (o reale ma irrazionale),  $z^\alpha$  indicherà infiniti numeri. Poniamo, per definizione:

$$z^\alpha = e^{\alpha \log z}$$

dove  $\log z$  indica il logaritmo nel campo complesso, definito in precedenza.

Chiameremo *valore principale* di  $z^\alpha$  quello che si ottiene da  $e^{\alpha \log z}$  scegliendo il logaritmo principale di  $z$ . Ad esempio, calcoliamo:

$$i^{\sqrt{2}} = e^{\sqrt{2} \log i}$$

Poiché  $i = e^{i\frac{\pi}{2}}$ ,  $\log i = i(\frac{\pi}{2} + 2k\pi)$ , e

$$i^{\sqrt{2}} = e^{\sqrt{2}i(\frac{\pi}{2} + 2k\pi)} = \cos\left(\frac{\pi}{\sqrt{2}} + 2\sqrt{2}k\pi\right) + i \sin\left(\frac{\pi}{\sqrt{2}} + 2\sqrt{2}k\pi\right)$$

Si osservi che, per l'irrazionalità di  $\sqrt{2}$ , i numeri  $2\sqrt{2}k\pi$  non sono mai multipli di  $2\pi$ , perciò i numeri scritti sono tutti diversi, al variare di  $k \in \mathbb{Z}$ . Il valore principale di  $i^{\sqrt{2}}$  è

$$\cos\frac{\pi}{\sqrt{2}} + i \sin\frac{\pi}{\sqrt{2}}$$

**Criterio di Dirichlet e somma di alcune serie trigonometriche**

Presentiamo ora un criterio per lo studio della convergenza di serie a termini complessi, che permette di trattare utilmente anche alcuni casi in cui la serie non converge assolutamente. Come vedremo, questo criterio avrà delle ricadute interessanti anche su certe serie a termini reali, ma di segno variabile, che non sarebbe possibile studiare con i criteri fin qui incontrati.

**TEOREMA 5.7 (CRITERIO DI DIRICHLET)** Siano  $\{a_n\}, \{b_n\}$  due successioni tali che:

- (1)  $\{a_n\}$  è a valori complessi e la sua successione delle somme parziali è limitata; in altre parole, esiste una costante  $M > 0$  tale che, detta  $A_n = \sum_{k=0}^n a_k$  risulta

$$|A_n| \leq M \quad \forall n.$$

- (2)  $\{b_n\}$  è a valori reali positivi e tende monotonamente a zero:

$$b_n > 0 \quad \forall n; \quad b_{n+1} \leq b_n \quad \forall n; \quad \lim_{n \rightarrow \infty} b_n = 0.$$

Allora la serie  $\sum a_n b_n$  è convergente.

Si presti attenzione all'ipotesi (1): non si richiede semplicemente che la successione  $\{a_n\}$  sia limitata, ma che la successione  $\{A_n\}$  delle sue somme parziali sia limitata (il che è più impegnativo, in generale). Mostriamo un'interessante applicazione:



Le due serie (a termini reali di segno variabile):

$$\sum_{n=1}^{\infty} \frac{\cos n}{n}; \quad \sum_{n=1}^{\infty} \frac{\sin n}{n}$$

convergono. Infatti: applichiamo il criterio di Dirichlet alle successioni:

$$a_n = e^{in}; \quad b_n = \frac{1}{n}$$

Ovviamente  $b_n$  è reale, positiva, e tende monotonamente a zero. Quanto ad  $a_n$ , scrivendo:

$$A_n = \sum_{k=0}^n a_k = \sum_{k=0}^n \left(e^i\right)^k = \frac{1 - e^{i(n+1)}}{1 - e^i}$$

si ha

$$|A_n| = \left| \frac{1 - e^{i(n+1)}}{1 - e^i} \right| \leq \frac{2}{|1 - e^i|}.$$

Pertanto  $A_n$  è limitata, e possiamo affermare che converge la serie a termini complessi

$$\sum_{n=1}^{\infty} a_n b_n = \sum_{n=1}^{\infty} \frac{e^{in}}{n} = \sum_{n=1}^{\infty} \frac{\cos n + i \sin n}{n}.$$

Separando parte reale e parte immaginaria, si conclude che convergono le due serie a termini reali  $\sum_{n=1}^{\infty} \frac{\cos n}{n}$ ;  $\sum_{n=1}^{\infty} \frac{\sin n}{n}$ .

Si rifletta sul fatto che la convergenza di queste serie non si sarebbe potuta ottenere in base ai criteri studiati nel par. 1. Infatti queste sono serie a termini di segno variabile, ma non a segni alterni; l'unico criterio visto nel par. 1 in questi casi è quello della convergenza assoluta, ma non si vede perché le serie  $\sum \frac{|\cos n|}{n}$ ;  $\sum \frac{|\sin n|}{n}$  dovrebbero essere convergenti.

Il risultato appena discusso sarà molto utile nello studio della convergenza delle *serie di Fourier*, di cui ci occuperemo nel secondo volume.



Scrivere in forma algebrica  $(a + ib)$  i seguenti numeri complessi:



$$e^{z^2}, \text{ con } z = x + iy, x, y \in \mathbb{R}$$



$$e^{\bar{z}}, \text{ con } z = x + iy, x, y \in \mathbb{R}$$



$$ie^z, \text{ con } z = x + iy, x, y \in \mathbb{R}$$



$$e^{2z+3i}, \text{ con } z = x + iy, x, y \in \mathbb{R}$$



$e^{-iz}$ , con  $z = x + iy$ ,  $x, y \in \mathbb{R}$



Sia  $x \in \mathbb{R}$ . Calcolare la parte reale del numero complesso:

$$\frac{1}{2+i} e^{(3-i)x}$$



Sia  $x \in \mathbb{R}$ . Calcolare la parte immaginaria del numero complesso:

$$(5+2i)e^{-3x+2ix}$$

Risolvere le seguenti equazioni nel campo complesso, ponendo  $z = x + iy$  e riscrivendo in forma esponenziale ambo i membri dell'equazione:

$$e^z = -1$$

$$e^z = 2i$$



Calcolare nel campo complesso i logaritmi dei seguenti numeri:

$$-1 \quad 2 \quad 1+i\sqrt{3}$$



Calcolare nel campo complesso le seguenti potenze, riscrivendo in forma algebrica il risultato:

$$2^i \quad 2^{\sqrt{2}} \quad (-2)^{\sqrt{3}} \quad (1-i)^{3/5} \quad (1+i)^{2+3i} \quad i^i$$



Scrivere parte reale e parte immaginaria del valore principale della funzione

$$f(x) = x^{a+ib}, \text{ per } x, a, b \in \mathbb{R}$$



Utilizzando opportunamente lo sviluppo in serie di  $\log(1+x)$ , calcolare la somma della serie

$$\sum_{n=1}^{\infty} \frac{1}{n2^n}$$



Utilizzando opportunamente lo sviluppo in serie di  $e^x$ , calcolare la somma della serie

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n!2^n}.$$

## COMPLEMENTI



Sfruttando la somma della serie geometrica nel campo complesso, dimostrare che:

$$\sum_{n=0}^{\infty} \rho^n \cos(n\theta) = \frac{1 - \rho \cos \theta}{1 + \rho^2 - 2\rho \cos \theta}$$

$$\sum_{n=1}^{\infty} \rho^n \sin(n\theta) = \frac{\rho \sin \theta}{1 + \rho^2 - 2\rho \cos \theta}$$

per ogni  $\rho < 1$ , ogni  $\theta$ .

*Suggerimento:* scrivere  $z = \rho(\cos \theta + i \sin \theta)$ , e separare parte reale e immaginaria nella somma della serie geometrica  $\sum z^n$ .

Calcolare quindi la somma delle serie

$$\sum_{n=0}^{\infty} \frac{\cos(n\theta)}{2^n}; \quad \sum_{n=1}^{\infty} \frac{\sin(2n\theta)}{3^n}.$$

**2** Dimostrare il criterio della convergenza assoluta per le serie a termini complessi (v. par. 2.2).

*Suggerimento:* ragionare in modo analogo alla dimostrazione data nel par. 1.3 per le serie a termini reali, introducendo le quattro successioni di somme parziali  $s_n^{\text{Re}+}, s_n^{\text{Re}-}, s_n^{\text{Im}+}, s_n^{\text{Im}-}$  (con l'ovvio significato dei simboli).

**3** Vogliamo dimostrare il seguente teorema, enunciato nel par. 2.2: ponendo  $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ , risulta:

$$e^{z_1+z_2} = e^{z_1} \cdot e^{z_2} \quad \forall z_1, z_2 \in \mathbb{C}.$$

(a) Consideriamo la successione delle somme parziali della serie che definisce  $e^{z_1+z_2}$ :

$$s_n = \sum_{k=0}^n \frac{(z_1 + z_2)^k}{k!}.$$

Sviluppando la potenza  $(z_1 + z_2)^k$  con la formula del binomio di Newton e applicando opportunamente le proprietà delle sommatorie, provare che:

$$(2.2) \quad \sum_{k=0}^n \frac{(z_1 + z_2)^k}{k!} = \sum_{k=0}^n \frac{z_1^k}{k!} \sum_{h=0}^{n-k} \frac{z_2^h}{h!}.$$

(b) Se al posto del secondo membro dell'ultima identità avessimo

$$\sum_{k=0}^n \frac{z_1^k}{k!} \sum_{h=0}^n \frac{z_2^h}{h!}$$

sarebbe ora sufficiente passare al limite per  $n \rightarrow \infty$ . Poiché

$$\sum_{k=0}^n \frac{z^k}{k!} \rightarrow e^z \text{ per } n \rightarrow \infty,$$

otterremmo proprio  $e^{z_1+z_2} = e^{z_1} \cdot e^{z_2}$ . La complicazione nasce dal fatto che, invece, nel secondo membro della (2.2), la sommatoria interna dipende anche dall'indice  $k$ . Si può allora spezzare la sommatoria al modo seguente (supponiamo per semplicità che  $n$  sia pari):

$$(2.3) \quad \sum_{k=0}^n \frac{z_1^k}{k!} \sum_{h=0}^{n-k} \frac{z_2^h}{h!} = \sum_{k=0}^{n/2} \frac{z_1^k}{k!} \sum_{h=0}^{n/2} \frac{z_2^h}{h!} + \sum_{k=0}^{n/2} \frac{z_1^k}{k!} \sum_{h=n/2+1}^{n-k} \frac{z_2^h}{h!} + \sum_{k=n/2+1}^n \frac{z_1^k}{k!} \sum_{h=0}^{n-k} \frac{z_2^h}{h!} \\ \equiv A_n + B_n + C_n.$$

Ora il termine  $A_n$  ha la forma giusta per poter affermare che  $A_n \rightarrow e^{z_1} \cdot e^{z_2}$ . Si tratta di dimostrare che  $B_n$  e  $C_n$  tendono a zero. Per far questo, si ragiona così:

$$|B_n| \leq \sum_{k=0}^{n/2} \frac{|z_1|^k}{k!} \sum_{h=n/2+1}^{n-k} \frac{|z_2|^h}{h!} \leq \sum_{k=0}^{\infty} \frac{|z_1|^k}{k!} \sum_{h=n/2+1}^{\infty} \frac{|z_2|^h}{h!} = e^{|z_1|} \cdot \sum_{h=n/2+1}^{\infty} \frac{|z_2|^h}{h!} \rightarrow 0$$

perché l'ultima serie scritta è il resto di una serie convergente. In modo analogo si ragiona con  $C_n$ . Questo conclude la dimostrazione.

Per capire l'idea geometrica con cui si è eseguito il passaggio (2.3) si rappresenti nel piano l'insieme delle coppie  $(k, h)$  su cui si somma l'espressione  $\sum_{k=0}^n \frac{z_1^k}{k!} \sum_{h=0}^{n-k} \frac{z_2^h}{h!}$ . L'idea è di vedere questo triangolo di punti a coordinate intere come unione di un quadrato (che dà l'espressione  $A_n$  che vorremmo avere) e due triangoli più piccoli, che dovranno dare somme infinitesime con  $n$ . Si ha:

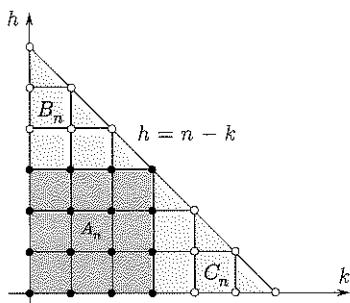


Figura 5.5.

**4** Un modo alternativo per introdurre l'esponenziale complesso senza usare le serie è il seguente. Per  $z = x + iy$ ,  $x, y \in \mathbb{R}$ , poniamo per definizione:

$$(2.4) \quad \exp(x + iy) = e^x (\cos y + i \sin y).$$

Sfruttando le formule di De Moivre per i numeri complessi scritti in forma trigonometrica, dimostrare in base alla (2.4) che:

$$(2.5) \quad \begin{aligned} e^{z_1+z_2} &= e^{z_1} \cdot e^{z_2} \\ (e^{z_1})^n &= e^{nz_1} \end{aligned}$$

$\forall z_1, z_2 \in \mathbb{C}, n \in \mathbb{N}$ . Inoltre, ovviamente, se  $z \in \mathbb{R}$  (cioè se  $z = x$  e  $y = 0$ ) si riottiene che  $\exp(z) = e^x$ . Perciò la funzione di variabile complessa appena definita può chiamarsi “esponenziale complesso”. Si noti che le formule di Eulero sono tutte conseguenze delle proprietà (2.5) dell'esponenziale.

**5** Discutere, al variare del parametro reale  $x$ , la convergenza delle serie

$$\sum \frac{\sin(nx)}{n}; \sum \frac{\cos(nx)}{n}.$$

*Suggerimento:* applicare il criterio di Dirichlet, adattando l'esempio che abbiamo presentato subito dopo il criterio.

**6** Dare condizioni sulla successione  $\{a_n\}$  affinché le serie

$$\sum a_n \cos(nx); \sum a_n \sin(nx)$$

risultino convergenti.

*Suggerimento:* applicare il criterio di Dirichlet.

**7** Il criterio di Leibniz per le serie a segni alterni si può dedurre anche dal criterio di Dirichlet. Come?

# 6

## Calcolo integrale per funzioni di una variabile

### 1 INTRODUZIONE AL CALCOLO INTEGRALE

Come abbiamo fatto nel capitolo precedente per il calcolo differenziale, introduciamo anche il calcolo integrale partendo da alcuni dei problemi che storicamente hanno portato alla sua nascita.

#### Che cosa vuol dire misurare l'area di una figura piana a contorno curvilineo?

Questo è il più antico problema di calcolo infinitesimale che sia stato affrontato e risolto, almeno in qualche caso particolare.

L'idea elementare di misura di un'area è: fissare un quadratino unità di misura e contare quante volte questo può essere riportato nella figura da misurare. Per questa via, si riescono però a misurare solo rettangoli e figure scomponibili in rettangoli. Andando oltre, la geometria elementare postula che se due figure sono equiscomponibili (cioè possono essere divise in un numero *finito* di parti rispettivamente sovrapponibili), allora sono equivalenti (cioè hanno la stessa area). Inoltre, si conviene che l'area sia *additiva*, cioè se una figura  $T$  è unione di due (o più) figure  $T_1, T_2$  (disgiunte), l'area di  $T$  è la somma dell'area di  $T_1$  e di  $T_2$ . L'utilizzo di questi principi permette di calcolare le aree di *poligoni* qualsiasi.

Comunque, per questa via non si arriverà mai a dar senso al concetto di *area del cerchio*, per esempio (né a calcolarla): infatti il cerchio non è scomponibile in un numero finito di triangoli e/o rettangoli.

#### Che cosa vuol dire misurare la lunghezza di una curva?

Il problema è analogo al precedente: non è possibile riportare un segmento unitario su una curva, e dunque “contare quante volte esso è contenuto”. Cos'è dunque la lunghezza di una curva?

Entrambi questi problemi vengono risolti, nella geometria elementare di Euclide (attorno al 300 a.C.), nel caso particolare del cerchio (area del cerchio, lunghezza della circonferenza), con un ragionamento che fa ricorso all'idea di approssimazioni successive, sempre più accurate: si calcola l'area, o il perimetro, dei poligoni regolari inscritti e circoscritti, e ci si chiede cosa accade quando il numero dei lati cresce indefinitamente. Ecco dunque un'idea di *procedimento infinito* e di *limite*, che si innesta nella geometria elementare come un'idea nuova.

Archimede di Siracusa (287-212 a.C.) applicò ingegnosamente idee simili (il suo *metodo di esaustione*) al calcolo dell'area del segmento di parabola e ad alcuni altri casi specifici. Nel 1600, vari matematici trovarono altri metodi per calcolare l'area sottesa al grafico di semplici funzioni:  $x^\alpha$  ( $\alpha > -1$ ) (Fermat, 1636),  $1/x$  (Mercator, 1668). Tuttavia, fu solo con l'invenzione del calcolo infinitesimale (fine 1600-inizio 1700) che mediante il concetto di *integrale*, inteso come opportuno *limite di somme*, venne data una definizione abbastanza generale di area di una figura piana e di lunghezza di una curva, e contemporaneamente venne fornito un algoritmo per il calcolo effettivo utilizzabile non solo in casi molto particolari: il famoso "Teorema fondamentale del calcolo integrale" (v. par. 4) di Newton, Leibniz, Giovanni Bernoulli, che ricondusse tale problema, almeno in molti casi, a un esercizio di routine: la ricerca di una primitiva (o antiderivata) di una funzione.

## 2 L'INTEGRALE COME LIMITE DI SOMME

### 2.1 La definizione di integrale

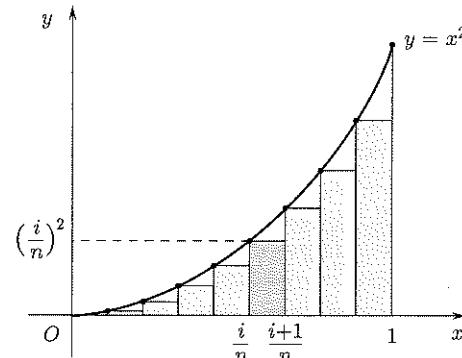


Figura 6.1.

$$\text{Area} \approx S_n = \sum_{i=1}^{n-1} \underbrace{\frac{1}{n}}_{\text{base}} \cdot \underbrace{\left(\frac{i}{n}\right)^2}_{\text{altezza}} = \frac{1}{n^3} \sum_{i=1}^{n-1} i^2 = \frac{(n-1)n(2n-1)}{6n^3}$$

(per l'ultima uguaglianza, v. esercizio 5, capitolo 1, par.2)

A questo punto si può far tendere  $n$  a  $+\infty$ , ottenendo

$$\frac{(n-1)n(2n-1)}{6n^3} \xrightarrow{n \rightarrow +\infty} \frac{1}{3}$$

e concludere così che

$$\text{Area} = \frac{1}{3}$$

La definizione di integrale che daremo segue questi stessi passi.

Sia  $f : [a, b] \rightarrow \mathbb{R}$ , *limitata*. Notiamo esplicitamente che  $f$  potrebbe non essere continua; ad ogni modo, almeno per il momento, assumiamo che sia almeno *definita* in ogni punto di  $[a, b]$  (e limitata).

Consideriamo la suddivisione di  $[a, b]$ , individuata dai punti

$$a = x_0, \quad x_1, x_2, \dots, x_{n-1}, \quad x_n = b$$

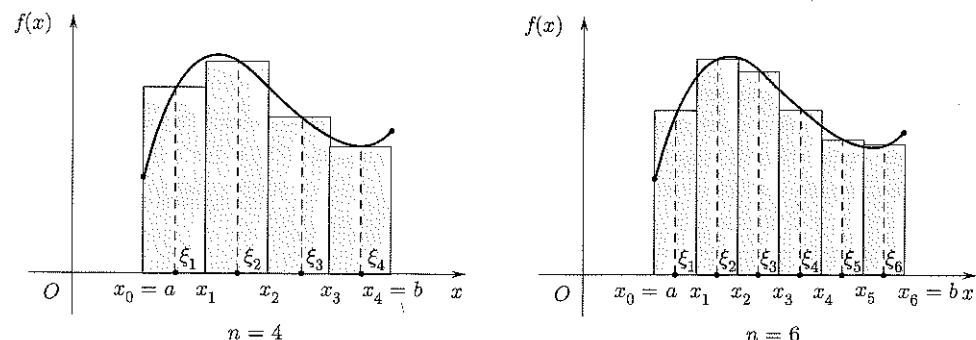
con

$$x_j = a + jh, \quad h = \frac{b-a}{n}, \quad j = 0, \dots, n$$

escegliamo in ciascuno degli  $n$  intervalli  $[x_{j-1}, x_j]$ , un punto arbitrario  $\xi_j$  ( $j = 1, 2, \dots, n$ ).

Costruiamo la somma (detta *somma di Cauchy-Riemann*):

$$(2.1) \quad S_n = \sum_{j=1}^n f(\xi_j) \cdot (x_j - x_{j-1}) = \frac{b-a}{n} \sum_{j=1}^n f(\xi_j)$$

Figura 6.2. Una particolare scelta dei punti  $\xi_j$  per  $n = 4$  e  $n = 6$ .

Si rifletta su questa costruzione. Ad ogni passo della costruzione, l'intervallo  $[a, b]$  viene diviso in  $n$  intervallini, e in ciascuno di questi viene scelto un punto  $\xi_j$ . Si noti che, ad esempio, i punti  $\xi_1, \xi_2, \dots, \xi_6$  scelti al passo  $n = 6$  (v. figura 6.2) sono in generale *tutti diversi* dai punti  $\xi_1, \dots, \xi_4$  scelti al passo  $n = 4$ . Ad ogni passo della costruzione, in generale, *tutti* gli addendi della somma  $S_n$  cambiano, diventando via via più numerosi e più piccoli in valore assoluto.

A questo punto si passa al limite per  $n \rightarrow +\infty$  e... si spera di ottenere un risultato significativo; che cosa questo significhi esattamente, è espresso dalla prossima definizione:

**DEFINIZIONE 6.1** Diciamo che la funzione  $f : [a, b] \rightarrow \mathbb{R}$ , limitata, è *integrabile* in  $[a, b]$ , se, detta  $S_n$  una sua qualsiasi successione di Cauchy-Riemann, esiste finito il limite di  $S_n$ , e tale limite non dipende da come abbiamo scelto i punti  $\xi_j$  ad ogni passo della costruzione iterativa. In tal caso, si pone

$$\lim_{n \rightarrow \infty} S_n = \int_a^b f(x) dx.$$

Il simbolo

$$\int_a^b f(x) dx$$

si legge “integrale da  $a$  a  $b$  di  $f(x)$  in  $dx$ ”. In esso, il segno  $\int$  (una esse allungata) è una deformazione del simbolo di somma; la scrittura  $f(x) dx$  ricorda il prodotto del valore di  $f(x)$  per la lunghezza di un piccolo intervallo sull’asse  $x$ . Tutta la scrittura, quindi, ricorda il procedimento con cui l’integrale è stato definito. La variabile  $x$  si dice *variabile d’integrazione* ed è una *variabile muta*; infatti la scrittura  $\int_a^b f(t) dt$  ha lo stesso significato di  $\int_a^b f(x) dx$ , esattamente come, ad esempio,  $\sum_{j=1}^{10} a_j = \sum_{n=1}^{10} a_n$ .

Ancora, notiamo che l’integrale di una funzione su un intervallo fissato è un *numero*, non una funzione.

Possiamo esprimere sinteticamente la definizione data dicendo che *l’integrale è un (particolare) limite di somme*, che può avere molte interpretazioni diverse (geometriche, fisiche...). Vediamone qualcuna.

#### Interpretazione geometrica (cfr. figura 6.2)

Sia  $f \geq 0$ , e sia  $f$  continua. Ogni addendo della (2.1) rappresenta l’area del rettangolo avente come base il segmento  $[x_{j-1}, x_j]$  e come altezza  $f(\xi_j)$ . La somma  $S_n$  rappresenta dunque un’approssimazione dell’area della parte di piano compreso tra l’asse  $x$ ,  $a \leq x \leq b$ , e il grafico di  $f$  (*trapezioide individuato da  $f$* ).

Passando al limite per  $n \rightarrow +\infty$ , si ha

$$S_n \rightarrow \int_a^b f(x) dx = \text{area del trapezioide}$$

Più precisamente, è l’integrale che costituisce una definizione precisa dell’area del trapezioide individuato da una curva  $y = f(x)$ , e non viceversa. Perciò, è il calcolo infinitesimale che permette di dar senso all’idea di *area di una figura piana* in generale.

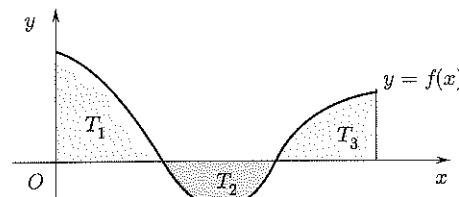


Figura 6.3.

Si noti che, nel caso in cui  $f$  cambia segno, l’integrale rappresenta una somma di *aree con segno*. Nel caso in figura 6.3 si ha:

$$\begin{aligned} \int_a^b f(x) dx &= \\ &= \text{area } (T_1) - \text{area } (T_2) + \text{area } (T_3) \end{aligned}$$

Ad esempio, per simmetria si ha:

$$\int_0^{2\pi} \sin x dx = 0$$

Si noti anche il caso elementare: se  $c$  è una costante,  $\int_a^b c dx = c(b-a)$  (area del rettangolo di altezza  $c$  e base  $(b-a)$ ).

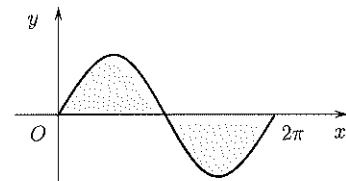


Figura 6.4.

#### Interpretazione cinematica

Supponiamo che un punto materiale si muova lungo una traiettoria fissata, con velocità  $v(t)$  variabile nel tempo, e chiediamoci quanto spazio è percorso dal punto nell’intervallo di tempo  $[0, T]$ . Suddividiamo  $[0, T]$  in  $n$  intervallini  $[t_{j-1}, t_j]$  sufficientemente piccoli da poter pensare che in ciascuno di essi la velocità vari di poco. Dunque, se  $\xi_j$  è un qualsiasi punto di  $[t_{j-1}, t_j]$ ,  $v(\xi_j)$  sarà circa uguale alla velocità media del punto nell’intervallo di tempo  $[t_{j-1}, t_j]$ ; perciò lo spazio percorso in questo intervallino di tempo sarà  $v(\xi_j)(t_j - t_{j-1})$ , e lo spazio totale percorso sarà

$$\sum_{j=1}^n v(\xi_j)(t_j - t_{j-1})$$

Al crescere di  $n$  questa approssimazione dello spazio percorso si fa sempre migliore. Perciò il valore vero dello spazio percorso è il limite per  $n \rightarrow \infty$  di questa espressione, ossia:

$$\text{Spazio percorso in } [0, T] = \int_0^T v(t) dt$$

Lo spazio percorso (o più precisamente lo spostamento netto) è l’integrale della velocità rispetto al tempo.

#### Interpretazione meccanica

Supponiamo che un sistema fisico caratterizzato da pressione (=  $p$ ) e volume (=  $V$ ) evolva da uno stato  $(p_A, V_A)$  ad uno stato  $(p_B, V_B)$ , a temperatura costante. Graficamente, nel piano  $V, p$  si ha:

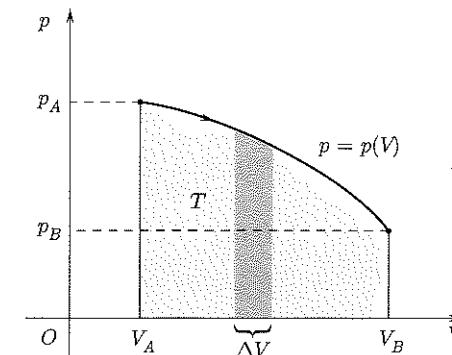


Figura 6.5.

In questo caso

$$\int_{V_A}^{V_B} p(V) dV = \text{area } (T) = \text{lavoro effettuato nella trasformazione da } (p_A, V_A) \text{ a } (p_B, V_B)$$

Infatti, ogni addendo della somma  $S_n = \sum_{j=1}^n p(\xi_j)(V_j - V_{j-1})$  costituisce il lavoro effettuato in corrispondenza di una variazione di volume  $\Delta V = V_j - V_{j-1}$  (dimensionalmente:  $[p] = [\text{Forza}] \cdot [\text{lunghezza}]^{-2}$ ,  $[V] = [\text{lunghezza}]^3$  e perciò  $[p \cdot \Delta V] =$

[Forza] · [lunghezza]). Passando al limite per  $n \rightarrow +\infty$  si ottiene l'interpretazione richiesta.

Alla stessa conclusione si può arrivare con un procedimento, spesso usato nelle applicazioni dell'integrale, come il seguente.

In corrispondenza alla variazione di volume da  $V$  a  $V + dV$ , con  $dV$  molto piccolo, si può considerare la pressione costante ( $= p(V)$ ) e quindi il lavoro elementare  $dL$  effettuato durante la trasformazione è dato da

$$dL = p(V)dV$$

Sommendo tutti i contributi da  $V_A$  a  $V_B$  e cioè integrando da  $V_A$  a  $V_B$  si ottiene

$$L = \int_{V_A}^{V_B} dL = \int_{V_A}^{V_B} p(V)dV$$

## 2.2 Classi di funzioni integrabili

La definizione di integrale che abbiamo dato lascia aperta finora una domanda naturale: come possiamo stabilire, in pratica, se una funzione è integrabile? Valgono in proposito i seguenti teoremi:

**TEOREMA 6.1** *Se  $f : [a, b] \rightarrow \mathbb{R}$  è continua, allora è integrabile.*

**TEOREMA 6.2** *Se  $f : [a, b] \rightarrow \mathbb{R}$  è monotona e limitata, allora è integrabile.*

In base a questi teoremi, conosciamo dunque alcune classi significative di funzioni che risultano certamente integrabili. Inoltre, a partire da funzioni integrabili se ne possono costruire altre integrabili con varie procedure. Una di queste è espressa dal prossimo teorema:

**TEOREMA 6.3** *Se  $f_1 : [a, b] \rightarrow \mathbb{R}$  e  $f_2 : [b, c] \rightarrow \mathbb{R}$  sono integrabili, allora la funzione*

$$f(x) = \begin{cases} f_1(x) & \text{se } x \in [a, b] \\ f_2(x) & \text{se } x \in (b, c] \end{cases}$$

(*f definita in qualsiasi modo in b*) è integrabile in  $[a, c]$ .

Applicando più volte questo teorema, il risultato si estende a funzioni ottenute "incollando tra loro" un numero finito di funzioni integrabili su intervalli adiacenti. Ad esempio, una funzione continua e limitata in un intervallo  $[a, c]$ , ad eccezione di un numero finito di punti di discontinuità a salto, risulta integrabile. Riprenderemo nel seguito questa affermazione, con opportuni esempi (si veda il par.5.6).

I tre teoremi appena enunciati sono piuttosto delicati da dimostrare; proveremo solo il primo di questi (integrabilità delle funzioni continue), ma per non spezzare il discorso, rimandiamo questa dimostrazione all'ultimo paragrafo di questo capitolo.

È bene notare che *non tutte* le funzioni limitate sono integrabili, come mostra il prossimo esempio:



Sia  $f : [0, 1] \rightarrow \mathbb{R}$  così definita:

$$f(x) = \begin{cases} 1 & \text{se } x \in \mathbb{Q} \\ 0 & \text{se } x \notin \mathbb{Q} \end{cases}$$

La funzione precedente (nota come "funzione di Dirichlet") vale 1 nei punti razionali e 0 nei punti irrazionali dell'intervallo  $[0, 1]$ . Un attimo di riflessione mostra che questa funzione è discontinua in ogni punto! Proviamo a calcolare  $S_n$ . Qualunque sia l'intero  $n$ , in ogni intervallino  $[x_{j-1}, x_j]$  possiamo scegliere  $\xi_j$  razionale o irrazionale, a nostro piacere. Supponiamo di scegliere  $\xi_j$  razionale in tutti gli intervallini e per tutti gli  $n$ . Allora sarà

$$S_n = \sum_{j=1}^n f(\xi_j)(x_j - x_{j-1}) = \sum_{j=1}^n 1 \cdot (x_j - x_{j-1}) = x_n - x_0 = 1.$$

Per questa costruzione di  $S_n$ , dunque, risulta  $\lim S_n = 1$ . Se invece scegliamo, in tutti gli intervallini e per tutti gli  $n$ , dei punti  $\xi_j$  irrazionali, si avrà

$$S_n = \sum_{j=1}^n f(\xi_j)(x_j - x_{j-1}) = \sum_{j=1}^n 0 \cdot (x_j - x_{j-1}) = 0$$

e in questo caso  $\lim S_n = 0$ . Poiché il limite delle successioni di Cauchy-Riemann dipende dalle scelte dei punti  $\xi_j$ , la funzione non è integrabile.

## 3 PROPRIETÀ DELL'INTEGRALE

Direttamente dalla definizione si possono dimostrare le seguenti proprietà dell'integrale.

**TEOREMA 6.4** *Siano  $f, g$  integrabili in  $[a, b]$ . Valgono allora le seguenti proprietà*

1. **Linearità dell'integrale.** Se  $\alpha, \beta$  sono costanti, allora anche la funzione  $\alpha f(x) + \beta g(x)$  è integrabile, e vale l'identità

$$\int_a^b [\alpha f(x) + \beta g(x)] dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx$$

2. **Additività dell'integrale rispetto all'intervallo di integrazione.** Se  $a \leq r \leq b$  allora  $f$  è integrabile anche su  $[a, r]$  e  $[r, b]$ , e vale l'identità:

$$(3.1) \quad \int_a^b f(x) dx = \int_a^r f(x) dx + \int_r^b f(x) dx$$

Convenzione: se  $a < b$ , si pone

$$\int_b^a f(x) dx = - \int_a^b f(x) dx$$

allora la (3.1) vale qualunque sia l'ordinamento dei punti  $a, b, r$ .

### 3. Positività e monotonia dell'integrale.

$$f \geq 0 \text{ in } [a, b] \implies \int_a^b f(x) dx \geq 0 \quad (a < b)$$

$$f \geq g \text{ in } [a, b] \implies \int_a^b f(x) dx \geq \int_a^b g(x) dx$$

In particolare

$$(3.2) \quad \left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx$$

#### DIMOSTRAZIONE.

- Dimostriamo la linearità dell'integrale. Sia  $S_n$  una somma di Cauchy-Riemann  $n$ -esima relativa ad  $(\alpha f + \beta g)$ :

$$\begin{aligned} S_n &= \frac{b-a}{n} \sum_{j=1}^n [\alpha f(\xi_j) + \beta g(\xi_j)] = \\ &= \alpha \left[ \frac{b-a}{n} \sum_{j=1}^n f(\xi_j) \right] + \beta \left[ \frac{b-a}{n} \sum_{j=1}^n g(\xi_j) \right] \end{aligned}$$

Facciamo ora tendere  $n$  a  $+\infty$ . Il secondo membro tende a

$$\alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx;$$

inoltre, questo limite non dipende da come sono scelti i punti  $\xi_j$  (perché per ipotesi  $f$  e  $g$  sono integrabili); dunque anche il primo membro avrà un limite indipendente dalla scelta dei punti  $\xi_j$ , e questo limite per definizione è

$$\int_a^b [\alpha f(x) + \beta g(x)] dx.$$

Questo mostra che  $\alpha f(x) + \beta g(x)$  è effettivamente integrabile e che vale l'uguaglianza delle due espressioni, ossia la tesi.

Si riflette sull'idea elementare della dimostrazione precedente: l'integrale è un limite di somme; ma la sommatoria è lineare e l'operazione di limite è lineare: perciò l'integrale è un'operazione lineare.

- Non dimostriamo l'additività rispetto all'intervallo di integrazione, che è un po' delicata. Ci limitiamo a segnalare che, nel caso di una funzione continua e non negativa, il significato geometrico dell'additività dell'integrale è semplicemente quello di additività dell'area, come mostra la figura 6.6.
- Sia  $f(x) \geq 0$  in  $[a, b]$ , e consideriamo una sua successione di Cauchy-Riemann:

$$S_n = \sum_{k=1}^n f(\xi_k) \left( \frac{b-a}{n} \right) \geq 0$$

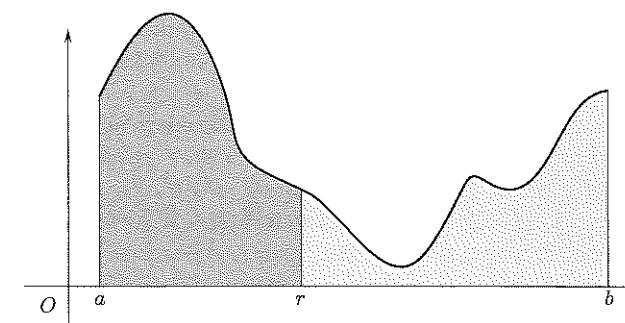


Figura 6.6. Additività dell'area.

perché  $f(\xi_k) \geq 0$  per ogni  $k$ . Passando al limite per  $n \rightarrow \infty$ , il teorema di permanenza del segno per le successioni garantisce che

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} S_n \geq 0.$$

La proprietà di positività appena dimostrata implica quella di monotonia: se ora  $f \geq g$  in  $[a, b]$ , allora  $f - g \geq 0$  in  $[a, b]$ , che per quanto appena dimostrato implica

$$\int_a^b [f(x) - g(x)] dx \geq 0.$$

Per la linearità dell'integrale, questo implica a sua volta

$$\int_a^b f(x) dx \geq \int_a^b g(x) dx$$

che è la proprietà di monotonia. Infine, essendo  $f \leq |f|$  e  $-f \leq |f|$  in  $[a, b]$  si ha:

$$\int_a^b f(x) dx \leq \int_a^b |f(x)| dx \quad \text{e} \quad \int_a^b -f(x) dx \leq \int_a^b |f(x)| dx$$

che insieme equivalgono alla (3.2). ◆

La prossima proprietà non vale per tutte le funzioni integrabili, ma specificamente per quelle *continue*:

**TEOREMA 6.5 (TEOREMA DELLA MEDIA)** *Sia  $f : [a, b] \rightarrow \mathbb{R}$  continua. Allora esiste  $c \in [a, b]$  tale che*

$$(3.3) \quad \frac{1}{b-a} \int_a^b f(x) dx = f(c).$$

**DIMOSTRAZIONE.** Essendo  $f$  continua in  $[a, b]$ , per il Teorema di Weierstrass (cap. 3, par. 4.1), essa è dotata di massimo ( $= M$ ) e minimo ( $= m$ ). Dalla proprietà di monotonia si ha

$$m = \frac{1}{b-a} \int_a^b m dx \leq \frac{1}{b-a} \int_a^b f(x) dx \leq \frac{1}{b-a} \int_a^b M dx = M$$

Quindi il valore  $\frac{1}{b-a} \int_a^b f(x) dx$  è compreso tra il minimo ed il massimo di  $f$ . Per la proprietà dei valori intermedi delle funzioni continue (ancora cap. 3, par. 4.1) tale valore è uguale a  $f(c)$  per qualche  $c \in [a, b]$ . ◆

**Valor medio, valor efficace**

Chiamiamo:

$$\frac{1}{b-a} \int_a^b f(x)dx = \text{valor medio di } f \text{ su } [a, b] = f_M$$

$$\left( \frac{1}{b-a} \int_a^b [f(x)]^2 dx \right)^{1/2} = \text{valor efficace di } f \text{ su } [a, b] = f_E$$

Essendo

$$\frac{1}{b-a} \int_a^b f(x)dx = \lim_{n \rightarrow +\infty} \frac{1}{b-a} S_n = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=1}^n f(\xi_j)$$

si vede che  $f_M$  è una generalizzazione dell'ordinaria media aritmetica tra  $n$  numeri.

Il valore efficace  $f_E$  interviene, ad esempio, nella teoria dei circuiti elettrici. Sia  $I(t) = I \sin \omega t$ ,  $\omega = \frac{2\pi}{T}$ , l'intensità di una corrente sinusoidale. Abbiamo, prendendone il valore efficace su un periodo:

$$\begin{aligned} I_E^2 &= \frac{1}{T} \int_0^T I^2 (\sin \omega t)^2 dt = \left( \text{essendo } (\sin x)^2 = \frac{1 - \cos 2x}{2} \right) \\ &= \frac{I^2}{T} \int_0^T \left[ \frac{1}{2} - \frac{1}{2} \cos 2\omega t \right] dt = \frac{I^2}{T} \left\{ \frac{T}{2} - \int_0^T \frac{1}{2} \cos 2\omega t dt \right\} \end{aligned}$$

Poiché, per simmetria,  $\int_0^T \cos 2\omega t dt = 0$  (vedi figura 6.7) si ha:

$$I_E^2 = \frac{I^2}{2} \quad \text{e cioè} \quad I_E = \boxed{\frac{I}{\sqrt{2}}}$$

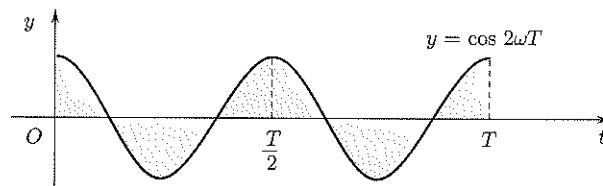


Figura 6.7.  $\cos 2\omega t$  è periodico di periodo  $\frac{T}{2}$  se  $\omega = \frac{2\pi}{T}$ .

## ■ 4 IL TEOREMA FONDAMENTALE DEL CALCOLO INTEGRALE

La definizione di integrale non si presta in generale al suo calcolo effettivo. Uno dei metodi più usati per il calcolo è quello per variazione di una primitiva, che qui illustriamo.

**DEFINIZIONE 6.2** Si dice che una funzione  $G$ , derivabile in un intervallo  $I$ , è una *primitiva* di  $f$  in  $I$ , se

$$G'(x) = f(x) \quad \text{per ogni } x \in I.$$

Ad esempio,  $G(x) = x^2$  è una primitiva di  $f(x) = 2x$  e  $G(x) = \sin x$  è una primitiva di  $f(x) = \cos x$ , su tutto l'asse reale.

Evidentemente se  $G$  è una primitiva di  $f$ , lo è anche  $G + c$  dove  $c$  è una costante; d'altra parte se  $G_1$  e  $G_2$  sono due primitive di  $f$  in  $I$  allora  $G'_1 - G'_2 = 0$  in  $I$  ossia  $(G_1 - G_2)' = 0$  e quindi  $G_1 - G_2 = \text{costante}$  (Il fatto che, se una funzione ha derivata nulla su un intervallo, allora è costante, è stato dimostrato come conseguenza del teorema di Lagrange, si veda il teorema 4.6 del capitolo 4).

Ne segue che se si conosce una primitiva  $G$  di  $f$ , tutte le altre sono della forma  $G + c$ ,  $c \in \mathbb{R}$ .

Tutte le funzioni possiedono primitiva? Non tutte; si può mostrare che se una funzione presenta punti di discontinuità a salto in un intervallo  $I$  allora *non* può avere primitiva.

Se però  $f$  è continua in  $I$ , proveremo in seguito (par. 9) che essa possiede certamente una primitiva (e quindi infinite, per quanto detto precedentemente). Fatte queste importanti premesse, enunciamo ora il risultato che ci permetterà di calcolare effettivamente gli integrali.

### TEOREMA 6.6 (TEOREMA FONDAMENTALE DEL CALCOLO INTEGRALE)

Se  $f : [a, b] \rightarrow \mathbb{R}$  è continua, e  $G$  è una sua primitiva su  $[a, b]$ , allora

$$\int_a^b f(x) dx = G(b) - G(a)$$

Per indicare  $G(b) - G(a)$  si usa comunemente il simbolo

$$[G(x)]_a^b$$

Ad esempio, si voglia calcolare

$$\int_0^1 x^2 dx$$

Poiché  $(\frac{x^3}{3})' = x^2$ ,  $G(x) = \frac{x^3}{3}$  è una primitiva di  $x^2$  in  $[0, 1]$ , perciò

$$\int_0^1 x^2 dx = \left[ \frac{x^3}{3} \right]_0^1 = \frac{1}{3}$$

Si confronti la semplicità di questo calcolo con il procedimento illustrato in apertura del capitolo, per calcolare questo integrale in base alla definizione.

Si esprime talvolta questo teorema dicendo che *integrazione e derivazione sono operazioni inverse*. È questa un'affermazione che non va equivocata: l'integrale è stato definito mediante un opportuno limite di somme, che (apparentemente!) non ha niente a che vedere col concetto di derivata (limite del rapporto incrementale). Tuttavia, grazie al teorema precedente, il calcolo effettivo di un integrale è ricondotto alla ricerca di una primitiva, che è l'operazione inversa della derivazione. (Un sinonimo di *primitiva* talvolta usato è *antiderivata*).

**DIMOSTRAZIONE.** Siano  $a = x_0, x_1, \dots, x_n = b$  punti che suddividono l'intervallo  $[a, b]$  in  $n$  intervallini di ugual ampiezza. Allora, aggiungendo e togliendo  $G(x_j)$  per  $j = 1, 2, \dots, n-1$  si

ha:

$$\begin{aligned} G(b) - G(a) &= G(x_n) - G(x_0) = [G(x_n) - G(x_{n-1})] + [G(x_{n-1}) - G(x_{n-2})] + \cdots \\ &\quad \cdots + [G(x_2) - G(x_1)] + [G(x_1) - G(x_0)] = \\ &= \sum_{j=1}^n [G(x_j) - G(x_{j-1})] \end{aligned}$$

Applichiamo ora il teorema di Lagrange alla funzione  $G(x)$  su ciascuno degli intervalli  $[x_{j-1}, x_j]$ . Esiste allora  $\xi_j \in (x_{j-1}, x_j)$  tale che

$$G(x_j) - G(x_{j-1}) = (x_j - x_{j-1})G'(\xi_j) = (x_j - x_{j-1})f(\xi_j)$$

perché per ipotesi  $G$  è una primitiva di  $f$  e perciò  $G'(\xi_j) = f(\xi_j)$ . Ne segue che

$$G(b) - G(a) = \sum_{j=1}^n (x_j - x_{j-1})f(\xi_j) = S_n$$

dove  $S_n$  è una somma  $n$ -esima di Cauchy-Riemann di  $f$ . L'identità scritta vale per ogni  $n$ ; possiamo allora far tendere  $n$  a  $+\infty$ , trovando

$$G(b) - G(a) = \int_a^b f(x) dx$$

Si osservi che questa dimostrazione prova che *una certa* successione di Cauchy-Riemann di  $f$  tende a  $G(b) - G(a)$ . Poiché però sappiamo già che  $f$  è integrabile, in quanto continua (per il Teorema 6.1, che proveremo in seguito), questo è sufficiente a concludere che ogni altra successione di Cauchy-Riemann converge allo stesso limite, e quindi vale la formula enunciata.

Nel prossimo paragrafo affronteremo il problema di come si determina una primitiva di  $f(x)$ , almeno per certe classi di funzioni  $f(x)$ .

Va tenuto presente, comunque, che:

1. A dispetto del fatto che ogni funzione continua su  $[a, b]$  ha una primitiva, anche per funzioni continue abbastanza semplici succede che tale primitiva non sia esprimibile come composizione di funzioni elementari. Ciò significa che in molti casi non si sa scrivere un'espressione analitica della primitiva di  $f(x)$  utilizzabile per il calcolo effettivo dell'integrale.
2. Anche quando una primitiva di  $f(x)$  non è nota, è possibile calcolare un valore approssimato dell'integrale di una funzione continua su un intervallo, con i metodi dell'analisi numerica, che si basano, in sostanza, sulla definizione di integrale come limite di somme, e non sul teorema fondamentale del calcolo integrale. Ecco quindi un caso in cui la "scomoda" definizione di integrale fornisce il modo migliore per calcolarlo. Alla discussione di uno di questi metodi dedicheremo il successivo par. 7.

## 5 METODI ELEMENTARI PER LA RICERCA DI UNA PRIMITIVA. CALCOLO DI INTEGRALI INDEFINITI E DEFINITI

### 5.1 Integrali immediati, per scomposizione, per sostituzione

Come abbiamo visto, se  $f$  è una funzione continua e  $G$  è una sua primitiva, tutte le primitive di  $f$  avranno la forma

$$G(x) + c \quad (c \in \mathbb{R})$$

L'insieme di tutte le primitive di  $f$  prende il nome di *integrale indefinito* di  $f$ , e si indica col simbolo

$$\int f(x) dx$$

Lo stesso simbolo si usa talvolta per indicare una particolare primitiva di  $f$ . Ricordiamo ancora che questo simbolo ha un significato ben diverso da quello di

$$\int_a^b f(x) dx$$

che indica l'integrale di  $f$  sull'intervallo  $[a, b]$ , detto anche *integrale definito*. L'integrale definito è un numero, l'integrale indefinito è un insieme di funzioni (o una qualunque di esse).

Ci occuperemo ora dei *metodi di integrazione*, ossia dei metodi per trovare una primitiva di una funzione data (integrazione indefinita) e quindi per calcolarne il suo integrale definito, mediante il teorema fondamentale del calcolo integrale, almeno per funzioni di tipo abbastanza semplice.

Queste tecniche di integrazione saranno usate in molti argomenti trattati nel secondo volume (equazioni differenziali ordinarie, integrali di linea, integrali doppi e di superficie, serie di Fourier...).

Leggendo la tabella delle derivate delle funzioni elementari "in senso inverso", si ottiene la prima tabella di primitive:

Funzione	Primitiva	Funzione	Primitiva
$k$	$kx$	$e^x$	$e^x$
$x^\alpha$	$\frac{x^{\alpha+1}}{\alpha+1} \quad (\alpha \neq -1)$	$a^x$	$\frac{a^x}{\log a} \quad (a > 0, a \neq 1)$
$\frac{1}{x}$	$\log x $	$\frac{1}{1+x^2}$	$\operatorname{arctg} x$
$\sin x$	$-\cos x$	$\frac{1}{\sqrt{1-x^2}}$	$\arcsin x$
$\cos x$	$\sin x$	$\operatorname{Sh} x$	$\operatorname{Ch} x$
$\frac{1}{(\cos x)^2} = 1 + (\operatorname{tg} x)^2$	$\operatorname{tg} x$	$\operatorname{Ch} x$	$\operatorname{Sh} x$
$\frac{1}{(\sin x)^2}$	$-\cotg x$	$\frac{1}{(\operatorname{Ch} x)^2}$	$\operatorname{Th} x$
		$\frac{1}{(\operatorname{Sh} x)^2}$	$-\operatorname{Coth} x$

Nella tabella precedente, quella fornita è *una* delle primitive. Per scrivere l'integrale indefinito, come insieme di *tutte* le primitive, occorre aggiungere sempre una costante arbitraria, ad esempio:

$$\int x^2 dx = \frac{x^3}{3} + c$$

Le regole elementari con cui, a partire da certi integrali, se ne calcolano altri, sono:

### Integrazione per scomposizione

$$(5.1) \quad \int [\alpha f(x) + \beta g(x)] dx = \alpha \int f(x) dx + \beta \int g(x) dx$$

(segue dalla linearità della derivata);

### Integrazione per sostituzione

Sia  $G$  una primitiva di  $f$  in un intervallo  $I$ , cioè  $G'(t) = f(t)$  per ogni  $t \in I$ . Sia ora  $t = \varphi(x)$  una funzione derivabile con continuità su un intervallo  $[a, b]$  tale che  $\varphi([a, b]) \subset I$ . Dal teorema di derivazione delle funzioni composte abbiamo allora:

$$\frac{d}{dx} G(\varphi(x)) = G'(\varphi(x)) \cdot \varphi'(x) = f(\varphi(x)) \varphi'(x) \quad \forall x \in [a, b]$$

e cioè:

$$\begin{array}{ccc} G(t) & & \Phi(x) = G(\varphi(x)) \\ \text{primitiva di} & \iff & \text{primitiva di} \\ f(t) & & f(\varphi(x)) \cdot \varphi'(x). \end{array}$$

Ne segue la formula (di integrazione per sostituzione)

$$(5.2) \quad \int f(\varphi(x)) \varphi'(x) dx = \int f(t) dt \quad (\varphi(x) = t).$$

È facile ricordare la (5.2) col seguente procedimento formale: si sostituisce  $t = \varphi(x)$  nell'integrale a primo membro, e si calcola  $dt = \varphi'(x) dx$ . Così facendo si ottiene proprio la (5.2) (si tenga presente che questo passaggio è puramente formale: il motivo per cui la (5.2) è vera è stato spiegato in precedenza).

La versione (5.2) per l'integrale definito è la seguente

$$(5.3) \quad \int_a^b f(\varphi(x)) \varphi'(x) dx = \int_{\varphi(a)}^{\varphi(b)} f(t) dt \quad \varphi(x) = t.$$

Di nuovo, è facile ricordare la formula precedente con i passaggi formali già visti: si pone  $t = \varphi(x)$  nell'integrale a primo membro, e si calcola  $dt = \varphi'(x) dx$ ; inoltre, si modificano gli estremi di integrazione tenendo conto che

$$x \in [a, b] \implies t = \varphi(x) \in [\varphi(a), \varphi(b)].$$

dove la scrittura  $[\varphi(a), \varphi(b)]$  indica che  $\varphi(a), \varphi(b)$  sono, in quest'ordine, gli estremi di integrazione del nuovo integrale, anche se, eventualmente,  $\varphi(a) > \varphi(b)$ .

$$\begin{aligned} \int (4x^2 + 5x + 1) dx &= (\text{per la (5.1)}) = \\ &= 4 \int x^2 dx + 5 \int x dx + \int 1 dx = \frac{4}{3} x^3 + \frac{5}{2} x^2 + x + c \end{aligned}$$

$$\int_0^1 \frac{1}{e^x + e^{-x}} dx$$

Poniamo  $e^x = t$  ossia  $x = \log t$ ,  $dx = \frac{1}{t} dt$ .

Si ha inoltre:  $x = 0 \Rightarrow t = 1$ ;  $x = 1 \Rightarrow t = e$ ; perciò dalla (5.3) abbiamo:

$$\int_0^1 \frac{1}{e^x + e^{-x}} dx = \int_1^e \frac{1}{t + \frac{1}{t}} \frac{1}{t} dt = \int_1^e \frac{1}{1+t^2} dt = [\arctg t]_1^e = \arctg e - \frac{\pi}{4}.$$

$$\begin{aligned} \int \tg x dx &= \int \frac{\sin x}{\cos x} dx = [\text{per la (5.3): } \cos x = t; -\sin x dx = dt] = \\ &= \int -\frac{dt}{t} = -\log |t| + c = -\log |\cos x| + c \end{aligned}$$

L'esempio precedente è anche un caso particolare della seguente formula:

$$\int \frac{f'(x)}{f(x)} dx = \log |f(x)| + c$$

come si verifica con la sostituzione  $f(x) = t$ ,  $f'(x) dx = dt$ .

$$\int [f(x)]^\alpha f'(x) dx = \frac{[f(x)]^{\alpha+1}}{\alpha+1} + c \quad \alpha \neq -1$$

Basta porre  $t = f(x)$ ,  $dt = df(x) = f'(x) dx$ ; si ha:

$$\int [f(x)]^\alpha f'(x) dx = \int t^\alpha dt = \frac{t^{\alpha+1}}{\alpha+1} + c = \frac{[f(x)]^{\alpha+1}}{\alpha+1} + c$$

Ad esempio:

$$\begin{aligned} \int x^2 \sqrt[3]{2+x^3} dx &= \frac{1}{3} \int \sqrt[3]{2+x^3} 3x^2 dx = \frac{1}{3} \int (\underbrace{2+x^3}_{f(x)})^{1/3} \underbrace{3x^2}_{f'(x)} dx = \\ &= \frac{1}{3} (2+x^3)^{4/3} \cdot \frac{3}{4} + c = \frac{1}{4} (2+x^3)^{4/3} + c \end{aligned}$$

### Simmetrie

Facciamo ora un'osservazione sull'*integrale di una funzione simmetrica (pari o dispari) su un intervallo simmetrico  $[-k, k]$* . Si osservino le figure 6.8a e 6.8b.

Pensando l'integrale come area con segno sottesa al grafico di  $f$ , si vede subito che se  $f$  è una funzione continua su  $[-k, k]$  e

$$f \text{ è pari, allora } \int_{-k}^k f(x) dx = 2 \int_0^k f(x) dx$$

$$f \text{ è dispari, allora } \int_{-k}^k f(x) dx = 0$$

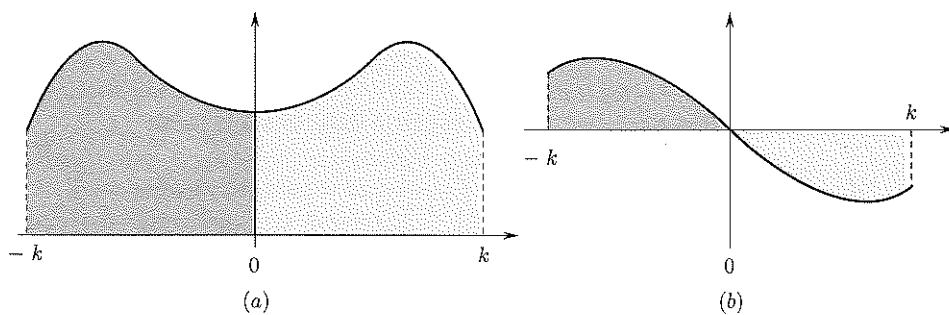


Figura 6.8.

Ad esempio:

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos x dx = (\text{integranda pari}) = 2 \int_0^{\frac{\pi}{2}} \cos x dx = 2[\sin x]_0^{\frac{\pi}{2}} = 2$$

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin x dx = (\text{integranda dispari}) = 0$$

Si osservi anche il prossimo esempio:

$$\int_{-1}^1 e^{-|x|} dx = 2 \int_0^1 e^{-x} dx = 2 \left[ -e^{-x} \right]_0^1 = 2 - \frac{2}{e}$$

In questo caso la simmetria ha consentito di ridursi a un integrale sull'intervallo  $[0, 1]$ ; a sua volta, questo ha consentito di togliere il valore assoluto (per  $x \in [0, 1]$ ,  $e^{-|x|} = e^{-x}$ ), permettendoci quindi di trovare la primitiva.

#### Valori assoluti

L'ultimo esempio suggerisce un'idea generale con cui si può calcolare l'integrale definito di una funzione contenente un valore assoluto. Si consideri il prossimo esempio:

$$\int_0^2 |(x-1)(x+3)| dx.$$

Per  $x \in [0, 2]$ , abbiamo

$$|(x-1)(x+3)| = \begin{cases} (x-1)(x+3) & \text{se } x \in [1, 2] \\ (1-x)(x+3) & \text{se } x \in [0, 1] \end{cases}$$

(si osservi che non ci interessano i casi in cui  $x \notin [0, 2]$ ). Quindi, per l'additività dell'integrale abbiamo:

$$\begin{aligned} \int_0^2 |(x-1)(x+3)| dx &= \int_0^1 (1-x)(x+3) dx + \int_1^2 (x-1)(x+3) dx = \\ &= \left[ -\frac{x^3}{3} - x^2 + 3x \right]_0^1 + \left[ \frac{x^3}{3} + x^2 - 3x \right]_1^2 = \\ &= \left( \frac{5}{3} - 0 \right) + \left( \frac{2}{3} + \frac{5}{3} \right) = \frac{10}{3}. \end{aligned}$$

In generale, volendo calcolare l'integrale definito di una funzione contenente uno o più valori assoluti, si spezzerà l'integrale nella somma di integrali su sottointervalli, in modo che in ciascun sottointervallo l'argomento di ciascun valore assoluto abbia segno costante, e quindi si possa esprimere l'integranda senza usare valori assoluti.

## 5.2 Integrazione delle funzioni razionali

Diamo ora un'idea schematica di come si calcoli in generale la primitiva di una funzione razionale:

$$\int \frac{P_n(x)}{Q_m(x)} dx$$

(abbiamo indicato con  $P_n, Q_m$  polinomi di grado  $n, m$ , rispettivamente).

Anzitutto: se il grado del numeratore è  $\geq$  del grado del denominatore, per prima cosa si esegue la divisione di polinomi; questo porta a riscrivere l'integranda come somma di un polinomio (che si integra immediatamente), più una funzione razionale con lo stesso denominatore di quella di partenza, e il numeratore di grado inferiore. (Questo non conclude il calcolo, ma è un primo passo che semplifica le cose).



$$\int \frac{x^3 + x}{x^2 + x + 1} dx$$

La divisione di polinomi dà:

$$x^3 + x = (x^2 + x + 1)(x - 1) + (x + 1)$$

quindi:

$$\int \frac{x^3 + x}{x^2 + x + 1} dx = \int (x - 1) dx + \int \frac{x + 1}{x^2 + x + 1} dx$$

In base a questa osservazione, possiamo d'ora in poi supporre che il grado del numeratore sia già minore del grado del denominatore (altrimenti ci si riduce a questo caso mediante divisione).

Se il denominatore è di **primo grado** il suo integrale si calcola facilmente, mediante logaritmo:



$$\int \frac{2}{3x+5} dx = \frac{2}{3} \log |3x+5| + c$$

La formula generale è la seguente:

$$\int \frac{dx}{ax+b} = \frac{1}{a} \log |ax+b| + c.$$

Se il denominatore è di **secondo grado** (e il numeratore di grado  $\leq 1$ ), occorre distinguere 3 casi, a seconda del segno del discriminante del denominatore.

**A.** Il denominatore ha due radici distinte. La frazione si scomponete in fratti semplici e si integra poi mediante somma di logaritmi:



$$\int \frac{x+2}{(x-2)(x+3)} dx$$

Scriviamo:

$$\frac{x+2}{(x-2)(x+3)} = \frac{a}{x-2} + \frac{b}{x+3}$$

per opportuni coefficienti  $a, b$  da determinarsi. Per determinarli, si esegue la somma a secondo membro facendo denominatore comune, e si impone che la frazione trovata sia uguale a quella a primo membro. Si trova:

$$\frac{x(a+b) + (3a-2b)}{(x-2)(x+3)} = \frac{x+2}{(x-2)(x+3)}$$

che è vero se e solo se:

$$\begin{cases} a+b=1 \\ 3a-2b=2 \end{cases} \quad \begin{cases} a=\frac{4}{5} \\ b=\frac{1}{5} \end{cases}$$

Quindi l'integrale di partenza è uguale a:

$$\int \left( \frac{\frac{4}{5}}{x-2} + \frac{\frac{1}{5}}{x+3} \right) dx = \frac{4}{5} \log|x-2| + \frac{1}{5} \log|x+3| + c$$

$$\begin{aligned} \int \frac{dx}{x^2-4} &= \int \frac{1}{(x-2)(x+2)} dx = \int \frac{1}{4} \left[ \frac{1}{x-2} - \frac{1}{x+2} \right] dx = \\ &= \frac{1}{4} [\log|x-2| - \log|x+2|] + c \end{aligned}$$

**B.** Il denominatore è un quadrato perfetto. Mediante sostituzione ci si riconduce alla somma di potenze (con esponente positivo o negativo), che si integra immediatamente:

$$\begin{aligned} \int \frac{x+1}{(3x+2)^2} dx &= [3x+2=t, 3dx=dt, x=\frac{t-2}{3}] \\ &= \int \frac{\frac{t-2}{3}+1}{t^2} \frac{dt}{3} = \frac{1}{9} \int \frac{t+1}{t^2} dt = \frac{1}{9} \int \left( \frac{1}{t} + \frac{1}{t^2} \right) dt = \\ &= \frac{1}{9} \left( \log|t| - \frac{1}{t} \right) + c = \frac{1}{9} \left( \log|3x+2| - \frac{1}{3x+2} \right) + c \end{aligned}$$

**C.** Il denominatore non si annulla mai. Vediamo su un paio di esempi come si procede in generale in questo caso:

$$\int \frac{dx}{x^2+3} = \frac{1}{3} \int \frac{dx}{(x/\sqrt{3})^2+1} = \frac{\sqrt{3}}{3} \operatorname{arctg}\left(\frac{x}{\sqrt{3}}\right) + c$$

In generale:

$$(5.4) \quad \int \frac{1}{x^2+a^2} dx = \frac{1}{a} \operatorname{arctg}\frac{x}{a} + c$$

come si verifica immediatamente calcolando la derivata del secondo membro. Questo esempio illustra il caso più semplice in cui il denominatore non si annulla mai. La situazione più generale è descritta nel prossimo esempio.

$$\int \frac{x}{x^2+2x+4} dx = \frac{1}{2} \int \frac{2x+2}{x^2+2x+4} dx - \int \frac{1}{x^2+2x+4} dx$$

Si osservi com'è costruita l'identità: il primo addendo deve avere a numeratore la derivata del denominatore, in questo caso  $2x+2$ ; d'altro canto il secondo addendo non deve contenere  $x$  a numeratore; il "trucco" è perciò scrivere:

$$x = \frac{1}{2}(2x+2) - 1$$

e *non*, ad esempio,  $x = (2x+2) - (x+2)$ , che sarebbe vero ma inconcludente, in quanto il secondo integrale avrebbe ancora la  $x$  a numeratore. Ora l'espressione trovata è uguale a:

$$\frac{1}{2} \log(x^2+2x+4) - \int \frac{1}{(x+1)^2+3} dx = \frac{1}{2} \log(x^2+2x+4) - \frac{1}{\sqrt{3}} \operatorname{arctg}\frac{x+1}{\sqrt{3}} + c$$

Nell'ultimo integrale si è usata la formula vista per la primitiva di  $\frac{1}{x^2+a^2}$ , più la sostituzione  $x+1=t$ ,  $dx=dt$ .

In generale, se il denominatore ha discriminante negativo, ossia è irriducibile, si spezza la frazione in due addendi, di cui una a numeratore presenta la derivata del denominatore, e porta ad un integrale logaritmico; l'altra ha numeratore costante, e si riporta all'integrale di una funzione arcotangente. Il procedimento generale è sintetizzato dalla formula seguente:

$$\begin{aligned} \int \frac{x+d}{(x+b)^2+a^2} dx &= \frac{1}{2} \int \frac{2x+2b}{x^2+2bx+b^2+a^2} dx + \int \frac{d-b}{(x+b)^2+a^2} dx = \\ &= \frac{1}{2} \log(x^2+2bx+b^2+a^2) + \frac{d-b}{a} \operatorname{arctg}\left(\frac{x+b}{a}\right) + c. \end{aligned}$$

**Se il denominatore è un polinomio di grado maggiore di due** è sempre possibile (teoricamente!) scomporlo in un prodotto di (potenze di) fattori di primo grado, oppure di secondo grado irriducibili. Fatto questo, si scomponete la frazione in fratti semplici, a cui si applicano i discorsi precedenti. I prossimi esempi illustrano solo alcuni casi semplici in cui questo accade. Se il grado del denominatore è molto maggiore di due... procurarsi un buon software!!



$$\int \frac{2x^2+3x-1}{x^3+2x^2+x+2} dx$$

Osserviamo il denominatore:

$$x^3+2x^2+x+2 = x^2(x+2)+(x+2) = (x+2)(x^2+1)$$

Scomposto il denominatore, si scomponete il quoziente in fratti semplici:

$$\frac{2x^2 + 3x - 1}{(x+2)(x^2+1)} = \frac{a}{x+2} + \frac{bx+c}{x^2+1}$$

con  $a, b, c$  da determinarsi (il criterio con cui abbiamo scritto i numeratori incogniti è che il grado del numeratore sia minore di uno del grado del denominatore). Mettendo a denominatore comune, si trova:

$$\frac{a}{x+2} + \frac{bx+c}{x^2+1} = \frac{a(x^2+1) + (bx+c)(x+2)}{(x+2)(x^2+1)} = \frac{2x^2 + 3x - 1}{(x+2)(x^2+1)}$$

se e solo se:

$$\begin{cases} a+b=2 \\ 2b+c=3 \\ a+2c=-1 \end{cases} \quad \text{sistema che risolto dà: } \begin{cases} a=\frac{1}{5} \\ b=\frac{9}{5} \\ c=-\frac{3}{5} \end{cases}$$

Perciò:

$$\begin{aligned} \int \frac{2x^2 + 3x - 1}{x^3 + 2x^2 + x + 2} dx &= \frac{1}{5} \int \frac{1}{x+2} dx + \frac{1}{5} \int \frac{9x-3}{x^2+1} dx = \\ &= \frac{1}{5} \log|x+2| + \frac{9}{10} \int \frac{2x}{x^2+1} dx - \frac{3}{5} \int \frac{dx}{x^2+1} = \\ &= \frac{1}{5} \log|x+2| + \frac{9}{10} \log(x^2+1) - \frac{3}{5} \arctg x + c \end{aligned}$$

$$\int \frac{x+1}{x^2(x+3)} dx$$

Si scomponete:

$$\frac{x+1}{x^2(x+3)} = \frac{ax+b}{x^2} + \frac{c}{x+3}$$

e si procede al solito modo. Si noti che

$$\int \frac{ax+b}{x^2} dx = \int \left( \frac{a}{x} + \frac{b}{x^2} \right) dx$$

che è immediato.

$$\int \frac{x^2+1}{(x+1)^2(x^2-x+1)} dx$$

Il fattore  $(x^2 - x + 1)$  è irriducibile. Si scomponete:

$$\frac{x^2+1}{(x+1)^2(x^2-x+1)} = \frac{ax+b}{(x+1)^2} + \frac{cx+d}{x^2-x+1}$$

e si imposta un sistema di quattro equazioni in quattro incognite che risolto dà la scomposizione in fratti semplici. Si trovano così due integrali di tipi che abbiamo già trattato.

### Funzioni razionali di $e^x$

Dovendo integrare una funzione razionale di  $e^x$ , si pone  $e^x = t$ ;  $x = \log t$ ;  $dx = \frac{dt}{t}$ , e ci si riconduce a una funzione razionale di  $t$ . Occorre, alla fine, tornare alla variabile  $x$ .



$$\begin{aligned} \int \frac{1}{Chx} dx &= 2 \int \frac{dx}{e^x + e^{-x}} = \left[ e^x = t; x = \log t; dx = \frac{dt}{t} \right] = \\ &= 2 \int \frac{1}{t + \frac{1}{t}} \frac{dt}{t} = 2 \int \frac{dt}{1+t^2} = 2 \arctg t + c = 2 \arctg(e^x) + c \end{aligned}$$



$$\int \frac{x^5}{1+x^2} dx$$

$$\int \frac{x+2}{x^2+4x+3} dx$$

$$\int \frac{1+2x^2}{x^4-1} dx$$

$$\int_0^1 \frac{1}{x^3+1} dx$$

$$\int \frac{x^3+2}{x^2+1} dx$$

$$\int \frac{2x+3}{(3x+2)^2} dx$$

$$\int \frac{1-x^6}{1-x} dx$$

$$\int \frac{2x-1}{x^2+2x+4} dx$$

$$\int \frac{x+5}{x^2(x+1)} dx$$

$$\int \frac{1-2x}{(x+1)^2(x^2+1)} dx$$

$$\int \frac{1+e^{-x}}{Chx} dx$$

$$\int_0^{\sqrt{2}} \frac{x+2}{2x^2+1} dx$$

$$\int_2^5 \frac{2x+3}{x^2+x-2} dx$$

$$\int \frac{3x+1}{x^2+4x+4} dx$$

### 5.3 Integrazione per parti

Se  $f$  e  $g$  sono derivabili in  $[a, b]$  si ha

$$(fg)' = f'g + fg'$$

ossia

$$fg' = (fg)' - f'g.$$

Prendendo l'integrale indefinito di entrambi i membri ed osservando che  $\int (fg)' dx = fg$  si trova la formula di *integrazione per parti*

$$(5.5) \quad \boxed{\int f(x)g'(x) dx = f(x)g(x) - \int f'(x)g(x) dx.}$$

Usando la notazione differenziale, la (5.5) si può scrivere nella forma seguente (più sintetica)

$$(5.6) \quad \underbrace{\int f}_{\substack{\text{fattore} \\ \text{finito}}} \underbrace{dg}_{\substack{\text{fattore} \\ \text{differenziale}}} = fg - \int g df$$

Le (5.5) e (5.6) hanno una versione per l'*integrale definito*:

$$\int_a^b f(x)g'(x) dx = [f(x)g(x)]_a^b - \int_a^b f'(x)g(x)dx$$

Nei casi concreti occorre saper scegliere correttamente il fattore finito e quello differenziale.



$$\int x \sin x dx$$

Sappiamo calcolare sia la derivata che la primitiva di entrambe le funzioni  $x, \sin x$ ; tuttavia la scelta migliore è quella di derivare  $x$ , il che semplifica l'integrale:

$$\int \begin{matrix} x \sin x dx \\ \downarrow \quad \downarrow \\ f \quad g' \end{matrix} = -x \cos x - \int (-\cos x)dx = -x \cos x + \sin x + c$$

La scelta inversa avrebbe portato

$$\int \begin{matrix} x \sin x dx \\ \downarrow \quad \downarrow \\ g' \quad f \end{matrix} = \frac{x^2}{2} \sin x - \int \frac{x^2}{2} \cos x dx = \text{più complicato dell'integrale di partenza}$$

Questo semplice esempio contiene un'idea molto generale: dovendo calcolare un integrale

$$\int x^n f(x) dx \quad \text{con} \quad f(x) = \cos x, \sin x, e^x, \operatorname{Sh} x, \operatorname{Ch} x$$

si integra per parti, derivando  $x^n$  (che si abbassa di grado) e integrando  $f(x)$  (che dà una funzione "simile":  $\sin x, -\cos x, e^x, \operatorname{Ch} x, \operatorname{Sh} x$ , rispettivamente).

Per  $n = 1$  questo porta un integrale immediato; per  $n > 1$  si itera il procedimento finché la derivata della potenza scompare del tutto. Per la linearità dell'integrale indefinito, il procedimento permette di calcolare un integrale del tipo

$$\int P_n(x) f(x) dx$$

con  $P_n(x)$  polinomio di grado  $n$ , e  $f(x)$  come sopra.



$$\int e^x \sin x dx$$

Entrambe le funzioni  $e^x, \sin x$  hanno derivata e primitiva altrettanto semplici: la scelta di quale fattore integrare e quale derivare sembra essere indifferente (e lo è). Procediamo ad esempio così:

$$\int \begin{matrix} e^x \sin x dx \\ \downarrow \quad \downarrow \\ g' \quad f \end{matrix} = e^x \sin x - \int e^x \cos x dx =$$

Applichiamo un'altra integrazione per parti, scegliendo  $f, g'$  coerentemente alla prima scelta fatta, ossia  $g'(x) = e^x, f(x) = \cos x$ :

$$= e^x \sin x - \{e^x \cos x - \int e^x (-\sin x) dx\}$$

Se chiamiamo  $I$  l'integrale di partenza, abbiamo trovato:

$$(5.7) \quad I = e^x \sin x - e^x \cos x - I$$

Apparentemente, la doppia integrazione per parti non ha semplificato nulla; in realtà, però, l'integrale  $I$  si ottiene ora risolvendo l'equazione elementare (5.7):

$$2I = e^x \sin x - e^x \cos x \implies I = \frac{1}{2} e^x (\sin x - \cos x) + c$$

Con lo stesso metodo dell'esempio precedente si calcolano le seguenti primitive (si lasciano i calcoli per esercizio):

$$\int (\operatorname{Ch} x)^2 dx = \frac{1}{2} (\operatorname{Sh} x \operatorname{Ch} x + x) + c; \quad \int (\operatorname{Sh} x)^2 dx = \frac{1}{2} (\operatorname{Sh} x \operatorname{Ch} x - x) + c$$

In generale: dovendo calcolare un integrale del tipo:

$$\int f(x)g(x)dx \quad \text{con} \quad f(x) = \begin{cases} e^{ax} \\ \operatorname{Sh}(ax) \\ \operatorname{Ch}(ax) \end{cases} \quad g(x) = \begin{cases} \cos(bx) \\ \sin(bx) \end{cases}$$

si eseguono due integrazioni per parti successive; nella prima, la scelta della funzione da derivare o da integrare è indifferente; nella seconda, però, la scelta dev'essere coerente alla prima. Dopo due integrazioni per parti si trova un'espressione del tipo  $I = h(x) - \left(\frac{b}{a}\right)^2 I$ , da cui si ricava  $I$ .



$$\int \log x dx$$

Il logaritmo è una funzione elementare di cui ancora non conosciamo la primitiva. Integriamo per parti, ponendo  $f(x) = \log x, g'(x) = 1$ :

$$\int \log x dx = x \log x - \int x \cdot \frac{1}{x} dx = x \log x - \int 1 dx = x \log x - x + c$$

Più in generale, con questo metodo si può calcolare

$$\begin{aligned} \int \begin{matrix} x^m (\log x)^n dx \\ \downarrow \quad \downarrow \\ g' \quad f \end{matrix} &= \frac{x^{m+1}}{m+1} (\log x)^n - \int \frac{x^{m+1}}{m+1} n \frac{(\log x)^{n-1}}{x} dx = \\ &= \frac{x^{m+1}}{m+1} (\log x)^n - \frac{n}{m+1} \int x^m (\log x)^{n-1} dx. \end{aligned}$$

L'ultimo integrale ottenuto è simile a quello di partenza, con il logaritmo elevato ad  $(n-1)$  anziché a  $n$ . Questo significa che, iterativamente, con  $n$  integrazioni per parti si può calcolare l'integrale. Più in generale, questo metodo consente allora di calcolare l'integrale di funzioni del tipo

$$P_m(x) Q_n(\log x)$$

dove  $P_m, Q_n$  sono polinomi di grado  $m, n$ , rispettivamente.



$$\int \arctg x dx$$

L'arcotangente è una funzione elementare di cui ancora non conosciamo la primitiva. Integriamo per parti, ponendo  $f(x) = \arctg x$ ,  $g'(x) = 1$ :

$$\begin{aligned}\int \arctg x dx &= x \arctg x - \int \frac{x}{1+x^2} dx = \\ &= x \arctg x - \frac{1}{2} \log(1+x^2) + c,\end{aligned}$$

dove si è integrata coi metodi noti la funzione razionale  $\frac{x}{1+x^2}$ .

Più in generale, si può calcolare:

$$\int \underset{g'}{\downarrow} x^n \underset{f}{\arctg} x dx = \frac{x^{n+1}}{n+1} \arctg x - \int \frac{x^{n+1}}{n+1} \cdot \frac{dx}{1+x^2}.$$

L'ultimo integrale ottenuto è quello di una funzione razionale: se  $n \geq 1$  occorrerà per prima cosa eseguire la divisione di polinomi  $x^{n+1} : 1+x^2$  per riscrivere

$$\frac{x^{n+1}}{1+x^2} = P_{n-1}(x) + \frac{ax+b}{1+x^2}$$

con  $P_{n-1}$  polinomio di grado  $n-1$ , che si integra direttamente, mentre la funzione razionale  $\frac{ax+b}{1+x^2}$  si integra coi metodi visti, riportandosi a funzioni logaritmo e arcotangente.



Si presti attenzione ai prossimi integrali: in quali casi è utile effettuare una sostituzione? (se non è utile, si può comunque sviluppare il cubo e integrare il polinomio termine a termine)

$$\int (3x+1)^3 dx \quad \int (3x^2+1)^3 dx \quad \int x(3x^2+1)^3 dx$$

$$\textcircled{1} \quad \int (x^{5/6} + 2x^{-2} - 3x^{-1} + 2) dx$$

$$\textcircled{2} \quad \int_0^{\frac{3\pi}{2}} (\sin x)^2 dx$$

$$\textcircled{3} \quad \int \cot g gx dx \quad \int \operatorname{Th} x dx$$

$$\textcircled{4} \quad \int_0^{\pi/4} \cos 2x \sin x dx$$

$$\textcircled{5} \quad \int \left(\sqrt{x} + \sqrt[3]{x+2} + \frac{1}{x^2}\right) dx$$

$$\textcircled{6} \quad \int_0^1 (3x^3 + 1)^2 x^2 dx$$

$$\textcircled{7} \quad \int (1 + (\sin x)^2)^3 \sin 2x dx$$

$$\textcircled{8} \quad \int_e^{e^2} \frac{1}{x(\log x)^2} dx$$

$$\textcircled{9} \quad \int (\operatorname{tg} x)^2 dx \quad \int \operatorname{Th}^2 x dx$$

$$\textcircled{10} \quad \int_{-1}^1 \frac{x}{1+x^2} dx$$

$$\textcircled{11} \quad \int \left(\frac{e^{2x} - e^{-x}}{3}\right) dx$$

$$\textcircled{12} \quad \int_{-1}^1 e^{-|x|} dx$$

$$\textcircled{13} \quad \int e^{-x^2} x dx$$

$$\textcircled{14} \quad \int (x^2 + 5x + 4) e^x dx$$

$$\textcircled{15} \quad \int \frac{dx}{x \log x}$$

$$\textcircled{16} \quad \int_0^1 x e^{3x} dx$$

$$\textcircled{17} \quad \int \sqrt{2 + \sin x} \cos x dx$$

$$\textcircled{18} \quad \int_1^3 x^2 \log x dx$$

$$\textcircled{19} \quad \int \frac{1}{x(1+(\log x)^2)} dx$$

$$\textcircled{20} \quad \int_0^\pi x \cos x dx$$

$$\textcircled{21} \quad \int \frac{1}{x(x+1)} dx$$

$$\textcircled{22} \quad \int e^x (\sin x)^2 dx$$

$$\textcircled{23} \quad \int \frac{1}{(x+1)^2} dx$$

$$\textcircled{24} \quad \int e^{-x^2} x^3 dx$$

$$\textcircled{25} \quad \int_0^{\frac{\pi}{4}} (\cos x)^2 dx$$

$$\textcircled{26} \quad \int (\operatorname{Ch} x)^2 (\operatorname{Sh} x)^2 dx$$

#### 5.4 Integrazione delle funzioni trigonometriche

In questo paragrafo faremo una breve panoramica di metodi per integrare una funzione trigonometrica.

##### Integrali del tipo

$$\int f(\sin x) \cos x dx : \text{ si pone } \sin x = t, \cos x dx = dt$$

$$\int f(\cos x) \sin x dx : \text{ si pone } \cos x = t, -\sin x dx = dt$$



$$\int_0^{\pi/3} (\sin x)^3 (\cos x)^2 dx$$

Si osserva che:

$$(\sin x)^3 (\cos x)^2 = \sin x \cdot (\sin x)^2 \cdot (\cos x)^2 = \sin x (1 - (\cos x)^2) (\cos x)^2 = f(\cos x) \sin x$$

$$\int_0^{\pi/3} \sin x (1 - (\cos x)^2) (\cos x)^2 dx = \left[ \cos x = t; -\sin x dx = dt; t \in \left(1, \frac{1}{2}\right) \right] =$$

$$= \int_1^{1/2} -(1-t^2)t^2 dt = \int_{1/2}^1 (t^2 - t^4) dt =$$

$$= \left[ \frac{t^3}{3} - \frac{t^5}{5} \right]_{1/2}^1 = \frac{1}{3} - \frac{1}{5} - \frac{1}{24} + \frac{1}{160}$$

In generale: dovendo calcolare

$$\int (\sin x)^n (\cos x)^m dx$$

con almeno uno degli esponenti dispari, è sempre possibile riscrivere l'integrandi in una delle forme  $f(\sin x) \cos x$  o  $f(\cos x) \sin x$ , sfruttando la relazione  $(\sin x)^2 + (\cos x)^2 = 1$ . A questo punto l'integrale si calcola con la sostituzione  $\sin x = t$ , o  $\cos x = t$ , rispettivamente.

Se invece entrambi gli esponenti sono pari, si possono usare le formule trigonometriche per l'abbassamento del grado:

$$(\cos x)^2 = \frac{1}{2}(1 + \cos 2x) \quad (\sin x)^2 = \frac{1}{2}(1 - \cos 2x)$$

$\int (\sin x)^2 (\cos x)^4 dx$

Si verifichi che dalle formule precedenti segue:

$$\int (\sin x)^2 (\cos x)^4 dx = \frac{1}{8} \int (-(\cos 2x)^3 - (\cos 2x)^2 + \cos 2x + 1) dx$$

Ora: gli addendi  $\cos 2x + 1$  hanno integrale immediato;  $(\cos 2x)^2$  si integra come visto in precedenza (attenzione alla presenza di  $2x$  anziché  $x$ );  $(\cos 2x)^3$  rientra nel caso in cui uno degli esponenti è dispari: si riscrive  $(\cos 2x)^3 = \cos 2x(1 - (\sin 2x)^2)$ , e si integra ponendo  $\sin 2x = t$ . Si trova (verificare i calcoli per esercizio):

$$\int (\sin x)^2 (\cos x)^4 dx = \frac{12x + 3 \sin 2x - 3 \sin 4x - \sin 6x}{192} + c$$

(o una delle molte espressioni trigonometriche equivalenti a questa!)

Calcoliamo, in particolare, due integrali notevoli di funzioni trigonometriche:

$$\int (\sin x)^2 dx, \int (\cos x)^2 dx$$

Sfruttando le formule trigonometriche per l'abbassamento del grado (v. sopra) si ha:

$$\int (\sin x)^2 dx = \frac{1}{2} \int (1 - \cos 2x) dx = \frac{1}{2} \left( x - \frac{1}{2} \sin 2x \right) + c = \frac{x - \sin x \cos x}{2} + c$$

$$\int (\cos x)^2 dx = \frac{1}{2} \int (1 + \cos 2x) dx = \frac{1}{2} \left( x + \frac{1}{2} \sin 2x \right) + c = \frac{x + \sin x \cos x}{2} + c$$

Sfruttando il risultato dell'esempio precedente si possono calcolare, ad esempio, i seguenti integrali definiti:

$$(5.8) \quad \int_0^{k\frac{\pi}{2}} (\sin x)^2 dx = \int_0^{k\frac{\pi}{2}} (\cos x)^2 dx = k \frac{\pi}{4} \quad \text{con } k \in \mathbb{Z}$$

La formula precedente, che esprime due *integrali notevoli* che si presentano spesso, è facile da ricordare: l'integrale di  $(\cos x)^2, (\sin x)^2$  su un intervallo del tipo  $[0, k\frac{\pi}{2}]$  vale la metà dell'area del rettangolo di base  $[0, k\frac{\pi}{2}]$  e altezza 1. Questa "coincidenza" ha un significato geometrico; si osservi la figura 6.9:

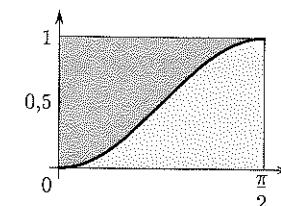


Figura 6.9.

Le due regioni disegnate hanno la stessa area; quella inferiore è l'integrale di  $(\sin x)^2$  su  $[0, \frac{\pi}{2}]$ , quella superiore è l'integrale di  $(\cos x)^2$ , perché  $(\cos x)^2 = 1 - (\sin x)^2$ ; ognuno dei due integrali vale dunque  $\frac{1}{2} \cdot \frac{\pi}{2}$ . Per le simmetrie e periodicità dei grafici di  $(\cos x)^2, (\sin x)^2$ , ne segue la formula generale.

### Integrali del tipo

$$\int \cos \alpha x \sin \beta x dx, \int \cos \alpha x \cos \beta x dx, \int \sin \alpha x \sin \beta x dx$$

Si possono utilizzare le formule di prostaferesi (v. appendice A), che riconducono alla somma di due integrali immediati.



Applicando il metodo visto, si verifica che:

$$\int_0^{2\pi} \cos nx \sin mx dx = 0 \text{ per ogni coppia di interi } n, m$$

$$\int_0^{2\pi} \sin nx \sin mx dx = 0 \text{ per ogni coppia di interi } n, m, n \neq m$$

$$\int_0^{2\pi} \cos nx \cos mx dx = 0 \text{ per ogni coppia di interi } n, m, n \neq m$$

Questi integrali sono importanti nello studio delle serie di Fourier (v. secondo volume), assieme ai seguenti (che non si calcolano con le formule di prostaferesi, ma sostituendo  $nx = t$  e sfruttando la (5.8)):

$$\int_0^{2\pi} (\cos nx)^2 dx = \int_0^{2\pi} (\sin nx)^2 dx = \pi \quad \text{per ogni intero positivo } n$$

### Funzioni razionali di $\sin x, \cos x$

L'integrale di una funzione razionale di  $\sin x, \cos x$ , può sempre essere ricondotto all'integrale di una funzione razionale, mediante la seguente sostituzione:

$$t = \tan \frac{x}{2} \quad x = 2 \arctan t \quad dx = \frac{2}{1+t^2} dt$$

La sostituzione è utile in forza delle seguenti identità trigonometriche:

$$\begin{cases} \cos x = \frac{1-t^2}{1+t^2} \\ \sin x = \frac{2t}{1+t^2} \end{cases} \quad \text{se } t = \tan \frac{x}{2}$$

Questo metodo porta, in genere, a calcoli laboriosi, e va utilizzato perciò solo quando non sembra esservi una via più semplice o un buon software a portata di mano.



$$\int \frac{\sin x - 5 \cos x}{3 + \sin x} dx$$

L'integrale si scomponete nella somma di due:

$$I = \int \frac{\sin x}{3 + \sin x} dx - 5 \int \frac{\cos x}{3 + \sin x} dx = I_1 - 5I_2$$

Il secondo è immediato:

$$I_2 = \log(3 + \sin x) + c$$

Il primo non è immediato, ma può essere trattato con la sostituzione appena vista. Può essere utile un'ultima semplificazione preliminare:

$$I_1 = \int \frac{\sin x}{3 + \sin x} dx = \int \left(1 - \frac{3}{3 + \sin x}\right) dx = x - 3 \int \frac{1}{3 + \sin x} dx = x - 3I_3$$

Calcoliamo ora

$$\begin{aligned} I_3 &= \left[ t = \operatorname{tg} \frac{x}{2}; \sin x = \frac{2t}{1+t^2}; dx = \frac{2dt}{1+t^2} \right] = \\ &= \int \frac{1}{3 + \frac{2t}{1+t^2}} \frac{2dt}{1+t^2} = 2 \int \frac{1}{3t^2 + 2t + 3} dt \end{aligned}$$

che è l'integrale di una funzione razionale di  $t$ . Scritta la primitiva di questa, occorre ritornare alla variabile  $x$ . I calcoli (noiosi ma ormai di routine) danno:

$$I_3 = \frac{1}{\sqrt{2}} \operatorname{arctg} \left( \frac{3}{2\sqrt{2}} \left( \operatorname{tg} \frac{x}{2} + \frac{1}{3} \right) \right) + c$$

e in definitiva si ha:

$$\begin{aligned} I &= I_1 - 5I_2 = x - 3I_3 - 5 \log(3 + \sin x) = \\ &= x - \frac{3}{\sqrt{2}} \operatorname{arctg} \left( \frac{3}{2\sqrt{2}} \left( \operatorname{tg} \frac{x}{2} + \frac{1}{3} \right) \right) - 5 \log(3 + \sin x) + c \end{aligned}$$



$$\textcircled{1} \quad \int (\sin x)^3 dx$$

$$\textcircled{2} \quad \int_0^{\pi/2} \frac{\sin x + 2 \cos x}{2 + \sin x} dx$$

$$\textcircled{3} \quad \int (\sin x)^4 dx$$

$$\textcircled{4} \quad \int (\sin x)^2 \sin 2x dx$$

$$\textcircled{5} \quad \int_0^{\pi/8} (\sin x)^2 (\cos x)^2 dx$$

$$\textcircled{6} \quad \int \sin 2x \cos 3x dx$$

## 5.5 Integrazione delle funzioni irrazionali

Vediamo alcuni metodi di integrazioni applicabili a funzioni irrazionali di tipo molto semplice.

### Sostituzioni tipiche per trattare funzioni razionali di $x$ e una delle seguenti funzioni

$$\sqrt{a^2 - x^2}; \sqrt{a^2 + x^2}; \sqrt{x^2 - a^2}$$

Se l'integranda è una funzione razionale di  $x$  e di una sola delle precedenti radici quadrate, si effettua una sostituzione standard:

- A.  $\sqrt{a^2 - x^2}$ , con  $a > 0$ . Si pone:  $x = a \sin t, dx = a \cos t dt$ ;

$$\sqrt{a^2 - x^2} = \sqrt{a^2(1 - (\sin t)^2)} = a |\cos t|$$

(Se si sta calcolando un integrale definito, si sa dove varia  $x$ , e quindi  $t$ ; perciò si può togliere il modulo mettendo il segno opportuno). L'integrale è quindi ricondotto a quello di una funzione razionale di  $\sin t, \cos t$ . A sua volta, questo è sempre riconducibile all'integrale di una funzione razionale, mediante la sostituzione  $\operatorname{tg} \frac{t}{2} = u$  (o un'altra via più semplice, se possibile).

- B.  $\sqrt{a^2 + x^2}$ , con  $a > 0$ . Si pone:  $x = a \operatorname{Sh} t, dx = a \operatorname{Ch} t dt$ ;

$$\sqrt{a^2 + x^2} = \sqrt{a^2(1 + (\operatorname{Sh} t)^2)} = a \operatorname{Ch} t$$

(Si osservi che  $\operatorname{Ch} t$  è sempre positivo). L'integrale è quindi ricondotto a quello di una funzione razionale di  $\operatorname{Sh} t, \operatorname{Ch} t$ . A sua volta, questo è sempre riconducibile all'integrale di una funzione razionale, mediante la sostituzione  $e^t = u; t = \log u; dt = \frac{du}{u}$  (oppure, l'integrale si calcola per un'altra via più semplice).

- C.  $\sqrt{x^2 - a^2}$ , con  $a > 0$ . Se  $x \geq a$ , si pone:  $x = a \operatorname{Ch} t, dx = a \operatorname{Sh} t dt$ ;

$$\sqrt{x^2 - a^2} = \sqrt{a^2((\operatorname{Ch} t)^2 - 1)} = a |\operatorname{Sh} t|$$

(Se si sta calcolando un integrale definito, si sa dove varia  $x$ , e quindi  $t$ ; perciò si può togliere il modulo mettendo il segno opportuno). L'integrale è quindi ricondotto a quello di una funzione razionale di  $\operatorname{Sh} t, \operatorname{Ch} t$  (v. caso B).

Se  $x \leq -a$ , si pone invece  $x = -a \operatorname{Cht}, dx = -a \operatorname{Sh} t dt$ .

Osserviamo che quando si effettuano le sostituzioni B e C, nel ritornare alla variabile  $x$  occorre usare le *funzioni iperboliche inverse*:

$$x = a \operatorname{Cht} \Rightarrow t = \operatorname{SettCh} \left( \frac{x}{a} \right) = \log \left( \frac{x}{a} + \sqrt{\frac{x^2}{a^2} - 1} \right)$$

$$x = a \operatorname{Sht} \Rightarrow t = \operatorname{SettSh} \left( \frac{x}{a} \right) = \log \left( \frac{x}{a} + \sqrt{\frac{x^2}{a^2} + 1} \right)$$

A questo proposito, è utile anche notare che:

$$\operatorname{Sh}(\operatorname{SettCh} \alpha) = \sqrt{\alpha^2 - 1} \quad \operatorname{Ch}(\operatorname{SettSh} \alpha) = \sqrt{\alpha^2 + 1}$$



Calcoliamo l'area dell'ellisse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Per simmetria, quest'area è quattro volte quella compresa nel primo quadrante, che può calcolarsi come integrale su  $[0, a]$  della funzione:

$$y = b\sqrt{1 - \frac{x^2}{a^2}}$$

$$\begin{aligned} \text{Area ellisse} &= 4 \int_0^a b\sqrt{1 - \frac{x^2}{a^2}} dx = \left[ x = a \sin t; dx = a \cos t dt; t = \arcsin \frac{x}{a} \in \left(0, \frac{\pi}{2}\right) \right] \\ &= 4ab \int_0^{\frac{\pi}{2}} (\cos t)^2 dt = 4ab \cdot \frac{\pi}{4} = \pi ab \end{aligned}$$

risultato che generalizza in modo naturale la formula per l'area del cerchio,  $\pi r^2$ .

$$\int_{\sqrt{2}}^2 \frac{1}{x^2\sqrt{x^2-1}} dx =$$

$$[x = \operatorname{Cht}; dx = \operatorname{Sht} dt; t = \operatorname{SettCh} x \in (\operatorname{SettCh}\sqrt{2}, \operatorname{SettCh} 2)]$$

$$= \int_{\operatorname{SettCh}\sqrt{2}}^{\operatorname{SettCh} 2} \frac{\operatorname{Sht}}{(\operatorname{Ch} t)^2 \operatorname{Sht}} dt = \int_{\operatorname{SettCh}\sqrt{2}}^{\operatorname{SettCh} 2} \frac{1}{(\operatorname{Ch} t)^2} dt = \left[ \frac{\operatorname{Sht}}{\operatorname{Ch} t} \right]_{\operatorname{SettCh}\sqrt{2}}^{\operatorname{SettCh} 2} = \frac{\sqrt{3}}{2} - \frac{1}{\sqrt{2}}$$

#### Integrale di una funzione razionale di $x, x^{n_1/m_1}, x^{n_2/m_2}, \dots$

Si pone  $x = t^n$  con  $n = \minimo$  comune multiplo di  $m_1, m_2, \dots$ . Si ha quindi  $dx = nt^{n-1}dt$ , e si ottiene una funzione razionale di  $t$ . Integrata questa, si torna alla variabile  $x$ .



$$\int \frac{\sqrt{x}}{2x^{1/3} + 3} \frac{dx}{x}$$

La funzione integranda dipende da due diverse potenze a esponente razionale di  $x$ :  $x^{1/2}, x^{1/3}$ . Se poniamo:

$$x = t^6, dx = 6t^5 dt$$

otteniamo una funzione razionale di  $t$ , perché  $x^{1/2} = t^3, x^{1/3} = t^2$ :

$$\begin{aligned} \int \frac{\sqrt{x}}{2x^{1/3} + 3} \frac{dx}{x} &= \int \frac{t^3}{2t^2 + 3} \cdot \frac{6t^5}{t^6} dt = \\ &= \int \frac{6t^2}{2t^2 + 3} dt = \int \left(3 - \frac{9}{2t^2 + 3}\right) dt = \\ &= 3t - 3\sqrt{\frac{3}{2}} \operatorname{arctg} \sqrt{\frac{2}{3}} t + c = 3x^{1/6} - 3\sqrt{\frac{3}{2}} \operatorname{arctg} \left(\sqrt{\frac{2}{3}} x^{1/6}\right) + c \end{aligned}$$

#### 5.6 Integrazione di funzioni discontinue

Se una funzione  $f : [a, b] \rightarrow \mathbb{R}$  è limitata in  $[a, b]$  e continua salvo un numero finito di punti di discontinuità a salto nei punti  $r_1, \dots, r_k$ , la funzione  $f$  è continua (o meglio prolungabile con continuità fino agli estremi) in ogni intervallino  $[r_{i-1}, r_i]$ ; pertanto è integrabile su ogni intervallino  $[r_{i-1}, r_i]$  e dunque (per il Teorema 6.3 visto nel par. 2.2) anche in tutto  $[a, b]$ . Inoltre, per la proprietà di additività dell'integrale, si può scrivere:

$$(5.9) \quad \int_a^b f(x) dx = \int_a^{r_1} f(x) dx + \int_{r_1}^{r_2} f(x) dx + \dots + \int_{r_k}^b f(x) dx$$

Questo è utile perché ciascuno degli integrali scritti a secondo membro è ora l'integrale di una funzione continua, di cui possiamo cercare la primitiva (o calcolare il valore numerico approssimato dell'integrale). In ogni caso, il problema viene ricondotto a un numero finito di problemi più semplici.

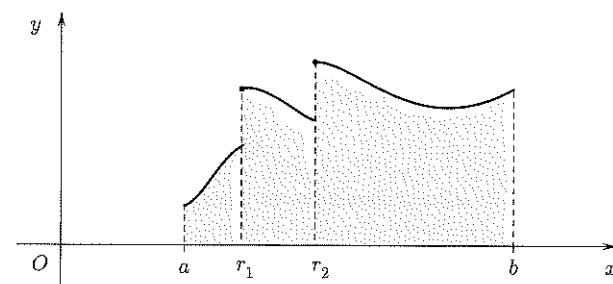


Figura 6.10. Integrazione di una funzione con due discontinuità a salto.

Si noti che, nel ragionamento precedente, ciò che conta è che su ciascun intervallino  $(r_k, r_{k+1})$  la funzione sia continua, e i limiti ai due estremi esistano finiti; non ha importanza, invece, il valore effettivo di  $f$  nei due estremi. In altre parole, alterare il valore della funzione integranda in un numero finito di punti è del tutto irrilevante ai fini del calcolo dell'integrale. (Si pensi al significato geometrico: se un rettangolo viene alterato togliendogli un segmento e sostituendolo con un altro di diversa lunghezza, la sua area non cambia).



Le seguenti funzioni sono integrabili (lo studente si renda conto in dettaglio del motivo):

$$\frac{\sin x}{x}, \text{ su } [-1, 1] \quad \frac{x}{|x|}, \text{ su } [-1, 2]$$

$$\operatorname{arctg} \frac{1}{x}, \text{ su } [-2, 3] \quad e^{-1/x}, \text{ su } [0, 1] \text{ (ma non su } [-1, 0]!)$$



$$\int \frac{x^3}{\sqrt{2-x^2}} dx$$

$$\int_0^1 \frac{x^2}{\sqrt{1+x^2}} dx$$

1)  $\int \frac{\sqrt{3+x^2}}{x^2} dx$

2)  $\int \frac{x^2}{\sqrt{x^2-1}} dx$

3)  $\int \frac{1-x}{\sqrt{1+x}} dx$

4)  $\int_0^1 (1-x^2)^{3/2} dx$

5)  $\int x \sqrt{4-x^2} dx$

6)  $\int_2^3 \frac{\sqrt{x^2-2}}{x^2} dx$

7)  $\int_1^2 \frac{\sqrt{x^2-2}}{x} dx$

8)  $\int \frac{\sqrt{x} + \sqrt[3]{x} + 1}{\sqrt[4]{x}} dx$

9)  $\int \frac{\sqrt{x} - \sqrt[3]{x}}{x(1+\sqrt[4]{x})} dx$

10)  $\int_0^{\frac{\pi}{2}} \sqrt{1+(\sin x)^2} \cos x dx$

I prossimi esercizi mostrano esempi di situazioni in cui, con opportune integrazioni per parti, ci si riconduce all'integrale di una funzione razionale o una funzione irrazionale dei tipi trattati in precedenza.

11)  $\int x \arcsin x dx$

12)  $\int_1^2 x(\log x)^2 dx$

13)  $\int_0^1 \operatorname{arctg} x dx$

14)  $\int x \operatorname{arctg}(x+1) dx$



a) Verificare che

$$S_n = \sum_{k=1}^n \frac{n}{n^2 + k^2} \quad (n \geq 1)$$

è una somma di Cauchy-Riemann per la funzione  $f(x) = \frac{1}{1+x^2}$ , relativa all'intervallo  $[0, 1]$ .

b) Calcolare

$$\lim_{n \rightarrow +\infty} S_n$$

Sia  $f(t) = a_0 + a_1 \cos \omega t + b_1 \sin \omega t$ .

Calcolare il valor medio e il valore efficace di  $f$ . Generalizzare a

$$f(t) = a_0 + \sum_{k=1}^n \{a_k \cos k \omega t + b_k \sin k \omega t\}$$

c) Calcolare i seguenti integrali indefiniti:

a)  $\int \frac{x}{x^2-4} dx$       b)  $\int e^{-x}(x^2+x) dx$       c)  $\int e^x \sin 2x dx$

d)  $\int \frac{1}{\sqrt{x}(1+x)} dx$       e)  $\int \sqrt{2x+1} \frac{1}{x} dx$

d) Calcolare i seguenti integrali definiti:

a)  $\int_0^\pi x \sin 2x dx$       b)  $\int_1^2 \frac{1}{x(\log x+1)} dx$

c)  $\int_1^2 \frac{\sqrt{x+1}}{x} dx$       d)  $\int_0^9 \sqrt{1+\sqrt{x}} dx$  (porre  $\sqrt{1+\sqrt{x}} = t$ )

e) Calcolare per ogni  $n$  intero positivo

$$a_n = \int_0^n \frac{x-1}{(x+1)^3} dx$$

Discutere il carattere della serie

$$\sum_{n=1}^{\infty} a_n$$

## ■ 6 ALCUNE APPLICAZIONI FISICHE E GEOMETRICHE

Oltre a quelle già incontrate nel par. 2.1, l'integrale ha numerose altre importanti applicazioni.

### Energia potenziale di una forza elastica di richiamo. Lavoro di una forza

Una particella di massa  $M$  è soggetta ad una forza elastica di richiamo, in direzione dell'asse  $x$ , cioè ad una forza direttamente proporzionale allo spostamento da un punto fisso determinato (posizione a riposo) con verso tale da ridurre lo spostamento. Il caso tipico è quello prodotto da una molla soggetta a spostamenti sufficientemente piccoli dalla posizione a riposo.

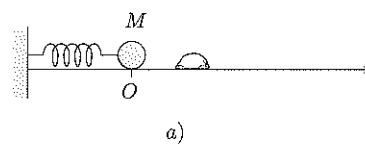
Se si scelgono le coordinate sull'asse in modo che la posizione di riposo per la particella corrisponda ad  $x = 0$  la forza avrà un'intensità data dalla formula

(6.1)

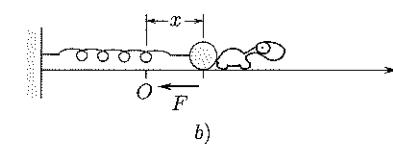
$$F = -kx$$

dove  $k > 0$  è una costante (detta costante elastica) che dipende dalla molla. La (6.1) si chiama *legge di Hooke*. Per calcolare l'energia potenziale associata alla forza  $F$ , calcoliamo il lavoro  $L$  che una forza applicata,  $F_{\text{app}}$ , deve compiere per spostare la massa  $M$ , inizialmente a riposo, in un punto di ascissa  $x_f$ . L'energia potenziale sarà data da  $E_{\text{pot}} = -L$ . In base alla (6.1), la forza applicata è funzione della posizione ed è in ogni punto uguale ed opposta a  $F$ :

$$F_{\text{app}} = -F = kx$$



a)



b)

Figura 6.11. a) Molla a riposo. b) Lo spostamento  $x$  provoca una forza  $F = -kx$ .

Il lavoro elementare  $dL$ , relativo a uno spostamento da  $x$  a  $x+dx$  ( $|dx|$  molto piccolo) nel quale si può considerare  $F_{\text{app}} = \text{costante} = kx$ , è dato da  $dL = kx dx$ . Integrando tra 0 e  $x_f$  si ottiene

$$L = \int_0^{x_f} kx dx = k \left[ \frac{1}{2} x^2 \right]_0^{x_f} = \frac{1}{2} kx_f^2$$

e infine  $E_{\text{pot}} = -\frac{1}{2} kx_f^2$ .

Più in generale, se una forza agisce lungo una retta, che sceglieremo come asse  $x$ , spostando un punto materiale da  $a$  a  $b$ , il lavoro della forza  $F(x)$  è dato per definizione dall'integrale:

$$L = \int_a^b F(x) dx$$

Naturalmente, da un punto di vista fisico, è molto restrittivo assumere che il moto si svolga lungo una retta. Questa è una delle motivazioni che porteranno a sviluppare ulteriormente la teoria dell'integrazione, definendo gli *integrali di linea* (si veda il secondo volume), mediante i quali si darà senso al concetto di lavoro lungo una curva qualsiasi dello spazio.

#### Lunghezza di un grafico

Sia  $y = f(x)$  una funzione continua con la sua derivata prima  $f'(x)$ , in un intervallo  $[a, b]$ . Vogliamo trovare una formula per la lunghezza del grafico di  $f$ . A tale scopo suddividiamo l'intervallo  $[a, b]$  in  $n$  parti uguali, mediante i "soliti" punti  $x_0 = a, x_1, x_2, \dots, x_n = b$ ; consideriamo ora la linea spezzata che ha per estremi i punti  $P_i \equiv (x_i, f(x_i))$ , che stanno sul grafico di  $f$ .

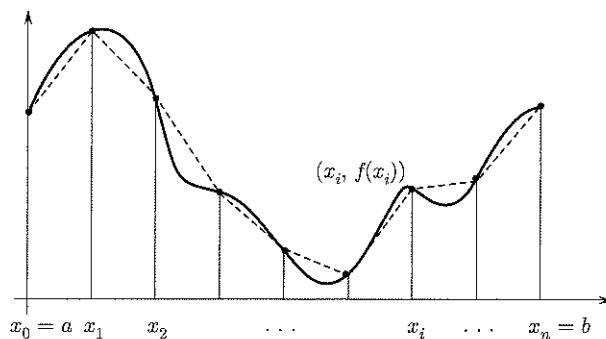


Figura 6.12.

In base alla nostra idea intuitiva di lunghezza di una curva, la lunghezza  $\ell_n$  di tale spezzata dovrebbe essere una *approssimazione per difetto* della lunghezza del grafico. Se esiste finito il limite di  $\ell_n$  per  $n \rightarrow \infty$ , sarà naturale assumere questo limite come definizione della lunghezza del grafico. Vediamo ora di mettere in pratica questa idea.

Anzitutto, è possibile calcolare  $\ell_n$  al modo seguente:

$$\ell_n = \sum_{i=1}^n |P_i - P_{i-1}| = \sum_{i=1}^n \sqrt{(x_i - x_{i-1})^2 + (f(x_i) - f(x_{i-1}))^2}.$$

D'altro canto, per il teorema di Lagrange applicato ad  $f$  su ciascun intervallino  $[x_{i-1}, x_i]$  si ha che esiste  $\xi_i \in [x_{i-1}, x_i]$  per cui risulta

$$f(x_i) - f(x_{i-1}) = f'(\xi_i)(x_i - x_{i-1})$$

e quindi

$$\begin{aligned} \ell_n &= \sum_{i=1}^n \sqrt{(x_i - x_{i-1})^2 + (f'(\xi_i)(x_i - x_{i-1}))^2} = \\ &= \sum_{i=1}^n (x_i - x_{i-1}) \sqrt{1 + f'(\xi_i)^2}. \end{aligned}$$

Si riconosce quindi che la lunghezza della  $n$ -esima poligonale che approssima il grafico non è altro che una somma di Cauchy-Riemann  $n$ -esima della funzione (continua, nelle nostre ipotesi)  $\sqrt{1 + f'(x)^2}$ . Perciò si ha:

$$\text{lunghezza del grafico} = \lim_{n \rightarrow \infty} \ell_n = \int_a^b \sqrt{1 + f'(x)^2} dx.$$

Ecco come il calcolo infinitesimale permette di risolvere un altro problema "geometrico" che con metodi elementari non è affrontabile. (Si confronti con quanto discusso nel par. 1).



Calcoliamo la lunghezza dell'arco di curva  $y = x^{3/2}$  per  $x \in [0, a]$ . Essendo  $y' = \frac{3}{2}\sqrt{x}$ , si ha:

$$L = \int_0^a \sqrt{1 + \frac{9}{4}x} dx = \frac{8}{27} [(1 + \frac{9}{4}x)^{3/2}]_0^a = \frac{8}{27} \left[ \left( 1 + \frac{9}{4}a \right)^{3/2} - 1 \right]$$



Calcoliamo la lunghezza dell'arco di parabola  $y = x^2$  per  $x \in [0, a]$ . Essendo  $y' = 2x$ , si ha:

$$\begin{aligned} L &= \int_0^a \sqrt{1 + 4x^2} dx = \left[ 2x = \text{Sht}; dx = \frac{1}{2} \text{Cht} dt; t \in (0, \text{SettSh2}a) \right] \\ &= \int_0^{\text{SettSh2}a} \frac{1}{2} (\text{Ch} t)^2 dt = (\text{ricordando la primitiva di } (\text{Ch} t)^2, \text{ calcolata nell'esempio 5.18}) \\ &= \frac{1}{2} \left[ \frac{1}{2} (\text{Sht} \text{Cht} + t) \right]_0^{\text{SettSh2}a} = \frac{1}{4} (2a\sqrt{1 + 4a^2} + \log(2a + \sqrt{1 + 4a^2})) \end{aligned}$$

Come si vede, anche in un caso elementare come quello della parabola, l'espressione trovata per la lunghezza dell'arco ha una certa complessità. Questo è un fatto abbastanza generale. Lo studente provi, a titolo d'esempio, a impostare il calcolo della lunghezza dell'arco di ellisse:

$$y = b \sqrt{1 - \frac{x^2}{a^2}} \quad x \in [0, a]$$

(cioè un quarto dell'ellisse). Troverà che, salvo il caso banale  $a = b$  (circonferenza!) non è possibile determinare una primitiva della funzione integranda con i metodi noti.

**OSSERVAZIONE (GRAFICI DI LUNGHEZZA INFINTA)** Notiamo ora che per ogni funzione  $f : [a, b] \rightarrow \mathbb{R}$ , continua, è possibile costruire la successione  $\ell_n$  delle lunghezze delle poligonalì di vertici  $(x_i, f(x_i))$ . Abbiamo inoltre dimostrato che, se  $f$  è derivabile con continuità, esiste certamente  $\lim_{n \rightarrow \infty} \ell_n$ , che è calcolabile mediante un certo integrale. Se  $f$  è continua ma non è derivabile con continuità, può accadere però che il  $\lim_{n \rightarrow \infty} \ell_n$  non esista finito: in questo caso la lunghezza della curva risulterà infinita.



La funzione

$$f(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & \text{per } x \in (0, 1] \\ 0 & \text{per } x = 0 \end{cases}$$

è continua in  $[0, 1]$  (ma la sua derivata non è continua in 0 e non è limitata in  $[0, 1]$ ). Si può dimostrare che il suo grafico ha lunghezza infinita:

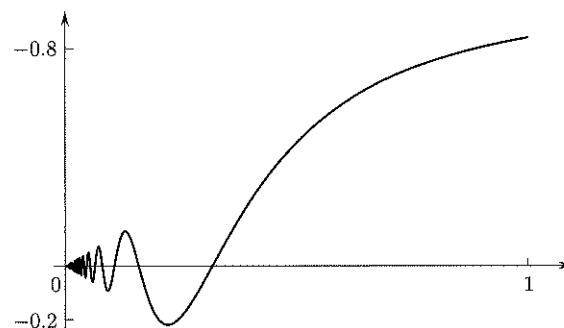


Figura 6.13.

#### Volume di un solido di rotazione attorno a un asse

Consideriamo un solido ottenuto per rotazione attorno all'asse  $x$  del trapezioide individuato dalla funzione continua  $y = f(x)$ ,  $a \leq x \leq b$  (cfr. fig. 6.14). Il volume si ottiene sommando i contributi  $dV$  dei volumi dei cilindretti aventi altezza  $dx$  e raggio

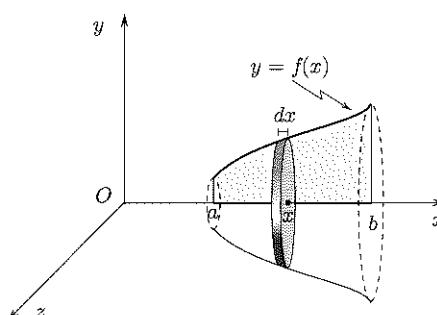


Figura 6.14.

di base  $f(x)$ . Essendo  $dV = \pi f(x)^2 dx$ , integrando tra  $a$  e  $b$  si trova la formula per il volume:

(6.2)

$$\text{vol} = \pi \int_a^b f(x)^2 dx$$



Calcoliamo il volume dell'ellissoide di rotazione, cioè del solido che si ottiene facendo ruotare attorno all'asse  $x$  la parte superiore dell'ellisse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

(L'equazione cartesiana della superficie dell'ellissoide trovato è

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

ovvero un ellissoide con due semiassi su tre uguali tra loro). In questo caso

$$f(x) = b\sqrt{1 - \frac{x^2}{a^2}}, \quad \text{con } x \in [-a, a]$$

e il volume del solido è

$$\begin{aligned} \text{vol} &= \pi \int_{-a}^a (b\sqrt{1 - \frac{x^2}{a^2}})^2 dx = 2\pi b^2 \int_0^a (1 - \frac{x^2}{a^2}) dx = \\ &= 2\pi b^2 [x - \frac{x^3}{3a^2}]_0^a = 2\pi b^2 \cdot \frac{2}{3}a = \frac{4}{3}\pi ab^2 \end{aligned}$$

formula che generalizza in modo naturale quella del volume della sfera,  $\frac{4}{3}\pi r^3$ , e che a sua volta suggerisce che il volume di un generico ellissoide (non di rotazione), di semiassi  $a, b, c$ , sia  $\text{vol} = \frac{4}{3}\pi abc$  (come dimostreremo nel secondo volume).



Calcolare il lavoro della forza  $F(x) = \frac{k}{x}$  nello spostare un punto da  $a$  a  $b$  ( $0 < a < b$ ).

Calcolare il volume del solido che si ottiene facendo ruotare intorno all'asse  $y$  l'arco di parabola  $y = x^2$ ,  $x \in [0, r]$ , e chiudendo la figura ottenuta col piano  $y = r^2$ . (Attenzione: poiché l'asse di rotazione è l'asse  $y$ , occorre vedere la curva come funzione  $x = f(y)$ ).

Calcolare il volume dei solidi che si ottengono dalla rotazione della curva  $\sqrt[3]{x}$  attorno ai due assi cartesiani, per  $0 \leq x \leq 1$ .

## 7 CALCOLO NUMERICO APPROXIMATO DI UN INTEGRALE

Come abbiamo già osservato, anche per funzioni continue abbastanza semplici può capitare di non essere in grado di scrivere esplicitamente una primitiva, in termini di funzioni elementari. Ad esempio, la semplice funzione  $e^{-x^2}$ , che come vedremo gioca un ruolo importante in molte questioni applicate, ha una primitiva non esprimibile mediante composizione di funzioni elementari. Quindi, per una funzione di questo tipo, il calcolo esatto dell'integrale definito su un certo intervallo, mediante il teorema fondamentale del calcolo integrale, risulta in pratica impossibile. In questi e in molti altri casi può essere utile avere comunque una valutazione approssimata del valore numerico di un integrale definito. Il modo più semplice di procedere è usare la definizione stessa:

sia data una funzione continua  $f : [a, b] \rightarrow \mathbb{R}$ , e costruiamo una successione  $s_n$  di Cauchy-Riemann per  $f$ . Poiché sappiamo che

$$s_n \rightarrow \int_a^b f(x) dx \text{ per } n \rightarrow \infty,$$

ci aspettiamo che il valore numerico di  $s_n$  per  $n$  "molto grande" sia una buona approssimazione numerica dell'integrale; d'altro canto  $s_n$  può essere calcolato facilmente mediante un computer. Il problema di questo procedimento è che non sappiamo, così facendo, *di quanto sbagliamo*. Questa prima idea, ingenua, diventa però un metodo efficace non appena abbiamo un risultato che ci permetta di stimare a priori l'errore commesso in questa approssimazione. Questo è il significato del prossimo teorema:

**TEOREMA 6.7 (CALCOLO NUMERICO DI UN INTEGRALE COL METODO DEL PUNTO MEDIO)** Sia  $f : [a, b] \rightarrow \mathbb{R}$  una funzione due volte derivabile con continuità, e sia  $K$  una costante tale che  $|f''(x)| \leq K$  per ogni  $x \in [a, b]$  (poiché  $f''$  è continua su  $[a, b]$ , è anche limitata, perciò una tale costante esiste certamente). Consideriamo ora, per un qualsiasi intero positivo  $n$ , la somma di Cauchy-Riemann  $s_n$  per  $f$ , costruita scegliendo come punto  $\xi_i \in [x_{i-1}, x_i]$  il punto medio di  $[x_{i-1}, x_i]$ , per ogni  $i = 1, 2, \dots, n$ :

$$s_n = \sum_{i=1}^n f\left(\frac{x_{i-1} + x_i}{2}\right) \frac{(b-a)}{n}.$$

Allora vale la stima:

$$\left| \int_a^b f(x) dx - s_n \right| \leq \frac{K}{24} \cdot \frac{(b-a)^3}{n^4}.$$

**DIMOSTRAZIONE.** Consideriamo un singolo intervallino  $[x_{i-1}, x_i]$ ; chiamiamo  $\xi_i$  il suo punto medio e  $h$  la sua ampiezza. Dunque risulta

$$[x_{i-1}, x_i] = \left[\xi_i - \frac{h}{2}, \xi_i + \frac{h}{2}\right]; \quad h = \frac{b-a}{n};$$

$$s_n = h \sum_{i=1}^n f(\xi_i);$$

Scriviamo ora lo sviluppo di Taylor al prim'ordine con resto secondo Lagrange, per  $f(x)$  con centro  $\xi_i$  e  $x \in [\xi_i - \frac{h}{2}, \xi_i + \frac{h}{2}]$ . Sarà:

$$f(x) = f(\xi_i) + f'(\xi_i)(x - \xi_i) + \frac{f''(c)}{2}(x - \xi_i)^2$$

per qualche  $c \in [\xi_i - \frac{h}{2}, \xi_i + \frac{h}{2}]$ . Poiché  $|f''(x)| \leq K$  per ogni  $x \in [a, b]$  possiamo scrivere allora:

$$f(\xi_i) + f'(\xi_i)(x - \xi_i) - \frac{K}{2}(x - \xi_i)^2 \leq f(x) \leq f(\xi_i) + f'(\xi_i)(x - \xi_i) + \frac{K}{2}(x - \xi_i)^2$$

per ogni  $x \in [\xi_i - \frac{h}{2}, \xi_i + \frac{h}{2}]$ . Integrando le diseguaglianze precedenti in  $[\xi_i - \frac{h}{2}, \xi_i + \frac{h}{2}]$  otteniamo:

$$\begin{aligned} \int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} f(x) dx &\leq \int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} \left[ f(\xi_i) + f'(\xi_i)(x - \xi_i) + \frac{K}{2}(x - \xi_i)^2 \right] dx = \\ &= f(\xi_i)h + f'(\xi_i) \int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} (x - \xi_i) dx + \frac{K}{2} \int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} (x - \xi_i)^2 dx. \end{aligned}$$

Notiamo ora che per simmetria

$$\int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} (x - \xi_i) dx = \int_{-h/2}^{h/2} x dx = 0$$

mentre

$$\int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} (x - \xi_i)^2 dx = \int_{-h/2}^{h/2} x^2 dx = 2 \left[ \frac{x^3}{3} \right]_0^{\frac{h}{2}} = \frac{h^3}{12}.$$

Quindi

$$\int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} f(x) dx \leq f(\xi_i)h + \frac{Kh^3}{24}.$$

Analogamente

$$\int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} f(x) dx \geq f(\xi_i)h - \frac{Kh^3}{24}.$$

quindi

$$\left| \int_{\xi_i - \frac{h}{2}}^{\xi_i + \frac{h}{2}} f(x) dx - f(\xi_i)h \right| \leq \frac{Kh^3}{24}.$$

Possiamo quindi stimare:

$$\begin{aligned} \left| \int_a^b f(x) dx - s_n \right| &= \left| \int_a^b f(x) dx - \sum_{i=1}^n f(\xi_i)h \right| = \\ &= \left| \sum_{i=1}^n \left( \int_{x_{i-1}}^{x_i} f(x) dx - f(\xi_i)h \right) \right| \leq \sum_{i=1}^n \left| \int_{x_{i-1}}^{x_i} f(x) dx - f(\xi_i)h \right| \leq \\ &\leq \sum_{i=1}^n \frac{Kh^3}{24} = \frac{K}{24} \cdot \frac{h^3}{n} = \frac{K}{24} \cdot \frac{(b-a)^3}{n^4}. \end{aligned}$$

## 8 INTEGRALI GENERALIZZATI

In molte situazioni concrete (calcolo delle probabilità e statistica ne sono esempi) si presentano integrali di funzioni illimitate e integrali estesi a intervalli illimitati. Ci occuperemo ora di queste situazioni, mostrando come la teoria precedente si possa estendere in maniera naturale per tenere conto anche di questi casi.

### 8.1 Integrazione di funzioni non limitate

Consideriamo il caso tipico illustrato in figura 6.15, in cui  $f : [a, b) \rightarrow \mathbb{R}$  è continua e

$$\lim_{x \rightarrow b^-} f(x) = +\infty$$

(Del tutto analogo è il caso:  $\lim_{x \rightarrow b^-} f(x) = -\infty$ ).

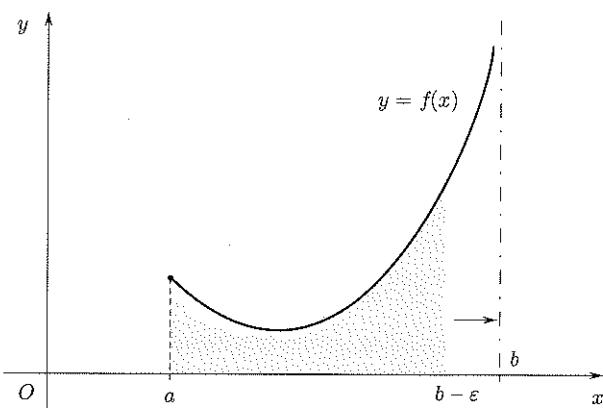


Figura 6.15.  $\lim_{x \rightarrow b^-} f(x) = +\infty$ .

Per definire l'integrale di  $f$  in  $[a, b]$ , l'idea è molto semplice: si integra tra  $a$  e  $b - \varepsilon$  ( $\varepsilon > 0$ ) e poi si passa al limite per  $\varepsilon \rightarrow 0^+$ . In simboli, si pone

$$(8.1) \quad \int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \int_a^{b-\varepsilon} f(x) dx$$

**DEFINIZIONE 6.3** Se il limite in (8.1) esiste finito allora  $f$  si dice *integrabile* in  $[a, b]$  oppure che l'integrale  $\int_a^b f(x) dx$  è *convergente*.

Se il limite in (8.1) è  $+\infty$  oppure  $-\infty$ , l'integrale si dirà *divergente*.

Se il limite non esiste allora l'integrale *non esiste*.

Analoghe definizioni si hanno se  $f : (a, b] \rightarrow \mathbb{R}$ , con  $f$  continua e

$$\lim_{x \rightarrow a^+} f(x) = \pm\infty$$

Si pone:

$$(8.2) \quad \int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dx$$

### ESEMPIO DI APPLICAZIONE

#### Calcolo dell'integrale

$$\int_a^b \frac{dx}{(b-x)^\alpha} \quad (\alpha > 0)$$

Caso  $\alpha = 1$ . Si ha:

$$\int_a^{b-\varepsilon} \frac{dx}{(b-x)} = [-\log(b-x)]_a^{b-\varepsilon} = -\log \varepsilon + \log(b-a)$$

Dunque

$$\int_a^b \frac{dx}{(b-x)} = \lim_{\varepsilon \rightarrow 0^+} [-\log \varepsilon + \log(b-a)] = +\infty$$

Quindi l'integrale è *divergente*.

Caso  $\alpha \neq 1$ . Si ha:

$$\int_a^{b-\varepsilon} \frac{dx}{(b-x)^\alpha} = \frac{1}{1-\alpha} [-(b-x)^{1-\alpha}]_a^{b-\varepsilon} = \frac{1}{1-\alpha} [-\varepsilon^{1-\alpha} + (b-a)^{1-\alpha}]$$

Dunque

$$\int_a^b \frac{dx}{(b-x)^\alpha} = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{1-\alpha} [-\varepsilon^{1-\alpha} + (b-a)^{1-\alpha}] = \begin{cases} +\infty & \text{se } \alpha > 1 \\ \frac{(b-a)^{1-\alpha}}{1-\alpha} & \text{se } \alpha < 1 \end{cases}$$

Riassumendo abbiamo:

$$(8.3) \quad \int_a^b \frac{dx}{(b-x)^\alpha} \quad \text{è} \quad \begin{cases} \text{divergente a } +\infty & \text{se } \alpha \geq 1 \\ \text{convergente} = \frac{(b-a)^{1-\alpha}}{1-\alpha} & \text{se } \alpha < 1 \end{cases}$$

Un risultato perfettamente analogo vale per l'integrale

$$\int_a^b \frac{dx}{(x-a)^\alpha}$$

### 8.2 Criteri di integrabilità al finito

Siano  $f, g : [a, b] \rightarrow \mathbb{R}$ , continue, con

$$\lim_{x \rightarrow b^-} f(x) = \lim_{x \rightarrow b^-} g(x) = +\infty$$

I seguenti criteri permettono di decidere se un integrale è convergente o divergente, senza calcolarlo:

**Confronto**

Se  $0 \leq f(x) \leq g(x)$  in  $[a, b]$ , allora

$$\begin{array}{ll} g \text{ integrabile} & \Rightarrow f \text{ integrabile} \\ f \text{ non integrabile} & \Rightarrow g \text{ non integrabile} \end{array}$$

Infatti, per la proprietà di monotonia dell'integrale, si ha:

$$0 \leq \int_a^{b-\varepsilon} f(x) dx \leq \int_a^{b-\varepsilon} g(x) dx$$

e, passando al limite per  $\varepsilon \rightarrow 0^+$ , si prova la tesi.

**Confronto asintotico**

Se  $f > 0$ ,  $g > 0$  e  $f \sim g$  per  $x \rightarrow b^-$  allora

$$f \text{ integrabile} \iff g \text{ integrabile}$$

Analoghi criteri valgono se  $f, g \rightarrow +\infty$  per  $x \rightarrow a^+$ , o se  $f, g \rightarrow -\infty$ .

In quest'ultimo caso, le disuguaglianze del criterio del confronto devono valere tra i moduli di  $f$  e  $g$ .



Consideriamo gli integrali

$$I_1 = \int_0^1 \frac{dx}{\sqrt[3]{1-x^2}}, \quad I_2 = \int_1^3 \frac{dx}{x^2-5x+4}$$

$I_1$ : La funzione  $f(x) = \frac{1}{\sqrt[3]{(1-x^2)}} = \frac{1}{(1-x)^{1/3}(1+x)^{1/3}}$  è continua e positiva in  $[0, 1]$  e tende a  $+\infty$  per  $x \rightarrow 1^-$ . Inoltre  $f(x) \sim \frac{1}{\sqrt[3]{2}} \frac{1}{(1-x)^{1/3}}$  per  $x \rightarrow 1^-$ . D'altra parte la funzione  $g(x) = \frac{1}{\sqrt[3]{2(1-x)^{1/3}}}$  è positiva e integrabile (dalla (8.3) con  $\alpha = 1/3$ ) e quindi anche  $f$  risulta integrabile, in base al confronto asintotico.

L'integrale  $I_1$  è perciò convergente.

$I_2$ : La funzione  $f(x) = \frac{1}{x^2-5x+4} = \frac{1}{(x-1)(x-4)}$  è continua e negativa in  $[1, 3]$  e tende a  $-\infty$  per  $x \rightarrow 1^+$ . Il segno non costituisce un problema in quanto possiamo riferirci a  $-f(x)$  che è positiva.

Osserviamo che  $-f(x) \sim \frac{1}{3(x-1)}$  e che  $g(x) = \frac{1}{3(x-1)}$  è (positiva) e non integrabile (sempre dalla (8.3) con  $\alpha = 1$ ). Ne segue che anche  $-f$  ed  $f$  non sono integrabili.

L'integrale  $I_2$  è perciò divergente a  $-\infty$ .

Analogamente a quanto accade con i criteri di convergenza per le serie numeriche (v. cap. 5), quando la convergenza di un integrale viene stabilita mediante confronto o confronto asintotico tra due funzioni, il valore numerico dei due integrali sarà in generale diverso.

Una funzione  $f(x)$  potrebbe essere illimitata per  $x \rightarrow a^+$  senza avere segno definitivamente costante. Ad esempio,

$$f(x) = \frac{1}{\sqrt{x}} \sin \frac{1}{x}$$

è illimitata per  $x \rightarrow 0^+$ , ma non tende a  $\pm\infty$ , e il suo segno è variabile in ogni intorno destro di 0. Per funzioni di questo tipo i criteri del confronto e del confronto asintotico non sono applicabili. Vale invece l'implicazione seguente:

**TEOREMA 6.8**

$$\int_a^b |f(x)| dx \text{ convergente} \implies \int_a^b f(x) dx \text{ convergente}$$

Se  $|f|$  è integrabile in  $[a, +\infty)$  si dice che  $f$  è *assolutamente integrabile* in  $[a, +\infty)$ . Perciò il teorema afferma che una funzione assolutamente integrabile è anche integrabile.

**DIMOSTRAZIONE.** La dimostrazione è simile a quella dell'analogia proprietà delle serie numeriche (si veda il Teorema 5.3 cap. 5). Per definizione,

$$\int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0} \int_{a+\varepsilon}^b f(x) dx.$$

Separiamo ora la funzione  $f$  nella sua *parte positiva e negativa*:

$$f(x) = f^+(x) - f^-(x),$$

dove, per definizione:

$$f^+(x) = \begin{cases} f(x) & \text{se } f(x) \geq 0 \\ 0 & \text{se } f(x) < 0 \end{cases} \quad f^-(x) = \begin{cases} -f(x) & \text{se } f(x) \leq 0 \\ 0 & \text{se } f(x) > 0 \end{cases}$$

Le funzioni  $f^+, f^-$  sono  $\geq 0$  e soddisfano le disuguaglianze:

$$\begin{aligned} \int_{a+\varepsilon}^b f^+(x) dx &\leq \int_{a+\varepsilon}^b |f(x)| dx; \\ \int_{a+\varepsilon}^b f^-(x) dx &\leq \int_{a+\varepsilon}^b |f(x)| dx. \end{aligned}$$

Poiché per ipotesi esiste finito  $\lim_{\varepsilon \rightarrow 0} \int_{a+\varepsilon}^b |f(x)| dx$  (in quanto  $\int_a^b |f(x)| dx$  è convergente), ne segue che le funzioni

$$\varepsilon \mapsto \int_{a+\varepsilon}^b f^+(x) dx; \quad \varepsilon \mapsto \int_{a+\varepsilon}^b f^-(x) dx$$

sono limitate. Inoltre, queste funzioni sono anche monotone decrescenti, perché

$$\varepsilon_1 < \varepsilon_2 \implies (a + \varepsilon_1, b) \supset (a + \varepsilon_2, b) \implies \int_{a+\varepsilon_1}^b f^+(x) dx \geq \int_{a+\varepsilon_2}^b f^+(x) dx$$

perché  $f^+ \geq 0$  (e lo stesso vale per  $f^-$ ). Perciò, ancora per il Teorema 3.31, esistono finiti

$$\lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f^+(x) dx, \quad \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f^-(x) dx$$

e quindi anche

$$\lim_{\varepsilon \rightarrow 0} \int_{a+\varepsilon}^b f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f^+(x) dx - \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f^-(x) dx.$$

Di conseguenza  $f$  è integrabile.  $\blacksquare$



### L'integrale

$$\int_0^1 \frac{1}{\sqrt{x}} \sin \frac{1}{x} dx$$

converge, perché

$$\left| \frac{1}{\sqrt{x}} \sin \frac{1}{x} \right| \leq \frac{1}{\sqrt{x}}$$

che ha integrale convergente.

### 8.3 Integrazione su intervalli illimitati

Sia  $f : [a, +\infty) \rightarrow \mathbb{R}$ , continua. Poniamo

$$(8.4) \quad \int_a^{+\infty} f(x) dx = \lim_{\omega \rightarrow +\infty} \int_a^\omega f(x) dx$$

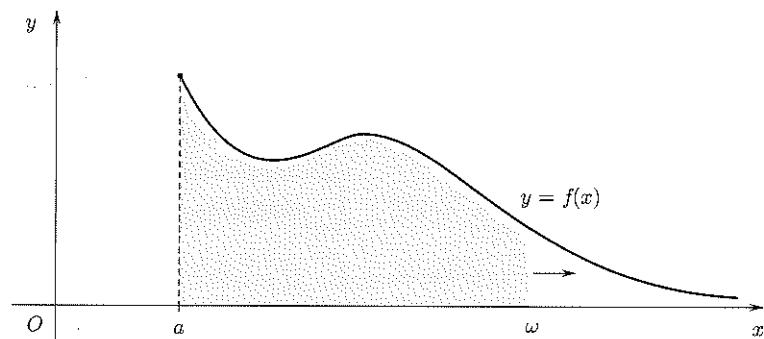


Figura 6.16. L'integrale su  $[a, +\infty)$  è definito come  $\lim_{\omega \rightarrow +\infty} \int_a^\omega f(x) dx$ .

**DEFINIZIONE 6.4** Se il limite in (8.4) esiste finito allora  $f$  si dice *integrabile* in  $[a, +\infty)$  oppure che l'integrale  $\int_a^{+\infty} f(x) dx$  è *convergente*. Se il limite in (8.4) è  $+\infty$  oppure  $-\infty$ , l'integrale si dirà *divergente*. Se infine il limite non esiste allora l'integrale *non esiste*.  $\blacksquare$

Analogamente se  $f : (-\infty, b] \rightarrow \mathbb{R}$  è continua, si pone

$$\int_{-\infty}^b f(x) dx = \lim_{\omega \rightarrow -\infty} \int_\omega^b f(x) dx$$

ed infine, se  $f : (-\infty, +\infty) \rightarrow \mathbb{R}$  è continua, si pone

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^c f(x) dx + \int_c^{+\infty} f(x) dx$$

dove  $c$  è un punto qualunque.

Nel seguito, per semplicità, ci riferiremo all'intervallo  $[a, +\infty)$ .



### Calcolo dell'integrale

$$\int_1^{+\infty} \frac{1}{x^\alpha} dx \quad \alpha > 0$$

Caso  $\alpha = 1$ . Si ha

$$\int_1^\omega \frac{1}{x} dx = [\log x]_1^\omega = \log \omega$$

Poiché  $\lim_{\omega \rightarrow +\infty} \log \omega = +\infty$ , l'integrale è divergente.

Caso  $\alpha \neq 1$ . Si ha:

$$\int_1^\omega \frac{1}{x^\alpha} dx = \frac{1}{1-\alpha} [x^{1-\alpha}]_1^\omega = \frac{1}{1-\alpha} (\omega^{1-\alpha} - 1)$$

Dunque

$$\int_1^{+\infty} \frac{1}{x^\alpha} dx = \lim_{\omega \rightarrow +\infty} \frac{1}{1-\alpha} (\omega^{1-\alpha} - 1) = \begin{cases} +\infty & \text{se } \alpha < 1 \\ \frac{1}{\alpha-1} & \text{se } \alpha > 1 \end{cases}$$

Riassumendo,

$$(8.5) \quad \int_1^{+\infty} \frac{1}{x^\alpha} dx \quad \begin{cases} \text{divergente a } +\infty & \text{se } \alpha \leq 1 \\ \text{convergente} & = \frac{1}{\alpha-1} \quad \text{se } \alpha > 1 \end{cases}$$

### Divergenza della serie armonica

Nel capitolo 5, par. 1.1 abbiamo affermato che la serie armonica

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

è divergente. L'uso degli integrali generalizzati permette una rapida giustificazione dell'affermazione.

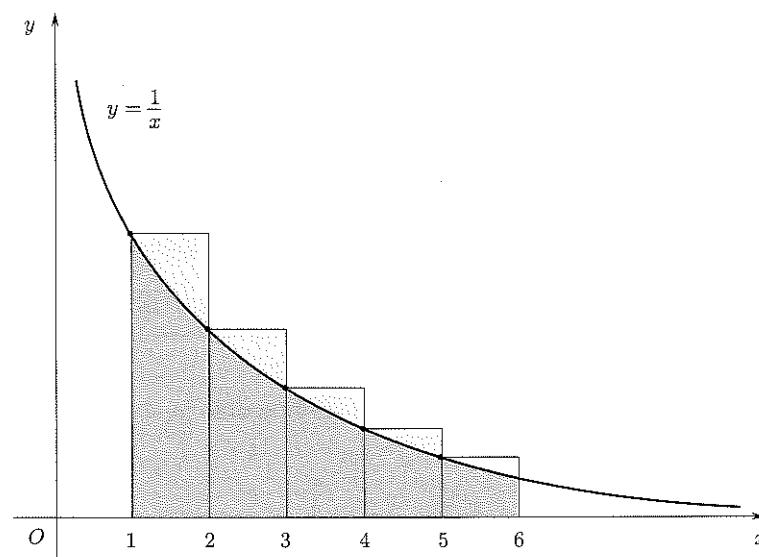


Figura 6.17. L'area ombreggiata è uguale a  $\int_1^6 \frac{1}{x} dx$  mentre la somma delle aree dei rettangoli è uguale a  $\sum_{n=1}^5 \frac{1}{n}$ .

Dalla figura 6.17 è facile convincersi che fissato  $N$ , intero, vale sempre la diseguaglianza

$$(8.6) \quad \int_1^N \frac{1}{x} dx < \sum_{n=1}^{N-1} \frac{1}{n}$$

Poiché  $\int_1^N \frac{1}{x} dx = \log N \rightarrow +\infty$  se  $N \rightarrow +\infty$ , anche  $\sum_{n=1}^N \frac{1}{n} \rightarrow +\infty$  se  $N \rightarrow +\infty$ , che dimostra la divergenza della serie armonica.

#### Convergenza della serie armonica generalizzata per $\alpha > 1$

Abbiamo visto (cap. 5, par. 1.2) che la serie

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$$

converge per  $\alpha > 1$  (e diverge per  $\alpha \leq 1$ ). L'affermazione è stata dimostrata per  $\alpha \geq 2$  (confronto con la serie di Mengoli). Siamo ora in grado di dimostrarla per qualunque  $\alpha > 1$ . Il ragionamento è analogo a quello sulla serie armonica, con le diseguaglianze in senso inverso. Dalla figura 6.18 si vede che per ogni intero  $N$  vale la diseguaglianza

$$\sum_{n=2}^N \frac{1}{n^\alpha} \leq \int_1^N \frac{dx}{x^\alpha}$$

Poiché per  $\alpha > 1$  e  $N \rightarrow \infty$  l'integrale converge (per quanto già visto), anche la serie converge (la successione delle somme parziali è crescente e superiormente limitata).

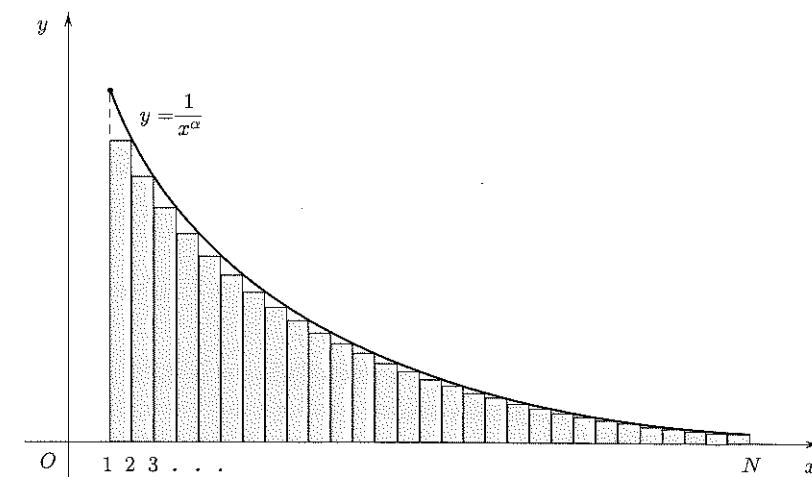


Figura 6.18.

#### 8.4 Criteri di integrabilità all'infinito

Siano  $f, g : [a, +\infty) \rightarrow \mathbb{R}$ , continue.

Per decidere se un integrale è convergente o meno, valgono criteri analoghi a quelli per l'integrale di funzioni illimitate.

##### Confronto

Se  $0 \leq f(x) \leq g(x)$  in  $[a, +\infty)$  allora

$g$ integrabile	$\Rightarrow$	$f$ integrabile
$f$ non integrabile	$\Rightarrow$	$g$ non integrabile

##### Confronto asintotico

Se  $f > 0$ ,  $g > 0$  e  $f \sim g$  per  $x \rightarrow +\infty$ , allora

$f$ integrabile	$\Leftrightarrow$	$g$ integrabile
-----------------	-------------------	-----------------



L'integrale

$$\int_0^{+\infty} e^{-x^2} dx$$

è convergente.

Infatti, si può scrivere

$$\int_0^{+\infty} e^{-x^2} dx = \int_0^1 e^{-x^2} dx + \int_1^{+\infty} e^{-x^2} dx$$

Osserviamo ora che per  $x > 1$  si ha  $x^2 > x$  e quindi  $e^{-x^2} < e^{-x}$ .

D'altra parte

$$\int_1^{+\infty} e^{-x} dx = \lim_{\omega \rightarrow +\infty} \int_1^{\omega} e^{-x} dx = \lim_{\omega \rightarrow +\infty} \left[ -e^{-\omega} + \frac{1}{e} \right] = \frac{1}{e}$$

Per confronto si deduce che anche  $\int_1^{+\infty} e^{-x^2} dx$  è convergente.

$I_1 = \int_0^{+\infty} \frac{x^2 + 1}{x^4 + x + 1} dx, \quad I_2 = \int_0^{+\infty} \frac{1}{\sqrt{x^2 + 1}} dx$

$I_1$ :  $f(x) \sim \frac{1}{x^2}$  per  $x \rightarrow +\infty$  e quindi è integrabile ((8.5), con  $\alpha = 2$ ). L'integrale  $I_1$  è pertanto convergente.

$I_2$ :  $f(x) \sim \frac{1}{x}$  per  $x \rightarrow +\infty$  e quindi non è integrabile ((8.5) con  $\alpha = 1$ ).

L'integrale  $I_2$  è pertanto divergente a  $+\infty$ .

Per funzioni di segno qualunque si ha ancora il:

### TEOREMA 6.9

$$\int_a^{+\infty} |f(x)| dx \text{ convergente} \implies \int_a^{+\infty} f(x) dx \text{ convergente}$$

In altre parole, se  $f$  è assolutamente integrabile in  $[a, +\infty)$ , è anche integrabile. La dimostrazione è molto simile a quella fatta nel caso degli integrali generalizzati su intervalli limitati.



Stabilire quali dei seguenti integrali esistono (eventualmente in senso generalizzato). In caso affermativo, calcolarlo quando è possibile.

$\int_0^1 x \log x dx$

$\int_0^{+\infty} \frac{1}{\sqrt{x}} e^{-\sqrt{x}} dx$

$\int_0^1 \frac{\sin x}{x^2} dx$

$\int_0^{+\infty} e^{-2x} \sin(e^{-x}) dx$

$\int_2^4 \operatorname{arctg} \left( \frac{x}{x-3} \right) dx$

$\int_2^{+\infty} \frac{1}{x \log x} dx$

$\int_0^1 \frac{1-e^{-x}}{x^{3/2}} dx$

$\int_2^{+\infty} \frac{1}{x(\log x)^2} dx$

$\int_0^{1/2} \frac{1}{x \log x} dx$

$\int_0^{+\infty} \frac{x}{(1+x^2)^2} dx$

$\int_0^{1/2} \frac{1}{x(\log x)^2} dx$

$\int_0^{+\infty} \sin(x^2) dx$

L'ultimo integrale è detto integrale di Fresnel, ed interviene in ottica.

*Suggerimento:* su  $[0, 1]$  la funzione  $\sin(x^2)$  è integrabile. Per studiare l'integrale su  $[1, +\infty)$ , valutare l'integrale su  $[1, a]$  eseguendo prima la sostituzione  $x^2 = t$  e poi un'integrazione per parti. Cosa succede per  $a \rightarrow +\infty$ ?

## 9 FUNZIONI INTEGRALI

Sia  $f$  una funzione integrabile (eventualmente in senso generalizzato) in un intervallo  $I$ , limitato o illimitato. Consideriamo l'integrale di  $f$  su un intervallo  $J$  contenuto in  $I$ . Ovviamente, al variare dell'intervallo  $J$  varia il valore di tale integrale. Se  $J$  è l'intervallo che ha un estremo  $x_0$  fissato una volta per tutte, e l'altro estremo  $x$  variabile, l'integrale di  $f$  su tale intervallo  $[x_0, x]$  diventa una funzione di  $x$ . Tale funzione si dice *funzione integrale* di  $f$ , e si indica con:

$$F(x) = \int_{x_0}^x f(t) dt$$

Si noti che la variabile d'integrazione  $t$  (variabile muta), ha un nome diverso dalla variabile  $x$ , estremo mobile dell'intervallo di integrazione: infatti  $t$  varia tra  $x_0$  e  $x$ .

Non sarà inutile a questo punto riepilogare i diversi concetti che sono stati introdotti a proposito dell'integrale:

1. L'integrale (definito) di  $f$  su  $[a, b]$ ,

$$\int_a^b f(x) dx$$

è un numero, definito come opportuno limite di somme;

2. La funzione integrale di  $f$ ,

$$F(x) = \int_{x_0}^x f(t) dt$$

è, appunto, una funzione: ad ogni  $x$  associa il numero ottenuto calcolando un integrale definito;

3. L'integrale indefinito di  $f$ ,

$$\int f(x) dx$$

è invece un simbolo che denota un insieme di funzioni, precisamente l'insieme di tutte le primitive di  $f$ , ossia delle funzioni  $G$  tali che  $G'(x) = f(x)$  per ogni  $x$ .

Come si vede, sono tre concetti ben diversi, che non devono essere confusi.

Le funzioni integrali, di cui ci occupiamo, compaiono in modo naturale in varie applicazioni:

1. *Meccanica.* Lo spazio percorso fino all'istante  $t$  da una particella che si muove di velocità variabile  $v(t)$  lungo una traiettoria assegnata, è la funzione integrale della velocità:

$$s(t) = \int_0^t v(\tau) d\tau$$

2. *Probabilità.* La "funzione degli errori"

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

descrive qual è la probabilità che una certa "variabile casuale" assuma un valore  $\leq x$ ; la variabile in questione descrive, ad esempio, sotto certe ipotesi, l'errore

commesso nella misurazione di una grandezza fisica. Si noti che per ogni  $x$  quello che si calcola in questo caso è un integrale generalizzato, esteso all'intervallo illimitato  $(-\infty, x)$ .

Il grafico della funzione integranda e della funzione integrale, rispettivamente, sono riportati nella figura 6.19.

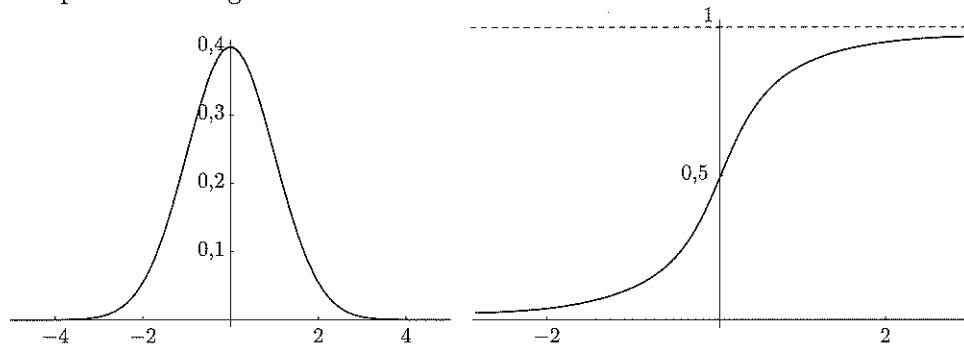


Figura 6.19.

3. *Ottica.* Nello studio di certi fenomeni legati all'interferenza luminosa, compaiono le *funzioni integrali di Fresnel*:

$$C(x) = c_1 \int_0^x \cos(t^2) dt \quad S(x) = c_2 \int_0^x \sin(t^2) dt$$

dove  $c_1, c_2$  sono opportune costanti di normalizzazione.

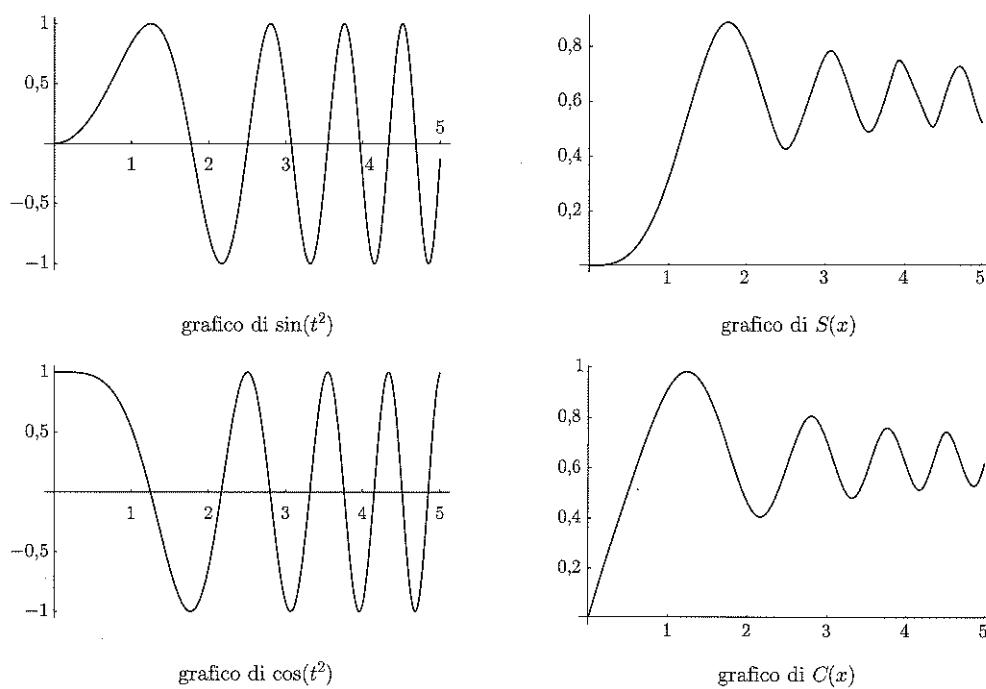


Figura 6.20.

Ma le funzioni integrali hanno anche una fondamentale importanza matematica, dovuta al:

**TEOREMA 6.10 (SECONDO TEOREMA FONDAMENTALE DEL CALCOLO INTEGRALE)** *Sia  $f : [a, b] \rightarrow \mathbb{R}$  una funzione integrabile (in senso proprio o generalizzato), sia  $x_0 \in [a, b]$  e sia*

$$F(x) = \int_{x_0}^x f(t) dt.$$

Allora:

1. La funzione  $F$  è continua in  $[a, b]$ .
2. Se inoltre  $f$  è continua in  $[a, b]$ , allora  $F$  è derivabile in  $[a, b]$ , e vale

$$F'(x) = f(x) \text{ per ogni } x \in [a, b].$$

**DIMOSTRAZIONE.** Proviamo il punto 1 prima nell'ipotesi in cui  $f$  è integrabile in senso proprio, e quindi, in particolare, è limitata. Perciò esiste  $k > 0$  per cui è

$$|f(t)| \leq k \forall t \in [a, b].$$

Per provare che  $F$  è continua in un generico punto  $\bar{x} \in [a, b]$ , consideriamo ora:

$$\begin{aligned} |F(\bar{x} + h) - F(\bar{x})| &= \left| \int_{x_0}^{\bar{x}+h} f(t) dt - \int_{x_0}^{\bar{x}} f(t) dt \right| = \\ &= \left| \int_{\bar{x}}^{\bar{x}+h} f(t) dt \right| \leq \left| \int_{\bar{x}}^{\bar{x}+h} |f(t)| dt \right| \leq \\ &\leq \left| \int_{\bar{x}}^{\bar{x}+h} k dt \right| = k|h|. \end{aligned}$$

(Abbiamo usato l'additività dell'integrale rispetto all'intervallo di integrazione, e la disuguaglianza del valore assoluto). Perciò per  $h \rightarrow 0$  si ha  $F(\bar{x} + h) \rightarrow F(\bar{x})$ , e  $F$  è continua in  $\bar{x}$ .

Supponiamo ora che  $f$  sia integrabile solo in senso generalizzato, e che  $\bar{x}$  sia proprio un punto nell'intorno del quale  $f$  è illimitata (altrimenti la precedente dimostrazione vale anche in questo caso). Abbiamo ancora:

$$F(\bar{x} + h) - F(\bar{x}) = \int_{\bar{x}}^{\bar{x}+h} f(t) dt.$$

Scriviamo ora:

$$\int_{\bar{x}}^b f(t) dt = \int_{\bar{x}}^{\bar{x}+h} f(t) dt + \int_{\bar{x}+h}^b f(t) dt$$

Poiché per ipotesi  $f$  è integrabile in senso generalizzato, si ha, per definizione:

$$\int_{\bar{x}+h}^b f(t) dt \rightarrow \int_{\bar{x}}^b f(t) dt \text{ per } h \rightarrow 0^+$$

e quindi per differenza

$$\int_{\bar{x}}^{\bar{x}+h} f(t) dt \rightarrow 0 \text{ per } h \rightarrow 0^+$$

ossia  $F(\bar{x} + h) \rightarrow F(\bar{x})$  per  $h \rightarrow 0^+$ . Analogamente si ragiona per  $h \rightarrow 0^-$ .

Proviamo ora il punto 2, partendo ancora dall'uguaglianza, già ottenuta,

$$F(\bar{x} + h) - F(\bar{x}) = \int_{\bar{x}}^{\bar{x}+h} f(t) dt.$$

Poiché ora per ipotesi  $f$  è continua, per il Teorema della media (Teorema 6.5 par. 3) si ha

$$\int_{\bar{x}}^{\bar{x}+h} f(t) dt = hf(x_h)$$

per qualche  $x_h \in [\bar{x}, \bar{x} + h]$  (supponendo, per fissare le idee,  $h > 0$ ). Ne segue

$$\frac{F(\bar{x} + h) - F(\bar{x})}{h} = f(x_h).$$

Per  $h \rightarrow 0$  si avrà  $x_h \rightarrow \bar{x}$  e quindi, ancora per la continuità di  $f$ ,  $f(x_h) \rightarrow f(\bar{x})$ . Pertanto esiste

$$\lim_{h \rightarrow 0} \frac{F(\bar{x} + h) - F(\bar{x})}{h} = f(\bar{x})$$

e il teorema è dimostrato.  $\diamond$

Il teorema ha varie conseguenze:

1. Anzitutto, se  $f$  è continua, allora  $F$  è derivabile con continuità, perché  $F' = f$  e  $f$  è continua.
2. Se  $f$  non è solo continua, ma è derivabile con continuità, allora anche  $F'$  è derivabile con continuità (perché è  $= f$ ), perciò  $F$  è due volte derivabile con continuità. Iterando: la funzione integrale ha sempre un grado di regolarità in più rispetto alla funzione integranda.
3. Dal teorema segue in particolare che ogni funzione continua su  $I$  ha ivi una primitiva: la sua funzione integrale. (Questa affermazione era stata fatta nel par. 4, ma non dimostrata). Possiamo scrivere quindi:

$$\int f(x) dx = \int_{x_0}^x f(t) dt + c$$

Sia

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

(Si noti che non si sa scrivere esplicitamente una primitiva di  $e^{-\frac{t^2}{2}}$ , e quindi la funzione integrale non si può riscrivere in forma più semplice). Allora:

$$F'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Poiché  $F'(x) > 0$  per ogni  $x$ ,  $F(x)$  è sempre crescente. Inoltre

$$F''(x) = -\frac{2x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \geq 0 \text{ per } x \leq 0$$

perciò  $F(x)$  ha un punto di flesso per  $x = 0$ , ed è concava verso l'alto per  $x < 0$ . Per  $x \rightarrow +\infty$ ,  $F(x)$  tende all'integrale generalizzato  $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ , che è finito. Perciò  $F(x)$  ha un asintoto orizzontale per  $x \rightarrow +\infty$  (si può dimostrare che questo asintoto è  $y = 1$ ). Infine, per come è definita, per  $x \rightarrow -\infty$   $F(x) \rightarrow 0$ . Tutti questi fatti si possono controllare sul grafico di  $F(x)$  riportato in precedenza (v. fig. 6.19).

Il punto 2 del teorema si può anche raffinare, al modo seguente. Se la funzione integranda  $f(t)$  non è continua su tutto  $I$ , ma è integrabile in senso generalizzato, in tutti i punti in cui l'integranda è continua,  $F(x)$  è derivabile, e  $F'(x) = f(x)$ . In sostanza, dove  $f$  è discontinua,  $F$  è ancora continua, ma non sarà derivabile.



Sia  $f(t) = \operatorname{sgn}(t)$ , e

$$F(x) = \int_0^x f(t) dt$$

Allora:

$$F(x) = \begin{cases} \text{se } x > 0 & \int_0^x 1 dt = x \\ \text{se } x < 0 & \int_0^x -1 dt = -x \end{cases}$$

dunque  $F(x) = |x|$ . Come si vede in figura 6.21,  $F$  è continua ovunque, ma presenta un punto angoloso (quindi di non derivabilità) esattamente dove la funzione integranda è discontinua.

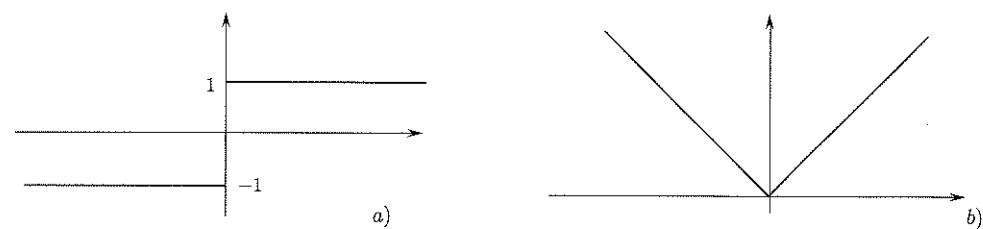


Figura 6.21. a) Grafico di  $f(t)$ ; b) grafico di  $F(x)$ .

Sia  $f(t) = \frac{1}{\sqrt[3]{t}}$ , e

$$F(x) = \int_0^x f(t) dt = \left[ \frac{3}{2} t^{2/3} \right]_0^x = \frac{3}{2} x^{2/3}$$

Come si vede in figura 6.22,  $F$  è continua ovunque (perché  $f$ , anche se illimitata, è comunque integrabile in senso generalizzato), ma presenta un punto di cuspide (quindi di non derivabilità) esattamente dove la funzione integranda è discontinua.

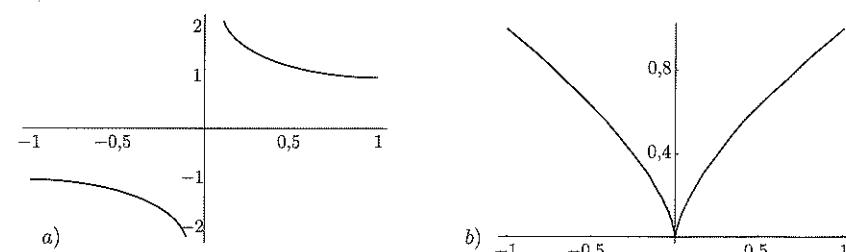


Figura 6.22. a) Grafico di  $f(t)$ ; b) grafico di  $F(x)$ .



Si considerino le funzioni di Fresnel  $C(x)$ ,  $S(x)$  definite negli esempi precedenti. Calcolare le loro derivate prima e seconda, e determinare i loro punti di massimo, minimo e flesso.

**Esercizio 6.10** Sia  $f(t) = |t|$ ,  $F(x) = \int_0^x f(t) dt$ . Calcolare esplicitamente  $F(x)$  e disegnarne il grafico. Quante volte la funzione  $F$  è derivabile su tutto  $\mathbb{R}$ ?

**Esercizio 6.11** Sia  $f(t) = e^{-t}$  per  $t > 0$ ,  $f(t) = 0$  per  $t \leq 0$ , e sia  $F(x) = \int_0^x f(t) dt$ . Calcolare esplicitamente  $F(x)$  e tracciarne il grafico.  $F$  è continua? È derivabile in tutto  $\mathbb{R}$ ?

**Esercizio 6.12** Sia  $f$  la funzione periodica di periodo 2 (cioè  $f(x+2) = f(x)$ , per ogni  $x$ ) che assume il valore 1 per  $0 \leq x < 1$  ed assume il valore zero per  $1 \leq x < 2$ .

- Disegnare il grafico di  $f$ .
- Calcolare  $F(x) = \int_0^x f(t) dt$  per  $x \geq 0$ .
- Disegnare il grafico di  $F$  per  $x \geq 0$ .

**Esercizio 6.13** Sia  $F(x) = \int_0^x f(t) dt$ . Provare che se  $f$  è pari allora  $F$  è dispari, se  $f$  è dispari allora  $F$  è pari.

**Esercizio 6.14** Studiare la seguente funzione integrale e tracciarne un grafico plausibile:

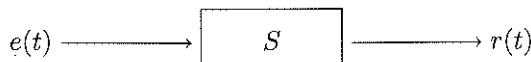
$$F(x) = \int_0^x \frac{\sin t}{1+t^2} dt$$

*Suggerimento:* non cercare una primitiva elementare dell'integranda: è inutile.

## 10 CONVOLUZIONE E SISTEMI FISICI LINEARI

Gli integrali estesi ad intervalli illimitati intervengono nel concetto di *convoluzione* fra due funzioni, che gioca un ruolo importante nella modellizzazione dei cosiddetti "sistemi lineari".

Precisamente, consideriamo un sistema fisico che operi nel modo descritto dal diagramma seguente:



cioè a un *ingresso*  $e(t)$  corrisponde una *risposta*  $r(t)$ , dipendente dall'ingresso mediante una funzione  $S$  caratterizzante il sistema. Si abbia cioè

$$r(t) = S[e](t)$$

Qui ingresso e risposta sono due grandezze variabili nel tempo, cioè due funzioni  $e(t), r(t)$ ;  $S$  è un operatore che trasforma la funzione  $e(t)$  nella funzione  $r(t)$ . Ciò significa, in particolare, che il valore  $r(t)$  a un istante particolare può dipendere da tutta la funzione  $e$ , e non solo dal valore  $e(t)$  in quell'istante.

Supponiamo che il sistema sia *lineare* ovvero che soddisfi il *principio di sovrapposizione* seguente:

**DEFINIZIONE 6.5** La risposta a una somma di ingressi è uguale alla somma delle risposte a ciascun ingresso.

Precisamente, se  $e_1(t), e_2(t), \dots, e_n(t)$  indicano  $n$  ingressi e  $\lambda_1, \dots, \lambda_n$  indicano numeri reali, si ha:

$$(10.1) \quad S \left[ \sum_{j=1}^n \lambda_j e_j \right] (t) = \sum_{j=1}^n \lambda_j S[e_j](t) = \sum_{j=1}^n \lambda_j r_j(t)$$

Supponiamo inoltre che il sistema abbia la seguente proprietà di *invarianza temporale*: se all'ingresso  $e(t)$  si sostituisce l'ingresso "ritardato"  $e(t-h)$  ( $h$  fissato), si ottiene la stessa risposta dell'ingresso  $e(t)$ , con uguale ritardo, cioè  $r(t-h)$ . In simboli:

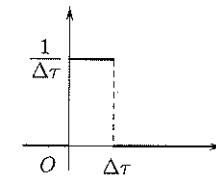
$$(10.2) \quad S[e(t-h)] = S[e](t-h)$$

Sotto queste ipotesi, la risposta a un qualunque ingresso  $e(t)$  può essere determinata mediante una particolare operazione, detta *convoluzione*, conoscendo la risposta a un ingresso speciale, l'*impulso di Dirac*.

Proviamo questa affermazione rimanendo tuttavia a un livello intuitivo.

L'impulso di Dirac concentrato in  $t = 0$ , e denotato con  $\delta(t)$  si può considerare come il "limite" per  $\Delta\tau \rightarrow 0$  dell'impulso unitario di durata  $\Delta\tau$ , definito dalla funzione

$$I_{\Delta\tau}(t) = \begin{cases} 1/\Delta\tau & 0 < t < \Delta\tau \\ 0 & \text{altrove} \end{cases}$$



L'impulso di Dirac concentrato in  $t = \tau$ , denotato con  $\delta(t-\tau)$ , sarà il "limite" per  $\Delta\tau \rightarrow 0$  dell'impulso unitario  $I_{\Delta\tau}(t-\tau)$ .

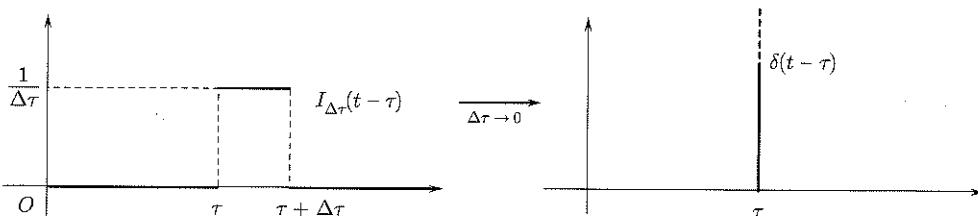


Figura 6.23. Si noti che  $I_{\Delta\tau}(t-\tau) \cdot \Delta\tau = 1$  per  $\tau < t < \tau + \Delta\tau$ .

Sia ora  $e = e(t)$  un ingresso, integrabile in  $(-\infty, +\infty)$ . Possiamo pensare di approssimare  $e(t)$  con la somma (generalmente infinita) di impulsi di durata  $\Delta\tau$ , relativi ad un intervallo temporale  $(\tau, \tau + \Delta\tau)$ , nel modo seguente (v. fig. 6.24).

Per ogni  $t$  tale che  $k\Delta\tau \leq t \leq (k+1)\Delta\tau$ , si può scrivere  $e(t) \approx e(k\Delta\tau) = e(k\Delta\tau)I_{\Delta\tau}(t-k\Delta\tau)\Delta\tau$ . Dunque, per ogni  $t$ :

$$(10.3) \quad e(t) \approx \sum_{k=-\infty}^{+\infty} e(k\Delta\tau)I_{\Delta\tau}(t-k\Delta\tau) \cdot \Delta\tau$$

(per ogni valore di  $t$  uno solo degli addendi nella somma (10.3) è diverso da zero!)

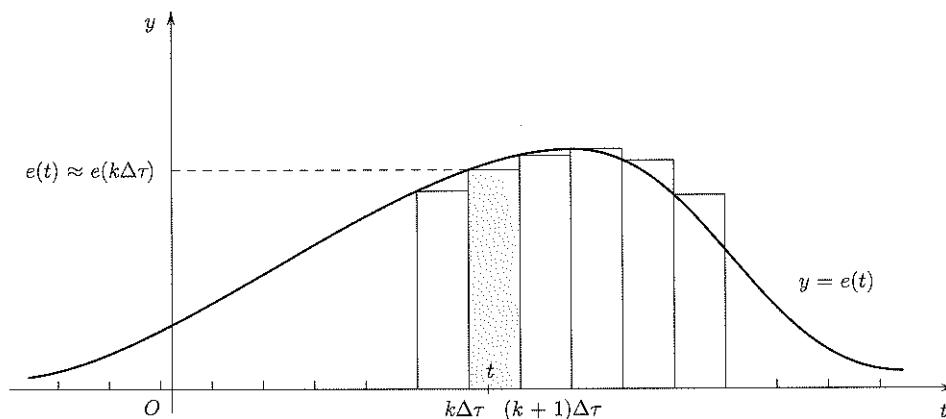


Figura 6.24.

Per il principio di sovrapposizione (10.1) si avrà allora

$$(10.4) \quad S[e](t) \approx \sum_{k=-\infty}^{+\infty} e(k\Delta\tau) S[I_{\Delta\tau}(t - k\Delta\tau)] \Delta\tau \\ (\text{se } \Delta\tau \approx 0) \quad \approx \sum_{k=-\infty}^{+\infty} e(k\Delta\tau) S[\delta(t - k\Delta\tau)] \Delta\tau$$

Ponendo  $h(t) = S[\delta](t)$  = risposta all'impulso di Dirac, e ricordando che, per la (10.2), si può anche scrivere  $S[\delta(t-\tau)] = h(t-\tau)$ , passando al limite nella (10.4) per  $\Delta\tau \rightarrow 0$  si ha:

$$(10.5) \quad S[e(t)] = \int_{-\infty}^{+\infty} e(\tau) h(t - \tau) d\tau$$

L'integrale nella (10.5) prende il nome di *convoluzione* di  $e$  e  $h$  e si indica col simbolo  $(e * h)(t)$ .

Dunque:

$$S[e](t) = (e * h)(t)$$

Ossia:

La risposta di un sistema lineare ad un qualunque ingresso  $e(t)$  si ottiene come convoluzione con la risposta all'impulso di Dirac.

In generale, date due funzioni  $f$  e  $g$  integrabili in  $\mathbb{R}$  si chiama *prodotto di convoluzione*, o semplicemente *convoluzione* di  $f$  e  $g$  la funzione definita dalla formula seguente

$$(10.6) \quad (f * g)(x) = \int_{-\infty}^{+\infty} f(y) g(x - y) dy$$

Il simbolo  $f * g$  si legge “ $f$  convoluzione  $g$ ”.

Frequentemente  $f$  e  $g$  sono nulle per  $x < 0$ ; in questo caso particolare si ha:

$$(f * g)(x) = \int_0^x f(y) g(x - y) dy$$

Infatti:

$$\int_{-\infty}^{+\infty} f(y) g(x - y) dy = \int_0^x f(y) g(x - y) dy + \int_{-\infty}^0 f(y) g(x - y) dy + \int_x^{+\infty} f(y) g(x - y) dy$$

ma il secondo integrale è nullo, essendo  $f = 0$  per  $y < 0$ , ed il terzo è nullo essendo  $g(x - y) = 0$  per  $x - y < 0$  cioè  $y > x$ .

### Proprietà della convoluzione

$$f * g = g * f \quad (\text{il prodotto } *) \text{ è commutativo}$$

$$(f * g) * h = f * (g * h) \quad (*) \text{ è associativo}$$

$$(f + g) * h = f * h + g * h \quad (*) \text{ è distributivo}.$$

#### ■ Esempio semplicissimo

*Convoluzione con l'impulso unitario di durata ε.*

$$\int_{-\infty}^{+\infty} I_\varepsilon(t) f(x - t) dt = \frac{1}{\varepsilon} \int_0^\varepsilon f(x - t) dt \xrightarrow{(x-t=u)} = \frac{1}{\varepsilon} \int_x^{x+\varepsilon} f(u) du$$

Perciò abbiamo:

$$(10.7) \quad (I_\varepsilon * f)(x) = \frac{1}{\varepsilon} \int_{x-\varepsilon}^x f(t) dt = \text{valor medio di } f \text{ su } [x - \varepsilon, x]$$

*Convoluzione con l'impulso di Dirac.*

Passando al limite per  $\varepsilon \rightarrow 0$  nella (10.7) si ha:

$$(I_\varepsilon * f)(x) \longrightarrow (\delta * f)(x)$$

e, usando il teorema della media per gli integrali

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{x-\varepsilon}^x f(t) dt = \lim_{c \rightarrow x} f(c) = f(x)$$

essendo  $x - \varepsilon < c < x$  ed  $f$  continua.

Abbiamo dunque l'importante formula:

$$(10.8) \quad (\delta * f)(x) = f(x)$$

La (10.8) indica che l'impulso di Dirac opera come l'elemento “unità” per il prodotto di convoluzione.



Sia  $I_1 = I_1(t)$  un impulso unitario di durata 1:

$$I_1(t) = \begin{cases} 1 & \text{se } 0 \leq t \leq 1 \\ 0 & \text{se } t < 0 \text{ oppure } t > 1 \end{cases}$$

Calcolare

$$(I_1 * I_1)(x) = \int_{-\infty}^{+\infty} I_1(y)I_1(x-y)dy$$

e disegnarne il grafico.

## COMPLEMENTI

### 11 TEOREMA DI BOLZANO-WEIERSTRASS, CONTINUITÀ UNIFORME E INTEGRABILITÀ DELLE FUNZIONI CONTINUE

In questo paragrafo dimostreremo il teorema, enunciato nel par. 2.2, che garantisce l'integrabilità delle funzioni continue su  $[a, b]$ . Questa dimostrazione costituisce la parte più delicata della teoria dell'integrazione, che come vedremo fa uso delle proprietà "fini" dell'insieme dei numeri reali, e richiederà l'introduzione di diverse nuove idee.

Anzitutto (par. 11.1) discuteremo certe proprietà delle successioni di numeri reali, che finora non erano state introdotte; quindi (par. 11.2) introduceremo una nozione simile alla continuità, ma più forte di questa, detta *uniforme continuità*, e proveremo un risultato che la riguarda; con questi strumenti, nel paragrafo 11.3 potremo provare l'integrabilità delle funzioni continue.

Segnaliamo che le nozioni introdotte nei paragrafi 11.1-11.2, pur molto importanti per i fondamenti dell'analisi matematica e per gli ulteriori sviluppi dell'analisi superiore, in questo corso saranno utilizzati esclusivamente in questo paragrafo.

#### 11.1 Alcuni risultati fondamentali per le successioni di numeri reali

Iniziamo con la seguente

**DEFINIZIONE 6.6** Data una successione  $\{a_n\}_{n=n_0}^{\infty}$  di numeri reali, chiamiamo *sottosuccessione* di  $\{a_n\}$  ogni successione estratta da questa, ossia ogni successione del tipo  $\{a_{n_k}\}_{k=1}^{\infty}$  dove  $n_1 < n_2 < n_3 < \dots$  è una successione crescente di indici interi.

Ad esempio, dalla successione

$$a_1, a_2, a_3, \dots, a_n \dots$$

potremmo estrarre le sottosuccessioni

$$a_2, a_4, a_6, \dots$$

oppure

$$a_1, a_3, a_5, \dots$$

o qualsiasi altra

$$a_{n_1}, a_{n_2}, a_{n_3}, \dots$$

pur di avere scelto una ben precisa successione (strettamente crescente) di indici interi  $\{n_k\}_{k=1}^{\infty}$ .

Si vede subito che se una successione  $\{a_n\}$  converge o diverge, cioè se  $a_n \rightarrow l$  per qualche  $l \in \mathbb{R}^*$ , allora ogni sottosuccessione estratta da  $\{a_n\}$  tenderà allo stesso  $l$ . Invece, in generale, dal fatto che una particolare sottosuccessione  $\{a_{n_k}\}$  tenda a un certo  $l$  non possiamo dedurre che lo stesso valga per la successione intera. Ad esempio, data la successione

$$a_n = (-1)^n,$$

se estraiamo la sottosuccessione

$$a_{2n} = (-1)^{2n} = 1$$

otteniamo una successione costante (e quindi convergente), mentre la successione di partenza è oscillante.

La situazione descritta da quest'esempio è più frequente di quanto si potrebbe credere: sovente una successione non è convergente, ma possiede una sottosuccessione convergente, e sapere questo può essere estremamente utile per certi scopi. Il prossimo teorema dà una condizione sufficiente perché questo accada:

**TEOREMA 6.11 (DI BOLZANO-WEIERSTRASS)** *Ogni successione di numeri reali contenuta in un intervallo limitato  $[a, b]$  possiede una sottosuccessione convergente in  $[a, b]$ .*

Per esempio, la successione  $\{\sin n\}_{n=1}^{\infty}$ , limitata ma estremamente irregolare, in base a questo teorema certamente possiede una sottosuccessione convergente (anche se non è così facile costruirla esplicitamente).

**DIMOSTRAZIONE.** Sia  $\{x_n\}$  una successione contenuta in  $[a, b]$ . Eseguiamo una costruzione con la tecnica di dicotomia, già utilizzata nella dimostrazione del teorema di Weierstrass e del teorema degli zeri (v. cap. 3, par. 4.1): dividiamo  $[a, b]$  in due sottointervalli di ugual ampiezza; almeno uno di essi dovrà contenere infiniti elementi  $x_n$  della successione (non necessariamente distinti: in altre parole, conterrà elementi  $x_n$  per un numero infinito di indici diversi). Chiamiamo  $[a_1, b_1]$  questo sottointervallo e scegliamo un elemento  $x_{n_1} \in [a_1, b_1]$ . Ora dividiamo  $[a_1, b_1]$  in due sottointervalli uguali; almeno uno di essi, che chiamiamo  $[a_2, b_2]$ , conterrà termini  $x_n$  per un numero infinito di indici diversi; in particolare, ne conterrà certamente uno del tipo  $x_{n_2}$  con  $n_2 > n_1$ . Procedendo iterativamente costruiamo una successione di intervalli  $[a_k, b_k]$  e una sottosuccessione  $\{x_{n_k}\}$  della successione di partenza, con le seguenti proprietà:

1.  $\{a_k\}$  è monotona crescente e superiormente limitata da  $b$ ;
2.  $\{b_k\}$  è monotona decrescente e inferiormente limitata da  $a$ ;
3.  $b_k - a_k = (b - a) / 2^k$
4.  $x_{n_k} \in [a_k, b_k]$

Per la (1) e (2),  $a_k$  e  $b_k$  convergono (per il teorema di monotonia, v. cap. 3 par. 1.2); per la (3), il limite a cui tendono  $a_k$  e  $b_k$  è lo stesso, chiamiamolo  $c$ ; per la (4) e il teorema del confronto, anche  $x_{n_k} \rightarrow c$ ; inoltre, poiché  $x_{n_k} \in [a_k, b_k] \subset [a, b]$ , per il teorema di permanenza del segno anche  $c \in [a, b]$ .

Il prossimo concetto che dobbiamo introdurre, riguardo alle successioni di numeri reali, è contenuto nella seguente

**DEFINIZIONE 6.7** Sia  $\{a_n\}$  una successione di numeri reali. Si dice che la successione soddisfa la *condizione di Cauchy*, o semplicemente che è  $\{a_n\}$  una *successione di Cauchy*, se avviene che

$$\forall \varepsilon > 0 \exists n_0 : \forall n, m \geq n_0 \text{ si ha } |a_n - a_m| < \varepsilon.$$

(Più brevemente:  $\forall \varepsilon > 0$  è  $|a_n - a_m| < \varepsilon$  definitivamente).

In sostanza, la condizione afferma che i termini della successione sono *sempre più vicini tra loro*. Si confronti con la definizione di successione convergente, che richiede invece che i termini della successione siano *sempre più vicini ad un certo numero reale  $\ell$* . È naturale pensare che una successione convergente soddisfi la condizione di Cauchy. Infatti, vale il prossimo:

### TEOREMA 6.12

- (i) Se una successione è convergente, allora è di Cauchy.
- (ii) Se una successione è di Cauchy, allora è limitata.

Si lascia al lettore la facile dimostrazione. Più interessante, e più delicato da dimostrare, è il fatto che la condizione di Cauchy sia non solo necessaria ma anche sufficiente per la convergenza:

**TEOREMA 6.13** *Sia  $\{a_n\}$  una successione di numeri reali. Se la successione è di Cauchy, allora ammette limite finito in  $\mathbb{R}$ .*

L'interesse del precedente teorema sta nel fatto che per provare la convergenza di una successione è sufficiente verificare la condizione di Cauchy che, come si vede dalla definizione, non richiede la previa conoscenza del *limite* a cui tende la successione.

**DIMOSTRAZIONE.** Sia  $\{a_n\}_{n=1}^{\infty}$  una successione di Cauchy. Per il teorema precedente,  $\{a_n\}$  è limitata. Allora, per il teorema di Bolzano-Weierstrass, contiene una sottosuccessione convergente:

$$\exists \{a_{n_k}\}_{k=1}^{\infty} \text{ ed } \exists \ell \in \mathbb{R} \text{ tale che } a_{n_k} \rightarrow \ell.$$

Mostriamo che, grazie alla condizione di Cauchy, allora, l'intera successione converge al limite  $\ell$ . Difatti:

$$|a_n - \ell| \leq |a_n - a_{n_k}| + |a_{n_k} - \ell|.$$

Fissato  $\varepsilon > 0$ ,  $|a_{n_k} - \ell| < \varepsilon$  definitivamente, perché  $a_{n_k} \rightarrow \ell$ ; d'altro canto, per la condizione di Cauchy è anche  $|a_n - a_m| < \varepsilon$  definitivamente, e quindi in particolare  $|a_n - a_{n_k}| < \varepsilon$  definitivamente. Perciò

$$|a_n - \ell| < 2\varepsilon$$

definitivamente, e il teorema è dimostrato.

Il teorema appena dimostrato esprime un'importante proprietà dell'insieme dei numeri reali che, ad esempio, non vale nell'insieme dei razionali. Per convincersene, si consideri

$$a_n = \left(1 + \frac{1}{n}\right)^n, \text{ per } n = 1, 2, \dots$$

Questa è una successione di numeri razionali, che converge (in  $\mathbb{R}$ ) al numero  $e$ . Di conseguenza la successione è di Cauchy (sia in  $\mathbb{R}$  che in  $\mathbb{Q}$ ). Tuttavia, in  $\mathbb{Q}$  la successione non è convergente (perché il numero  $e$  è irrazionale).

### 11.2 Continuità uniforme

La dimostrazione dell'integrabilità delle funzioni continue coinvolge in modo essenziale, come vedremo, una nozione simile alla continuità, ma più forte. Premettiamo quindi la seguente

**DEFINIZIONE 6.8** Si dice che  $f : I \rightarrow \mathbb{R}$  è *uniformemente continua* in  $I$  se

$$(11.1) \quad \forall \varepsilon > 0 \quad \exists \delta > 0 \text{ tale che } \forall x_1, x_2 \in I, \\ \text{se } |x_1 - x_2| < \delta \text{ allora } |f(x_1) - f(x_2)| < \varepsilon.$$

Notiamo la sottile differenza esistente tra le affermazioni “ $f$  è uniformemente continua in  $I$ ” e “ $f$  è continua in  $I$ ”. La seconda significa:

$$\forall x_1 \in I, \forall \varepsilon > 0, \quad \exists \delta > 0 \text{ tale che } \forall x_2 \in I, \\ \text{se } |x_1 - x_2| < \delta \text{ allora } |f(x_1) - f(x_2)| < \varepsilon.$$

Nella definizione di continuità, il numero  $\delta$  dipende da  $\varepsilon$  e anche da  $x_1$ , mentre in quella di uniforme continuità dipende solo da  $\varepsilon$ . (Questo si capisce facendo attenzione all'*ordine* in cui compaiono i vari quantificatori  $\forall, \exists$  nelle due affermazioni). Detto in modo meno formale: uniforme continuità significa che la quantità  $|f(x_1) - f(x_2)|$  può essere controllata conoscendo quanto è piccola la differenza  $|x_1 - x_2|$ , e questo senza bisogno di sapere dove si trovano i punti  $x_1, x_2$ .



**TEOREMA** La funzione  $f(x) = x^2$  è uniformemente continua su  $I = [0, 1]$ , ma non lo è su  $I = \mathbb{R}$ . Per provare la prima affermazione, scriviamo:

$$|f(x_1) - f(x_2)| = |x_1^2 - x_2^2| = |(x_1 - x_2)(x_1 + x_2)| \leq \\ \leq |x_1 - x_2|(|x_1| + |x_2|) \leq 2|x_1 - x_2|$$

perché  $x_1, x_2 \in [0, 1]$ . Quindi fissato  $\varepsilon > 0$ , se  $2|x_1 - x_2| < \varepsilon$ , ossia se  $|x_1 - x_2| < \varepsilon/2$ , sarà anche  $|f(x_1) - f(x_2)| < \varepsilon$ . In questo caso quindi basta scegliere  $\delta = \varepsilon/2$ , e la definizione di uniforme continuità è soddisfatta.

Consideriamo ora  $f$  su tutto  $\mathbb{R}$ . In questo caso, fissato  $\varepsilon > 0$  noi vorremmo trovare  $\delta > 0$  tale che  $|x_1 - x_2| < \delta$  implichi  $|x_1^2 - x_2^2| < \varepsilon$ . Questo non è possibile, infatti: comunque sceglieremo  $\varepsilon$  e  $\delta$ , sia  $x_1 = n$  e  $x_2 = n + \delta/2$  (con  $n$  intero che sceglieremo poi). Si ha:

$$|x_1 - x_2| = \delta/2 < \delta$$

ma

$$|x_1^2 - x_2^2| = |(x_1 - x_2)(x_1 + x_2)| = \left| \frac{\delta}{2} (n + \delta) \right| > n \frac{\delta}{2}.$$

Per quanto piccolo sia  $\delta$ , questa quantità, per  $n$  opportuno, sarà sempre maggiore di  $\varepsilon > 0$  fissato. Quindi  $f$  non è uniformemente continua su  $\mathbb{R}$ .

L'esempio precedente voleva illustrare come le nozioni di continuità e continuità uniforme siano effettivamente distinte. La continuità uniforme implica la continuità, ma è una condizione più forte, in generale.

Vale tuttavia il seguente importante

**TEOREMA 6.14 (DI CANTOR-HEINE)** *Se  $f : [a, b] \rightarrow \mathbb{R}$  è continua, allora è uniformemente continua in  $[a, b]$ .*

In altre parole, una funzione continua può violare la condizione di uniforme continuità solo se l'intervallo su cui è definita non è limitato o non è chiuso.

**DIMOSTRAZIONE.** Per assurdo, supponiamo che  $f$  sia continua in  $[a, b]$  ma non uniformemente continua. Questo significa che:

$$\exists \varepsilon > 0 \text{ tale che } \forall \delta > 0 \exists x_1, x_2 \in I \text{ tali che } |x_1 - x_2| < \delta \text{ ma } |f(x_1) - f(x_2)| \geq \varepsilon.$$

(si noti come abbiano costruito la *negazione logica* della definizione di uniforme continuità). Scegliendo allora  $\delta = 1/n$  successivamente per  $n = 1, 2, 3, \dots$ , possiamo costruire due successioni  $\{x_n\}, \{y_n\} \subset [a, b]$  tali che

$$|x_n - y_n| < \frac{1}{n} \text{ ma } |f(x_n) - f(y_n)| \geq \varepsilon.$$

Poiché la successione  $\{x_n\}$  è limitata (contenuta in  $[a, b]$ ), per il teorema di Bolzano-Weierstrass ammette una sottosuccessione  $\{x_{n_k}\}$  convergente a un certo  $c \in [a, b]$ . Consideriamo allora la sottosuccessione  $\{y_{n_k}\}$  corrispondente agli stessi indici; anche questa, essendo limitata, avrà una sottosuccessione  $\{y_{n_{k_h}}\}$  convergente. Possiamo allora concludere che le due sottosuccessioni

$$\{x_{n_{k_h}}\}, \{y_{n_{k_h}}\}$$

sono convergenti; anzi, poiché

$$|x_{n_{k_h}} - y_{n_{k_h}}| < \frac{1}{n_{k_h}} \rightarrow 0,$$

il limite è lo stesso  $c$ . D'altro canto la funzione  $f$  è continua, perciò

$$f(x_{n_{k_h}}) \rightarrow c \text{ e } f(y_{n_{k_h}}) \rightarrow c.$$

Ma questo è assurdo perché sappiamo che

$$|f(x_{n_{k_h}}) - f(y_{n_{k_h}})| \geq \varepsilon.$$

Perciò il teorema è dimostrato. ◆

### 11.3 Integrità delle funzioni continue

Abbiamo ora gli strumenti per provare il seguente:

**TEOREMA 6.15** *Se  $f : [a, b] \rightarrow \mathbb{R}$  è continua, allora è integrabile.*

**DIMOSTRAZIONE.** La dimostrazione è in due passi.

1. Mostriamo prima che una qualsiasi successione di Cauchy-Riemann  $s_n$  di  $f$  è di Cauchy, ossia che per ogni  $\varepsilon > 0$  risulta  $|s_n - s_m| < \varepsilon$  definitivamente. Consideriamo dunque, per due interi  $n, m$  fissati ( $n < m$ ), i due insiemi di punti, equispaziati in  $[a, b]$ :

$$x_0 = a, x_1, x_2, \dots, x_n = b$$

$$y_0 = a, y_1, y_2, \dots, y_m = b$$

e le corrispondenti somme di Cauchy-Riemann:

$$s_n = \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}); s_m = \sum_{j=1}^m f(\eta_j)(y_j - y_{j-1})$$

con

$$\xi_i \in [x_{i-1}, x_i], i = 1, 2, \dots, n;$$

$$\eta_j \in [y_{j-1}, y_j], j = 1, 2, \dots, m.$$

Per studiare la differenza  $s_n - s_m$  è utile esprimere entrambe le somme rispetto alla medesima suddivisione di  $[a, b]$ . Siano quindi

$$z_0 = a < z_1 < z_2 < \dots < z_k = b$$

(con  $k = n + m - 1$ ) tutti i punti  $x_i$  e  $y_j$ , disposti in ordine crescente. Consideriamo allora la somma  $s_n$ . Possiamo riscriverla come somma di  $k$  addendi, al modo seguente

$$s_n = \sum_{i=1}^k f(\xi_i^*) (z_i - z_{i-1})$$

a patto di definire i punti  $\xi_i^*$  a questo modo: per tutti gli intervallini  $[z_{i-1}, z_i]$  che compongono l'intervallo (più grande)  $[x_{j-1}, x_j]$ , si pone  $\xi_i^* = \xi_j$  (vedi figure 6.25-6.27).

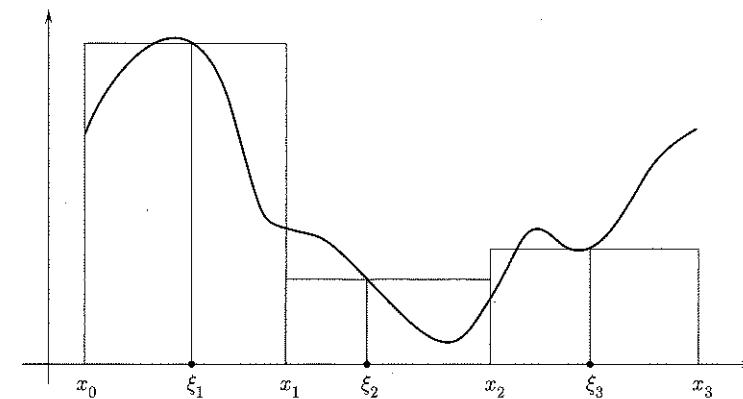


Figura 6.25. La somma  $s_3$ .

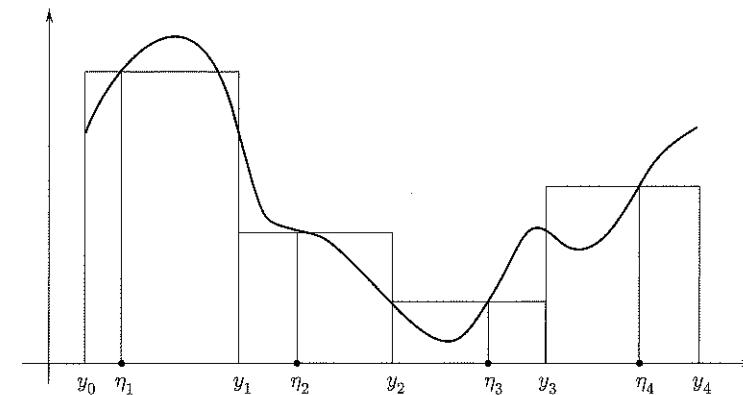


Figura 6.26. La somma  $s_4$ .

Analogamente riscriviamo la somma  $s_m$  come:

$$s_m = \sum_{i=1}^k f(\eta_i^*) (z_i - z_{i-1})$$

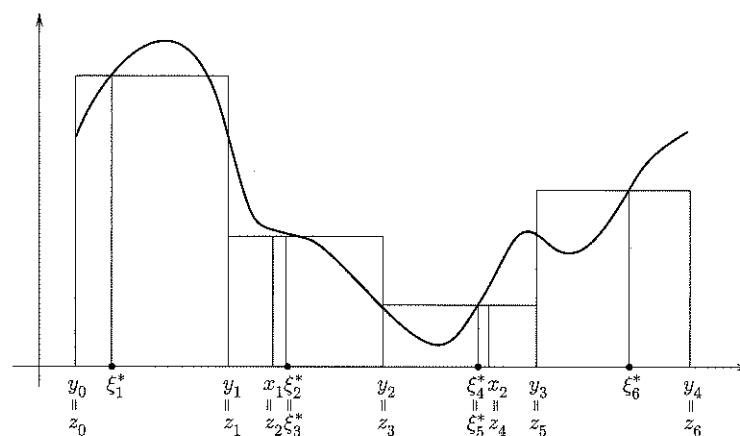


Figura 6.27. La somma  $s_3$  riscritta aggiungendo ai punti  $y_i$  anche i punti  $x_2, x_3$ : il punto  $\xi_2$  che si era scelto in  $[y_1, y_2]$  ora si usa sia per  $[y_1, x_1]$  sia per  $[x_1, y_2]$ , e si chiamerà  $\xi_2^*$  e  $\xi_3^*$ ; il punto  $\xi_3$  che si era scelto in  $[y_2, y_3]$  ora si usa sia per  $[y_2, x_2]$  sia per  $[x_2, y_3]$ , e si chiamerà  $\xi_4^*$  e  $\xi_5^*$ .

e possiamo quindi valutare la differenza:

$$s_n - s_m = \sum_{i=1}^k [f(\xi_i^*) - f(\eta_i^*)] (z_i - z_{i-1}).$$

Ragioniamo ora sul fatto che i due punti  $\xi_i^*, \eta_i^*$ , anche se non appartengono allo stesso intervallino  $[z_{i-1}, z_i]$ , si trovano ad una distanza che non supera  $\frac{2(b-a)}{n}$ . Poiché  $f$  è continua, e quindi uniformemente continua, fissato  $\varepsilon > 0$  esiste un  $\delta > 0$  tale che

$$|f(\xi_i^*) - f(\eta_i^*)| < \varepsilon \text{ purché } |\xi_i^* - \eta_i^*| < \delta,$$

e questo capita se  $\frac{2(b-a)}{n} < \delta$ , ossia per  $n > 2(b-a)/\delta$ . Allora se  $m > n > 2(b-a)/\delta$  si ha:

$$|s_n - s_m| \leq \sum_{i=1}^k |f(\xi_i^*) - f(\eta_i^*)| (z_i - z_{i-1}) \leq \varepsilon \sum_{i=1}^k (z_i - z_{i-1}) = \varepsilon (b - a).$$

Questo dimostra che *ogni successione di somme di Cauchy-Riemann di  $f$  è di Cauchy, e quindi convergente*.

2. Proviamo ora che, prese due diverse successioni di somme di Cauchy-Riemann di  $f$ , il loro limite (che già sappiamo esistere) è lo stesso. Per questo, considerate due successioni

$$\begin{aligned} s_n &= \sum_{i=1}^n f(t_i) (x_i - x_{i-1}); \\ s_n^* &= \sum_{i=1}^n f(t_i^*) (x_i - x_{i-1}) \end{aligned}$$

con  $t_i, t_i^* \in [x_{i-1}, x_i]$ , sarà sufficiente provare che  $s_n - s_n^* \rightarrow 0$ . Il ragionamento è simile a quello del primo punto, ma più semplice, perché in questo caso confrontiamo due somme relative alla stessa suddivisione. Fissiamo  $\varepsilon > 0$ ; poiché  $f$  è uniformemente continua e

$$|t_i - t_i^*| \leq \frac{b-a}{n},$$

per  $n$  abbastanza grande sarà  $|f(t_i) - f(t_i^*)| < \varepsilon$  e quindi:

$$\begin{aligned} |s_n - s_n^*| &= \left| \sum_{i=1}^n [f(t_i) - f(t_i^*)] (x_i - x_{i-1}) \right| \leq \\ &\leq \sum_{i=1}^n |f(t_i) - f(t_i^*)| |x_i - x_{i-1}| < \sum_{i=1}^n \varepsilon |x_i - x_{i-1}| = \varepsilon (b - a). \end{aligned}$$

Questo prova che  $s_n - s_n^* \rightarrow 0$ . Pertanto tutte le successioni di somme di Cauchy-Riemann di  $f$  tendono allo stesso limite, ed  $f$  è integrabile.

# 7

## Modelli dinamici discreti

### ■ 1 INTRODUZIONE ALLA MODELLISTICA

Da qualche decennio, la modellistica matematica, già tradizionalmente fondamentale in Fisica, è entrata pesantemente (forse anche inaspettatamente) in importanti rami dell'Ecologia, della Biologia, della Medicina e in modo ancor più rilevante in Economia e Finanza. Le ragioni di questo successo sono dovute alla speciale flessibilità e all'universalità degli strumenti matematici, spesso capaci (e a volte gli unici) di dare risposte soddisfacenti a problemi di crescente complessità, grazie alla potenza di calcolo di macchine sempre più sofisticate, oggi disponibili.

Che cos'è un *modello matematico*? Come si costruisce? Un modello matematico è un'interfaccia tra una parte del mondo reale e il mondo delle teorie matematiche. Supponiamo di aver isolato un fenomeno reale (per esempio il trasporto e/o la diffusione di un inquinante in un corso d'acqua) e di volerlo studiare allo scopo di fare previsioni. Se si vuole usare un modello matematico occorre anzitutto capire quali siano i fattori (tradotti in variabili matematiche) che riteniamo essenziali e caratteristici nell'evoluzione. Questa fase è probabilmente la più delicata. Per queste variabili si cercano poi le relazioni fondamentali che descrivono quantitativamente la dinamica. Solitamente si perviene a una o più equazioni di vario tipo che costituiscono il modello matematico. Si tratta poi di analizzare il modello, cercando di estrarne le informazioni desiderate.

Qual è la credibilità di un modello matematico? Dipende dalla sua bontà, determinata dalla dicotomia completezza/analizzabilità. La completezza, in generale, richiederebbe di tener conto di un numero troppo elevato di componenti e di relazioni tra esse il più realistico possibile; d'altra parte, quanto maggiore è il numero di variabili e quanto più realistiche sono le relazioni tra esse, tanto più complicato è il modello, e minore la capacità di analisi. Un buon modello costituisce un efficace compromesso, ma proprio per questo motivo, occorrono sempre test a posteriori, usando casi già noti, per garantirne la validità predittiva. In ogni caso, si deve tener sempre presente che un modello matematico (ma questo vale per qualunque modello descrittivo di fenomeni reali) non ha nessuna pretesa di universalità e va sempre usato *cum grano salis*. Alcuni modelli permettono analisi e previsioni quantitative, come per esempio i modelli in Fisica, Chimica, Medicina, Finanza; in altri, tipicamente quelli

in Economia, le analisi sono forzatamente qualitative e vanno usate unicamente per testare l'influenza di alcuni fattori sull'evoluzione di un dato sistema.

Perché usare modelli matematici? Alcune caratteristiche li rendono particolarmente appetibili. Il loro basso costo, la loro flessibilità, legata alla possibilità praticamente illimitata di simulazioni ripetute al computer per descrivere l'andamento di fenomeni complessi, altrimenti incomprensibili. A volte sono necessari, essendo l'unico strumento utilizzabile. Un esempio notevole in questo senso è costituito dalla simulazione della circolazione sanguigna nel cosiddetto *circolo di Willis*, che assicura l'irrorazione nei due emisferi cerebrali, usato per valutare l'effetto di un'eventuale occlusione in uno dei vasi. In questo caso, non è certo possibile fare esperimenti in loco!

Molta cautela e un atteggiamento sempre critico vanno impiegati nell'uso del computer, evitando il macroscopico errore, tipico dei principianti, di affermare "l'ha detto il computer e quindi deve essere così." Particolarmenete vero è questo fatto per i modelli di tipo economico.

In generale, un modello matematico si costruisce traducendo in termini matematici *leggi generali* di evoluzione accoppiate con *leggi specifiche (costitutive)* del fenomeno da descrivere. I modelli di cui ci occuperemo in questo capitolo appaiono spesso nelle scienze applicate e sono costituiti da *equazioni alle differenze*, che traducono il monitoraggio dell'evoluzione ad intervalli di tempo regolari. Per questa ragione si chiamano anche *modelli dinamici a tempo discreto*. Nel secondo volume esamineremo modelli *a tempo continuo*, che conducono ad *equazioni differenziali*, in cui le incognite sono funzioni che appaiono insieme ad alcune loro derivate.

Vediamo ora come si procede nel caso di alcuni esempi classici che inoltre ci permetteranno di motivare lo sviluppo della teoria.

## 1.1 Modello di Malthus

È storicamente il primo modello per la dinamica delle *popolazioni*, dovuto a Malthus<sup>1</sup> nel 1798. Consideriamo una popolazione<sup>2</sup> che evolve isolata e i cui unici fattori di evoluzione sono *fertilità* e *mortalità*. Indichiamo con  $x(t)$  il numero di individui presenti al tempo  $t$  e ci proponiamo di studiarne l'evoluzione a partire dall'istante (convenzionale)  $t = 0$ . Naturalmente  $x(t)$  è un valore medio (per esempio, numero di individui per unità di area), quindi non necessariamente intero, che possiamo senz'altro ritenere reale. Siano  $\lambda$  il tasso di fertilità e  $\mu$  il tasso di mortalità, cioè il numero di individui nati (rispettivamente, morti) nell'unità di tempo per ogni individuo già esistente. Ne segue che, in un intervallo di tempo di durata  $h$ , la percentuale di nuovi nati e di morti è, rispettivamente,  $\lambda h$  e  $\mu h$ . Si noti che, a priori,  $\lambda$  e  $\mu$  potrebbero dipendere da  $t$  e da  $x(t)$ .

La legge generale di evoluzione qui è semplice: *la variazione relativa del numero di individui in un intervallo di tempo di durata  $h$  è funzione di  $(\lambda - \mu) h$* .

La legge costitutiva indica *che tipo di funzione si intende adottare*. Malthus assume che questa funzione sia esattamente  $(\lambda - \mu) h$ , con  $\lambda$  e  $\mu$  costanti. Il tutto si traduce

<sup>1</sup>Thomas Robert Malthus (1766-1834), economista inglese.

<sup>2</sup>Non necessariamente di esseri umani: potrebbe, per esempio, trattarsi di una popolazione di batteri, o della quantità di atomi radioattivi in una massa di materiale fissile.

allora facilmente in termini matematici:

$$(1.1) \quad \frac{x(t+h) - x(t)}{x(t)} = (\lambda - \mu) h.$$

Il numero  $\varepsilon = \lambda - \mu$  si chiama *potenziale biologico*.

Pensiamo ora di monitorare l'evoluzione di  $x$  ad intervalli di tempo regolari. Questo modo di procedere corrisponde a considerare il tempo come variabile *discreta*. Prendendo allora come unità di misura temporale l'intervallo tra due osservazioni successive si ha  $t \in \mathbb{N}$ . Lo stato del sistema è quindi descritto dalla successione  $x(0)$ ,  $x(1)$ ,  $x(2)$ , ... . Si usa anche scrivere  $x_0$ ,  $x_1$ ,  $x_2$  e, in generale,  $x_n$  con  $n$  intero, al posto di  $x(t)$ . Sostituendo nella (1.1) e scegliendo  $h = 1$  (l'intervallo di tempo più piccolo possibile, che potrebbe essere un mese, un anno, un giorno...) si ha

$$(1.2) \quad x_{n+1} - x_n = \varepsilon x_n.$$

La (1.2) connette l'incremento  $x_{n+1} - x_n$  con lo stato al tempo  $n$  e per questa ragione si chiama *equazione (lineare) alle differenze, a un passo o del prim'ordine*. Spesso conviene scrivere la (1.2) nella forma

$$(1.3) \quad x_{n+1} = (1 + \varepsilon) x_n$$

dove lo stato al tempo  $n+1$  è espresso in funzione dello stato al tempo precedente  $n$ . La conoscenza della consistenza della popolazione all'istante iniziale  $n = 0$  permette di determinare un'unica soluzione, che è

$$(1.4) \quad x_n = M_0 (1 + \varepsilon)^n$$

dove  $x_0 = M_0$  è la condizione iniziale.

Infatti, sostituendo  $n = 0$  nella (1.3), si ottiene  $x_1 = (1 + \varepsilon) M_0$  e, successivamente, sostituendo  $n = 1$ , si ha

$$x_2 = (1 + \varepsilon) x_1 = (1 + \varepsilon)^2 M_0.$$

Procedendo in tal modo si ottiene

$$x_n = (1 + \varepsilon) x_{n-1} = (1 + \varepsilon)^2 x_{n-2} = (1 + \varepsilon)^3 x_{n-3} = \dots = (1 + \varepsilon)^n M_0.$$

Alla condizione iniziale  $M_0 = 0$  corrisponde la soluzione nulla. Essendo costante nel tempo si dice che è una *soluzione d'equilibrio (steady state)*. Se lasciamo variare la condizione iniziale  $M_0$  su tutti i numeri reali, si ottiene la totalità delle soluzioni.

Le informazioni più interessanti per le applicazioni si ricavano dal comportamento asintotico delle soluzioni rispetto alla soluzione di equilibrio. Tecnicamente si tratta di calcolare il limite per  $n \rightarrow +\infty$  di  $x_n$ .

Vediamo che cosa succede. Se  $\varepsilon > 0$  allora  $1 + \varepsilon > 1$  e  $(1 + \varepsilon)^n M_0 \rightarrow +\infty$  se  $n \rightarrow +\infty$ . In questo caso, dunque le soluzioni si allontanano dalla soluzione di equilibrio. Diremo allora che quest'ultima è *instabile*.

Se  $-2 < \varepsilon < 0$  allora  $-1 < 1 + \varepsilon < 1$  e  $(1 + \varepsilon)^n M_0 \rightarrow 0$  se  $n \rightarrow +\infty$  e la soluzione nulla si dice *asintoticamente stabile*; se  $\varepsilon = -2$ , allora  $1 + \varepsilon = -1$  e la soluzione oscilla tra i valori  $-M_0$  e  $M_0$  (soluzione nulla *stabile ma non asintoticamente*), mentre se  $\varepsilon < -2$ , allora  $1 + \varepsilon < -1$  e  $x_n$  è oscillante e illimitata (soluzione nulla *instabile*).

Preciseremo più avanti i concetti di instabilità, stabilità e stabilità asintotica.

## 1.2 Modello logistico

Il modello di Malthus è troppo semplicistico e irrealistico nella maggioranza dei casi. In tale modello si fa l'ipotesi che l'ambiente esterno non influenzi la crescita e, di conseguenza, che il tasso di crescita sia proporzionale alla popolazione presente, ossia che il tasso relativo di crescita sia costante. Una popolazione più numerosa comporta però minori risorse e ciò implica un minor tasso di crescita. Nel 1845 Verhulst<sup>3</sup> propose un modello nel quale la legge costitutiva di evoluzione prevede una *soglia di sopravvivenza*  $M$  che la popolazione non può superare, e prevede che la variazione relativa di  $x(t)$  nell'intervallo di tempo  $h$  decresca linearmente come funzione di  $x(t)$ . Sotto queste condizioni la legge di evoluzione si può scrivere nel modo seguente:

$$(1.5) \quad x(t+h) - x(t) = \varepsilon h x(t) \left[ 1 - \frac{x(t)}{M} \right]$$

Scegliendo come prima  $h = 1$ , ponendo  $t = n$ , intero, e  $x(n) = x_n$ , dalla (1.5) si ottiene

$$x_{n+1} = x_n + \varepsilon x_n \left( 1 - \frac{x_n}{M} \right) = (1 + \varepsilon) x_n \left( 1 - \frac{\varepsilon}{(1 + \varepsilon) M} x_n \right)$$

che è un' *equazione alle differenze* del primo ordine, *non lineare*, che prende il nome di *equazione logistica discreta*. Rendiamo più leggibile l'equazione ponendo

$$x_n = \frac{(1 + \varepsilon) M}{\varepsilon} y_n.$$

L'equazione per  $y_n$  è

$$(1.6) \quad y_{n+1} = r y_n (1 - y_n)$$

dove  $r = 1 + \varepsilon$ .

La successione  $y_n = 0$  per ogni  $n$  (cioè la successione di zeri) è *soluzione di equilibrio*. Che cosa possiamo dedurre dalla (1.6) riguardo alle altre soluzioni? Anzitutto: vi sono altre soluzioni di equilibrio? Una soluzione di equilibrio è costante, per cui  $y_{n+1} = y_n$ . Dalla (1.6) otteniamo allora

$$y_{n+1} = r y_n (1 - y_n) = y_n$$

da cui

$$r (1 - y_n) = 1$$

e successivamente

$$y_n = 1 - \frac{1}{r}.$$

Questa soluzione è tra 0 e 1 se  $r > 1$  e quindi, in questo caso vi sono due soluzioni di equilibrio. Ci si chiede se l'andamento delle altre soluzioni tenderà verso una delle due. Intanto, se il potenziale biologico  $\varepsilon$  è negativo ma non troppo, in modo che  $0 < r = 1 + \varepsilon < 1$ , e se  $y_n$  parte da  $y_0$  tra 0 e 1, si ha subito

$$0 < y_{n+1} = r y_n (1 - y_n) < r y_n < r^2 y_{n-1} < \dots < r^n y_0$$

e quindi  $y_n \rightarrow 0$  per  $n \rightarrow +\infty$ . La soluzione nulla è asintoticamente stabile.

<sup>3</sup>Pierre Francois Verhulst, 1804-1849.

E se  $\varepsilon > 0$ ? Allora  $r > 1$  ed i due fattori  $r y_n$  e  $(1 - y_n)$  entrano in competizione: la moltiplicazione per  $r$  fa aumentare rapidamente  $y_n$ , ma appena  $y_n$  si avvicina alla quota 1 il secondo fattore è molto piccolo e tende a diminuire il valore di  $y_n$ . Quale dei due fattori prevalga è molto difficile da stabilire *a priori*, senza contare il fatto che  $1 - y_n$  potrebbe diventare negativo e la successione perderebbe il suo significato di numero (medio) di individui di una popolazione.

Proviamo a ricorrere al computer e a disegnare nel piano  $(n, y_n)$  i punti generati dalla logistica discreta per vari valori del parametro  $r$ . In figura 7.1 si vede quello che si ottiene con  $y_1 = 0,2$ , dando a  $r$  i valori 0,7 (circolo), 2,6 (rombo), 3,4 (croce), 4 (asterisco).

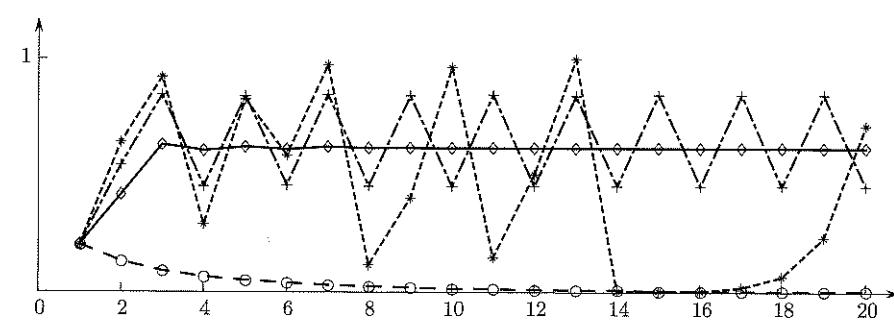


Figura 7.1. Sensibilità della logistica rispetto al parametro  $r$ .

Si osserva un'estrema sensibilità alle variazioni del parametro  $r$  ed una varietà, a priori insospettabile, di possibili comportamenti asintotici, non escluso quello che prende il nome di *comportamento caotico*. Per interpretare i grafici della figura 7.1 occorre un po' di teoria. Ce ne occuperemo più avanti.

## 1.3 Modello dell'acceleratore

Questo classico modello (P.A. Samuelson, 1938) costituisce un tentativo di spiegare elementarmente l'alternarsi di periodi d'espansione e di recessione nello sviluppo economico. Consideriamo un sistema economico osservato per un certo numero d'anni in sequenza. Indichiamo con  $Y_n$  il *prodotto nazionale* per l'anno  $n$  e con  $C_n$  il *consumo* nell'anno  $n$ . Nel modello vi sono tre leggi costitutive:

a)  $C_n$  è *funzione lineare (affine)*<sup>4</sup> di quanto prodotto nell'anno precedente:

$$(1.7) \quad C_n = a Y_{n-1} + k,$$

dove  $a$  è un coefficiente di propensione al consumo,  $0 < a < 1$  e  $k$  è un numero positivo che rappresenta un consumo fisso costante. Questa funzione è nota come *funzione del consumo*;

<sup>4</sup>Correttamente:  $y = ax$  è *lineare*,  $y = ax + b$  è *lineare affine*.

- b) i produttori investono un ammontare  $I_n$  costituito da una parte autonoma  $A > 0$  e da una parte proporzionale alla variazione dei consumi<sup>5</sup>:

$$I_n = A + b(C_n - C_{n-1}),$$

con  $b > 0$ .

- c) Il livello di  $Y_n$  (ossia la decisione sulla produzione) è scelto in modo da soddisfare la domanda di beni (consumo + investimento). Si avrà allora

$$Y_n = C_n + I_n.$$

Sostituendo in quest'ultima uguaglianza l'espressione precedente di  $I_n$  ed eliminiamo i consumi attraverso la (1.7); otteniamo:

$$(1.8) \quad Y_n - a(1+b)Y_{n-1} + abY_{n-2} = A + k.$$

La (1.8) è un'equazione alle differenze che connette il valore di  $Y_n$  con quelli di  $Y_{n-1}$  e  $Y_{n-2}$ ; inoltre, il primo membro è un polinomio di primo grado in  $Y_n$ ,  $Y_{n-1}$  e  $Y_{n-2}$ . Per questi motivi l'equazione si dice *lineare del second'ordine* (o *a due passi*).

Per le soluzioni di questo tipo di equazioni ricaveremo (v. par. 5) una formula generale da cui sarà facile dedurre il comportamento di  $Y_n$  per  $n \rightarrow +\infty$ .

## 2 GENERALITÀ SULLE EQUAZIONI ALLE DIFFERENZE

Abbiamo considerato modelli matematici che descrivono l'evoluzione di una variabile, detta *variabile di stato*, in funzione del tempo, misurato ad intervalli regolari. In tal caso, scegliendo opportunamente l'unità di misura temporale si può far in modo che la variabile temporale assuma i valori interi  $0, 1, 2, \dots$

L'unità di misura temporale può essere un anno, un semestre, un mese, un giorno, ecc. In genere si sceglie come unità di misura l'intervallo di tempo tra due "osservazioni" successive.

La funzione che descrive l'evoluzione della variabile di stato è allora una successione  $\{x_n\}$ :

$$n \mapsto x_n \quad n \in \mathbb{N}.$$

Gli ingredienti tipici di un modello dinamico discreto sono i seguenti. Dato  $k \geq 1$ , si assegna:

- a) la *dinamica*, ossia la *legge d'evoluzione di  $x$* , che esprime un legame tra il valore della variabile  $x$  al tempo  $n+k$ , in termini del valore di  $x$  ai tempi precedenti  $n, n+1, \dots, n+k-1$ .

In formule, si ha un'equazione del tipo

$$x_{n+k} = f(n, x_{n+k-1}, \dots, x_n)$$

che si chiama equazione *alle differenze di ordine  $k$  o a  $k$  passi*.

<sup>5</sup>L'idea sottostante è che se, per esempio, i consumi sono in crescita, per far fronte alla domanda che si manifestera nel periodo successivo, è opportuno espandere la capacità produttiva. Viceversa nel caso di consumi in contrazione.

- b) le *condizioni iniziali*, ossia i primi  $k$  valori di  $x$ ,cio è  $x_0, x_1, \dots, x_{k-1}$ .

Nei prossimi due paragrafi ci occuperemo delle equazioni alle differenze del prim'ordine o a un passo, del tipo

$$(2.1) \quad x_{n+1} = f(n, x_n), \quad n \in \mathbb{N}$$

con

$$(2.2) \quad x_0 \text{ assegnato.}$$

Se  $f$  non dipende esplicitamente da  $n$ , cioè

$$x_{n+1} = f(x_n), \quad n \in \mathbb{N}$$

l'equazione si dice *autonoma*.

Abbiamo assunto  $n \in \mathbb{N}$ , ma può succedere che l'equazione abbia senso da un certo intero  $n_0$  in poi, nel qual caso si assegna  $x_{n_0}$  come condizione iniziale.

Per alcuni tipi di equazione, sono disponibili formule esplicite per la totalità delle soluzioni, che chiameremo *soluzione generale*. Un caso importante è quello delle equazioni lineari a coefficienti costanti, che analizzeremo nel paragrafo 3.

Osserviamo subito che se la funzione  $f$  risulta ben definita in  $(n, x_n)$  per ogni  $n \geq 0$ , *esistenza ed unicità* della soluzione del problema (2.1), (2.2) sono assicurate.

L'analisi degli equilibri ed il comportamento asintotico per  $t \rightarrow +\infty$  possono invece presentare aspetti notevolmente complessi.

Il caso particolare delle equazioni *non lineari autonome*, ci servirà per introdurre gli importanti concetti di punto fisso o *di equilibrio*, di *orbite periodiche* e di *caos deterministico* (par. 4).

Infine, nel paragrafo 5 ci occuperemo delle equazioni alle differenze del 2° ordine, lineari a coefficienti costanti.

## 3 EQUAZIONI LINEARI DEL PRIM'ORDINE A COEFFICIENTI COSTANTI

Un'equazione *lineare del prim'ordine a coefficienti costanti* è del tipo

$$(3.1) \quad x_{n+1} = ax_n + b, \quad n \in \mathbb{N}$$

dove  $a, b \in \mathbb{R}$ . Se  $b = 0$  l'equazione si dice *omogenea*. Per queste equazioni è disponibile una formula esplicita per la soluzione generale.

Iniziamo con il caso dell'equazione *omogenea associata*

$$w_{n+1} = aw_n,$$

che corrisponde al modello di Malthus, con  $a = 1 + \varepsilon$ . Come abbiamo già visto, la soluzione generale è

$$w_n = ca^n, \quad c \in \mathbb{R}$$

dove il parametro  $c$  coincide con il valore iniziale della variabile di stato:  $c = w_0$ .

Il caso non omogeneo non è molto più complicato. Infatti, supponiamo di conoscere una soluzione particolare  $\{y_n\}$  della (3.1). Data un'altra soluzione qualsiasi  $\{x_n\}$  e posto  $w_n = x_n - y_n$ , si ha:

$$\begin{aligned} w_{n+1} &= x_{n+1} - y_{n+1} = ax_n + b - (ay_n + b) = \\ &= a(x_n - y_n) = aw_n. \end{aligned}$$

Pertanto,  $w_n$  è soluzione dell'equazione omogenea associata e quindi  $w_n = ca^n$ . Deduciamo che

$$x_n = ca^n + y_n$$

ossia:

**La soluzione generale della (3.1) si ottiene come somma della soluzione generale dell'omogenea associata e di una soluzione particolare della non omogenea.**

Il problema è così ricondotto alla ricerca di una soluzione particolare della (3.1).

Esendo  $b$  costante, una soluzione particolare si trova facilmente. Infatti, provando con una successione costante  $\bar{x}_n = k$ , deve essere, dato che  $\bar{x}_{n+1} = k$ ,

$$k = ak + b.$$

Se  $a \neq 1$  si trova

$$\bar{x}_n = k = \frac{b}{1-a}$$

che, essendo costante, è una *soluzione di equilibrio*. Di conseguenza, la soluzione generale dell'equazione

$$x_{n+1} = ax_n + b, \quad a, b \in \mathbb{R}, \quad a \neq 1$$

è

$$(3.2) \quad x_n = ca^n + \frac{b}{1-a}, \quad c \in \mathbb{R}$$

e la soluzione che "parte" da  $x_0$  è

$$(3.3) \quad x_n = \left( x_0 - \frac{b}{1-a} \right) a^n + \frac{b}{1-a}.$$

Infatti, sostituendo nella (3.2)  $n = 0$ , si trova

$$x_0 = c + \frac{b}{1-a}$$

e quindi

$$c = x_0 - \frac{b}{1-a}.$$

Se  $a = 1$  da

$$x_{n+1} = x_n + b$$

sostituendo  $n = 0, 1, 2, \dots$  si ha

$$x_1 = x_0 + b,$$

$$x_2 = x_1 + b = x_0 + 2b,$$

$$x_3 = x_2 + b = x_0 + 3b,$$

⋮

e si ricava facilmente

$$x_n = x_0 + nb$$



**Cerchiamo la soluzione dell'equazione alle differenze**

$$x_{n+1} = \frac{1}{2}x_n + 1,$$

che soddisfa la condizione iniziale  $x_0 = 1$ . La soluzione generale dell'equazione omogenea corrispondente è

$$z_n = c \left( \frac{1}{2} \right)^n$$

mentre la soluzione di equilibrio è (qui  $a = 1/2 \neq 1$ )  $\bar{x}_n = 2$  per cui la soluzione generale è

$$(3.4) \quad x_n = c \left( \frac{1}{2} \right)^n + 2, \quad c \in \mathbb{R}.$$

Per determinare la soluzione particolare che verifica la condizione iniziale assegnata determiniamo la costante  $c$ , sostituendo il valore 0 nella (3.4); si ha

$$x_0 = 1 = c + 2 \Rightarrow c = -1.$$

La soluzione cercata è quindi

$$x_n = 2 - \left( \frac{1}{2} \right)^n.$$

Si controlla facilmente che  $x_n \rightarrow 2$  per  $n \rightarrow +\infty$ .

**Capitalizzazione semplice e composta.** Un capitale  $C$ , investito in capitalizzazione semplice con tasso mensile d'interesse semplice  $r$ , ogni mese frutta interessi pari a  $rC$ . Indicando con  $C_n$  la somma disponibile dopo  $n$  mesi e con  $C_{n+1}$  la somma disponibile al mese successivo, si ha

$$(3.5) \quad C_{n+1} = C_n + rC.$$

Se la stessa somma fosse investita in capitalizzazione composta con tasso mensile d'interesse composto  $i$ , indicando sempre con  $C_n$  la somma disponibile dopo  $n$  mesi e con  $C_{n+1}$  la somma disponibile al mese successivo, si avrebbe, essendo gli interessi calcolati non sul capitale iniziale ma sul capitale maturato,

$$(3.6) \quad C_{n+1} = C_n + iC_n = (1+i)C_n.$$

La (3.5) e la (3.6) sono equazioni lineari a coefficienti costanti, la seconda è anche omogenea. Per le due equazioni (3.5) e (3.6), tenendo conto della condizione iniziale  $C_0 = C$  (il capitale in data 0 è  $C$ ), si ottengono, rispettivamente, le soluzioni

$$C_n = C(1 + rn), \quad C_n = C(1 + i)^n.$$

**Modello della ragnatela.** La quantità domandata e la quantità offerta d'un bene dipendono dal prezzo dello stesso. In un semplice modello che descrive la formazione del prezzo di mercato, si assume che domanda e offerta siano funzioni lineari del prezzo.

Consideriamo prima il caso statico, in cui il prezzo  $p$ , e quindi anche domanda e offerta siano costanti. Se  $D$  è la quantità domandata,  $O$  la quantità offerta e  $a, b, c, d$  parametri positivi<sup>6</sup>, si ha

$$D(p) = a - bp, \quad O(p) = -c + dp.$$

Il mercato è in equilibrio se  $D = O$  ovvero se

$$a - bp = -c + dp$$

da cui s'ottiene il prezzo  $p^*$  d'equilibrio

$$p^* = \frac{a + c}{b + d}.$$

Supponiamo ora che il prezzo e, quindi, la domanda e l'offerta siano funzioni del tempo. Mentre la domanda s'adegua immediatamente al prezzo, l'offerta richiede un po' di tempo. La quantità offerta dipende, in particolare, dal prezzo nel periodo temporale precedente. Sceglieremo come unità di tempo il "ritardo" nell'adeguamento dell'offerta. Per molti prodotti, per esempio agricoli, che vengono seminati un anno prima dell'immissione sul mercato, l'offerta dipende dal prezzo dell'anno precedente. In tal caso l'unità di tempo è un anno. Siano, quindi,

$$\begin{cases} D_n = a - bp_n \\ O_n = -c + dp_{n-1} \end{cases}$$

rispettivamente, la funzione di domanda e la funzione di offerta (con  $a, b, c, d$  parametri positivi). All'equilibrio sarà  $D_n = O_n$ , ovvero

$$a - bp_n = -c + dp_{n-1}.$$

Risolvendo rispetto a  $p_n$  s'ottiene

$$p_n = -\frac{d}{b}p_{n-1} + \frac{a+c}{b},$$

equazione alle differenze del primo ordine lineare a coefficienti costanti, la cui soluzione è data (poiché il primo coefficiente è diverso da 1) dalla formula (3.3):

$$p_n = \left(-\frac{d}{b}\right)^n (p_0 - p^*) + p^*,$$

dove  $p_0$  indica il prezzo iniziale e  $p^* = \frac{a+c}{b+d}$  è il prezzo di equilibrio del mercato in condizioni statiche.

Osserviamo che se  $p_0 = p^*$ , la soluzione è costante e coincide con il prezzo di equilibrio:  $p_n = p^*$  per ogni  $n$ . Se  $p_0 \neq p^*$ , abbiamo una dinamica oscillante: il prezzo  $p_n$  oscilla attorno a  $p^*$ .

<sup>6</sup>Il parametro  $a$  rappresenta il mercato potenziale. Il parametro  $c$  può anche essere nullo.

L'andamento per  $n \rightarrow +\infty$  dipende dal rapporto  $d/b$ . Se  $d > b$  le oscillazioni s'amplificano, se  $d < b$  le oscillazioni si smorzano. Se  $d = b$  l'ampiezza delle oscillazioni rimane costante. La dinamica converge verso l'equilibrio se e solo se  $d < b$ , cioè se la pendenza della curva d'offerta è minore del modulo della pendenza della curva di domanda:

$$\lim_{n \rightarrow +\infty} p_n = p^* \quad \text{se e solo se } d < b$$

La dinamica del sistema può essere illustrata graficamente: si tracciano le due rette d'equazione

$$D = a - bp \quad \text{e} \quad O = -c + dp,$$

che rappresentano, rispettivamente, la funzione di domanda e la funzione d'offerta. Si colloca  $p_0$  sull'asse  $p$  e si determina l'ordinata  $O_1$  sulla retta di offerta. Spostandosi orizzontalmente sulla retta di domanda si determina  $D_1$  (essendo  $O_1 = D_1$ ): l'ascissa corrispondente è  $p_1$ . Trovato  $p_1$  si ripete il percorso precedente e si determinano  $O_2$ , poi  $D_2 = O_2$  e quindi  $p_2$ . Procedendo in questo modo, s'ottengono le rappresentazioni sull'asse  $p$  dei termini della successione  $p_n$ .

Per visualizzare la dinamica del sistema, si può "automatizzare" la costruzione geometrica, procedendo nel modo seguente. Si parte da  $p_0$  sull'asse  $p$ , ci si sposta in verticale sulla retta  $O$ , poi in orizzontale sulla retta  $D$ , poi in verticale su  $O$ , poi in orizzontale su  $D$  e così via. Nel caso di convergenza, si traccia un cammino a "ragnatela", che giustifica il nome del modello (v. fig. 7.2).

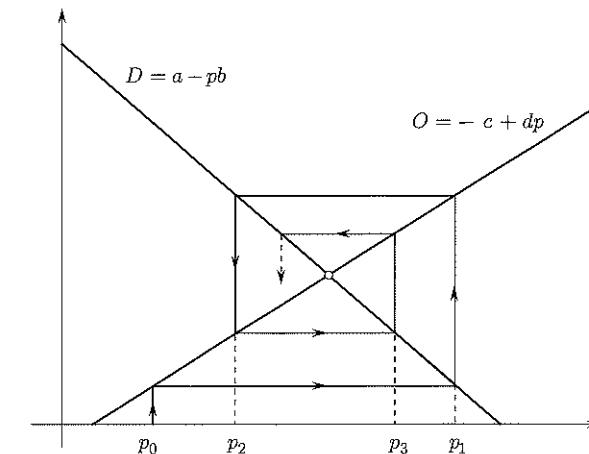


Figura 7.2. Modello della ragnatela.

## 4 EQUAZIONI AUTONOME NON LINEARI

### 4.1 Orbite, diagrammi a gradino e punti fissi o d'equilibrio

Tra i modelli discreti non lineari ci limitiamo a considerare equazioni alle differenze autonome del primo ordine del tipo:

$$(4.1) \quad x_{n+1} = f(x_n), \quad n \geq 0.$$

Assegnando  $x_0$ , attraverso la relazione di ricorrenza (4.1) è possibile calcolare  $x_1 = f(x_0)$ ,  $x_2 = f(x_1), \dots$  e ottenere così la successione  $\{x_n\}$ , che prende il nome di *orbita uscente da  $x_0$* .

La funzione  $f$  si chiama *generatrice* e, affinché la successione  $x_n$  sia ben definita, occorre che

$$f : I \rightarrow I$$

dove  $I \subseteq \mathbb{R}$  è un intervallo. Infatti, la (4.1) richiede ogni volta di calcolare  $f$  su uno dei suoi *output*, secondo la sequenza di calcolo raffigurata sotto:

$$x_0 \rightarrow \boxed{f} \rightarrow x_1 \rightarrow \boxed{f} \rightarrow x_2 \rightarrow \cdots \rightarrow x_n \rightarrow \boxed{f} \rightarrow x_{n+1} \cdots$$

Assumeremo che  $f$  sia *continua* in  $I$ . Dato che

$$\begin{aligned} x_1 &= f(x_0), \\ x_2 &= f(x_1) = f(f(x_0)) = (f \circ f)(x_0), \\ x_3 &= f(x_2) = f(f(f(x_0))) = (f \circ f \circ f)(x_0), \\ &\vdots \end{aligned}$$

indicando con  $f^n$  l'iterata  $n$ -esima di  $f$  cioè  $f^n = \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ volte}}$ , si ha

$$x_n = f^n(x_0)$$

che evidenzia il punto di partenza dell'orbita.

**OSSERVAZIONE** Sottolineiamo che il simbolo  $f^n$  indica l'iterata  $n$ -esima di  $f$ . Occorre evitare confusioni tra

$$f^n(x) \quad \text{e} \quad [f(x)]^n.$$

L'andamento di un'orbita si può efficacemente visualizzare tramite il cosiddetto *diagramma di fase* o *diagramma a gradino*. Vediamo come si costruisce tale diagramma per l'equazione

$$x_{n+1} = f(x_n).$$

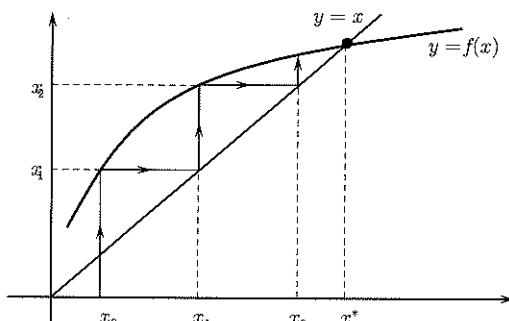


Figura 7.3. Diagramma a gradino.

Disegniamo il grafico di  $f$  e la bisettrice  $y = x$  e, in riferimento alla figura 7.3, cerchiamo di visualizzare l'orbita uscente da  $x_0$ .

Il punto sul grafico di  $f$  con ascissa  $x_0$  ha ordinata  $f(x_0) = x_1$ , il successivo valore assunto dalla variabile di stato. Se sfruttiamo la bisettrice possiamo "riportare"  $x_1$  sull'asse delle ascisse ed individuare  $x_2 = f(x_1)$ .

Procedendo in questo modo, connettendo ogni volta con un segmento i punti di coordinate

$$(x_0, x_1), (x_1, x_1), (x_1, x_2), \dots, (x_n, x_{n+1}), (x_{n+1}, x_{n+1}), \dots$$

si visualizza l'orbita mediante una spezzata.

Dopo un po' d'esercizio la costruzione diventa automatica: si prende un punto iniziale sull'asse  $x$ , si cerca la corrispondente ordinata sulla curva, da lì ci si sposta orizzontalmente verso la bisettrice, poi si va verticalmente alla curva, poi alla bisettrice e così via.

I punti come  $x^*$  in figura sono particolarmente importanti. Infatti se il punto iniziale  $x_0$  coincide con  $x^*$ , l'intera orbita coincide con  $x^*$  e genera perciò una successione costante, ossia una soluzione di equilibrio. Si dice anche che  $x^*$  è *punto di equilibrio* o *punto fisso*. L'ultima locuzione è dovuta al fatto che

$$f(x^*) = x^*$$

e cioè il punto  $x^*$  rimane *fisso* sotto l'azione di  $f$ .

Il diagramma ottenuto in tal modo, evidenziando i punti fissi e tracciando le spezzate che visualizzano l'andamento delle orbite, prende il nome di *diagramma a gradino*.



Nel modello logistico

$$x_{n+1} = rx_n(1 - x_n) \quad (r > 0)$$

abbiamo

$$f(x) = rx(1 - x)$$

il cui grafico è una parabola. I punti fissi si trovano risolvendo l'equazione

$$rx(1 - x) = x$$

da cui

$$x_1^* = 0 \quad \text{e} \quad x_2^* = 1 - \frac{1}{r}.$$

## 4.2 Punti fissi e stabilità

Dai diagrammi a gradino si possono ricavare informazioni utili sul comportamento asintotico delle orbite. Sempre nel caso in figura 7.3, si vede che l'orbita uscente da  $x_0$  converge al punto fisso  $x^*$ . Provando con altri punti di partenza, anche alla destra dell'equilibrio, s'osserva lo stesso fenomeno<sup>7</sup>:  $x^*$  è *asintoticamente stabile*. Precisamente, introduciamo i seguenti concetti.

<sup>7</sup>Invitiamo il lettore a controllare.

**DEFINIZIONE 7.1 (STABILITÀ, INSTABILITÀ E STABILITÀ ASINTOTICA)** Il punto di equilibrio  $x^*$  si dice:

- a) *stabile* se: per ogni  $\varepsilon > 0$  esiste  $\delta = \delta_\varepsilon$  tale che se  $|x_0 - x^*| < \delta$  e  $x_n$  è l'orbita uscente da  $x_0$ , si ha

$$|x_n - x^*| < \varepsilon \quad \text{per ogni } n \geq 0.$$

- b) *asintoticamente stabile* se è stabile e, inoltre, esiste  $\delta_1$  tale che, se  $|x_0 - x^*| < \delta_1$  e  $x_n$  è l'orbita uscente da  $x_0$ , si ha

$$x_n \rightarrow x^* \quad \text{per } n \rightarrow +\infty.$$

- c) *instabile* se non è stabile (se cioè non vale la condizione (a)). ■

Nel caso in cui  $x^*$  sia asintoticamente stabile, dalla definizione si deduce che esiste un intorno  $U$  di  $x^*$  tale che ogni orbita uscente da  $U$  converge a  $x^*$ . Il più grande insieme  $B \supseteq U$  con la proprietà che ogni orbita uscente da  $B$  converga a  $x^*$ , si chiama *bacino di attrazione* di  $x^*$ .



**Il caso lineare.** Nel caso lineare

$$x_{n+1} = ax_n + b, \quad a, b \in \mathbb{R}, a \neq 0,$$

se  $a \neq 1$ , l'unico punto di equilibrio è

$$x^* = \frac{b}{1-a}.$$

Dato che la soluzione generale è

$$x_n = ca^n + \frac{b}{1-a}, \quad c \in \mathbb{R}$$

si ricava facilmente che per ogni altra soluzione (non costante), si ha:

$$\lim_{n \rightarrow +\infty} x_n = \begin{cases} \frac{b}{1-a} & \text{se } |a| < 1 \\ \pm\infty (\text{col segno di } c) & \text{se } a > 1 \\ \text{non esiste} & \text{se } a \leq -1 \end{cases}$$

quindi il punto di equilibrio è (globalmente) asintoticamente stabile se e solo se  $|a| < 1$ .

Se  $a = 1$  e  $b \neq 0$  si ha

$$x_{n+1} = x_n + b$$

che non ha punti fissi.

Se  $a = 1$  e  $b = 0$  l'equazione è

$$x_{n+1} = x_n$$

ed ogni punto è d'equilibrio.

I diagrammi a gradino confermano i risultati trovati poco sopra (v. fig. 7.4).

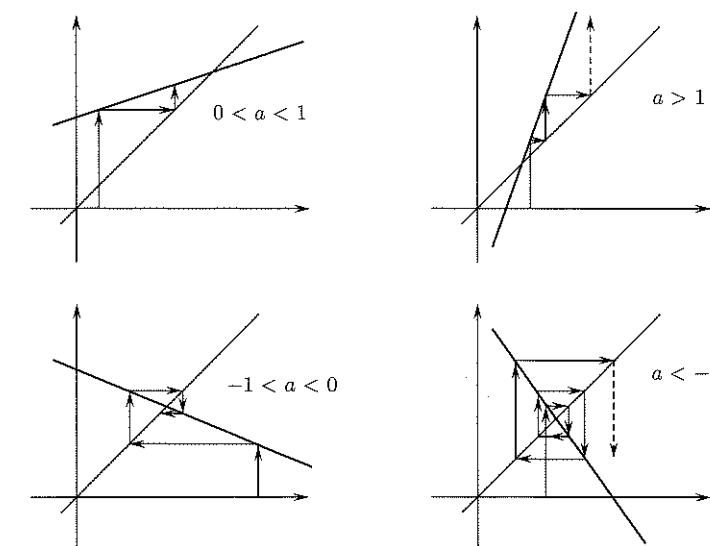


Figura 7.4. Diagrammi a gradino nel caso lineare.

### 4.3 Stabilità per linearizzazione

In qualche caso, attraverso il diagramma a gradino si può capire l'andamento asintotico di  $x_n$ , anche senza conoscere l'espressione analitica di  $f$ . A volte bastano proprietà qualitative del grafico per convincerci che se il grafico di  $f$  non è "tropo ripido" vicino al punto di equilibrio, quest'ultimo è asintoticamente stabile. "Molto ripido" o "poco ripido" s'intendono rispetto alla bisettrice, che ha pendenza unitaria, quindi è decisivo il confronto del modulo della pendenza del grafico di  $f$  con la pendenza 1. Sostanzialmente l'idea è riportarsi al caso lineare sostituendo il grafico di  $f$  con la retta tangente ad esso nel punto fisso  $(x^*, f(x^*))$ .

Questa tecnica prende perciò il nome di *linearizzazione*. Precisamente, vale il seguente criterio:

**TEOREMA 7.1** Se  $f : I \rightarrow \mathbb{R}$  è derivabile con continuità in  $I$ ,  $f(x^*) = x^*$  e  $|f'(x^*)| \neq 1$ , allora:

se  $|f'(x^*)| < 1$ ,  $x^*$  è punto di equilibrio asintoticamente stabile;  
se  $|f'(x^*)| > 1$ ,  $x^*$  è punto di equilibrio instabile.

**DIMOSTRAZIONE.** Se  $|f'(x^*)| < 1$ , essendo  $f'$  continua, per il teorema della permanenza del segno, in un intorno  $I_\delta = [x^* - \delta, x^* + \delta]$  di  $x^*$  si ha

$$|f'(x)| \leq k < 1.$$

Poiché  $f(I_\delta) \subseteq I_\delta$  (il lettore lo dimostri!), se  $x_0 \in I_\delta$ , anche  $x_n \in I_\delta$ . Valutiamo la distanza tra  $x_n$  e  $x^*$ ; dall'equazione

$$x_{n+1} = f(x_n)$$

e dal fatto che  $f(x^*) = x^*$  si ha

$$|x_{n+1} - x^*| = |f(x_n) - f(x^*)|.$$

Usando il teorema del valor medio di Lagrange e il fatto che  $|f'| \leq k$ , possiamo scrivere

$$|f(x_n) - f(x^*)| = |f'(c_n)(x_n - x^*)| \leq k|x_n - x^*|$$

da cui

$$|x_{n+1} - x^*| \leq k|x_n - x^*|.$$

Rifacendo gli stessi calcoli per valutare  $|x_n - x^*|$  si trova

$$|x_n - x^*| \leq k|x_{n-1} - x^*| \leq k^2|x_{n-2} - x^*| \dots \leq k^n|x_0 - x^*|.$$

Procedendo ancora in questo modo si arriva a

$$|x_n - x^*| \leq k|x_{n-1} - x^*| \leq k^2|x_{n-2} - x^*| \leq \dots \leq k^n|x_0 - x^*|.$$

Essendo  $k < 1$  si conclude che  $|x_n - x^*| \rightarrow 0$ , ossia  $x_n \rightarrow x^*$ , purché  $x_0 \in I$ .

Se  $|f'(x^*)| > 1$ , si ha, in un intorno  $I$ ,

$$|f'(x)| \geq \rho > 1$$

e quindi se  $x_0 \in I$ , si ha, applicando il teorema del valor medio,

$$|x_1 - x^*| = |f(x_0) - f(x^*)| = |f'(c)||x_0 - x^*| \geq \rho|x_0 - x^*|;$$

se anche  $x_1 \in I$ , con lo stesso ragionamento, si trova

$$|x_2 - x^*| \geq \rho^2|x_0 - x^*|$$

e dopo un numero finito di passi si esce da  $I$ . Cioè  $x^*$  è instabile<sup>8</sup>



Studiamo il carattere della successione per ricorrenza

$$\begin{cases} s_{n+1} = 3 - e^{-s_n} \\ s_0 = a. \end{cases}$$

La funzione generatrice della successione è

$$f(x) = 3 - e^{-x}.$$

Come mostra il grafico,  $f$  ha due punti fissi  $\alpha < 0$  e  $\beta$  ( $2 < \beta < 3$ ). Dato che  $f'(x) = e^{-x}$ , si ha

$$f'(\alpha) > 1, \quad 0 < f'(\beta) < 1.$$

Dunque  $\alpha$  è punto fisso instabile (repulsore),  $\beta$  è punto fisso (localmente) asintoticamente stabile.

Se  $a < \alpha$ , la successione  $s_n$  è decrescente<sup>9</sup> e tende a  $-\infty$ , se  $\alpha < a < \beta$  è crescente e tende a  $\beta$ , se  $a > \beta$  è decrescente e tende a  $\beta$ . Ovviamente se  $a = \alpha$  o  $a = \beta$ ,  $s_n$  è costante. La figura 7.5 mostra che il bacino d'attrazione del punto  $\beta$  è  $(\alpha, +\infty)$ .

<sup>8</sup>Prima o poi, tutte le orbite che hanno origine in  $I$  escono da  $I$ , allontanandosi da  $x^*$ . In tal caso si dice che  $x^*$  è repulsore.

<sup>9</sup>Se la funzione generatrice è crescente, la successione per ricorrenza è monotona (vedi l'esercizio 6 alla fine del par. 4).

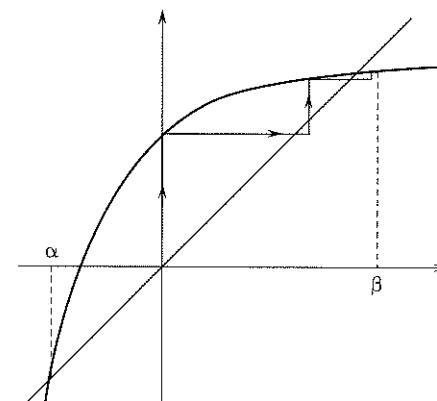


Figura 7.5.

Il criterio di stabilità/instabilità per linearizzazione vale solo per punti  $x^*$  di equilibrio in cui  $|f'(x^*)|$  è maggiore o minore di 1. Punti di equilibrio di questo tipo si chiamano iperbolici.

Se un punto di equilibrio  $x^*$  non è iperbolico, il lettore controlli attraverso i diagrammi di fase che si verificano i seguenti casi:

se  $f'(x^*) = 1$  ed  $f$  taglia la bisettrice da sopra a sotto,  $x^*$  è asintoticamente stabile, mentre se  $f$  taglia la bisettrice da sotto a sopra,  $x^*$  è instabile. Se il grafico di  $f$  si trova al di sopra (al di sotto) della bisettrice,  $x^*$  attrae solo le orbite che partono da punti a sinistra (a destra) di  $x^*$  stesso. In tal caso si può dire che il punto è semi-stabile (v. fig. 7.6).

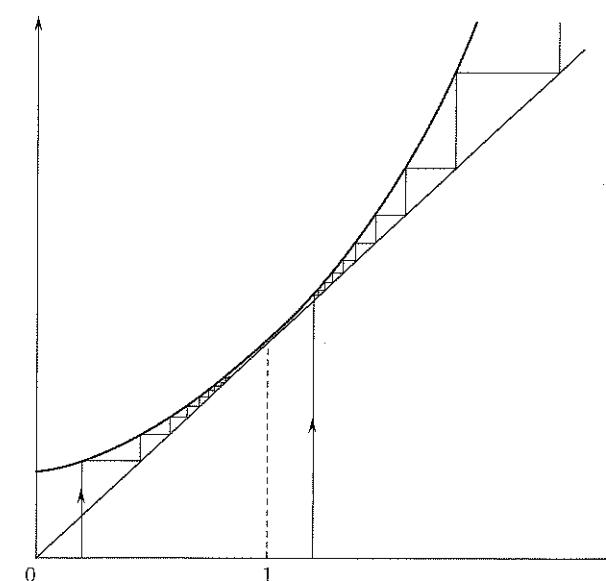


Figura 7.6. Punto semi-stabile.

#### 4.4 Orbite periodiche e stabilità

Le *orbite periodiche* sono del tipo

$$x_0, x_1, \dots, x_{p-1}, \boxed{x_p = x_0}, x_1, \dots, x_{p-1}, x_0, \dots$$

dove i primi  $p$  punti sono tutti diversi tra loro ( $x_0 \neq x_1 \neq \dots \neq x_{p-1}$ ), mentre  $x_p = x_0$  e si ricomincia da capo. I loro punti si dicono *punti periodici* e il numero  $p$  si chiama *periodo* dell'orbita. Osserviamo che

$$x_0 = x_p = f^p(x_0)$$

quindi  $x_0$  è *punto fisso per l'iterata  $f^p$* . Nello stesso modo si verifica che anche  $x_1, \dots, x_{p-1}$  sono punti fissi per  $f^p$ :

$$x_j = f^p(x_j) \quad \text{per ogni } j = 1, \dots, p-1.$$

Per esempio, un'*orbita di periodo 2* è un insieme di due punti

$$\{x^*, y^*\}$$

entrambi fissi per l'iterata seconda  $f^2$  e si ha

$$f(x^*) = y^*, \quad f(y^*) = x^*$$



**REMARK** La successione  $\left\{2, \frac{1}{2}, 2, \frac{1}{2}, 2, \frac{1}{2}, \dots\right\}$  è un'orbita di periodo 2 per l'equazione

$$x_{n+1} = \frac{1}{x_n}.$$

Il diagramma a gradino di un'orbita di periodo 2 è particolarmente semplice (v. fig. 7.7).

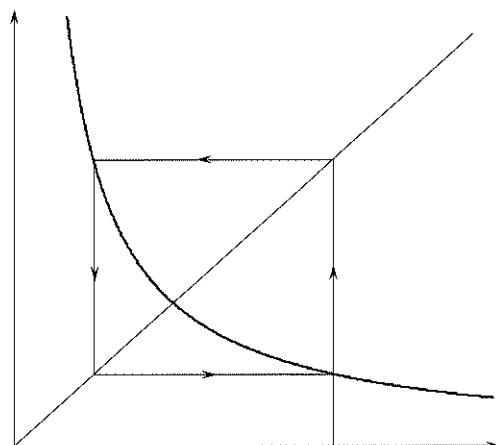


Figura 7.7. Orbita di periodo 2.

Anche per le orbite periodiche si può parlare di stabilità, stavolta con l'idea che se si parte abbastanza vicino ad uno dei suoi punti si continua a rimanere vicino, non necessariamente a quel punto, ma all'orbita.

Precisamente, consideriamo l'equazione (4.1) (con  $f$  derivabile con continuità in un intervallo  $I \subseteq \mathbb{R}$ ). Sia  $\Gamma_p$  un'orbita periodica di periodo  $p$ , costituita dai punti

$$w_0, w_1 = f(w_0), \dots, w_{p-1} = f^{p-1}(w_0).$$

Per *distanza di un punto  $z$  da  $\Gamma_p$*  intendiamo la più piccola tra le distanze di  $z$  da un punto di  $\Gamma$ :

$$\text{dist}(z, \Gamma_p) = \min \{|z - w_0|, |z - w_1|, \dots, |z - w_{p-1}|\}.$$

**DEFINIZIONE 7.2 (STABILITÀ, INSTABILITÀ E STABILITÀ ASINTOTICA)** L'orbita  $\Gamma_p$  si dice:

a) *stabile* se: per ogni  $\varepsilon > 0$  esiste un insieme aperto  $A$  che contiene  $\Gamma_p$  e tale che, se  $x_0 \in A$  e  $x_n$  è l'orbita uscente da  $x_0$ , si ha

$$\text{dist}(x_n, \Gamma_p) < \varepsilon \quad \text{per ogni } n \geq 0.$$

b) *asintoticamente stabile* se è stabile e, inoltre, esiste un insieme  $A_1$ , costituito da un'unione di intervalli aperti, tale che  $\Gamma_p \subset A_1$  e, se  $x_0 \in A_1$  si ha

$$\text{dist}(x_n, \Gamma_p) \rightarrow 0 \quad \text{per } n \rightarrow +\infty.$$

c) *instabile* se non è stabile (se cioè non vale la condizione (a)).

Usare la definizione in generale non è semplice. C'è però un modo piuttosto efficace per assicurarsi che un'orbita periodica sia asintoticamente stabile.

Sappiamo infatti che ciascuno dei punti  $w_j$  ( $j = 0, \dots, p-1$ ) è punto fisso per l'iterata  $f^p$ . Notiamo che la *derivata prima dell'iterata  $f^p$  è uguale in tutti i punti  $w_j$* , infatti si ha

$$(f^p)'(x) = f'(x) \cdot f'(f(x)) \cdots \cdot f'(f^{p-2}(x)) \cdot f'(f^{p-1}(x))$$

e in  $w_j$ :

$$(f^p)'(w_j) = f'(w_j) \cdot f'(w_{j+1}) \cdots \cdot f'(w_{j+p-1}) = \prod_{j=0}^{p-1} f'(w_j).$$

Si deduce che i punti dell'orbita sono *simultaneamente tutti iperbolici* oppure *tutti non iperbolici*. Pertanto se in uno dei punti  $w_j$  (basta controllarne uno qualsiasi!) si ha

$$|(f^p)'(w_j)| < 1$$

allora tutti i punti dell'orbita sono asintoticamente stabili per  $f^p$  e quindi l'*orbita stessa è asintoticamente stabile*. Analogamente, se in uno dei punti  $w_j$  si ha

$$|(f^p)'(w_j)| > 1$$

allora tutti i punti dell'orbita sono instabili e quindi l'*orbita stessa è instabile*.

## 4.5 Esistenza di orbite periodiche

La presenza di orbite periodiche rivela un andamento ciclico del sistema dinamico ed è piuttosto importante stabilire condizioni che ne garantiscano l'esistenza e, se possibile, la loro natura stabile/instabile. I due teoremi seguenti danno una risposta parziale al problema. Il primo si applica a funzioni generatrici cosiddette *unimodali*, che presentano cioè un unico punto stazionario nell'intervallo di definizione (come per la logistica).

**TEOREMA 7.2 (DI SINGER)** *Sia*

$$f : [0, 1] \rightarrow [0, 1]$$

*continua con le sue derivate fino alla terza inclusa. Se*

1.  $f(0) = f(1) = 0$ ,
2.  $f$  ha un solo punto  $x_0 \in (0, 1)$  tale che

$$f'(x_0) = 0,$$

3. la quantità

$$\{f, x\} = \frac{f'''(x)}{f'(x)} - \frac{3}{2} \left( \frac{f''(x)}{f'(x)} \right)^2$$

*è negativa in  $[0, 1]$ , allora l'equazione alle differenze*

$$x_{n+1} = f(x_n)$$

*ammette al massimo un'unica orbita periodica asintoticamente stabile, contenuta in  $(0, 1)$ . Se tale orbita esiste,  $x_0$  è nel suo bacino d'attrazione.*

La quantità  $\{f, x\}$  si chiama *derivata di Schwarz di f* (la definizione risale al 1869!). Il fatto che  $x_0$  appartenga al bacino di attrazione dell'eventuale orbita periodica è un'informazione preziosa qualora si disponga di un computer per individuare tale orbita. In particolare, il teorema di Singer si applica al caso dell'equazione logistica, come vedremo più avanti.

Il secondo teorema, interessante se non sorprendente per la sua originalità, richiede l'introduzione di uno speciale ordinamento dei numeri naturali, detto *ordinamento di Sharkowsky*. Prima si collocano i numeri dispari, poi il doppio di tutti i dispari, poi il quadruplo di tutti i dispari e così via. Infine si collocano le potenze di due in ordine decrescente. Se utilizziamo il simbolo  $\blacktriangleleft$  per indicare che, nella formula

$$m \blacktriangleleft n$$

*n segue m*, l'ordinamento si può scrivere nel modo seguente:

$$\begin{aligned} 3 &\blacktriangleleft 5 \blacktriangleleft 7 \blacktriangleleft \cdots \blacktriangleleft 2 \cdot 3 \blacktriangleleft 2 \cdot 5 \blacktriangleleft 2 \cdot 7 \blacktriangleleft \cdots \blacktriangleleft 2^2 \cdot 3 \blacktriangleleft 2^2 \cdot 5 \blacktriangleleft 2^2 \cdot 7 \blacktriangleleft \cdots \\ &\cdots \blacktriangleleft 2^3 \cdot 3 \blacktriangleleft 2^3 \cdot 5 \blacktriangleleft 2^3 \cdot 7 \blacktriangleleft \cdots \blacktriangleleft 2^4 \blacktriangleleft 2^3 \blacktriangleleft 2^2 \blacktriangleleft 2 \blacktriangleleft 1 \end{aligned}$$

**TEOREMA 7.3 (DI SHARKOWSKY)** *Sia  $f : \mathbb{R} \rightarrow \mathbb{R}$  continua. Se l'equazione alle differenze*

$$x_{n+1} = f(x_n)$$

*ammette un'orbita periodica di periodo  $m$ , allora, per ogni  $n$  che segue  $m$  nell'ordinamento di Sharkowsky esiste un'orbita periodica di periodo  $n$ .*

Non si può evitare di sottolineare una (clamorosa!) conseguenza del teorema: *se esiste un'orbita di periodo 3 esistono orbite di qualunque periodo*. Il periodo 3 è sinonimo di ... caos, come vedremo.

Non solo: si può esibire un esempio di funzione generatrice continua il cui sistema dinamico corrispondente ammette un'orbita di periodo 5 ma non di periodo 3, a conferma della profondità del teorema. Si può usare il teorema anche per mostrare che *non* possono esistere orbite di un certo periodo. Per esempio, consideriamo la logistica con

$$f(x) = 3,2x(1-x)$$

e disegniamo i grafici di  $f$ , dell'iterata seconda  $f^2$  e dell'iterata quarta  $f^4$  (v. fig. 7.8).

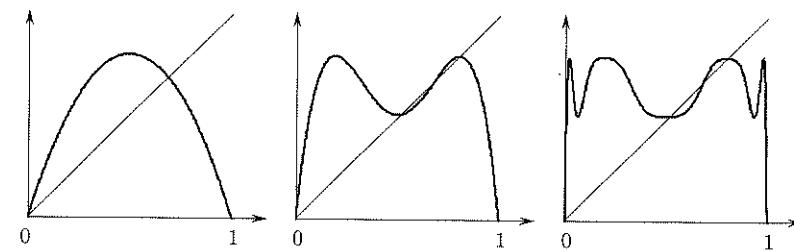


Figura 7.8. Grafici di  $f$ ,  $f^2$ ,  $f^4$ .

Il primo grafico mostra i due punti fissi 0 e 0,7.... Nel secondo, osserviamo ancora i due punti fissi di  $f$  e altri due punti fissi di  $f^2$  che costituiscono un'orbita periodica di periodo 2. Nel terzo i punti fissi coincidono con quelli di  $f^2$ . Ne segue che *non* vi sono orbite di periodo 4 e quindi, per il teorema di Sharkowsky, non possono esistere orbite periodiche di periodo maggiore di 2.

## 4.6 Comportamento caotico

Accenniamo brevemente al concetto di *caos deterministico*. La parola *caos* è collegata spesso a mancanza di regolarità, di ordine o di qualche correlazione. Da un punto di vista dinamico, caos significa impossibilità di misurare, osservare o prevedere le variazioni di un sistema in evoluzione, entrando, inevitabilmente, in una condizione di aleatorietà. In realtà vi sono vari tipi di caos e quello che tratteremo qui, brevemente, non si riferisce all'osservazione di fenomeni aleatori (come il lancio di un dado) bensì a sistemi che evolvono secondo *leggi deterministiche*, di cui le equazioni che stiamo considerando costituiscono un esempio.

Nel nostro caso parleremo di *caos deterministico* e della conseguente *dinamica caotica*. Uno dei problemi che si presenta è caratterizzare in modo rigoroso che cosa si debba intendere con ciò. La questione non è così semplice, tant'è vero che esistono varie definizioni più o meno soddisfacenti di comportamento caotico. Adottiamo quella di Li e Yorke.

Sia  $f : I \rightarrow I$ . L'idea è che lo stato caotico sia determinato dall'esistenza di un insieme non numerabile  $S \subset I$  con la proprietà seguente: per ogni coppia di punti  $p, q \in S$ , esistono due successioni di iterate  $f^{n_j}$  e  $f^{m_k}$  tali che

$$\lim_{n_j \rightarrow +\infty} |f^{n_j}(p) - f^{n_j}(q)| = 0$$

e

$$\lim_{m_k \rightarrow +\infty} |f^{m_k}(p) - f^{m_k}(q)| = \delta > 0.$$

In altri termini, le orbite  $f^n(p)$  e  $f^n(q)$ , uscenti da punti qualunque  $p$  e  $q$  in  $S$ , hanno un andamento erratico, nel senso che per qualche indice  $n$  sono molto vicine mentre per certi altri si mantengono a distanza positiva. Naturalmente ciò significa che la dinamica esibisce estrema sensibilità rispetto alle condizioni iniziali: una piccolissima variazione in queste ultime provoca, in generale, notevoli differenze nelle orbite corrispondenti. Se una dinamica è caotica, le previsioni sul suo andamento eseguite al computer non possono essere credibili!

Un risultato ormai classico è il seguente:

**TEOREMA 7.4 (DI LI E YORKE)** *Periodo tre implica caos.*

L'equazione logistica

$$x_{n+1} = 4x_n(1-x_n)$$

fornisce un importante e classico esempio di dinamica caotica. La esaminiamo in qualche dettaglio nella prossima sezione.

#### 4.7 Equazione logistica discreta

In questa sezione analizziamo il comportamento asintotico delle orbite dell'equazione logistica

$$(4.2) \quad x_{n+1} = rx_n(1-x_n), \quad r > 0.$$

La funzione generatrice è

$$f(x) = rx(1-x)$$

Se limitiamo, come faremo, i valori del parametro  $r$  nell'intervallo  $(0, 4]$ , si ha (il lettore controlli)  $f([0, 1]) \subseteq [0, 1]$  ossia  $f : [0, 1] \rightarrow [0, 1]$ .

Osserviamo subito che  $f$  soddisfa le ipotesi del teorema di Singer; infatti,  $f(0) = f(1) = 0$ , ha un solo punto stazionario  $x = 1/2$  e

$$\{f, x\} = \frac{f'''(x)}{f'(x)} - \frac{3}{2} \left( \frac{f''(x)}{f'(x)} \right)^2 = -\frac{3}{2} \left( \frac{-2}{1-2x} \right)^2 < 0$$

in  $[0, 1]$ . Deduciamo, in base al teorema di Singer, che per ogni valore di  $r \in (0, 4]$  esiste al massimo un'orbita periodica in  $(0, 1)$ , asintoticamente stabile e  $x_0 = 1/2$  è nel suo bacino d'attrazione.

La dinamica della (4.2) presenta una insospettabile dipendenza dal parametro  $r$  e un'interessante varietà di fenomeni. Come abbiamo visto nel paragrafo 1.2, i punti fissi, soluzioni dell'equazione

$$rx(1-x) = x$$

sono

$$x_1 = 0 \quad \text{e} \quad x_2 = 1 - \frac{1}{r}.$$

Distinguiamo ora vari casi.

$$0 < r \leq 1$$

Solo  $x_1 = 0$  sta nell'intervallo  $[0, 1]$  e risulta asintoticamente stabile con bacino d'attrazione  $[0, 1]$ , infatti per  $0 < r < 1$  si ha

$$f'(x) = r - 2rx$$

e

$$f'(0) = r \in (0, 1).$$

Si ha stabilità asintotica anche per  $r = 1$ , come mostra il diagramma a gradino (v. fig. 7.9).

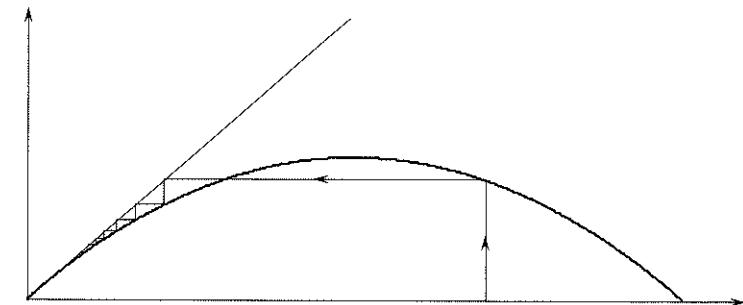


Figura 7.9. Mappa logistica per  $r = 1$ .

Si osservi che se  $0 < r < 1$  il punto  $x_1 = 0$  risulterebbe (localmente) asintoticamente stabile anche considerando  $f : \mathbb{R} \rightarrow \mathbb{R}$ , mentre per  $r = 1$  è *semistabile* (le orbite che partono a sinistra di 0, si allontanano).

$$1 < r \leq 3$$

Anche  $x_2 \in [0, 1]$  e si ha

$$f'\left(1 - \frac{1}{r}\right) = r - 2r\left(1 - \frac{1}{r}\right) = 2 - r$$

e

$$|2 - r| < 1 \quad \text{per} \quad 1 < r < 3.$$

Il punto  $x_1 = 0$  è instabile, mentre  $x_2$  è asintoticamente stabile con bacino di attrazione  $(0, 1)$ . Anche per  $r = 3$ ,  $x_2$  è asintoticamente stabile.

$$3 < r \leq 1 + \sqrt{6} = 3,449499\dots$$

Per  $r > 3$  i punti  $x_1$  e  $x_2$  sono instabili. Dal grafico di  $f$  non si vede di più.

Cerchiamo orbite periodiche di periodo 2 considerando l'iterata seconda di  $f$ :

$$f^2(x) = r^2x(1-x)(1-rx(1-x)).$$

I punti periodici di periodo 2 corrispondono ai punti fissi di  $f^2$  e si trovano risolvendo l'equazione

$$r^2x(1-x)(1-rx(1-x)) = x.$$

Oltre (ovviamente) a  $x_1 = 0$  e  $x_2 = 1 - \frac{1}{r}$ , si trovano i punti

$$x_{3,4} = \frac{1+r \pm \sqrt{r^2 - 2r - 3}}{2r}$$

soluzioni dell'equazione

$$r^2x^2 - r(r+1)x + r+1 = 0.$$

Per studiarne la stabilità calcoliamo

$$(f^2)'(x_3) = (f^2)'(x_4) = f'(x_3)f'(x_4) = -r^2 + 2r + 4.$$

Si ha che

$$|-r^2 + 2r + 4| < 1$$

per  $3 < r < 1 + \sqrt{6}$  e quindi, per tali valori del parametro, esiste un'orbita periodica asintoticamente stabile di periodo 2. Il bacino d'attrazione dell'orbita  $\{x_3, x_4\}$  è l'intervallo  $(0, 1)$ , tranne un'infinità numerabile di punti che arriva a  $x_2$  dopo un numero finito di iterazioni. Dal grafico di  $f^2$  si intuisce che lo stesso accade per  $r = r_2 = 1 + \sqrt{6}$ .

In conclusione, quando  $r$  supera il valore  $r_1 = 3$  il punto fisso  $x_2$  diventa instabile e si crea un'orbita periodica asintoticamente stabile di periodo 2. Questo cambio di comportamento delle orbite prende il nome di *biforcazione*. La figura 7.10 mostra,

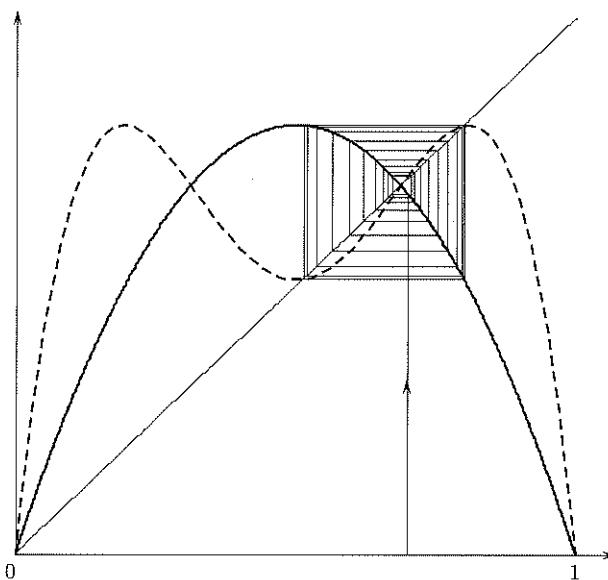


Figura 7.10. Mappa logistica per  $r = 3,2$ .

in corrispondenza del valore  $r = 3,2$ , l'orbita uscente dal punto  $x_0 = 0,7$ ; la curva tratteggiata è il grafico dell'iterata seconda  $f^2$ .

$$1 + \sqrt{6} < r \leq r_\infty = 3,5699456\dots$$

Superato il valore  $r_2 = 1 + \sqrt{6}$ , l'orbita di periodo 2 diventa instabile e si crea un'orbita stabile di periodo 4. Ciò si vede esaminando l'iterata quarta  $f^4$ . Per  $r = r_2$  si verifica quindi un'altra biforcazione con raddoppio del periodo. La figura 7.11 mostra, in corrispondenza del valore  $r = 3,52$ , l'orbita uscente dal punto  $x_0 = 0,6$ ; la curva tratteggiata è il grafico dell'iterata quarta  $f^4$ .

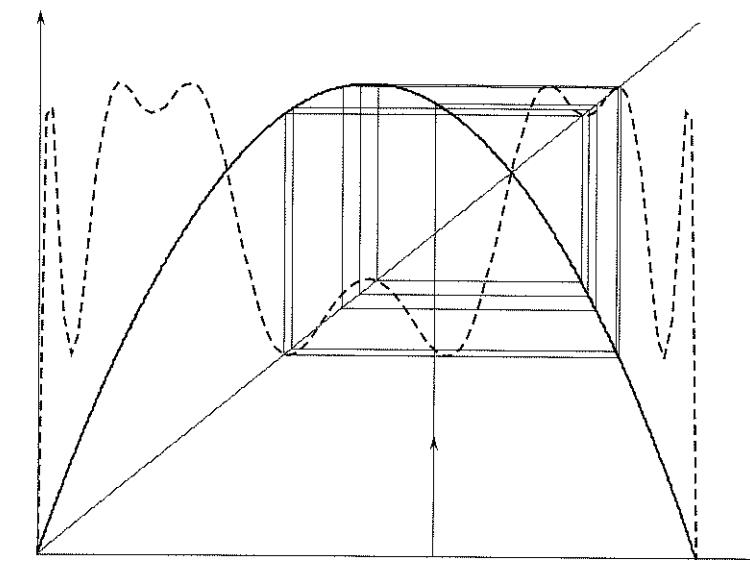


Figura 7.11. Mappa logistica per  $r = 3,52$ .

L'orbita di periodo 4 rimane stabile fino ad un valore  $r_3 = 3,544090\dots$ , in corrispondenza al quale si verifica un'altra biforcazione con raddoppio del periodo; si crea, cioè, un'orbita asintoticamente stabile di periodo 8.

Successivamente, al crescere di  $r$ , si verifica una sequenza infinita di biforcazioni con raddoppio del periodo in corrispondenza ad infiniti valori  $r_k$ . I primi otto sono

$$\begin{aligned} r_1 &= 3 & r_2 &= 3,449499\dots & r_3 &= 3,544090\dots & r_4 &= 3,564407\dots \\ r_5 &= 3,568759\dots & r_6 &= 3,569692\dots & r_7 &= 3,569891\dots & r_8 &= 3,569934\dots \end{aligned}$$

I valori  $r_k$  convergono ad un valore

$$r_\infty = 3,5699456\dots$$

Sperimentalmente, si è osservato che  $r_n = r_\infty - \text{costante} \cdot \delta^{-n}$  ( $n \gg 1$ ) dove  $\delta = 4,6692016091\dots$  è detta *costante di Feigenbaum*.

In corrispondenza a  $r_\infty$  esiste un insieme infinito  $F$  che attrae i punti di  $(0, 1)$ , tranne i punti periodici (instabili) e i punti che arrivano ai punti instabili dopo un numero finito di iterazioni. L'insieme  $F$  ha struttura *frattale*. L'insieme  $1 < r < r_\infty$  definisce il *regime di biforcazione* per la mappa logistica.

$$r_\infty < r \leq 4$$

Per  $r > r_\infty$  si entra nel *regime caotico* ed il comportamento è molto più complicato. Esistono orbite periodiche instabili di periodo  $2^n$  (per ogni  $n$  ce n'è una), orbite stabili (di periodo  $3 \cdot 2^n, 5 \cdot 2^n$  ecc.) e anche orbite aperiodiche.

Per  $r > 3,67857\dots$  cominciano a presentarsi orbite di periodo dispari. In particolare, per  $r = 1 + \sqrt{8}$  esiste un'unica orbita di periodo 3 asintoticamente stabile e, per il teorema di Sharkovsky, orbite di periodo qualunque.

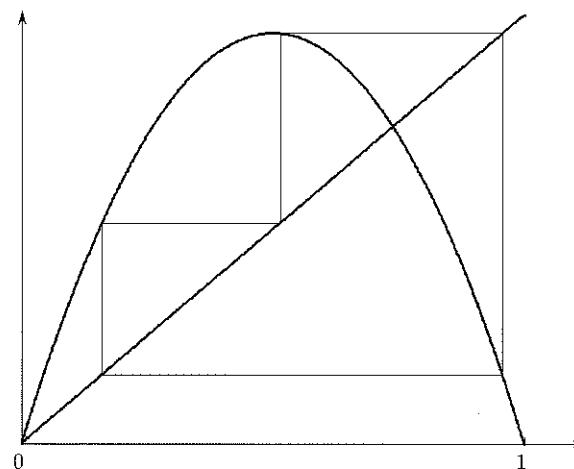


Figura 7.12. Orbita di periodo 3.

Come abbiamo già visto, il regime caotico presenta forte instabilità rispetto alle condizioni iniziali. La figura 7.13 mostra i grafici di due soluzioni dell'equazione logistica in corrispondenza del valore  $r = 4$ . Le due soluzioni, pur partendo da punti molto vicini ( $x_1 = 0,71$  e  $x_1 = 0,72$ ) hanno un andamento assai differente.

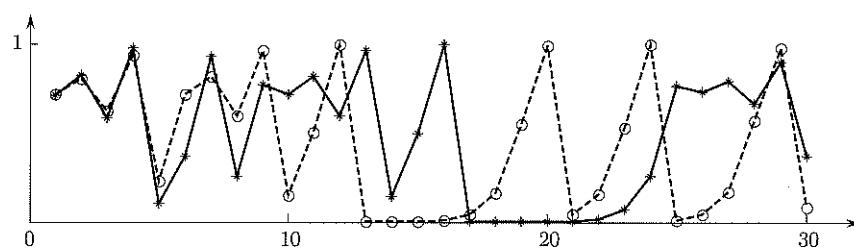


Figura 7.13. Grafici di due soluzioni dell'equazione logistica per  $r = 4$ .



**Esercizio** Scrivere la soluzione generale delle seguenti equazioni alle differenze lineari del primo'ordine.

$$s_{n+1} = -\frac{s_n}{3}, \quad s_{n+1} = -2s_n + 1.$$

Nei due casi discutere il carattere di  $s_n$ .

**Esercizio** Stabilire il carattere delle successioni per ricorrenza.

$$\text{a)} \quad \begin{cases} s_{n+1} = 2s_n - 1 \\ s_0 = 3. \end{cases} \quad \text{b)} \quad \begin{cases} s_{n+1} = s_n - 1 \\ s_0 = 4 \end{cases} \quad \text{c)} \quad \begin{cases} s_{n+1} = -s_n \\ s_0 = 4 \end{cases}$$

**Esercizio** Determinare una soluzione particolare per l'equazione

$$x_{n+1} = \frac{1}{2}x_n + b_n,$$

nei casi

$$b_n = n \quad \text{e} \quad b_n = 2^n.$$

**Esercizio** Sia  $f : [a, b] \rightarrow [a, b]$  continua. Dimostrare che  $f$  ammette un *punto fisso* in  $[a, b]$ , ossia l'equazione

$$f(x) = x$$

ha almeno una soluzione  $x^* \in [a, b]$ .

**Esercizio** Sia

$$\begin{cases} x_{n+1} = f(x_n) \\ x_0 \text{ (assegnato)}, \end{cases}$$

con  $f : [a, b] \rightarrow [a, b]$  continua. Dimostrare che se  $x_n \rightarrow l \in \mathbb{R}$ , allora  $l$  è punto fisso per  $f$ .

**Esercizio** Sia

$$\begin{cases} x_{n+1} = f(x_n) \\ x_0 \text{ (assegnato)}, \end{cases}$$

con  $f : [a, b] \rightarrow [a, b]$  crescente. Dimostrare che se  $x_1 > x_0$ , allora  $x_n$  è crescente e che se  $x_1 < x_0$ , allora  $x_n$  è decrescente, quindi, in ogni caso,  $x_n$  è monotona e, quindi, regolare.

**Esercizio** Studiare, al variare di  $a > 0$ , il comportamento asintotico della successione per ricorrenza

$$\begin{cases} s_0 = a \\ s_{n+1} = \sqrt{2 + s_n}. \end{cases}$$

**Esercizio** Sia

$$\begin{cases} s_{n+1} = f(s_n) \\ s_0 \text{ (assegnato)}, \end{cases}$$

con  $f : [a, b] \rightarrow [a, b]$  continua e decrescente. Dimostrare che

a) la successione<sup>10</sup>

$$\begin{cases} x_{n+1} = f^2(x_n) \\ x_0 = s_0 \end{cases}$$

coincide con  $s_{2n}$  e la successione

$$\begin{cases} y_{n+1} = f^2(y_n) \\ y_0 = s_1 \end{cases}$$

coincide con  $s_{2n+1}$ ;

<sup>10</sup>Ricordiamo che il simbolo  $f^2$  indica  $f \circ f$ , ossia l'iterata seconda di  $f$ .

- b) le successioni  $x_n$  e  $y_n$  sono monotone; in particolare  $x_n = s_{2n}$  è crescente se e solo se  $y_n = s_{2n+1}$  è decrescente e viceversa.
- c)  $x_n \rightarrow x^*$  se e solo se  $y_n \rightarrow y^*$  con  $x^* = f(y^*)$  e  $y^* = f(x^*)$ .
- d)  $s_n \rightarrow l \in \mathbb{R}$  se e solo se  $l = x^* = y^*$ .

Calcolare i punti di equilibrio stabile per l'equazione alle differenze

$$x_{n+1} = \frac{1}{2} \left( x_n + \frac{A}{x_n} \right), \quad A > 0.$$

La relazione indicata può servire per calcolare velocemente radici quadrate. Calcolare  $\sqrt{7}$  con tre cifre decimali esatte.

Studiare, al variare di  $a > 0$ , il comportamento asintotico della successione per ricorrenza

$$\begin{cases} s_0 = a \\ s_{n+1} = \frac{2}{1+s_n}. \end{cases}$$

Nel cuore, l'eccitazione generata da un normale *pacemaker* negli atrii, passa ai ventricoli causando contrazione del cuore e l'invio di sangue agli altri organi. L'eccitazione deve passare attraverso il nodo atrio-ventricolare, che connette elettricamente atrii e ventricoli. Il problema seguente è basato su un modello matematico per la conduzione atrio-ventricolare nei mammiferi.<sup>11</sup> Assumiamo che i tempi (in millisecondi) di conduzione atrio-ventricolare soddisfino la seguente legge di ricorrenza

$$x_{n+1} = \frac{375}{x_n - 90} + 100, \quad x_n > 90.$$

- a) Determinare i punti fissi e studiarne la stabilità.
- b) Analizzare il comportamento asintotico della soluzione che verifica la condizione iniziale  $x_0 = 150$ .

La seguente equazione alle differenze è stata proposta come modello di evoluzione per la densità di popolazione di insetti in anni successivi

$$x_{n+1} = f(x_n) = \alpha x_n e^{-\beta x_n^3},$$

dove  $\alpha, \beta$  sono parametri positivi. Per  $\alpha = e$  e  $\beta = 1/3$ ,

- a) tracciare il grafico di  $f$ ;
- b) determinare i punti di equilibrio e studiarne la stabilità.

Nello studio delle reti neurali<sup>12</sup> si incontra la seguente equazione alle differenze

$$x_{n+1} = f(x_n) = \frac{\alpha x_n}{1 + \beta x_n},$$

con  $\alpha$  e  $\beta$  parametri positivi e  $x_n \geq 0$ .

<sup>11</sup>Simson e altri, *American Journal of Physiology*, 1981.

<sup>12</sup>Glass e Pasternack, *Journal of math. biology*, 1978.

- a) Determinare al variare di  $\alpha$  e  $\beta$  i punti di equilibrio, la loro stabilità e se la dinamica in un intorno dei punti fissi è monotona o oscillatoria.
- b) Nel caso  $\alpha = \beta = 1$ , tracciare il diagramma di fase e studiare il comportamento asintotico di  $x_n$ .
- c) Esprimere  $x_n$  in funzione di  $x_0$  ( $> 0$  generico) e verificare il risultato ottenuto in (b).

## 5 EQUAZIONI LINEARI A COEFFICIENTI COSTANTI DEL SECOND'ORDINE

In questo capitolo analizziamo le equazioni lineari a coefficienti costanti del second'ordine:

**DEFINIZIONE 7.3** Un'equazione alle differenze lineare del second'ordine o a due passi a coefficienti costanti è un'equazione del tipo<sup>13</sup>

$$ax_{n+2} + bx_{n+1} + cx_n = f_n, \quad a, b, c \in \mathbb{R}, \quad a, c \neq 0.$$

Se la successione  $f_n$  è identicamente nulla (cioè  $f_n = 0$  per ogni  $n$ ) l'equazione si dice *omogenea*. Per esempio, è omogenea l'equazione  $x_{n+2} + x_n = 0$ , mentre non lo è l'equazione  $x_{n+2} + x_n = 3n$ .

### 5.1 Equazioni omogenee

Si può dimostrare che la *soluzione generale di un'equazione alle differenze del second'ordine lineare omogenea*

$$(5.1) \quad ax_{n+2} + bx_{n+1} + cx_n = 0$$

è costituita da tutte le combinazioni lineari di due soluzioni particolari  $\varphi_n$  e  $\psi_n$  linearmente indipendenti (ossia che non siano una multipla dell'altra) può, cioè, essere scritta nelle forme

$$x_n = C_1 \varphi_n + C_2 \psi_n, \quad C_1, C_2 \in \mathbb{R}.$$

La *soluzione generale* della (5.1) contiene due costanti arbitrarie. Per determinare una soluzione particolare è necessario assegnare una coppia di condizioni; di solito si assegnano due valori per la successione  $x_n$  in due tempi diversi, tipicamente per  $n = 0$  e  $n = 1$ .

Per scrivere la soluzione generale della (5.1), occorre, dunque, determinare le due soluzioni particolari  $\varphi_n$  e  $\psi_n$  opportune. Ricordiamo che la soluzione generale dell'equazione del primo ordine:

$$x_{n+1} = \alpha x_n,$$

è  $x_n = C\alpha^n$ . Proviamo quindi a cercare funzioni del tipo  $x_n = \lambda^n$ , con  $\lambda \neq 0$  da determinarsi in modo da ottenere una soluzione della (5.1). Se  $x_n = \lambda^n$ , si ha  $x_{n+1} = \lambda^{n+1}$  e  $x_{n+2} = \lambda^{n+2}$ . Sostituendo nella (5.1) troviamo

$$a\lambda^2 \lambda^n + b\lambda \lambda^n + c\lambda^n = 0,$$

<sup>13</sup>Chiediamo che i coefficienti  $a$  e  $c$  siano diversi da 0, altrimenti non sarebbe di second'ordine.

da cui, dividendo per  $\lambda^n$ , otteniamo l'equazione di secondo grado in  $\lambda$ :

$$(5.2) \quad a\lambda^2 + b\lambda + c = 0$$

La (5.2) si dice *equazione caratteristica* dell'equazione alle differenze (5.1). Dovremo allora distinguere tre casi.

$$b^2 - 4ac > 0$$

Se il discriminante dell'equazione è positivo la (5.2) ha le due soluzioni reali e distinte:

$$\lambda_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Le due successioni  $\lambda_1^n, \lambda_2^n$  verificano la (5.1) e sono linearmente indipendenti (controllare). In questo caso la soluzione generale dell'equazione alle differenze (5.1) è

$$x_n = C_1 \lambda_1^n + C_2 \lambda_2^n, \quad C_1, C_2 \in \mathbb{R}$$



Per l'equazione alle differenze

$$x_{n+2} - 6x_{n+1} + 5x_n = 0$$

l'equazione caratteristica è

$$(5.3) \quad \lambda^2 - 6\lambda + 5 = 0$$

che ha le due soluzioni  $\lambda_1 = 1$  e  $\lambda_2 = 5$ . Pertanto, la soluzione generale dell'equazione è

$$x_n = C_1 + C_2 5^n, \quad C_1, C_2 \in \mathbb{R}.$$

Supponiamo interessi trovare la soluzione particolare che soddisfa le condizioni  $x_0 = 2$  e  $x_1 = 1$ . Sostituendo nella soluzione generale  $n = 0$  e successivamente  $n = 1$  si trova il sistema

$$\begin{cases} C_1 + C_2 = 2 \\ C_1 + 5C_2 = 1. \end{cases}$$

che ha soluzione

$$C_1 = \frac{9}{4}, \quad C_2 = -\frac{1}{4}.$$

La soluzione particolare cercata per l'equazione alle differenze è, quindi,

$$x_n = \frac{9}{4} - \frac{1}{4} 5^n.$$

$$b^2 - 4ac = 0$$

L'equazione caratteristica ha una sola radice (*doppia*)  $\lambda = -\frac{b}{2a}$ ; essa genera la soluzione  $\lambda^n$ . Si può controllare che un'altra soluzione linearmente indipendente è data dalla successione  $n\lambda^n$ . Infatti da  $x_n = n\lambda^n$  si ricava

$$x_{n+1} = (n+1)\lambda^{n+1} \quad \text{e} \quad x_{n+2} = (n+2)\lambda^{n+2}.$$

Sostituendo nell'equazione alle differenze

$$a(n+2)\lambda^{n+2} + b(n+1)\lambda^{n+1} + cn\lambda^n = 0,$$

dopo aver diviso per  $\lambda^n$ , ci si riduce a

$$(a\lambda^2 + b\lambda + c)n + 2a\lambda^2 + b\lambda = 0$$

che è un'identità poiché  $\lambda = -\frac{b}{2a}$  è radice dell'equazione caratteristica. La soluzione generale per l'equazione alle differenze in questo caso è

$$x_n = C_1 \lambda^n + C_2 n \lambda^n = (C_1 + C_2 n) \lambda^n, \quad C_1, C_2 \in \mathbb{R}$$



Sia

$$x_{n+2} + 4x_{n+1} + 4 = 0.$$

L'equazione caratteristica è

$$\lambda^2 + 4\lambda + 4 = 0,$$

che possiede la sola soluzione  $\lambda = -2$ . La sua soluzione generale è

$$x_n = (-2)^n (C_1 + C_2 n),$$

con  $C_1, C_2$  arbitrari.

$$b^2 - 4ac < 0$$

L'equazione caratteristica ha due soluzioni complesse coniugate  $\lambda_{1,2} = \alpha \pm i\beta$ . Introduciamo il loro modulo

$$|\lambda_{1,2}| = \rho = \sqrt{\alpha^2 + \beta^2}$$

e (a meno del segno) l'argomento  $\theta \neq 0$ , definito da

$$\operatorname{tg} \theta = \frac{\beta}{\alpha}.$$

Si può controllare che le due successioni<sup>14</sup>

$$\rho^n \cos n\theta \quad \text{e} \quad \rho^n \sin n\theta$$

sono soluzioni indipendenti, per cui la soluzione generale in questo caso è

$$x_n = C_1 \rho^n \cos n\theta + C_2 \rho^n \sin n\theta, \quad C_1, C_2 \in \mathbb{R}$$

<sup>14</sup>Scrivendo  $\lambda_{1,2} = \rho(\cos \theta \pm i \sin \theta)$ , si ha, per la formula di De Moivre

$$\lambda_1^n = \rho^n (\cos n\theta + i \sin n\theta), \quad \lambda_2^n = \rho^n (\cos n\theta - i \sin n\theta).$$

Due soluzioni (reali) dell'equazione sono quindi

$$\frac{\lambda_1^n + \lambda_2^n}{2} \quad \text{e} \quad \frac{\lambda_1^n - \lambda_2^n}{2i}.$$



Sia

$$x_{n+2} + 4x_n = 0.$$

Il discriminante dell'equazione caratteristica è negativo: le due radici sono  $\pm 2i$  e, conseguentemente

$$\rho = 2 \quad \text{e} \quad \theta = \frac{\pi}{2}.$$

La soluzione generale è

$$x_n = C_1 2^n \cos\left(\frac{\pi}{2}n\right) + C_2 2^n \sin\left(\frac{\pi}{2}n\right),$$

con  $C_1$  e  $C_2$  costanti reali.

Scriviamo ora la soluzione dell'equazione alle differenze  $x_{n+2} + 4x_n = 0$ , che verifica le condizioni iniziali

$$x_0 = 1, \quad x_1 = 2.$$

Per determinare tale soluzione particolare dobbiamo determinare le costanti  $C_1$  e  $C_2$ . Per far questo sostituiamo i valori 0 e 1 nella soluzione generale; si ha

$$\begin{cases} x_0 = C_1 = 1 \\ x_1 = 2C_2 = 2 \end{cases} \Rightarrow \begin{cases} C_1 = 1 \\ C_2 = 1. \end{cases}$$

La soluzione cercata<sup>15</sup> è

$$x_n = 2^n \left( \cos\left(\frac{\pi}{2}n\right) + \sin\left(\frac{\pi}{2}n\right) \right).$$

## 5.2 I numeri di Fibonacci

Presentiamo un problema esaminato da Fibonacci<sup>16</sup> nel suo *Liber Abaci* (1202). Si tratta di un problema di ... allevamento di conigli. Ci chiediamo quante coppie di conigli si possono ottenere in un anno a partire da una sola coppia se formuliamo le seguenti ipotesi:

- la prima coppia è costituita da conigli appena nati;
- ogni coppia di conigli genera una nuova coppia ogni mese;
- ogni coppia di conigli diventa fertile dopo un mese di vita;
- i conigli non muoiono.

<sup>15</sup>Si noti che

$$\begin{aligned} \left\{ \cos \frac{n\pi}{2} \right\} &= \{1, 0, -1, 0, 1, 0, -1, 0, \dots\} \\ \left\{ \sin \frac{n\pi}{2} \right\} &= \{0, 1, 0, -1, 0, 1, 0, -1, \dots\} \end{aligned}$$

per cui

$$\left\{ \cos \frac{n\pi}{2} + \sin \frac{n\pi}{2} \right\} = \{1, 1, -1, -1, 1, 1, -1, -1, \dots\}.$$

La soluzione è periodica di periodo 4.

<sup>16</sup>Soprannome di Leonardo Pisano (1175-1240).

Indichiamo con  $F_k$  il numero di coppie dopo  $k$  mesi ( $k \geq 0$ ). All'inizio e anche dopo un mese avremo solo la prima coppia, quindi  $F_0 = F_1 = 1$ . Al secondo mese avremo un'altra coppia, quindi  $F_2 = 1 + 1 = 2$ , al terzo mese solo la prima coppia è in grado di generare un'altra e quindi  $F_3 = 2 + 1 = 3 \dots$  e così via.

Alla fine del generico mese  $k+2$  avremo oltre le coppie di conigli già presenti  $F_{k+1}$ , e le neonate, pari al numero di coppie fertili del mese precedente  $F_k$ , quindi

$$(5.4) \quad F_{k+2} = F_{k+1} + F_k.$$

La (5.4) è un'equazione alle differenze a due passi, lineare omogenea a coefficienti costanti.

Dalla (5.4), con la condizione (iniziale)  $F_0 = F_1 = 1$  è facile calcolare  $F_k$  (specialmente se si ha un computer!). I primi 12 termini sono

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144.$$

Risolviamo la (5.4). L'equazione caratteristica è

$$\lambda^2 - \lambda - 1 = 0$$

che ha le soluzioni

$$\lambda_1 = \frac{1 - \sqrt{5}}{2}, \quad \lambda_2 = \frac{1 + \sqrt{5}}{2},$$

per cui la soluzione generale è

$$F_k = c_1 \left( \frac{1 - \sqrt{5}}{2} \right)^k + c_2 \left( \frac{1 + \sqrt{5}}{2} \right)^k, \quad c_1, c_2 \in \mathbb{R}.$$

Determiniamo  $c_1$  e  $c_2$  imponendo le condizioni iniziali  $F_0 = F_1 = 1$ . Si ottiene il sistema

$$\begin{cases} c_1 + c_2 = 1 \\ c_1 \frac{1 - \sqrt{5}}{2} + c_2 \frac{1 + \sqrt{5}}{2} = 1 \end{cases}$$

che ha la soluzione

$$\begin{cases} c_1 = -\frac{1 - \sqrt{5}}{2\sqrt{5}} \\ c_2 = \frac{1 + \sqrt{5}}{2\sqrt{5}}. \end{cases}$$

Quindi

$$F_k = \frac{1}{\sqrt{5}} \left\{ \left( \frac{1 + \sqrt{5}}{2} \right)^{k+1} - \left( \frac{1 - \sqrt{5}}{2} \right)^{k+1} \right\}.$$

Si noti che  $F_k$  è un numero naturale per ogni  $k$ . Poiché

$$\frac{\sqrt{5} - 1}{2} < 1$$

si ha

$$\left(\frac{1-\sqrt{5}}{2}\right)^{k+1} \rightarrow 0 \quad \text{per } k \rightarrow +\infty$$

quindi

$$F_k \sim \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^{k+1}$$

e

$$\lim_{k \rightarrow +\infty} \frac{F_{k+1}}{F_k} = \frac{1+\sqrt{5}}{2}.$$

Incidentalmente, osserviamo che il numero  $\frac{\sqrt{5}-1}{2}$  è noto come *sezione aurea* (golden ratio).

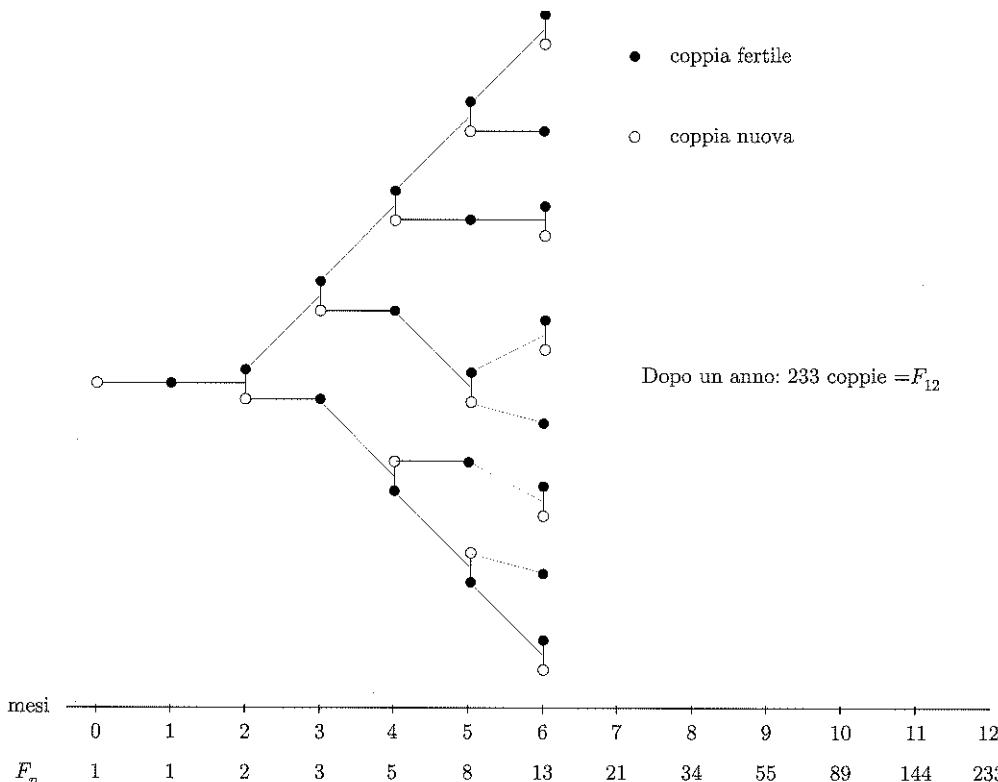


Figura 7.14. Il problema dei conigli e i numeri di Fibonacci.

### 5.3 Equazioni non omogenee

Consideriamo ora un'equazione alle differenze del tipo

$$(5.5) \quad ax_{n+2} + bx_{n+1} + cx_n = f_n,$$

con secondo membro non identicamente nullo. Si può dimostrare che la soluzione generale può scriversi come

$$(5.6) \quad x_n = C_1 \varphi_n + C_2 \psi_n + y_n, \quad C_1, C_2 \in \mathbb{R}$$

ove  $y_n$  è una soluzione particolare della (5.5) e  $C_1 \varphi_n + C_2 \psi_n$  è la soluzione generale dell'equazione omogenea associata

$$az_{n+2} + bz_{n+1} + cz_n = 0.$$

Il problema è quindi determinare una soluzione particolare. In alcuni casi si può utilizzare il *metodo di somiglianza*: se  $f_n$  è d'un certo tipo, allora anche una soluzione particolare è dello stesso tipo. Per esempio, se  $f_n$  è un polinomio allora anche una soluzione particolare  $y_n$  è un polinomio, se il termine noto è esponenziale allora anche una soluzione particolare è esponenziale e così via.



Scriviamo la soluzione generale dell'equazione alle differenze

$$(5.7) \quad 6x_{n+2} + 5x_{n+1} + x_n = 3.$$

Come prima cosa determiniamo le soluzioni dell'equazione omogenea associata

$$6y_{n+2} + 5y_{n+1} + y_n = 0.$$

L'equazione caratteristica

$$6\lambda^2 + 5\lambda + 1 = 0$$

ha le radici

$$\lambda = -\frac{1}{3} \quad \text{e} \quad \lambda = -\frac{1}{2},$$

quindi la soluzione generale dell'equazione omogenea è

$$y_n = c_1 \left(-\frac{1}{3}\right)^n + c_2 \left(-\frac{1}{2}\right)^n \quad c_1, c_2 \in \mathbb{R}.$$

Cerchiamo poi una soluzione dell'equazione non omogenea. Dato che il secondo membro dell'equazione completa è una costante, pensiamo come soluzione particolare dell'equazione completa a una successione costante  $u_n = k$ . Se  $u_n = k$  anche  $u_{n+1} = k$  e  $u_{n+2} = k$  e sostituendo nella (5.7), si ricava

$$6k + 5k + k = 3 \quad \Rightarrow \quad k = \frac{1}{4}.$$

La soluzione generale della (5.7) è, dunque,

$$x_n = c_1 \left(-\frac{1}{3}\right)^n + c_2 \left(-\frac{1}{2}\right)^n + \frac{1}{4} \quad c_1, c_2 \in \mathbb{R}.$$

## 5.4 Stabilità

Per l'equazione alle differenze omogenea la successione  $x_n = 0$  per ogni  $n$  rappresenta una *soluzione di equilibrio*. Se ogni altra soluzione dell'equazione converge a 0 per  $n \rightarrow +\infty$  la soluzione nulla si dice *globalmente asintoticamente stabile*.

Per le equazioni alle differenze non omogenee è importante sapere se la soluzione nulla è asintoticamente stabile per l'omogenea associata. In tal caso infatti può succedere che sia la soluzione particolare  $u_n$  a determinare il comportamento a regime del sistema, mentre l'integrale generale dell'omogenea assume un significato di regime transitorio.

Avendo a disposizione una formula esplicita per la soluzione generale possiamo dedurre direttamente il comportamento all'infinito.

**PROPOSIZIONE 7.1** *La soluzione nulla è*

- asintoticamente stabile se e solo se il modulo di ogni radice dell'equazione caratteristica è minore di 1;*
- stabile se e solo se le radici sono complesse con modulo uguale a 1 oppure le radici sono reali e distinte, una con modulo minore di 1 e l'altra uguale a  $\pm 1$ .*
- instabile se l'equazione caratteristica ha una radice di modulo maggiore di uno oppure se ha una radice doppia uguale a  $\pm 1$ .*

Esistono criteri di stabilità particolarmente comodi che non richiedono di risolvere l'equazione caratteristica. Ci limitiamo al seguente.

**PROPOSIZIONE 7.2** *Criterio per equazioni di secondo grado.*

*La soluzione nulla è asintoticamente stabile per l'equazione*

$$ax_{n+2} + bx_{n+1} + cx_n = 0$$

*se e solo se, dette  $\lambda_1$  e  $\lambda_2$  le radici dell'equazione caratteristica, si ha*

$$\lambda_1 \lambda_2 < 1 \quad e \quad |\lambda_1 + \lambda_2| < 1 + \lambda_1 \lambda_2$$

*ossia se e solo se*

$$\frac{c}{a} < 1 \quad e \quad \left| \frac{b}{a} \right| < 1 + \frac{c}{a}$$

*quindi se e solo se, posto  $a > 0$ ,*

$$c < a \quad e \quad |b| < a + c.$$

La dimostrazione non è difficile ed il lettore può cimentarsi (si veda l'esercizio 18).



Riprendiamo il *modello di Samuelson* considerato nell'introduzione, a cui rimandiamo per la derivazione. Si tratta dell'equazione del second'ordine

$$(5.8) \quad Y_n - a(1+b)Y_{n-1} + abY_{n-2} = A + k.$$

Interessa sapere se esiste una soluzione d'equilibrio e, in caso affermativo, studiarne la stabilità. La soluzione di equilibrio, cioè una soluzione costante  $\bar{Y}$ , si trova subito dall'equazione

$$\bar{Y} - a(1+b)\bar{Y} + ab\bar{Y} = A + k.$$

ed è

$$\bar{Y} = \frac{A + k}{1 - a}.$$

Questa soluzione risulta asintoticamente stabile se e solo se la soluzione nulla è asintoticamente stabile per l'equazione omogenea e cioè se e solo se le radici  $\lambda_1$  e  $\lambda_2$  dell'equazione

$$(5.9) \quad \lambda^2 - a(1+b)\lambda + ab = 0$$

sono in modulo minori di 1. In base al criterio di stabilità, ciò accade se

$$ab < 1 \quad e \quad a(1+b) < 1 + ab.$$

Essendo  $0 < a < 1$ , la seconda condizione è sempre verificata per cui ci si riduce a

$$(5.10) \quad ab < 1.$$

L'andamento delle soluzioni è differente a seconda dei valori dei parametri. Il discriminante dell'equazione (5.9) risulta positivo o nullo se

$$(5.11) \quad a^2(1+b)^2 - 4ab \geq 0 \quad \text{ossia se} \quad a \geq \frac{4b}{(1+b)^2},$$

quindi, se sono verificate entrambe le condizioni (5.10) e (5.11), la soluzione dell'equazione (5.8) tende all'equilibrio senza compiere oscillazioni; se è verificata la (5.10), ma non la (5.11), la soluzione dell'equazione (5.8) tende all'equilibrio oscillando. La figura 7.15 mostra il grafico una soluzione di questo tipo, ottenuta in corrispondenza dei valori dei parametri  $a = 1/4$  e  $b = 3$ .

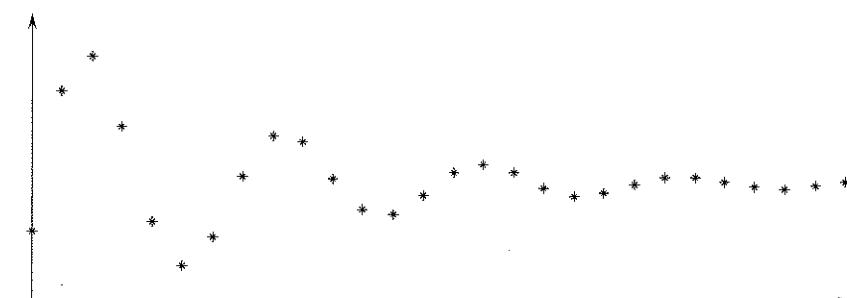


Figura 7.15. Una soluzione del modello di Samuelson.



Scrivere la soluzione dell'equazione alle differenze

$$x_{n+2} - 5x_{n+1} + 6x_n = 0$$

che soddisfa le condizioni  $x_0 = 0$  e  $x_1 = 1$ .

Scrivere la soluzione generale delle seguenti equazioni alle differenze lineari del secondo ordine.

$$(a) 2x_{n+2} - 3x_{n+1} + x_n = 0, \quad (b) x_{n+2} + x_n = 0, \quad (c) x_{n+2} - 4x_{n+1} + 4x_n = 1.$$

Dire se la soluzione costante è stabile.

Risolvere le equazioni alle differenze

$$(a) x_{n+2} - 3x_{n+1} + 2x_n = 3^n, \quad (b) x_{n+3} - x_{n+2} + x_{n+1} - x_n = -1.$$

Determinare la soluzione dell'equazione alle differenze

$$x_{n+3} + x_n = 0$$

che corrisponde ai dati iniziali  $x_0 = 1$ ,  $x_1 = 0$  e  $x_2 = 2$ . Controllare che tale soluzione è periodica di periodo 6.

Dimostrare la Proposizione 7.2 ossia che se  $\lambda_1$  e  $\lambda_2$  sono radici di un trinomio di secondo grado

$$|\lambda_1| < 1 \text{ e } |\lambda_2| < 1 \iff \lambda_1\lambda_2 < 1 \text{ e } |\lambda_1 + \lambda_2| < 1 + \lambda_1\lambda_2.$$

Sia

$$x_{n+2} = x_{n+1} + ax_n, \quad a \in \mathbb{R}.$$

a) Stabilire per quali valori del parametro la soluzione nulla è asintoticamente stabile.

b) Posto  $a = \frac{1}{8}$ , calcolare la soluzione che verifica la condizione iniziale  $x_0 = 0$ ,  $x_1 = 10$ .

# 8

## Elementi di geometria e algebra lineare

### 1 VETTORI NEL PIANO E NELLO SPAZIO

Il concetto di *vettore*, fondamentale sia in matematica che nelle applicazioni (fisiche ecc.), può essere introdotto a vari livelli di astrazione. In questa sezione ci occuperemo di *vettori nel piano e nello spazio*: in questo contesto, è possibile dare una definizione geometrica elementare di vettore; molte grandezze fisiche (come velocità, accelerazione, forza...) si rappresentano in questo modo. In seguito (par. 3) vedremo come la nozione di vettore si possa generalizzare in termini astratti, ottenendo un concetto più flessibile, che risulta molto utile per l'algebra, il calcolo infinitesimale e le loro applicazioni.

#### 1.1 Operazioni fondamentali sui vettori

Un vettore nel piano o nello spazio è individuato assegnando:

- a) un numero reale non negativo che esprime la sua *lunghezza* o *modulo* o *intensità*;
- b) una *direzione*, individuata da una retta (rette parallele individuano la stessa direzione);
- c) un *verso*.

Geometricamente, possiamo pensare ai vettori come a *segmenti orientati*, con la precisazione che due segmenti orientati che possano ottersi l'uno dall'altro per traslazione sono *lo stesso vettore*.

Se nello spazio è fissato un sistema di riferimento cartesiano di cui  $O$  è l'origine, possiamo anche vedere i vettori come *frecce uscenti* da  $O$  (fig. 8.1).



Figura 8.1. Vettori come frecce uscenti da  $O$ .  $\overrightarrow{OA}$  e  $\overrightarrow{PQ}$  rappresentano lo stesso vettore applicato in due punti diversi dello spazio.

La direzione ed il verso del vettore sono allora quelli indicati dalla freccia corrispondente mentre il suo modulo coincide con la lunghezza della freccia. Per indicare un vettore potremo usare una lettera in grassetto (come  $\mathbf{v}$ ,  $\mathbf{w}$ , etc.) oppure il simbolo  $\overrightarrow{OA}$  se  $A$  è la punta della freccia. Si dice anche che  $\overrightarrow{OA}$  è il vettore posizione del punto  $A$ .

Naturalmente si può pensare un vettore come freccia uscente da un qualsiasi altro punto dello spazio. Due punti fissati  $P$  e  $Q$ , nell'ordine, individuano un vettore con direzione coincidente con quella della retta passante per  $P$  e  $Q$ , con verso da  $P$  a  $Q$  e con modulo pari alla lunghezza del segmento  $PQ$ . Il vettore individuato dai due punti  $P$  e  $Q$  coincide col vettore rappresentato dalla freccia  $\overrightarrow{OA}$  che si ottiene traslando parallelamente il segmento  $PQ$  fino a far coincidere  $P$  con  $O$  e  $Q$  con  $A$  (v. fig. 8.1); per il vettore così rappresentato si può anche usare il simbolo  $\overrightarrow{PQ}$ . Sottolineiamo però che  $\overrightarrow{OA}$  e  $\overrightarrow{PQ}$  sono due rappresentazioni (geometriche) diverse dello stesso vettore  $\mathbf{v}$ , una volta applicato in  $O$ , un'altra applicato in  $P$ .

Il modulo di un vettore  $\mathbf{v}$  sarà indicato con  $|\mathbf{v}|$ . Il vettore di modulo 0 si chiama *vettore nullo* e sarà indicato con  $\mathbf{0}$ .

Sui vettori si possono definire varie operazioni. Cominciamo a introdurre le due fondamentali: la somma di vettori e il prodotto di un vettore per uno scalare.

- *Somma di vettori.* Dati due vettori  $\mathbf{v} = \overrightarrow{OA}$  e  $\mathbf{w} = \overrightarrow{OB}$ , la loro *somma* è definita al seguente modo: si applica  $\mathbf{w}$  in  $A$ , cioè si scrive  $\mathbf{w} = \overrightarrow{AC}$  con  $C$  opportuno; allora  $\mathbf{v} + \mathbf{w} = \overrightarrow{OC}$ . Ne segue che  $\mathbf{v} + \mathbf{w}$  si può ottenere con la ben nota regola del parallelogramma: il vettore  $\mathbf{v} + \mathbf{w}$  è la diagonale del parallelogramma costruito su  $\mathbf{v}$  e  $\mathbf{w}$ , nel modo indicato in figura.

Si noti che il segmento orientato associato a  $\mathbf{v} + \mathbf{w}$  è contenuto nel piano dei segmenti orientati associati a  $\mathbf{v}$  e  $\mathbf{w}$ .

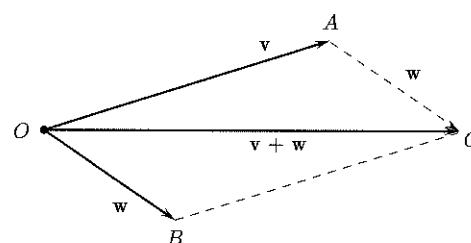


Figura 8.2. Somma di due vettori.

La somma è un'operazione *associativa*

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$

per cui si può scrivere semplicemente  $\mathbf{u} + \mathbf{v} + \mathbf{w}$ , e *commutativa*

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$

Inoltre, poiché il vettore nullo ha lunghezza zero, si ha:

$$\mathbf{v} + \mathbf{0} = \mathbf{v} \quad \text{per ogni vettore } \mathbf{v}$$

Si indica con  $-\mathbf{v}$  il vettore *opposto* a  $\mathbf{v}$ , avente stessa lunghezza e direzione ma *verso opposto*. Evidentemente

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$$

La somma  $\mathbf{v} + (-\mathbf{w})$  si scrive semplicemente  $\mathbf{v} - \mathbf{w}$ , definendo in tal modo la *differenza* tra due vettori (fig. 8.3).

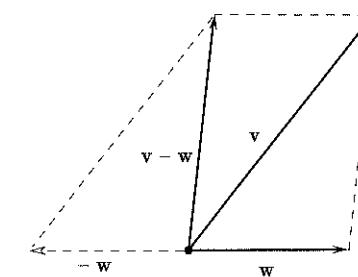


Figura 8.3. Differenza tra vettori.

Le proprietà della somma di vettori appena enunciate possono essere dimostrate geometricamente, in base alla definizione di somma.

Per esempio, la proprietà commutativa dipende dal fatto che in un parallelogramma i lati opposti sono uguali e paralleli. Da questo infatti discende che (v. fig. 8.2)  $\mathbf{v} = \overrightarrow{OA} = \overrightarrow{BC}$ ,  $\mathbf{w} = \overrightarrow{OB} = \overrightarrow{AC}$ , perciò

$$\mathbf{v} + \mathbf{w} = \overrightarrow{OA} + \overrightarrow{AC} = \overrightarrow{OC}; \mathbf{w} + \mathbf{v} = \overrightarrow{OB} + \overrightarrow{BC} = \overrightarrow{OC}$$

e quindi  $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ .

Con un ragionamento geometrico analogo (ma un po' più complesso) si può dimostrare la proprietà associativa. Le proprietà del vettore nullo e dell'opposto sono poi immediate.

- *Moltiplicazione di un vettore per uno scalare.* Se  $\mathbf{v}$  è un vettore e  $t$  è uno *scalare* (cioè un numero reale) il *prodotto* di  $t$  per  $\mathbf{v}$ , indicato con  $t\mathbf{v}$ , è definito come il vettore che ha lunghezza  $|t| \cdot |\mathbf{v}|$ , ha la direzione di  $\mathbf{v}$  e lo stesso verso se  $t > 0$ , verso opposto se  $t < 0$ .

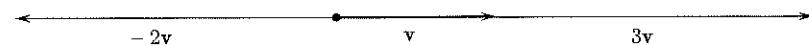
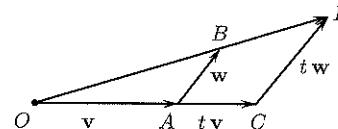


Figura 8.4. Moltiplicazione di  $\mathbf{v}$  per 3 e per  $-2$ .

La moltiplicazione di un vettore per uno scalare soddisfa le seguenti proprietà:

$$\begin{aligned} 1 \cdot \mathbf{v} &= \mathbf{v} \\ s(t\mathbf{v}) &= (st)\mathbf{v} \\ t(\mathbf{v} + \mathbf{w}) &= t\mathbf{v} + t\mathbf{w} \quad (s, t \in \mathbb{R}) \\ (s+t)\mathbf{v} &= s\mathbf{v} + t\mathbf{v} \end{aligned}$$

Di nuovo, queste proprietà hanno una dimostrazione geometrica. Per esempio, per provare la terza, si consideri la figura 8.5: i triangoli  $OAB$ ,  $OCD$  sono simili, per il Teorema di Talete, perché  $AB$  e  $CD$  sono paralleli (rappresentando i vettori  $\mathbf{w}$  e  $t\mathbf{w}$ ). Ne segue che  $OD : OB = OC : OA = t$ . Ma  $\overrightarrow{OD} = t\mathbf{v} + t\mathbf{w}$ ,  $\overrightarrow{OB} = \mathbf{v} + \mathbf{w}$ , perciò  $t\mathbf{v} + t\mathbf{w} = t(\mathbf{v} + \mathbf{w})$ .

Figura 8.5. La proprietà  $t(\mathbf{v} + \mathbf{w}) = t\mathbf{v} + t\mathbf{w}$ .

- **Versori.** Un vettore di modulo unitario si chiama *versore*.

Dato un vettore  $\mathbf{v}$  (non nullo), indichiamo con

$$\text{vers}(\mathbf{v}) = \frac{\mathbf{v}}{|\mathbf{v}|}$$

il versore ottenuto da  $\mathbf{v}$  dividendolo per il suo modulo, ossia, come si usa dire, *normalizzandolo*.

È utile ora passare dal punto di vista della geometria elementare (sintetica), che abbiamo adottato fin qui, a quello della geometria analitica, ossia introdurre un sistema di riferimento cartesiano. Questo renderà possibili calcoli analitici coi vettori e costituirà il punto di partenza per le generalizzazioni che saranno sviluppate in seguito (par. 3). Per maggior gradualità, considereremo separatamente il caso dei vettori nel piano e nello spazio.

### Vettori nel piano

Se introduciamo un sistema di riferimento cartesiano ortogonale nel piano, questo si può identificare con l'insieme  $\mathbb{R}^2$  delle coppie ordinate di numeri reali. All'origine  $O$  si può associare il vettore nullo  $\mathbf{0}$ ; ad un punto  $A(x, y)$  viene associato il vettore  $\mathbf{v} = \overrightarrow{OA}$ ; viceversa, ad ogni vettore  $\mathbf{v}$  è associata un'unica freccia che ha come primo estremo l'origine e come secondo un punto  $A = (x, y)$ . In questo modo è possibile identificare il punto di coordinate  $A = (x, y)$  con il vettore posizione  $\overrightarrow{OA}$ . Potremo allora scrivere

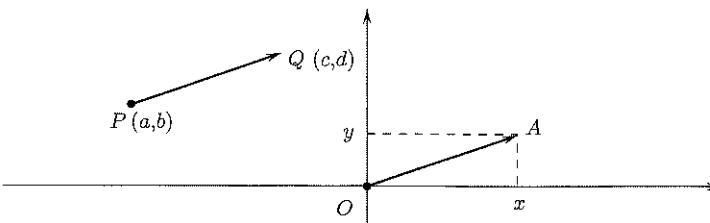
$$\mathbf{v} = (x, y) \quad \text{invece di} \quad \mathbf{v} = \overrightarrow{OA}$$

e i numeri  $x$  e  $y$  si dicono *componenti scalari* di  $\mathbf{v}$ . Si noti che

$$|\mathbf{v}| = \text{lunghezza di } \overrightarrow{OA} = \sqrt{x^2 + y^2}$$

(Teorema di Pitagora). Due punti  $P = (a, b)$  e  $Q = (c, d)$ , nell'ordine, individuano il vettore di componenti scalari  $x = c - a$ ,  $y = d - b$ ; cioè:

$$\overrightarrow{PQ} = (c - a, d - b)$$

Figura 8.6. Le frecce  $\overrightarrow{PO}$  e  $\overrightarrow{OA}$  rappresentano lo stesso vettore, le cui componenti scalari sono  $x = c - a$ ,  $y = d - b$ .

Se consideriamo due vettori  $\mathbf{u} = (x_1, y_1)$  e  $\mathbf{v} = (x_2, y_2)$  si verifica immediatamente che (fig. 8.7)

$$\mathbf{u} \pm \mathbf{v} = (x_1 \pm x_2, y_1 \pm y_2)$$

e che, se  $t \in \mathbb{R}$

$$t\mathbf{u} = (tx_1, ty_1)$$

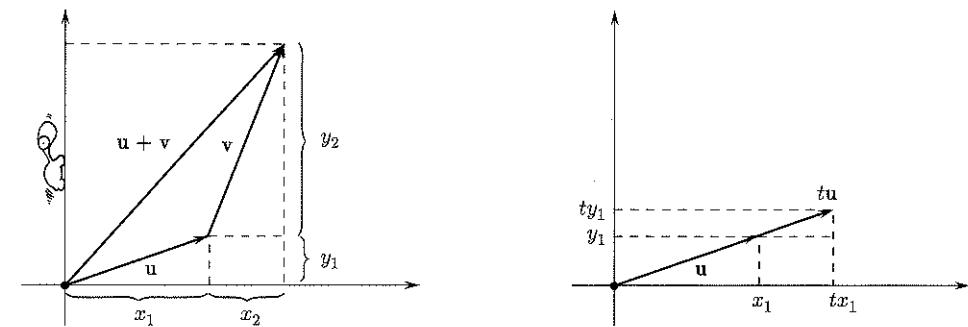


Figura 8.7. Somma e prodotto per uno scalare, componente per componente.

Le formule precedenti sono fondamentali in quanto permettono di eseguire le operazioni sui vettori per via analitica, senza necessità di costruzioni geometriche.

I vettori  $\mathbf{i} = (1, 0)$  e  $\mathbf{j} = (0, 1)$  sono diretti come gli assi coordinati, hanno lunghezza  $|\mathbf{i}| = 1 = |\mathbf{j}|$  (sono versori) e sono ortogonali tra loro. Inoltre ogni altro vettore  $\mathbf{v} = (x, y)$  si può esprimere nella forma

$$\mathbf{v} = x\mathbf{i} + y\mathbf{j}$$

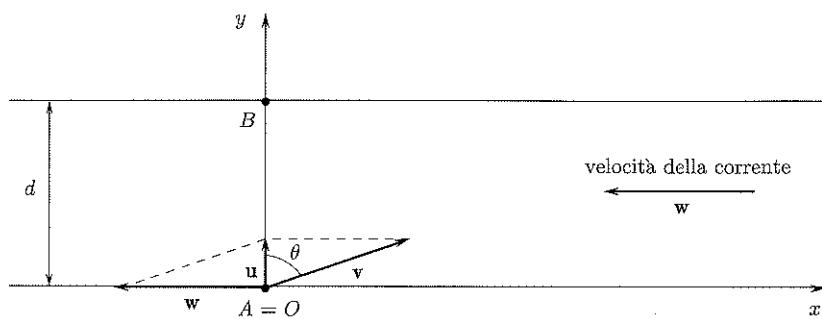
$\mathbf{i}$  e  $\mathbf{j}$  si dicono *versori fondamentali* nel piano.



Una motonave deve attraversare un fiume per portarsi da una località  $A$  ad una località  $B$ , opposta, a distanza  $d$ , sull'altra sponda. La motonave sviluppa una velocità (scalare) di  $v$  km/h e la corrente si muove ad una velocità (scalare)  $w$ .

In quale direzione si deve muovere la motonave per raggiungere  $B$  nel più breve tempo possibile?

Sceglieremo il riferimento cartesiano come indicato in figura, nel quale  $A$  è l'origine.

Figura 8.8.  $|\mathbf{v}| = v$ ,  $|\mathbf{w}| = w$ .

Per raggiungere lo scopo occorre scegliere l'angolo  $\theta$  in modo tale che  $\mathbf{u} = \mathbf{v} + \mathbf{w}$  punti nella direzione positiva dell'asse  $y$ .

Il vettore  $\mathbf{u}$  rappresenta la velocità effettiva della motonave nella direzione  $AB$ . Ora, si ha:

$$\mathbf{w} = -w\mathbf{i}, \quad \mathbf{v} = v \sin \theta \mathbf{i} + v \cos \theta \mathbf{j}$$

Affinché  $\mathbf{u}$  punti nella direzione dell'asse  $y$ , occorre dunque che la prima componente di  $\mathbf{u}$  si annulli; cioè deve essere

$$v \sin \theta = w$$

da cui

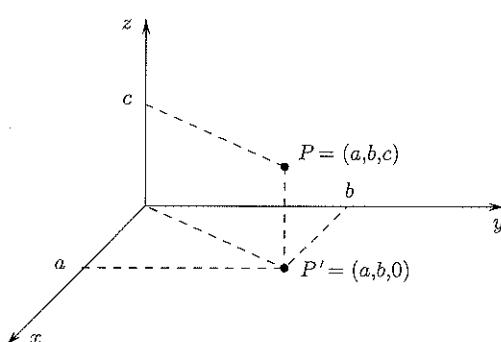
$$\sin \theta = \frac{w}{v} \quad \text{ossia} \quad \theta = \arcsin \frac{w}{v}$$

Per esempio, se  $w = 3$  km/h e  $v = 20$  km/h si trova  $\theta = \arcsin \frac{3}{20} \approx 0.15$  rad.

### Vettori nello spazio tridimensionale

Se introduciamo un sistema di riferimento cartesiano ortogonale nello spazio tridimensionale, con origine nel punto  $O$  di riferimento, questo si può identificare con l'insieme  $\mathbb{R}^3$  delle terne ordinate  $(x, y, z)$  di numeri reali.

Di solito si sceglie una terna di assi ortogonali con una *orientazione destrorsa*. Con ciò si intende che se indice e medio della mano destra puntano rispettivamente nel verso positivo degli assi  $x$  e  $y$ , il pollice punta nel verso positivo dell'asse  $z$ .

Figura 8.9. Sistema di riferimento destrorso in  $\mathbb{R}^3$ .

La formula della distanza tra due punti  $P = (a, b, c)$  e  $Q = (a', b', c')$  si estende subito dal caso bidimensionale:

$$\overline{PQ} = \sqrt{(a - a')^2 + (b - b')^2 + (c - c')^2}$$

Infatti  $\overline{PQ}$  è la diagonale del parallelepipedo indicato in figura 8.10.

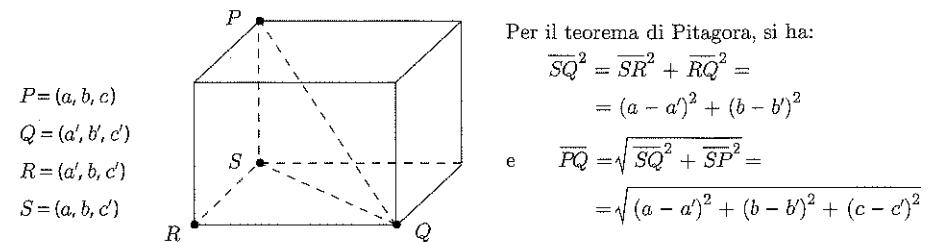


Figura 8.10. Distanza tra due punti nello spazio.

Il vettore  $\mathbf{v} = \overrightarrow{PQ}$  ha componenti scalari  $x = (a - a')$ ,  $y = (b - b')$ ,  $z = (c - c')$  e coincide con il vettore  $\overrightarrow{OA}$ , vettore posizione del punto  $A = (x, y, z)$ . Si può scrivere allora  $\mathbf{v} = (x, y, z)$  invece di  $\mathbf{v} = \overrightarrow{OA}$ .

La lunghezza di  $\mathbf{v}$  coincide con la lunghezza di  $\overline{PQ}$ , ossia

$$|\mathbf{v}| = \sqrt{x^2 + y^2 + z^2}$$

I vettori  $\mathbf{i} = (1, 0, 0)$ ,  $\mathbf{j} = (0, 1, 0)$ ,  $\mathbf{k} = (0, 0, 1)$  sono versori, mutuamente ortogonali, diretti nel verso positivo dei tre assi, rispettivamente. Ogni altro vettore  $\mathbf{v} = (x, y, z)$  si può scrivere nella forma

$$(1.1) \quad \mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

$\mathbf{i}, \mathbf{j}, \mathbf{k}$  si dicono *versori fondamentali* nello spazio. Le operazioni di somma di vettori e prodotto per uno scalare si possono eseguire componente per componente, come nel caso piano:

$$(x_1, y_1, z_1) + (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2);$$

$$t(x_1, y_1, z_1) = (tx_1, ty_1, tz_1)$$

### Combinazioni lineari di vettori. Vettori linearmente indipendenti

Consideriamo  $k$  vettori  $\mathbf{v}_1, \dots, \mathbf{v}_k$  (nel piano o nello spazio) e  $k$  scalari  $\alpha_1, \dots, \alpha_k$ . Il vettore

$$(1.2) \quad \mathbf{w} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \dots + \alpha_k\mathbf{v}_k$$

si chiama *combinazione lineare* dei vettori  $\mathbf{v}_1, \dots, \mathbf{v}_k$  con coefficienti  $\alpha_1, \dots, \alpha_k$ .

**DEFINIZIONE 8.1** I vettori  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  si dicono *linearmente dipendenti* se (almeno) uno di essi si può esprimere come combinazione lineare degli altri; in caso contrario si dicono *linearmente indipendenti*.

Equivalentemente,  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  sono indipendenti se

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_k \mathbf{v}_k = \mathbf{0}$$

implica  $\alpha_1 = \alpha_2 = \dots = \alpha_k = 0$ . (Il lettore è invitato a dimostrare tale equivalenza.)

Nel caso particolare di 2 vettori  $\mathbf{v}_1, \mathbf{v}_2$ , essi sono linearmente dipendenti se, per qualche scalare  $\alpha$ , si ha

$$\mathbf{v}_1 = \alpha \mathbf{v}_2$$

cioè, se uno è un multiplo dell'altro. Geometricamente, ciò significa che sono paralleli, o (pensando i vettori spiccati dall'origine del sistema di riferimento) si trovano sulla stessa retta.



Figura 8.11. (a) Vettori linearmente dipendenti; (b) vettori indipendenti.

Nel caso di tre vettori  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ , di cui per esempio  $\mathbf{v}_2$  e  $\mathbf{v}_3$  indipendenti, dire che  $\mathbf{v}_1$  è *combinazione lineare di*  $\mathbf{v}_2$  e  $\mathbf{v}_3$  significa geometricamente che  $\mathbf{v}_1$  giace nel piano individuato da  $\mathbf{v}_2$  e  $\mathbf{v}_3$  (quando i 3 vettori siano spiccati dallo stesso punto). Infatti, in tal caso  $\mathbf{v}_1$  è la diagonale del parallelogramma che ha per lati due multipli di  $\mathbf{v}_2, \mathbf{v}_3$ , quindi giace nel piano individuato da  $\mathbf{v}_2, \mathbf{v}_3$  (v. fig. 8.12)

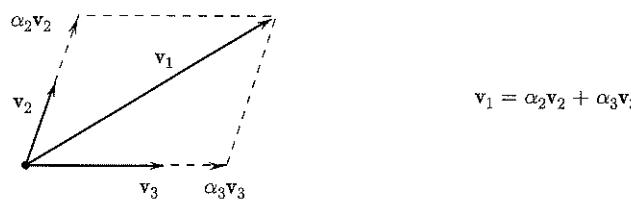


Figura 8.12. Vettori linearmente dipendenti.

Per esempio, nel piano, i vettori  $\mathbf{i} = (1, 0)$  e  $\mathbf{j} = (0, 1)$  sono linearmente indipendenti, come pure, nello spazio, i vettori

$$(1.3) \quad \mathbf{i} = (1, 0, 0) \quad \mathbf{j} = (0, 1, 0) \quad \mathbf{k} = (0, 0, 1)$$

La (1.1) indica che ogni altro vettore  $\mathbf{v}$  dello spazio può essere espresso come combinazione lineare dei vettori (1.3).

In realtà, data una qualunque terna di vettori nello spazio linearmente indipendenti,  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ , ed un quarto vettore  $\mathbf{v}$ , si può sempre scrivere

$$(1.4) \quad \mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3$$

con opportuni scalari  $\alpha_1, \alpha_2, \alpha_3$ .

Analogamente, data una qualunque coppia di vettori nel piano linearmente indipendenti,  $\mathbf{v}_1, \mathbf{v}_2$ , ogni altro vettore  $\mathbf{v}$  può sciversi come combinazione lineare dei primi:

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2$$

per opportuni scalari  $\alpha_1, \alpha_2$ .

Dimostriamo per esempio quest'ultima affermazione. Siano  $\mathbf{v}_1 = \overrightarrow{OA}$ ,  $\mathbf{v}_2 = \overrightarrow{OB}$ ,  $\mathbf{v} = \overrightarrow{OC}$  (v. fig. 8.13). Tracciando da  $C$  le rette parallele ad  $OA$  e  $OB$ , risultano individuati su tali rette due vettori, rispettivamente paralleli a  $\mathbf{v}_1, \mathbf{v}_2$ , la cui somma dà  $\mathbf{v}$  (per la regola del parallelogramma). Ma essendo paralleli a  $\mathbf{v}_1, \mathbf{v}_2$  tali vettori sono del tipo  $\alpha_1 \mathbf{v}_1, \alpha_2 \mathbf{v}_2$ , per opportuni  $\alpha_1, \alpha_2$ . Dunque abbiamo provato che per opportuni  $\alpha_1, \alpha_2$  si ha  $\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2$ .

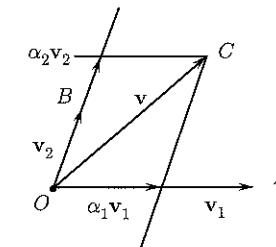


Figura 8.13.

Si lascia allo studente per esercizio la dimostrazione dell'analogia proprietà nello spazio.

Da queste considerazioni segue che nello spazio non vi possono essere più di tre vettori linearmente indipendenti. Analogamente, nel piano, non più di due.

## 1.2 Prodotto scalare e vettoriale

• **Prodotto scalare.** prodotto scalare Dati due vettori  $\mathbf{v}$  e  $\mathbf{w}$  nel piano o nello spazio, il loro *prodotto scalare* o *interno*, denotato con  $\mathbf{v} \cdot \mathbf{w}$  o con  $\langle \mathbf{v}, \mathbf{w} \rangle$ , è assegnato, per definizione, dalla formula seguente:

(1.5)

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}| \cdot |\mathbf{w}| \cdot \cos \alpha$$

dove  $\alpha$  è l'angolo che essi formano ( $0 \leq \alpha \leq \pi$ ). Si noti che il prodotto scalare di 2 vettori è un numero reale (non un vettore).

Il prodotto così definito è *commutativo*:

(1.6)

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$$

ed è *distributivo* rispetto alla somma:

(1.7)

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$$

Inoltre, per ogni  $t \in \mathbb{R}$ , si ha

(1.8)

$$(t\mathbf{v}) \cdot \mathbf{w} = t(\mathbf{v} \cdot \mathbf{w})$$

Si noti poi che

$$\mathbf{v} \cdot \mathbf{v} = |\mathbf{v}|^2$$

e che

$$\mathbf{v} \text{ è perpendicolare a } \mathbf{w} \text{ se e solo se } \mathbf{v} \cdot \mathbf{w} = 0$$

Tutte queste proprietà seguono dalla definizione (1.5) (il lettore rifletta sul perché).

- *Proiezioni.* La proiezione di un vettore  $\mathbf{v}$  su una retta  $r$ , orientata, si chiama *componente vettoriale* di  $\mathbf{v}$  rispetto all'asse  $r$  ed è data dal vettore:

$$(\mathbf{v} \cdot \mathbf{r})\mathbf{r}$$

dove  $\mathbf{r}$  è il *versore* lungo la retta.

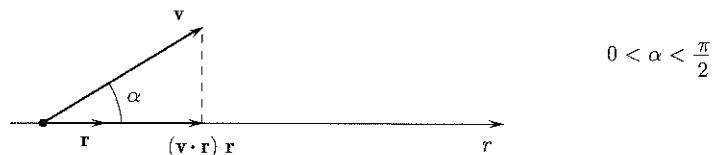


Figura 8.14. Proiezione di  $\mathbf{v}$  sulla retta  $r$ .

Infatti,  $\mathbf{v} \cdot \mathbf{r} = |\mathbf{v}| \cdot |\mathbf{r}| \cos \alpha = |\mathbf{v}| \cos \alpha$  e quindi  $|\mathbf{v} \cdot \mathbf{r}|$  dà la lunghezza del vettore proiezione; il segno di  $\mathbf{v} \cdot \mathbf{r}$  ( $> 0$  se  $\alpha < \frac{\pi}{2}$ ,  $< 0$  se  $\alpha > \frac{\pi}{2}$ ) determina il verso.

Nel piano, i versori  $\mathbf{i}, \mathbf{j}$  sono ortogonali e quindi  $\mathbf{i} \cdot \mathbf{j} = 0$ . In termini di componenti, dati due vettori nel piano

$$\mathbf{v} = x_1 \mathbf{i} + x_2 \mathbf{j} \quad \mathbf{w} = y_1 \mathbf{i} + y_2 \mathbf{j}$$

si ha, usando le proprietà (1.6)-(1.8) del prodotto scalare:

$$\begin{aligned} \mathbf{v} \cdot \mathbf{w} &= (x_1 \mathbf{i} + x_2 \mathbf{j}) \cdot (y_1 \mathbf{i} + y_2 \mathbf{j}) = \\ &= x_1 y_1 \mathbf{i} \cdot \mathbf{i} + x_1 y_2 \mathbf{i} \cdot \mathbf{j} + x_2 y_1 \mathbf{j} \cdot \mathbf{i} + x_2 y_2 \mathbf{j} \cdot \mathbf{j} = x_1 y_1 + x_2 y_2 \end{aligned}$$

da cui l'importante formula:

$$(1.9) \quad \boxed{\mathbf{v} \cdot \mathbf{w} = x_1 y_1 + x_2 y_2} \quad (\text{prodotto scalare nel piano})$$

che dà il prodotto scalare in termini di componenti.

Nello spazio, i versori  $\mathbf{i}, \mathbf{j}, \mathbf{k}$  della base canonica in  $\mathbb{R}^3$  sono mutuamente ortogonali. Con un calcolo perfettamente analogo, se  $\mathbf{v} = x_1 \mathbf{i} + x_2 \mathbf{j} + x_3 \mathbf{k}$ ,  $\mathbf{w} = y_1 \mathbf{i} + y_2 \mathbf{j} + y_3 \mathbf{k}$ , si trova la formula

$$(1.10) \quad \boxed{\mathbf{v} \cdot \mathbf{w} = x_1 y_1 + x_2 y_2 + x_3 y_3} \quad (\text{prodotto scalare nello spazio})$$

Le formule (1.9)-(1.10) risultano molto comode per il calcolo effettivo del prodotto scalare di due vettori di cui si conoscono le componenti, in quanto non richiedono il calcolo esplicito dell'angolo formato dai vettori.



**Lavoro di una forza.** Il lavoro compiuto da una forza costante  $\mathbf{F}$  che sposta il suo punto di applicazione lungo un segmento da  $A$  a  $B$  è definito come il prodotto della lunghezza del segmento  $AB$  per la lunghezza della componente (vettoriale) di  $\mathbf{F}$  lungo la retta  $AB$ . In altri termini, introdotto il vettore spostamento  $\mathbf{s} = \overrightarrow{AB}$ , il lavoro  $L$  è dato dalla formula

$$L = \mathbf{F} \cdot \mathbf{s} = |\mathbf{F}| \cdot |\mathbf{s}| \cdot \cos \alpha$$

dove  $\alpha$  è l'angolo tra la direzione di  $\mathbf{F}$  e quella della retta  $AB$ . Si noti che il lavoro è nullo se  $\mathbf{F}$  è ortogonale alla retta  $AB$ .

- *Prodotto vettoriale nello spazio tridimensionale.* prodotto vettoriale Dati due vettori  $\mathbf{v}$  e  $\mathbf{w}$ , il loro *prodotto vettoriale*, denotato con  $\mathbf{v} \times \mathbf{w}$  (si legge  $\mathbf{v}$  vettor  $\mathbf{w}$ ), è il vettore caratterizzato dalle seguenti proprietà:

- 1) la lunghezza di  $\mathbf{v} \times \mathbf{w}$  è assegnata dalla formula

$$(1.11) \quad \boxed{|\mathbf{v} \times \mathbf{w}| = |\mathbf{v}| \cdot |\mathbf{w}| \sin \alpha}$$

dove  $\alpha$  è l'angolo formato dai due vettori ( $0 \leq \alpha \leq \pi$ );

- 2)  $\mathbf{v} \times \mathbf{w}$  è perpendicolare al piano di  $\mathbf{v}$  e  $\mathbf{w}$ ;
- 3)  $\mathbf{v}, \mathbf{w}$  e  $\mathbf{v} \times \mathbf{w}$ , nell'ordine, formano una terna destrorsa di vettori.

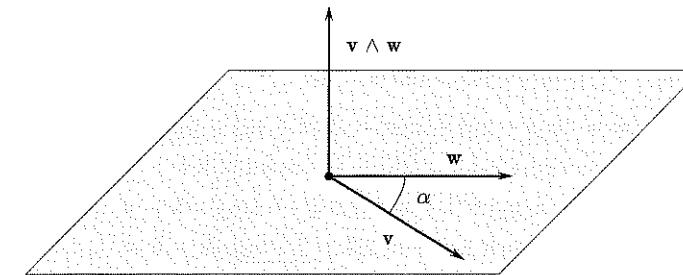


Figura 8.15. Prodotto vettoriale.

Il prodotto vettoriale così definito è *anticommutativo*

$$\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$$

e *distributivo* rispetto alla somma

$$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$$

Inoltre, per ogni  $t \in \mathbb{R}$

$$(t\mathbf{v}) \times \mathbf{w} = t(\mathbf{v} \times \mathbf{w})$$

Si noti poi che

$$\mathbf{v} \times \mathbf{v} = \mathbf{0} \quad (\text{essendo } \alpha = 0)$$

e che

$$\mathbf{v} \text{ è parallelo a } \mathbf{w} \text{ se e solo se } \mathbf{v} \times \mathbf{w} = \mathbf{0}$$

Tutte le proprietà del prodotto vettoriale sopra enunciate seguono dalla definizione; a titolo d'esercizio, il lettore è invitato a verificare questi fatti.

La formula (1.11) indica che la lunghezza di  $\mathbf{v} \times \mathbf{w}$  è uguale all'area del parallelogramma costruito su  $\mathbf{v}$  e  $\mathbf{w}$ .



Figura 8.16. L'area del parallelogramma costruito su  $\mathbf{v}$  e  $\mathbf{w}$  è  $|\mathbf{v}| \cdot |\mathbf{w}| \cdot \sin \alpha = |\mathbf{v} \times \mathbf{w}|$ .

### ESEMPIO

I vettori della base canonica per  $\mathbb{R}^3$ ,  $\mathbf{i}, \mathbf{j}, \mathbf{k}$  formano una terna ortogonale destrorsa e pertanto si ha:

$$(1.12) \quad \begin{aligned} \mathbf{i} \times \mathbf{i} &= 0 & \mathbf{j} \times \mathbf{j} &= 0 & \mathbf{k} \times \mathbf{k} &= 0 \\ \mathbf{i} \times \mathbf{j} &= \mathbf{k} & \mathbf{j} \times \mathbf{k} &= \mathbf{i} & \mathbf{k} \times \mathbf{i} &= \mathbf{j} \end{aligned}$$

Le (1.12) e le proprietà del prodotto vettoriale permettono di scrivere il prodotto vettoriale di due vettori

$$\mathbf{v} = x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k} \quad \mathbf{w} = y_1\mathbf{i} + y_2\mathbf{j} + y_3\mathbf{k}$$

in termini delle loro componenti. Infatti si ha

$$\begin{aligned} \mathbf{v} \times \mathbf{w} &= (x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}) \times (y_1\mathbf{i} + y_2\mathbf{j} + y_3\mathbf{k}) = \\ &= x_1y_1\mathbf{i} \times \mathbf{i} + x_1y_2\mathbf{i} \times \mathbf{j} + x_1y_3\mathbf{i} \times \mathbf{k} + \\ &\quad + x_2y_1\mathbf{j} \times \mathbf{i} + x_2y_2\mathbf{j} \times \mathbf{j} + x_2y_3\mathbf{j} \times \mathbf{k} + \\ &\quad + x_3y_1\mathbf{k} \times \mathbf{i} + x_3y_2\mathbf{k} \times \mathbf{j} + x_3y_3\mathbf{k} \times \mathbf{k} = \\ &= (x_2y_3 - x_3y_2)\mathbf{i} + (x_3y_1 - x_1y_3)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k} \end{aligned}$$

da cui l'importante formula

$$(1.13) \quad \mathbf{v} \times \mathbf{w} = (x_2y_3 - x_3y_2)\mathbf{i} + (x_3y_1 - x_1y_3)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k}$$

Nel paragrafo 4.3 vedremo come questa formula si possa riscrivere in modo più sintetico e facile da ricordare, usando la nozione di *determinante di una matrice*.

### ESEMPIO

**Momento di una forza rispetto ad un punto.**

Sia  $\mathbf{F}$  una forza con punto di applicazione  $A$ . Il momento di  $\mathbf{F}$  rispetto ad un punto  $O$  diverso da  $A$  è definito dal vettore

$$\mathbf{M} = \mathbf{r} \times \mathbf{F}$$

dove  $\mathbf{r} = \overrightarrow{OA}$ .

### Forza in un campo elettromagnetico.

Assegnare un campo di forze significa assegnare in ogni punto dello spazio fisico un vettore forza (che può variare da punto a punto). In un punto fissato, perciò, un campo non è altro che un vettore. (Il concetto di campo vettoriale sarà ripreso in Analisi Matematica 2). Con questa premessa, si consideri questo esempio fisico.

La forza esercitata su una carica  $e$  da un campo elettrico  $\mathbf{E}$  è data da  $\mathbf{F}_e = e\mathbf{E}$ . Se la carica  $e$  si muove con velocità  $\mathbf{v}$  in un campo magnetico  $\mathbf{H}$  è soggetta ad una forza pari a

$$\mathbf{F}_m = \frac{e}{c}(\mathbf{v} \times \mathbf{H}) \quad (c = \text{velocità della luce})$$

La forza totale che agisce sulla carica  $e$  in moto con velocità  $\mathbf{v}$  in un campo elettromagnetico è pertanto

$$\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m = e\mathbf{E} + \frac{e}{c}(\mathbf{v} \times \mathbf{H})$$

- **Prodotto misto nello spazio tridimensionale.** prodotto!misto Se  $\mathbf{u}, \mathbf{v}$  e  $\mathbf{w}$  sono tre vettori nello spazio tridimensionale il loro *prodotto misto* è definito dal numero reale

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$$

La parentesi è in realtà superflua, in quanto  $(\mathbf{u} \cdot \mathbf{v}) \times \mathbf{w}$  non ha senso. Si può perciò scrivere semplicemente  $\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}$ . Mostreremo nel paragrafo 4.3 che il prodotto misto non varia permutando ciclicamente i tre vettori  $\mathbf{u}, \mathbf{v}$  e  $\mathbf{w}$ :

$$\mathbf{u} \cdot \mathbf{v} \times \mathbf{w} = \mathbf{w} \cdot \mathbf{u} \times \mathbf{v} = \mathbf{v} \cdot \mathbf{w} \times \mathbf{u}$$

Geometricamente il valore assoluto del prodotto misto rappresenta il volume del parallelepipedo costruito sui vettori  $\mathbf{v}, \mathbf{w}$  e  $\mathbf{u}$

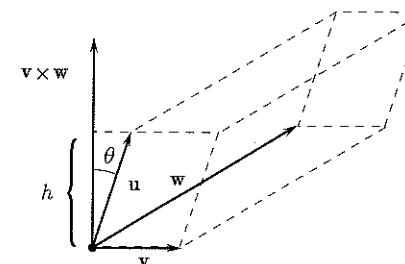


Figura 8.17.  $|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})| = \text{volume del parallelepipedo}$ .

Infatti si ha: volume = altezza · area di base

L'area di base è  $|\mathbf{v} \times \mathbf{w}|$ , mentre l'altezza  $h$  è uguale alla lunghezza della componente di  $\mathbf{u}$  nella direzione di  $\mathbf{v} \times \mathbf{w}$ , perpendicolare alla base. Se l'angolo  $\theta$  tra  $\mathbf{u}$  e  $\mathbf{v} \times \mathbf{w}$  è acuto, si ha  $h = |\mathbf{u}| \cos \theta$  mentre se  $\theta$  è ottuso si ha  $h = -|\mathbf{u}| \cos \theta$ . Dunque: volume parallelepipedo =  $|\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}| = |\mathbf{v} \times \mathbf{w}| \cdot |\mathbf{u}| \cos \theta$

Se il prodotto misto è nullo allora  $\mathbf{u}$  è perpendicolare a  $\mathbf{v} \times \mathbf{w}$  e pertanto giace nel piano individuato da  $\mathbf{v}$  e  $\mathbf{w}$ . Il viceversa è ovvio. Conclusione:

$$\mathbf{u} \cdot \mathbf{v} \times \mathbf{w} = 0 \quad \text{se e solo se } \mathbf{u}, \mathbf{v}, \mathbf{w} \text{ sono complanari}$$

ossia se e solo se  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  sono linearmente dipendenti.

Questo fatto ci fornisce un metodo analitico per verificare se 3 vettori dello spazio siano indipendenti oppure no, mediante un calcolo puramente meccanico. Infatti, siano:

$$\mathbf{u} = (u_1, u_2, u_3) \quad \mathbf{v} = (x_1, x_2, x_3) \quad \mathbf{w} = (y_1, y_2, y_3)$$

Allora, sfruttando le (1.13) e (1.10),abbiamo:

$$\mathbf{u} \cdot \mathbf{v} \times \mathbf{w} = u_1(x_1y_3 - x_3y_1) + u_2(x_3y_1 - x_1y_3) + u_3(x_1y_2 - x_2y_1)$$

e dunque i 3 vettori sono indipendenti se e solo se l'ultima espressione scritta è diversa da zero.



**ESEMPIO** Verificare se i vettori  $(1, 2, 3), (0, 2, 2), (2, 1, 2)$  sono indipendenti oppure no. Calcoliammo:

$$(0, 2, 2) \cdot (1, 2, 3) \times (2, 1, 2) = 2(3 \cdot 2 - 1 \cdot 2) + 2(1 \cdot 1 - 2 \cdot 2) = \\ = 2 \cdot 4 + 2 \cdot (-3) = 2 \neq 0.$$

I vettori sono perciò *indipendenti*.

## 2 GEOMETRIA ANALITICA LINEARE NELLO SPAZIO

Mediante il calcolo vettoriale introdotto nei paragrafi precedenti è facile scrivere, in un dato sistema di riferimento, vari tipi di equazioni per rette e piani.

### Equazione della retta

Una retta nello spazio è individuata da:

- a) un punto e un vettore direzionale, oppure
- b) due punti, oppure
- c) due piani non paralleli di cui è intersezione.

Cominciamo col caso a). Consideriamo un punto  $P_0 = (x_0, y_0, z_0)$  ed un vettore  $\mathbf{v} = (a, b, c)$  non nullo; ci proponiamo di scrivere l'equazione della retta *passante* per  $P_0$ , parallela a  $\mathbf{v}$ .

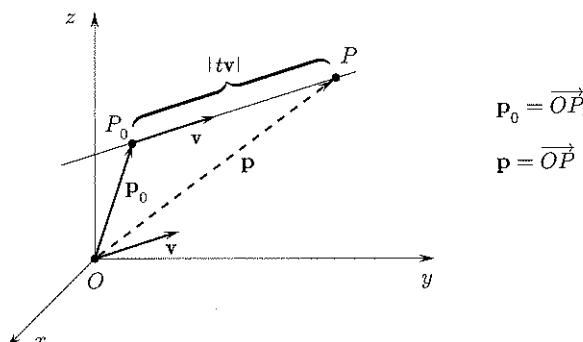


Figura 8.18.

Dalla figura si vede che un generico punto  $P = (x, y, z)$ , corrente sulla retta, si trova aggiungendo al vettore posizione  $\mathbf{p}_0 = \overrightarrow{OP_0}$  un opportuno multiplo  $t\mathbf{v}$  di  $\mathbf{v}$ .

Abbiamo quindi per il vettore posizione  $\mathbf{p} = \overrightarrow{OP}$  del generico punto sulla retta:

(2.1)

$$\mathbf{p} = \mathbf{p}_0 + t\mathbf{v} \quad t \in \mathbb{R}$$

che prende il nome di *equazione parametrica vettoriale* della retta;  $t$  è una "coordinata" sulla retta, corrispondente alla scelta di  $P_0$  come origine. Le *equazioni parametriche scalari* si ricavano dalla (2.1) scrivendo componente per componente:

$$(2.2) \quad \begin{cases} x = x_0 + ta \\ y = y_0 + tb \\ z = z_0 + tc \end{cases} \quad t \in \mathbb{R}$$

Il significato delle equazioni parametriche (2.2) (o analogamente dell'equazione parametrica vettoriale (2.1)) è il seguente: al variare del *parametro*  $t$  in  $\mathbb{R}$ , il punto  $(x, y, z)$  si muove sulla retta, descrivendo tale linea.

Se  $a \neq 0, b \neq 0, c \neq 0$ , si può eliminare il parametro  $t$  nelle (2.2) ottenendo

$$\frac{x - x_0}{a} = t \quad \frac{y - y_0}{b} = t \quad \frac{z - z_0}{c} = t$$

ossia

$$(2.3) \quad \frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

che sono le *equazioni cartesiane* della retta.

Se uno o due tra i parametri  $a, b, c$  è nullo, occorre modificare la (2.3). Per esempio se  $a = 0, b \neq 0, c \neq 0$  si troverebbe

$$x = x_0 \quad \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

Nel caso in cui  $a^2 + b^2 + c^2 = 1$ , cioè nel caso in cui  $\mathbf{v}$  è un *versore*, i numeri  $a, b, c$  prendono il nome di *coseni direttori della retta*, in quanto rappresentano i coseni degli angoli che la retta forma con ciascuno degli assi coordinati (v. fig. 8.19).

Nel caso b) si vuole scrivere l'equazione della retta passante per i punti

$$P_0 = (x_0, y_0, z_0) \quad \text{e} \quad P_1 = (x_1, y_1, z_1)$$

Riconduciamoci al caso a) scegliendo uno dei due punti, per esempio  $P_0$ , e il vettore direzionale

$$\mathbf{v} = (x_1 - x_0, y_1 - y_0, z_1 - z_0)$$

Le equazioni parametriche della retta risultano allora

$$\begin{cases} x = x_0 + t(x_1 - x_0) \\ y = y_0 + t(y_1 - y_0) \\ z = z_0 + t(z_1 - z_0) \end{cases}$$

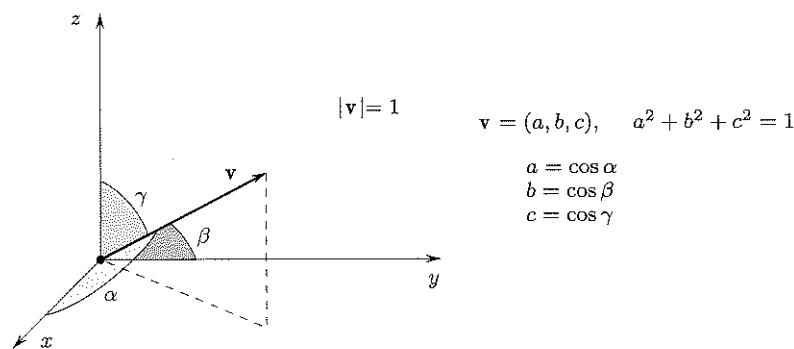


Figura 8.19.

- Condizioni di parallelismo e ortogonalità tra due rette.** Due rette sono *parallele* se i loro vettori direzionali sono paralleli (ossia linearmente dipendenti). Precisamente, le due rette

$$\begin{cases} x = x_0 + ta \\ y = y_0 + tb \\ z = z_0 + tc \end{cases} \quad \text{e} \quad \begin{cases} x = x_1 + ta' \\ y = y_1 + tb' \\ z = z_1 + tc' \end{cases}$$

sono *parallele* se esiste un numero  $\lambda \neq 0$  tale che

$$(a', b', c') = \lambda(a, b, c)$$

Le due rette sono *ortogonali* se lo sono i loro vettori direzionali e cioè se

$$(a', b', c') \cdot (a, b, c) = aa' + bb' + cc' \equiv 0$$

Si noti che per essere ortogonali non è necessario che le due rette siano incidenti (cioè si intersechino in un punto).



Si voglia calcolare l'equazione della retta passante per i due punti  $P_0 = (3, -1, 1)$ ,  $P_1 = (4, 2, -3)$ . Come vettore direzionale si può prendere  $\mathbf{v} = \overrightarrow{P_0 P_1} = (1, 3, -4)$ .

L'equazione parametrica vettoriale è dunque

$$(x, y, z) = (3, -1, 1) + t(1, 3, -4) \quad t \in \mathbb{R}$$

Le equazioni scalari sono

$$\begin{cases} x = 3 + t \\ y = -1 + 3t \\ z = 1 - 4t \end{cases} \quad t \in \mathbb{R}$$

Infine, quelle cartesiane sono

$$\frac{x-3}{1} = \frac{y+1}{3} = \frac{z-1}{-4}$$

Questa retta è parallela alla retta

$$\begin{cases} x = 2 + 4t \\ y = 1 + 12t \\ z = -16t \end{cases}$$

in quanto  $(4, 12, -16) = 4(1, 3, -4)$ , e ortogonale alla retta

$$\begin{cases} x = 2 - 2t \\ y = 2t \\ z = 5 + t \end{cases}$$

in quanto  $(-2, 2, 1) \cdot (1, 3, -4) = -2 + 6 - 4 = 0$ .

Prima di vedere come si scrive l'equazione di una retta come intersezione di due piani (caso c), vediamo come si scrive l'equazione di un piano.

#### Equazione del piano

Un piano è individuato da:

- un punto e un vettore ortogonale al piano stesso, oppure
- tre punti, oppure
- due rette incidenti

Cominciamo col caso a). Consideriamo un punto  $P_0 = (x_0, y_0, z_0)$  ed un vettore  $\mathbf{n} = (a, b, c)$  non nullo; ci proponiamo di determinare l'equazione del piano passante per  $P_0$  e ortogonale a  $\mathbf{n}$ .

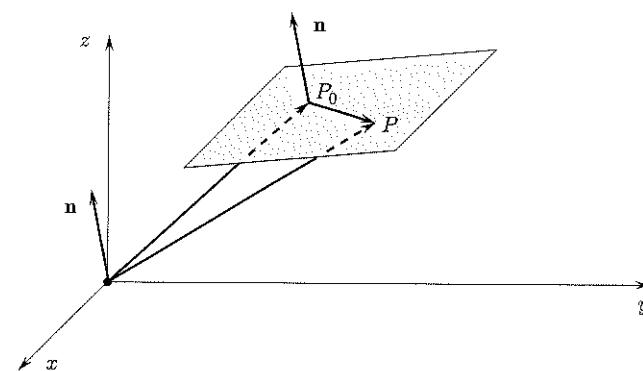


Figura 8.20.

Dalla figura si vede che se  $P = (x, y, z)$  è il generico punto del piano, allora il vettore  $\overrightarrow{P_0 P}$  è parallelo al piano e quindi perpendicolare a  $\mathbf{n}$ . Si ottiene allora l'equazione vettoriale:

(2.4)

$$\mathbf{n} \cdot \overrightarrow{P_0 P} = 0$$

L'equazione cartesiana del piano si ottiene passando dai vettori alle coordinate, cioè:

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

ossia

$$(2.5) \quad \boxed{\begin{aligned} ax + by + cz &= d \\ \text{dove } d &= \mathbf{n} \cdot \mathbf{p}_0 = ax_0 + by_0 + cz_0 \end{aligned}}$$

Facciamo qualche osservazione sulla (2.5).

1. Se  $a^2 + b^2 + c^2 = 1$ , cioè  $\mathbf{n}$  è un versore,  $a, b, c$  sono detti *coseni direttori* del piano.
2. Se  $d = 0$ , il piano passa per l'origine  $(0, 0, 0)$ .
3. Se uno dei coefficienti  $a, b$ , o  $c$  è nullo, il piano è parallelo ad uno degli assi coordinati. Per esempio se  $c = 0$  il piano è parallelo all'asse  $z$ , ossia perpendicolare al piano  $(x, y)$  (figura 8.21a).
- Si noti che l'equazione  $ax + by = d$ , che nel piano rappresenta una retta, nello spazio rappresenta un piano perpendicolare al piano  $x, y$ .
- Analoghe considerazioni valgono nei casi  $a = 0$  oppure  $b = 0$ .
4. Se due dei coefficienti sono nulli, allora il piano è parallelo ad uno dei piani coordinati. Per esempio, se  $a = b = 0$ , il piano ha equazione  $z = k$  ed è parallelo al piano  $x, y$  (figura 8.21b).
5. Se  $d \neq 0$ , si può dividere la (2.5) per  $d$ , ottenendo l'equazione del piano nella forma

$$(2.6) \quad \frac{x}{\alpha} + \frac{y}{\beta} + \frac{z}{\gamma} = 1$$

In tal caso  $\alpha, \beta, \gamma$  rappresentano le intercette del piano con gli assi coordinati (figura 8.21c).

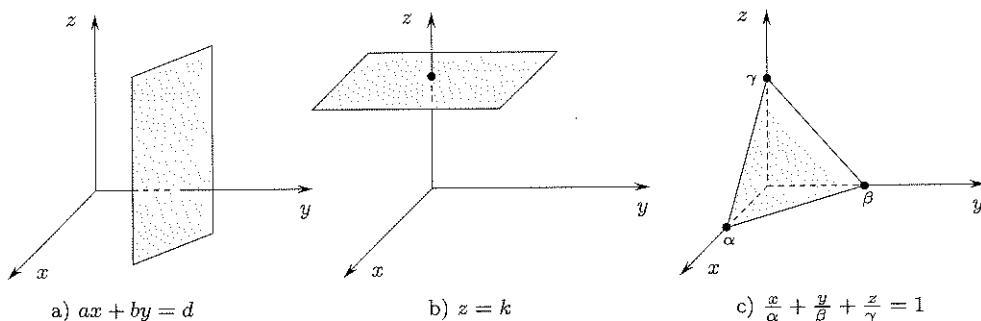


Figura 8.21.



Si voglia scrivere l'equazione del piano passante per il punto  $(1, 2, -1)$  e normale al vettore  $\mathbf{v} = (2, 3, 0)$ . La (2.4) diventa:

$$(x - 1, y - 2, z + 3) \cdot (2, 3, 0) = 0$$

cioè

$$\begin{aligned} 2(x - 1) + 3(y - 2) &= 0 \\ 2x + 3y - 8 &= 0. \end{aligned}$$

Nel caso b) si vuole scrivere l'equazione del piano passante per i punti

$$P_0 = (x_0, y_0, z_0) \quad P_1 = (x_1, y_1, z_1) \quad P_2 = (x_2, y_2, z_2)$$

Per ricondursi al caso a) individuiamo un vettore ortogonale al piano.

I vettori  $\mathbf{v} = \overrightarrow{P_0 P_1} = (x_1 - x_0, y_1 - y_0, z_1 - z_0)$  e  $\mathbf{w} = \overrightarrow{P_0 P_2} = (x_2 - x_0, y_2 - y_0, z_2 - z_0)$  sono paralleli al piano e perciò il loro prodotto vettoriale sarà ortogonale al piano stesso. Poniamo

$$\mathbf{n} = \mathbf{v} \times \mathbf{w}$$

Scegliendo poi, per esempio,  $P_0$ , l'equazione del piano è

$$\mathbf{n} \cdot \overrightarrow{P_0 P} = 0$$



Si voglia calcolare l'equazione del piano passante per i tre punti  $A = (1, 0, -1)$ ,  $B = (2, 1, 0)$ ,  $C = (0, 1, -2)$ . Per usare la formula (2.4) occorre individuare un vettore normale al piano. Poiché i vettori  $\mathbf{v} = \overrightarrow{AB} = (1, 1, 1)$ ,  $\mathbf{w} = \overrightarrow{AC} = (-1, 1, -1)$  sono paralleli al piano, il loro prodotto vettoriale sarà ortogonale al piano stesso. Sceglieremo perciò

$$\mathbf{n} = \mathbf{v} \times \mathbf{w} = -2\mathbf{i} + 2\mathbf{k}$$

Abbiamo quindi, scegliendo  $P_0 = A$ :

$$-2(x - 1) + 2(z + 1) = 0$$

ossia

$$-x + z + 2 = 0$$

Nel caso c), per scrivere l'equazione del piano contentente le due rette  $\mathbf{p} = \mathbf{p}_0 + t\mathbf{v}$ ,  $\mathbf{p} = \mathbf{p}_0 + t\mathbf{w}$ , incidenti in  $P_0$  ma distinte ( $\mathbf{v}, \mathbf{w}$  indipendenti), è sufficiente applicare il ragionamento del caso b, ottenendo l'equazione:

$$(\mathbf{p} - \mathbf{p}_0) \cdot (\mathbf{v} \times \mathbf{w}) = 0.$$



Stabilire se le seguenti rette

$$P = (4, -2, 2) + t(1, -1, 0)$$

$$P = (1, 1, 2) + t(2, 1, 3)$$

sono complanari o sghembe e, nel primo caso, scrivere l'equazione del piano che le contiene.

I vettori  $(1, -1, 0)$ ,  $(2, 1, 3)$  evidentemente non sono paralleli, perciò le rette sono complanari se e solo se sono incidenti. Stabiliamo se lo sono cercando se esistono valori dei parametri  $t, u$  per cui si abbia

$$(4, -2, 2) + t(1, -1, 0) = (1, 1, 2) + u(2, 1, 3).$$

Questo porta un sistema di 3 equazioni in 2 incognite  $t, u$ :

$$\begin{cases} t - 2u = -3 \\ -t - u = 3 \\ 3u = 0 \end{cases}$$

che dà  $u = 0, t = -3$ . Per questi valori dei parametri le rette si incontrano in  $P_0 = (1, 1, 2)$ , dunque le rette sono incidenti e le loro equazioni si possono riscrivere, evidenziando il punto comune  $P_0$ , nella forma:

$$\begin{aligned} P &= (1, 1, 2) + t(1, -1, 0) \\ &= (1, 1, 2) + t(2, 1, 3). \end{aligned}$$

A questo punto la determinazione del piano che le contiene si può fare come illustrato al punto c):

$$(x - 1, y - 1, z - 2) \cdot ((1, -1, 0) \times (2, 1, 3)) = 0$$

che dà

$$-3(x - 1) - 3(y - 1) + 3(z - 2) = 0$$

e quindi

$$z - x - y = 0.$$

• *Parallelismo e ortogonalità tra due piani.* Due piani sono paralleli se lo sono i loro vettori ortogonali. Precisamente, i piani

$$ax + by + cz = d \quad a'x + b'y + c'z = d'$$

sono paralleli se esiste  $\lambda \neq 0$  tale che

$$(a, b, c) = \lambda(a', b', c')$$

Sono ortogonali se lo sono i loro vettori normali:

$$(a, b, c) \cdot (a', b', c') = aa' + bb' + cc' = 0$$

Possiamo ora trattare il caso rimasto in sospeso nella discussione dei modi di scrivere l'equazione di una retta.

• *Retta intersezione di due piani.* Siano

$$ax + by + cz = d$$

e

$$a'x + b'y + c'z = d'$$

le equazioni di due piani non paralleli.

Essi si intersecano lungo una retta rappresentata dal sistema

$$\begin{cases} ax + by + cz = d \\ a'x + b'y + c'z = d' \end{cases}$$

Un vettore direzionale della retta deve essere ortogonale ad entrambi i vettori

$$\mathbf{n} = (a, b, c) \quad \text{e} \quad \mathbf{n}' = (a', b', c')$$

Si può perciò scegliere come vettore direzionale della retta

$$\mathbf{v} = \mathbf{n} \times \mathbf{n}' = (bc' - b'c, a'c - ac', ab' - a'b)$$

In alternativa, si possono scrivere le equazioni parametriche della retta col seguente procedimento: si considera il sistema di 2 equazioni nelle 3 incognite  $x, y, z$ , che definisce la retta; se in tale sistema si riescono ad esprimere due variabili in funzione della terza, si può assumere la terza variabile come parametro.



Scriviamo le equazioni parametriche della retta intersezione dei piani:

$$x + y + z = 1; \quad x - 2y - z = 2.$$

Dal sistema:

$$\begin{cases} x + y + z = 1 \\ x - 2y - z = 2 \end{cases}$$

ricaviamo:

$$\begin{cases} x + y = 1 - z \\ x - 2y = 2 + z \end{cases}$$

e quindi, risolvendo in  $x, y$  rispetto a  $z$ :

$$\begin{cases} y = \frac{-1 - 2z}{3} \\ x = \frac{4 - z}{3} \end{cases}$$

da cui, ponendo  $z = t$  otteniamo le equazioni parametriche della retta:

$$\begin{cases} x = \frac{4}{3} - \frac{1}{3}t \\ y = -\frac{1}{3} - \frac{2}{3}t \\ z = t \end{cases}$$

• *Distanza di un punto da un piano.* Si voglia calcolare la distanza  $\delta$  di un punto  $P_1 = (x_1, y_1, z_1)$  dal piano passante per  $P_0 = (x_0, y_0, z_0)$  e ortogonale al vettore  $\mathbf{n} = (a, b, c)$ . Dalla figura si vede che  $\delta$  è la lunghezza del segmento  $P_2P_1$ , uguale alla lunghezza della proiezione di  $P_0P_1$  lungo la direzione  $\mathbf{n}$ .

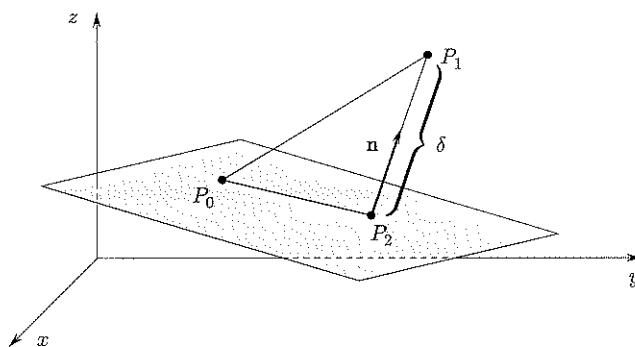


Figura 8.22.

Se si introduce il versore  $\mathbf{N} = \frac{1}{\sqrt{a^2+b^2+c^2}}(a,b,c)$ , la lunghezza della proiezione è il modulo del prodotto scalare tra  $\mathbf{N}$  ed il vettore  $\overrightarrow{P_0P_1}$ . Cioè:

$$\delta = \left| \mathbf{N} \cdot \overrightarrow{P_0P_1} \right| = \frac{|a(x_1 - x_0) + b(y_1 - y_0) + c(z_1 - z_0)|}{\sqrt{a^2 + b^2 + c^2}}$$

ossia, ricordando che abbiamo posto  $d = ax_0 + by_0 + cz_0$ :

$$(2.7) \quad \boxed{\delta = \frac{|ax_1 + by_1 + cz_1 - d|}{\sqrt{a^2 + b^2 + c^2}}}$$



Considerare i seguenti gruppi di vettori e dire, per ognuno, se i vettori che lo compongono sono linearmente indipendenti. In caso negativo, esprimere uno di questi come combinazione lineare degli altri due.

- a)  $(1, 0)$     $(0, 1)$     $(1, -1)$
- b)  $(2, -5, 1)$     $(2, -5, 0)$     $(0, -5, 1)$
- c)  $(1, 1, 1)$     $(2, -1, 4)$     $(4, -1, 6)$

Dati i vettori  $(1, 1, -1)$  e  $(0, 2, 1)$  verificare se il vettore  $(1, 1, 1)$  si può ottenere come combinazione lineare di essi.

Dati i vettori:

$$\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} - \mathbf{k} \quad \mathbf{u} = -\mathbf{i} + 2\mathbf{k}$$

calcolare  $\mathbf{u} \cdot \mathbf{v}$ ,  $\mathbf{u} \times \mathbf{v}$ ,  $\mathbf{j} \cdot \mathbf{u} \times \mathbf{v}$ ,  $\text{vers}(\mathbf{u} \times \mathbf{v})$ .

Calcolare il volume del parallelepipedo costruito sui vettori  $\mathbf{i} + \mathbf{j}$ ,  $\mathbf{i}$ ,  $\mathbf{j} + \mathbf{k}$ .

Tra i seguenti vettori, individuare eventuali coppie di vettori paralleli o perpendicolari:

$$(2, -1, 1) \quad (1, 2, -1) \quad (1, 4, 2) \quad (2, 4, -2)$$

Dati i vettori:

$$\mathbf{v} = (1, 2, 3) \quad \mathbf{w} = (-2, 0, 1)$$

calcolare: il vettore  $\mathbf{v} + \mathbf{w}$ ; il vettore  $2\mathbf{v} - 3\mathbf{w}$ ; il modulo del vettore  $\mathbf{v} + \mathbf{w}$ ; il versore del vettore  $2\mathbf{v} - 3\mathbf{w}$ .

Normalizzare i seguenti vettori del piano o dello spazio:

$$(3, 4) \quad (-1, 3, 2) \quad \left(2, \sqrt{2}, -\sqrt{2}\right)$$

Dimostrare la proprietà associativa della somma di vettori nel piano, con un ragionamento geometrico basato sulla definizione di somma e le proprietà dei parallelogrammi.

Dati i vettori:

$$\mathbf{v} = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k} \quad \mathbf{w} = -2\mathbf{i} + \mathbf{k}$$

calcolare il prodotto scalare  $\mathbf{v} \cdot \mathbf{w}$ . Qual è l'angolo formato tra i due vettori? Calcolare poi il prodotto vettoriale  $\mathbf{v} \times \mathbf{w}$  e il prodotto misto  $(\mathbf{i} + \mathbf{j}) \cdot \mathbf{v} \times \mathbf{w}$ .

Scrivere l'equazione parametrica vettoriale, e le equazioni parametriche scalari, delle seguenti rette:

- a) La retta  $r_1$  passante per i punti  $(1, 3, 2)$  e  $(0, 4, 1)$
- b) La retta  $r_2$  passante per il punto  $(1, 3, 2)$  e avente la direzione del vettore  $\mathbf{v} = \mathbf{i} - 3\mathbf{k}$ .
- c) La retta  $r_3$  intersezione dei piani  $x + 3y - 2z = 1$  e  $2x - y + z = 0$ .
- d) La retta  $r_4$  passante per  $(2, 1, 0)$  e parallela alla retta  $r_1$ .

Scrivere l'equazione cartesiana dei seguenti piani:

- a) Il piano  $\pi_1$  passante per  $(1, 2, -1)$  e ortogonale al vettore  $\mathbf{i} + 2\mathbf{j}$ .
- b) Il piano  $\pi_2$  passante per i punti  $(1, 2, -1)$ ,  $(0, 1, 2)$ ,  $(0, 0, 1)$ .
- c) Il piano  $\pi_3$  contenente le 2 rette:

$$r : \begin{cases} x = 1 + 2t \\ y = 3t \\ z = 2 - t \end{cases} \quad s : \begin{cases} x = 3 - 2u \\ y = 3 + u \\ z = 1 + 2u \end{cases}$$

- d) Il piano  $\pi_4$  passante per  $(1, 0, -1)$  e parallelo al piano  $\pi_1$ .

Stabilire quale tra i seguenti punti ha distanza minima dal piano di equazione

$$x - 2y + 3z = 1$$

$$A = (0, 0, 0); B = (1, 1, 1); C = (2, 1, -1).$$

**E** Nel piano  $xy$  scrivere l'equazione

- dell'asse del segmento<sup>1</sup> avente per estremi i punti  $P_1(0, 2)$  e  $P_2(1, 1)$
- della retta perpendicolare all'asse trovato e passante per l'origine
- della retta parallela all'asse trovato e passante per  $P_1$ .

**E** Scrivere l'equazione della retta

- congiungente i punti  $(0, 0, 0)$  e  $(1, -1, 1)$
- congiungente i punti  $(2, 0, -1)$  e  $(3, -5, 2)$
- passante per il punto  $(1, -2, 3)$  e parallela alla retta di equazione:  $x = y = z$
- passante per il punto  $(3, -1, -1)$  e perpendicolare al piano di equazione  $x + y + z = 2$ .

**E** Scrivere, in forma cartesiana, l'equazione del piano

- determinato dai punti  $(0, 0, 0)$ ,  $(1, 1, 1)$ ,  $(3, -2, -1)$
- passante per il punto  $(2, 1, 3)$  e parallelo al piano determinato in i).

**E** Sia  $P_0 = (-1, 4, 2)$ . Scrivere in forma cartesiana l'equazione della retta passante per  $P_0$  e parallela alla retta

$$\begin{cases} x = t \\ y = 2t \\ z = 3t \end{cases}$$

**E** Stabilire se le due rette

$$r_1 : \begin{cases} x = 1 + 2t \\ y = 3 - t \\ z = -t \end{cases} \quad r_2 : \begin{cases} x = -\frac{1}{2}s \\ y = 2s \\ z = -1 + s \end{cases}$$

sono incidenti. In caso affermativo calcolare le coordinate del punto di intersezione e l'angolo tra esse.

**E** Sia  $r$  la retta di equazione

$$4x = 2y - 1 = -2z$$

Determinare

- il piano che contiene  $r$  e il punto  $P = (1, 0, 2)$
- il piano che contiene  $r$  ed è ortogonale al piano  $x + 2z = 0$ .

**E** Scrivere la condizione di parallelismo e di perpendicolarità tra la retta

$$\begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = z_0 + ct \end{cases}$$

ed il piano di equazione

$$a'x + b'y + c'z = d'$$

<sup>1</sup>Ricordiamo che l'asse di un segmento è la retta perpendicolare al segmento e passante per il suo punto medio.

### 3 SPAZI VETTORIALI

In questo paragrafo iniziamo lo studio dei primi elementi di *algebra lineare*, che proseguiremo nei paragrafi 4-6. Si tratta di una disciplina che viene utilizzata sia in altri rami della matematica (geometria, analisi, calcolo numerico...) che nelle discipline applicative vere e proprie. In particolare, lo studente utilizzerà questi concetti nel corso di Analisi Matematica 2 quando affronterà lo studio delle equazioni differenziali, il calcolo differenziale per funzioni di più variabili e le serie di Fourier. L'idea centrale di questo capitolo, come si vedrà, è quella di *linearità*.

Generalizzando via via la nozione elementare di vettore nel piano e nello spazio, che è stata introdotta nel paragrafo 1, intordurremo anzitutto (par. 3) la nozione di *spazio vettoriale* astratto, idea centrale in matematica, che permette di descrivere unitariamente molte situazioni diverse, e definiremo le *trasformazioni lineari* tra spazi vettoriali. Studieremo quindi il *calcolo matriciale* (par. 4), strumento basilare per lo studio di queste trasformazioni. Infine applicheremo questi concetti allo studio dei *sistemi lineari* (par. 5) e al problema della *diagonalizzazione di una trasformazione* (par. 6). Come si vedrà subito, nello studio dell'algebra lineare è spesso presente, ed è utile ricercare, una certa intuizione di tipo geometrico.

#### 3.1 Vettori $n$ -dimensionali: lo spazio $\mathbb{R}^n$ . Spazi vettoriali astratti

Nel paragrafo 1 abbiamo visto come i vettori nel piano e nello spazio si possano identificare, previa la scelta di un sistema di riferimento cartesiano, con coppie o terne ordinate di numeri reali, rispettivamente. Una volta fatta questa identificazione, è possibile eseguire le operazioni fondamentali sui vettori (somma e prodotto per uno scalare) operando direttamente su queste coppie (o terne):

$$(v_1, v_2) + (u_1, u_2) = (v_1 + u_1, v_2 + u_2) \\ \lambda(v_1, v_2) = (\lambda v_1, \lambda v_2)$$

Questi fatti suggeriscono la possibilità di considerare  $n$ -uple ordinate di numeri reali come vettori di uno spazio astratto a  $n$  dimensioni.

#### Lo spazio $\mathbb{R}^n$

Consideriamo dunque l'insieme  $\mathbb{R}^n$  di tutte le  $n$ -uple ordinate di numeri reali:

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) : x_i \in \mathbb{R}\}$$

(Ricordiamo che il simbolo  $\mathbb{R}^n$  è una abbreviazione del prodotto cartesiano  $\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$   $n$  volte.) Potremo indicare un elemento di  $\mathbb{R}^n$  con  $\mathbf{x}, \mathbf{y}, \dots$  (o  $\underline{x}, \underline{y}, \dots$ ). Se, per esempio,  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , diremo che  $x_i$  sono le *componenti* del vettore  $\mathbf{x}$ . È possibile definire in modo naturale la somma di due vettori e la moltiplicazione per uno scalare:

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

$$\lambda(x_1, x_2, \dots, x_n) = (\lambda x_1, \lambda x_2, \dots, \lambda x_n)$$

Si noti il diverso punto di vista che stiamo adottando, rispetto al paragrafo 1: lì, dopo aver definito la somma di due vettori nel piano per via geometrica ("regola del parallelogramma"), si dimostrava che la somma si poteva calcolare, rappresentando i vettori

come coppie ordinate di numeri reali, sommando componente per componente. Qui invece, non essendo visualizzabili i vettori  $n$ -dimensionali (quando  $n > 3$ ), l'analoga formula della somma componente per componente è la *definizione* stessa di somma di due  $n$ -uple. Possiamo anche dire che i vettori  $n$ -dimensionali sono enti algebrici, che però condividono molte proprietà degli enti geometrici elementari studiati nel piano e nello spazio. Pertanto, nello studio di  $\mathbb{R}^n$ , la terminologia e le intuizioni geometriche legate allo spazio ordinario risultano spesso utili.

Le operazioni di somma di due vettori e di prodotto per uno scalare godono delle stesse proprietà formali che abbiamo evidenziato nel paragrafo 1, come si verifica immediatamente in base alla definizione:

- *Proprietà della somma.*

- proprietà associativa:

$$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$$

- proprietà commutativa:

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$

- esiste un vettore  $\mathbf{0}$  tale che:

$$\mathbf{v} + \mathbf{0} = \mathbf{v}$$

- per ogni vettore  $\mathbf{v}$  esiste un vettore  $-\mathbf{v}$  tale che

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}.$$

Precisamente: il vettore nullo è  $\mathbf{0} = (0, 0, \dots, 0)$ , e l'opposto del vettore  $\mathbf{v} = (x_1, x_2, \dots, x_n)$  è il vettore  $-\mathbf{v} = (-x_1, -x_2, \dots, -x_n)$ .

- *Proprietà del prodotto per uno scalare:*

$$\begin{aligned} 1 \cdot \mathbf{v} &= \mathbf{v} \\ s(tv) &= (st)\mathbf{v} \\ t(\mathbf{v} + \mathbf{w}) &= tv + tw \\ (s+t)\mathbf{v} &= sv + tv \end{aligned}$$

L'insieme dei vettori del piano e dello spazio possono essere identificati, rispettivamente, con  $\mathbb{R}^2$  e  $\mathbb{R}^3$ .

### Spazi vettoriali astratti

Generalizzando ancora, si può dare una definizione astratta di spazio vettoriale, che si rivela utile in matematica perché consente di trattare unitariamente molte situazioni diverse:

**DEFINIZIONE 8.2** Si dice spazio vettoriale su un campo numerico  $\mathbb{K}$  (che per noi sarà  $\mathbb{R}$  o  $\mathbb{C}$ ) un insieme  $V$  di elementi a priori generici, per i quali sono definite:

- un'operazione di somma che associa ad ogni coppia di elementi di  $V$  un altro (unico) elemento di  $V$ ;

– un'operazione di prodotto che associa ad ogni coppia formata da un elemento di  $V$  e da un numero appartenente a  $\mathbb{K}$  un altro (unico) elemento di  $V$ .

Le operazioni di somma e prodotto così definite devono possedere tutte le proprietà che abbiamo elencato per le operazioni sui vettori di  $\mathbb{R}^n$ . Gli elementi di  $V$  si chiamano vettori, gli elementi di  $\mathbb{K}$  si chiamano scalari.

Naturalmente, l'insieme dei vettori nel piano, l'insieme dei vettori nello spazio, lo spazio  $\mathbb{R}^n$ , sono esempi di spazi vettoriali sul campo  $\mathbb{R}$ . L'esempio più naturale di spazio vettoriale sul campo  $\mathbb{C}$ , invece, è lo spazio  $\mathbb{C}^n$  delle  $n$ -uple ordinate di numeri complessi: due  $n$ -uple di numeri complessi si possono sommare tra loro componente per componente; una  $n$ -upla si può moltiplicare per un numero complesso.

Per semplicità, nel seguito ci riferiremo, salvo avviso contrario, a spazi vettoriali su  $\mathbb{R}$ . Vi sono spazi vettoriali di natura molto diversa da  $\mathbb{R}^n$  o  $\mathbb{C}^n$ , come già mostra il prossimo esempio.



Sia  $V$  l'insieme dei polinomi in una variabile. Questo insieme, munito dell'usuale operazione di somma di polinomi e dell'usuale operazione di prodotto per un numero reale, soddisfa tutte le proprietà richieste: è uno spazio vettoriale.

Come si vede, i suoi elementi sono ben diversi dai vettori nel piano o nello spazio (per esempio, non si parla di "direzione" e "verso" di un polinomio!); tuttavia, le proprietà algebriche che riguardano le due operazioni specificate sono le stesse.

Si noti che in uno spazio vettoriale "astratto" non sono definite in generale tutte le operazioni che possiamo fare tra due vettori del piano o dello spazio (come il prodotto scalare e il prodotto vettoriale), ma solo la somma di vettori e il prodotto di un vettore per uno scalare.

**DEFINIZIONE 8.3** Sia  $V$  uno spazio vettoriale e  $V_1$  un sottoinsieme di  $V$ . Se  $V_1$ , munito delle stesse operazioni definite in  $V$ , risulta essere anch'esso uno spazio vettoriale, diremo che  $V_1$  è un sottospazio vettoriale (o più brevemente un sottospazio) di  $V$ .

• *Criterio di riconoscimento dei sottospazi.* Per verificare che un sottoinsieme  $V_1$  di uno spazio vettoriale  $V$  è un sottospazio non è necessario verificare che le operazioni abbiano le proprietà richieste (associativa...): se queste proprietà valgono in tutto  $V$ , a maggior ragione varranno in  $V_1$ . Invece, occorre verificare che eseguendo tali operazioni su elementi di  $V_1$  non si esca da  $V_1$ , ovvero che:

$$\text{per ogni } \mathbf{v}_1, \mathbf{v}_2 \in V_1, \lambda \in \mathbb{R} \quad \text{si ha che } \mathbf{v}_1 + \mathbf{v}_2 \in V_1 \text{ e } \lambda\mathbf{v}_1 \in V_1$$



Sia  $V$  lo spazio vettoriale dei polinomi di una variabile, considerato nell'esempio precedente, e sia  $V_1$  l'insieme dei polinomi di grado  $\leq 2$  (compreso il polinomio identicamente nullo). Poiché la somma di due polinomi di grado  $\leq 2$  e il prodotto di una costante per un polinomio di grado  $\leq 2$  sono ancora polinomi di grado  $\leq 2$ ,  $V_1$  risulta un sottospazio vettoriale di  $V$ .

**Indipendenza lineare, base e dimensione**

In uno spazio vettoriale qualunque si può definire la nozione di *combinazione lineare di vettori*, e quella di *indipendenza lineare*, in modo del tutto analogo a quanto visto per i vettori nel piano o nello spazio.

Si dice combinazione lineare di  $n$  vettori  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  ogni vettore del tipo

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_n \mathbf{v}_n$$

con  $\alpha_i \in \mathbb{R}$ . Si dice che  $n$  vettori  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  sono linearmente dipendenti se esiste una loro combinazione lineare, a coefficienti non tutti nulli, che dà il vettore nullo; viceversa, si dice che sono indipendenti se l'identità

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_n \mathbf{v}_n = \mathbf{0}$$

implica  $\alpha_i = 0$  per  $i = 1, 2, \dots, n$ .

**DEFINIZIONE 8.4** Sia  $V$  uno spazio vettoriale, e supponiamo che esistano  $n$  vettori  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  tali che:

- 1)  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  sono linearmente indipendenti;
- 2) ogni altro vettore di  $V$  può scriversi come combinazione lineare di questi, ossia: per ogni  $\mathbf{v} \in V$  esistono  $n$  coefficienti reali  $v_1, v_2, \dots, v_n$  tali che

$$\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}_i$$

Allora si dice che  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  costituiscono una base di  $V$ .

Si dimostra che, se  $V$  ha una base di  $n$  vettori, ogni altra base di  $V$  è costituita da  $n$  vettori. Diremo allora che  $V$  ha *dimensione n*.

Può accadere, però (come si vedrà dagli esempi) che non esista alcun  $n$  per cui  $V$  ha una base di  $n$  vettori; in tal caso si dice che  $V$  ha *dimensione infinita*. Quindi: o  $V$  ha *dimensione infinita*, oppure la sua dimensione è un numero  $n$  univocamente determinato.

La scomposizione di un vettore  $\mathbf{v}$  come combinazione lineare dei vettori di una base è *unica*. Infatti se

$$\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}_i \quad \text{e} \quad \mathbf{v} = \sum_{i=1}^n w_i \mathbf{e}_i$$

si ha, sottraendo,  $\mathbf{0} = \sum_{i=1}^n (v_i - w_i) \mathbf{e}_i$  che implica, per l'indipendenza dei vettori  $\mathbf{e}_i$ ,  $v_i = w_i$  per ogni  $i = 1, 2, \dots, n$ .

I coefficienti  $v_1, v_2, \dots, v_n$  si chiamano *componenti scalari* di  $\mathbf{v}$  rispetto alla base  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ . Se la base si ritiene fissata si può individuare il vettore  $\mathbf{v}$  scrivendo

$$\mathbf{v} = (v_1, v_2, \dots, v_n) \quad \text{o} \quad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

Chiameremo il primo *vettore riga* e il secondo *vettore colonna*. Naturalmente, cambiando la base, cambiano le componenti.



In  $\mathbb{R}^2$  la coppia  $\mathbf{i}, \mathbf{j}$  costituisce una *base* detta *canonica*. Analogamente la terna  $\mathbf{i}, \mathbf{j}, \mathbf{k}$  è la base canonica di  $\mathbb{R}^3$ . Più in generale, in  $\mathbb{R}^n$  consideriamo la  $n$ -upla di vettori:

$$\mathbf{e}_1 = (1, 0, 0, \dots, 0)$$

$$\mathbf{e}_2 = (0, 1, 0, \dots, 0)$$

.....

$$\mathbf{e}_n = (0, 0, 0, \dots, 1)$$

che costituiscono la *base canonica*. In particolare, dato  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , si ha:

$$\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n$$

In  $\mathbb{R}^n$ , salvo caso contrario, si sceglie la base canonica.

Sia  $V$  l'insieme dei polinomi in una variabile, di grado  $\leq 2$ . Si verifica subito che è uno spazio vettoriale (v. esempio 3.2). Inoltre, poiché il generico polinomio di grado  $\leq 2$  si può scrivere nella forma:

$$p = ax^2 + bx + c \quad \text{con} \quad a, b, c \in \mathbb{R}$$

ne segue che i tre polinomi

$$\mathbf{e}_1 = x^2 \quad \mathbf{e}_2 = x \quad \mathbf{e}_3 = 1$$

che sono evidentemente indipendenti, costituiscono una base di  $V$ ; pertanto  $V$  ha dimensione tre.

Sia  $V$  l'insieme di *tutti* i polinomi in una variabile. Come già osservato,  $V$  è uno spazio vettoriale. Ci si convince facilmente, tuttavia, che non esiste un numero *finito* di polinomi tale che *ogni* polinomio sia combinazione lineare di questi. Infatti l'idea naturale sarebbe di scegliere come base i polinomi:

$$1, x, x^2, x^3, \dots$$

ma il problema è: dove fermarsi? Qualunque numero finito di questi polinomi noi sceglieremo, ci sarà sempre un polinomio che non è combinazione lineare di questi. Dunque  $V$  è uno spazio vettoriale di *dimensione infinita*.

- **Sottospazi e loro dimensione.** Se  $V$  è uno spazio vettoriale di dimensione  $n$  e  $V_1$  un sottospazio vettoriale di  $V$ , sarà  $\dim V_1 < \dim V$  (tranne il caso banale  $V_1 = V$ ).



Il vettore nullo  $\mathbf{0}$  di  $\mathbb{R}^n$  costituisce da solo un sottospazio di dimensione zero.

**ESEMPIO** Se in  $\mathbb{R}^2$  (spazio vettoriale di dimensione 2) consideriamo una retta passante per l'origine (ossia l'insieme dei vettori di tipo  $tv$  con  $v$  fissato e  $t$  variabile in  $\mathbb{R}$ ), otteniamo un sottospazio vettoriale di dimensione 1. Invece, una retta non passante per l'origine (ossia un insieme di vettori del tipo  $tv + u$  con  $v, u$  fissati e  $t$  variabile in  $\mathbb{R}$ ), è un sottoinsieme di  $\mathbb{R}^2$  ma non un sottospazio. Analogamente, in  $\mathbb{R}^3$ , rette e piani passanti per l'origine sono sottospazi di dimensione 1, 2, rispettivamente.

**ESEMPIO** Siano  $V$  uno spazio vettoriale e  $v_1, v_2, \dots, v_k$  vettori di  $V$ . L'insieme  $S$  di tutte le possibili combinazioni lineari di  $v_1, v_2, \dots, v_k$  costituisce un sottospazio di  $V$ , che si chiama sottospazio *generato* dai vettori  $v_1, v_2, \dots, v_k$ . In simboli

$$S = \left\{ v : v = \sum_{j=1}^k \alpha_j v_j, \alpha_j \in \mathbb{R}, j = 1, 2, \dots, k \right\}$$

Si scrive anche, talvolta:  $S = \langle v_1, v_2, \dots, v_k \rangle$ . Se i vettori  $v_1, v_2, \dots, v_k$  sono indipendenti, allora costituiscono una base per  $S$ , che ha perciò dimensione  $k$ .

Lasciamo al lettore la verifica di queste affermazioni.

**ESEMPIO** Sia  $V$  lo spazio vettoriale dei polinomi in una variabile reale, e  $V_n$  lo spazio dei polinomi di grado  $\leq n$ . Abbiamo visto (esempio 3.5) che  $V$  ha dimensione infinita;  $V_n$ , invece, è un sottospazio vettoriale di  $V$ , di dimensione finita: precisamente,  $\dim V_n = n + 1$  (lo studente verifichi questa affermazione).

### Spazi di funzioni

Una classe di spazi vettoriali particolarmente importanti nello studio dell'Analisi Matematica è quella degli *spazi di funzioni*. Sia  $I \subset \mathbb{R}$  un intervallo e indichiamo con  $\mathcal{F}_I$  l'insieme di tutte le funzioni  $f : I \rightarrow \mathbb{R}$ . Tra queste funzioni si possono definire, al solito modo, un'operazione di somma e un'operazione di prodotto per uno scalare, ponendo, per  $f, g : I \rightarrow \mathbb{R}$  e  $\lambda \in \mathbb{R}$ ,

$$\begin{aligned} (f+g)(x) &= f(x) + g(x) \\ (\lambda f)(x) &= \lambda f(x). \end{aligned}$$

È immediato verificare che, munito di queste operazioni,  $\mathcal{F}_I$  risulta essere uno spazio vettoriale su  $\mathbb{R}$ . Ci si convince facilmente che questo spazio ha dimensione infinita (tranne il caso banale in cui l'intervallo  $I$  è ridotto a un solo punto!), in quanto non esiste un numero finito di funzioni  $f_i : I \rightarrow \mathbb{R}$ ,  $i = 1, 2, \dots, n$ , tali che ogni altra funzione si possa esprimere come loro combinazione lineare.

Spesso è più interessante considerare, anziché l'insieme di tutte le funzioni possibili, l'insieme delle funzioni con qualche proprietà importante, come il fatto di essere continue, o integrabili, o derivabili, ecc. Ad esempio, indicheremo con  $C(I)$  l'insieme di tutte le funzioni  $f : I \rightarrow \mathbb{R}$  continue in  $I$ . Poiché ogni combinazione lineare di funzioni continue è ancora una funzione continua, in base al criterio di riconoscimento dei sottospazi che abbiamo citato possiamo concludere che  $C(I)$  è un sottospazio di  $\mathcal{F}_I$ , in particolare  $C(I)$  è a sua volta uno spazio vettoriale. Un analogo ragionamento mostra che l'insieme di tutte le funzioni derivabili in  $(a, b)$  e l'insieme di tutte le funzioni integrabili in  $[a, b]$  risultano essere spazi vettoriali.

Sono tutti esempi di spazi vettoriali di dimensione infinita, che vengono detti *spazi di funzioni* appunto perché i loro elementi sono funzioni. Lo studente ne vedrà l'importanza nel corso di Analisi Matematica 2.

### 3.2 Prodotto scalare in $\mathbb{R}^n$

Torniamo ora ad occuparci specificamente dello spazio  $\mathbb{R}^n$ . Tra i vettori del piano e dello spazio, abbiamo visto che sono definite, oltre alle due operazioni caratteristiche di ogni spazio vettoriale, altre due operazioni: il prodotto scalare e, per vettori dello spazio, anche il prodotto vettoriale. Mentre il prodotto vettoriale è un'operazione tipica di  $\mathbb{R}^3$ , il *prodotto scalare* può essere definito anche in  $\mathbb{R}^n$ :

$$(x_1, x_2, \dots, x_n) \cdot (y_1, y_2, \dots, y_n) = \sum_{i=1}^n x_i y_i$$

Questo prodotto scalare ha le stesse proprietà formali che aveva nel caso di  $\mathbb{R}^2$  o  $\mathbb{R}^3$ , come si verifica facilmente in base alla definizione:

$$(3.1) \quad \begin{array}{lll} \text{i)} & \mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v} \\ \text{ii)} & \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \\ \text{iii)} & (t\mathbf{v}) \cdot \mathbf{w} = t(\mathbf{v} \cdot \mathbf{w}) \\ \text{iv)} & \mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^n v_i^2 \geq 0 \text{ per ogni } \mathbf{v}, \text{ e } \mathbf{v} \cdot \mathbf{v} = 0 \iff \mathbf{v} = \mathbf{0} \end{array}$$

In analogia a quanto accade nel piano o nello spazio, si dirà che in  $\mathbb{R}^n$ :

- due vettori  $\mathbf{u}, \mathbf{v}$  sono *ortogonali* se  $\mathbf{u} \cdot \mathbf{v} = \mathbf{0}$ ;
- due vettori  $\mathbf{u}, \mathbf{v}$  sono *paralleli* se  $\mathbf{u} = \lambda \mathbf{v}$  per qualche  $\lambda \in \mathbb{R}$ .
- Definiamo *modulo* o *norma* di un vettore  $\mathbf{v}$  il numero  $|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \left( \sum_{i=1}^n v_i^2 \right)^{\frac{1}{2}}$

Un vettore di modulo unitario si dice *versore*. Dato un vettore  $\mathbf{v}$  non nullo, si definisce

$$\text{vers}(\mathbf{v}) = \frac{\mathbf{v}}{|\mathbf{v}|}.$$

Accade spesso di voler sostituire un vettore  $\mathbf{v}$  con il corrispondente versore  $\text{vers}(\mathbf{v})$ ; questa operazione si dice *normalizzazione del vettore*  $\mathbf{v}$  e  $\text{vers}(\mathbf{v})$  si dice anche *vettore normalizzato*.

**TEOREMA 8.1** Il modulo di un vettore soddisfa le proprietà seguenti per ogni  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ,  $\lambda \in \mathbb{R}$ :

*positività:*

$$|\mathbf{v}| \geq 0 \text{ e } |\mathbf{v}| = 0 \iff \mathbf{v} = \mathbf{0}$$

*omogeneità:*

$$|\lambda \mathbf{v}| = |\lambda| |\mathbf{v}|$$

*disugualanza triangolare:*

$$(3.2) \quad |\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$$

*disugualanza di Cauchy-Schwarz:*

$$(3.3) \quad |\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}| |\mathbf{v}|$$

**DIMOSTRAZIONE.** Le prime due proprietà sono immediate. Per provare la (3.3), consideriamo il vettore  $\mathbf{u} + t\mathbf{v}$ , con  $t \in \mathbb{R}$  qualunque. Per le proprietà del prodotto scalare si ha:

$$0 \leq (\mathbf{u} + t\mathbf{v}) \cdot (\mathbf{u} + t\mathbf{v}) = \mathbf{u} \cdot \mathbf{u} + 2t\mathbf{u} \cdot \mathbf{v} + t^2\mathbf{v} \cdot \mathbf{v}$$

Ricordando che  $\mathbf{u} \cdot \mathbf{u}$ ,  $\mathbf{u} \cdot \mathbf{v}$ ,  $\mathbf{v} \cdot \mathbf{v}$  sono tre costanti, mentre  $t$  è generico, si può rileggere la precedente diseguaglianza dicendo che il trinomio di secondo grado in  $t$  non è mai negativo. Ciò è possibile solo se il suo discriminante non è positivo, ovvero se

$$(\mathbf{u} \cdot \mathbf{v})^2 - (\mathbf{u} \cdot \mathbf{u})(\mathbf{v} \cdot \mathbf{v}) \leq 0$$

cioè

$$|\mathbf{u} \cdot \mathbf{v}| \leq \sqrt{(\mathbf{u} \cdot \mathbf{u})(\mathbf{v} \cdot \mathbf{v})} = |\mathbf{u}| \cdot |\mathbf{v}|$$

che è la (3.3).

La (3.2) segue dalle relazioni:

$$\begin{aligned} |\mathbf{u} + \mathbf{v}|^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{u} \cdot \mathbf{u} + 2(\mathbf{u} \cdot \mathbf{v}) + \mathbf{v} \cdot \mathbf{v} \leq \\ &\leq |\mathbf{u}|^2 + 2|\mathbf{u}||\mathbf{v}| + |\mathbf{v}|^2 = (|\mathbf{u}| + |\mathbf{v}|)^2 \end{aligned}$$

da cui la (3.2). (L'ultima diseguaglianza scritta è un'applicazione della (3.3).)

La nozione di modulo consente di definire la *distanza* tra due vettori:

$$d(\mathbf{v}, \mathbf{w}) = |\mathbf{v} - \mathbf{w}|$$

Le proprietà della distanza così definita sono:

- a)  $d(\mathbf{v}, \mathbf{w}) \geq 0$  e  $d(\mathbf{v}, \mathbf{w}) = 0$  se e solo se  $\mathbf{v} = \mathbf{w}$  (*positività e annullamento*).
- b)  $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$  (*simmetria*).
- c)  $d(\mathbf{v}, \mathbf{w}) \leq d(\mathbf{v}, \mathbf{u}) + d(\mathbf{u}, \mathbf{w})$  (*diseguaglianza triangolare*).

Tali proprietà seguono immediatamente dalle proprietà del modulo.

### 3.3 Spazi vettoriali con prodotto scalare

Facciamo ora un altro salto di astrazione, e consideriamo la possibilità di introdurre, in certi spazi vettoriali diversi da  $\mathbb{R}^n$ , un'operazione che goda delle stesse proprietà formali del prodotto scalare.

**DEFINIZIONE 8.5** Sia  $V$  uno spazio vettoriale su  $\mathbb{R}$ , e supponiamo che sia definita un'operazione che ad ogni coppia di vettori  $\mathbf{u}, \mathbf{v} \in V$  associa uno scalare  $\mathbf{u} \cdot \mathbf{v}$ , in modo che siano soddisfatte le 4 proprietà (3.1). Diremo allora che l'operazione  $\cdot$  è un *prodotto scalare*, o *prodotto interno*, in  $V$ , e che  $V$  è uno *spazio vettoriale con prodotto scalare*. Due vettori  $\mathbf{u}, \mathbf{v} \in V$  si diranno *ortogonalni* se  $\mathbf{u} \cdot \mathbf{v} = 0$ . Si definisce *modulo* (o *norma*) del vettore  $\mathbf{v}$  il numero:

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Si definisce la *distanza tra due vettori* al modo seguente:

$$d(\mathbf{u}, \mathbf{v}) = |\mathbf{u} - \mathbf{v}|$$

Si noti che la definizione di modulo ha senso per la (3.1), iv), che garantisce che il radicando non sia negativo. Continua a valere il Teorema 8.1 (con la stessa dimostrazione): in altre parole, dagli assiomi di prodotto scalare discendono le proprietà del modulo enunciate dal Teorema 8.1. Analogamente, dalle proprietà del modulo discendono le proprietà della distanza a), b), c) richiamate in precedenza.



In  $\mathbb{R}^3$  si consideri l'operazione così definita:

$$(v_1, v_2, v_3) \cdot (u_1, u_2, u_3) = 2v_1u_1 + v_2u_2 + 3v_3u_3$$

Si verifica facilmente che gli assiomi di prodotto scalare sono soddisfatti. Questo dà quindi ad  $\mathbb{R}^3$  una struttura di spazio vettoriale con prodotto scalare, diversa da quella "euclidea" che abbiamo considerato in precedenza. Rispetto a questa struttura, si ha, per esempio:

$$|(v_1, v_2, v_3)| = \sqrt{2v_1^2 + v_2^2 + 3v_3^2}$$

Nello spazio di funzioni  $C([a, b])$  si consideri l'operazione definita da

$$f \cdot g = \int_a^b f(x)g(x)dx.$$

Si verifica che gli assiomi di prodotto scalare sono soddisfatti. Quindi  $C([a, b])$  assume una struttura di spazio vettoriale con prodotto scalare, in cui due funzioni  $f, g$  sono ortogonali quando

$$\int_a^b f(x)g(x)dx = 0.$$

Nello studio delle *serie di Fourier*, in Analisi Matematica 2, si vedrà come questo prodotto scalare e questa nozione astratta di ortogonalità siano estremamente utili nello studio di certi problemi di approssimazione di funzioni.

Quando gli elementi di uno spazio vettoriale con prodotto scalare sono funzioni, si preferisce di solito indicare il prodotto scalare con la scrittura

$$\langle f, g \rangle \text{ anziché } f \cdot g,$$

che si potrebbe facilmente confondere con l'usuale prodotto di funzioni.

- *Complemento ortogonale.* complemento!ortogonale Se  $V$  è uno spazio vettoriale con prodotto interno e  $V_1$  è un sottospazio di  $V$ , definiamo

$$V_1^\perp = \{\mathbf{v} \in V : \mathbf{v} \cdot \mathbf{u} = 0 \quad \forall \mathbf{u} \in V_1\}$$

In altre parole,  $V_1^\perp$  è l'insieme dei vettori che sono ortogonali a tutti i vettori di  $V_1$ . Si verifica facilmente che tale insieme è un sottospazio di  $V$ , che prende il nome di *complemento ortogonale* di  $V_1$ .

Più in generale, due sottospazi  $U$  e  $W$  di  $V$  dicono *ortogonali* se ogni vettore di  $U$  è ortogonale ad ogni vettore di  $W$ :

$$\mathbf{u} \cdot \mathbf{w} = 0 \quad \text{per ogni } \mathbf{u} \in U, \mathbf{w} \in W$$



In  $\mathbb{R}^3$  si consideri un piano passante per l'origine (sottospazio di dimensione 2), sia esso  $V_1$ . Allora  $V_1^\perp$  è il sottospazio unidimensionale individuato dalla retta passante per l'origine e perpendicolare al piano, sia esso  $V_2$ :  $V_2 = V_1^\perp$ . Si noti che vale anche la relazione reciproca:  $V_2^\perp = V_1$ .

In  $V = \mathbb{R}^4$  sia  $U$  il sottospazio (iperpiano) generato dai vettori

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ e } \mathbf{u}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

e  $W$  la retta (sottospazio unidimensionale) generata dal vettore

$$\mathbf{w} = \begin{pmatrix} 0 \\ 3 \\ 0 \\ 2 \end{pmatrix}.$$

Allora  $U$  e  $W$  sono ortogonali. Si noti però che non sono l'uno il complemento ortogonale dell'altro: esistono vettori ortogonali a  $U$  che non stanno in  $W$ , come il vettore

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

### Basi ortonormali

Consideriamo uno spazio vettoriale di dimensione finita, dotato di prodotto scalare. Fissata una base  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  calcoliamo il prodotto scalare di due vettori

$$\mathbf{u} = \sum_{i=1}^n u_i \mathbf{e}_i \quad \mathbf{v} = \sum_{j=1}^n v_j \mathbf{e}_j$$

Applicando le proprietà del prodotto scalare si ha:

$$\mathbf{u} \cdot \mathbf{v} = \left( \sum_{i=1}^n u_i \mathbf{e}_i \right) \cdot \left( \sum_{j=1}^n v_j \mathbf{e}_j \right) = \sum_{i,j=1}^n u_i v_j \mathbf{e}_i \cdot \mathbf{e}_j$$

È sufficiente quindi conoscere gli  $n^2$  prodotti scalari  $\mathbf{e}_i \cdot \mathbf{e}_j$  per poter calcolare il prodotto scalare di due vettori qualsiasi. Il calcolo risulta particolarmente comodo quando gli  $n$  vettori della base risultano *ortonormali*, ossia:

- 1) sono a due a due ortogonali:  $\mathbf{e}_i \cdot \mathbf{e}_j = 0$  per  $i \neq j$ ;
- 2) ogni vettore ha modulo unitario:  $|\mathbf{e}_i|^2 = \mathbf{e}_i \cdot \mathbf{e}_i = 1$  per  $i = 1, \dots, n$ .

Una tale base si dice *base ortonormale*, e gioca un ruolo fondamentale nello studio degli spazi con prodotto scalare. In tal caso infatti si ottiene:

$$\mathbf{u} \cdot \mathbf{v} = \left( \sum_{i=1}^n u_i \mathbf{e}_i \right) \cdot \left( \sum_{j=1}^n v_j \mathbf{e}_j \right) = \sum_{i=1}^n u_i v_i$$

ovvero: il prodotto scalare si ottiene dalle componenti dei due vettori con la stessa formula che vale nel caso del prodotto scalare euclideo in  $\mathbb{R}^n$ . In particolare, per  $\mathbf{u} = \mathbf{v}$  si ha:

$$|\mathbf{u}|^2 = \mathbf{u} \cdot \mathbf{u} = \sum_{i=1}^n u_i^2$$

ovvero anche il modulo di un vettore si calcola mediante le sue componenti con la stessa formula che vale nel caso euclideo. L'ultima proprietà si può anche esprimere così: se  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  sono vettori a due a due ortogonali (di modulo qualsiasi) si ha:

$$(3.4) \quad \left| \sum_{i=1}^n \mathbf{v}_i \right|^2 = \sum_{i=1}^n |\mathbf{v}_i|^2$$

che è una sorta di Teorema di Pitagora in versione astratta.

Per esempio, in  $\mathbb{R}^n$  la base canonica è anche una base ortonormale, rispetto al prodotto scalare euclideo.

Si può dimostrare che uno spazio vettoriale di dimensione finita, dotato di prodotto scalare, ha *sempre* una base ortonormale:

**TEOREMA 8.2 (PROCEDIMENTO DI ORTONORMALIZZAZIONE DI GRAM-SCHMIDT)** *Sia  $V$  uno spazio vettoriale di dimensione finita,  $n$ , dotato di prodotto interno. Allora:*

1.  *$V$  ha una base ortonormale;*
2. *In particolare, se  $V_1$  è un sottospazio di  $V$  di dimensione  $m < n$ , si può costruire una base ortonormale di  $V$  del tipo*

$$\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$$

*dove i primi  $m$  vettori  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$  sono una base ortonormale di  $V_1$  e gli ultimi  $n - m$  vettori  $\mathbf{u}_{m+1}, \mathbf{u}_2, \dots, \mathbf{u}_n$  sono una base ortonormale di  $V_1^\perp$ .*

Invece di presentare la dimostrazione del teorema in astratto, illustriamo su un esempio numerico il procedimento di ortonormalizzazione.

Consideriamo, in  $\mathbb{R}^3$ , i due vettori indipendenti:

$$\mathbf{v}_1 = (1, 0, -1) \quad \mathbf{v}_2 = (0, 1, -1)$$

Detto  $V$  il sottospazio vettoriale di  $\mathbb{R}^3$  generato da  $\mathbf{v}_1, \mathbf{v}_2$  (che, come si verifica, non è altro che il piano passante per l'origine di equazione  $x + y + z = 0$ ), proponiamoci di costruire una base ortonormale di  $V$ , ossia di ortonormalizzare la base  $\mathbf{v}_1, \mathbf{v}_2$ . I passi sono i seguenti:

1. Normalizziamo  $\mathbf{v}_1$ , definendo:

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|} = \left( \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right).$$

2. Calcoliamo la componente di  $\mathbf{v}_2$  nella direzione di  $\mathbf{u}_1$ , data da:

$$(\mathbf{v}_2 \cdot \mathbf{u}_1) \mathbf{u}_1 = \frac{1}{\sqrt{2}} \left( \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right) = \left( \frac{1}{2}, 0, -\frac{1}{2} \right).$$

3. Sottraiamo a  $\mathbf{v}_2$  la sua componente nella direzione di  $\mathbf{u}_1$ , ottenendo così un vettore ortogonale a  $\mathbf{u}_1$  e che insieme ad esso genera  $V$ :

$$\mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1) \mathbf{u}_1 = (0, 1, -1) - \left( \frac{1}{2}, 0, -\frac{1}{2} \right) = \left( -\frac{1}{2}, 1, -\frac{1}{2} \right)$$

4. Normalizziamo quest'ultimo vettore, definendo:

$$\mathbf{u}_2 = \frac{\left( -\frac{1}{2}, 1, -\frac{1}{2} \right)}{\sqrt{\frac{1}{4} + 1 + \frac{1}{4}}} = \sqrt{\frac{2}{3}} \left( -\frac{1}{2}, 1, -\frac{1}{2} \right) = \left( -\frac{1}{\sqrt{6}}, \sqrt{\frac{2}{3}}, -\frac{1}{\sqrt{6}} \right)$$

Il lettore verifichi che  $\mathbf{u}_1, \mathbf{u}_2$  costituiscono effettivamente una base ortonormale di  $V$ .

Se ora volessimo costruire (come afferma il punto 2 del teorema) una base ortonormale di  $\mathbb{R}^3$  che abbia come primi due elementi  $\mathbf{u}_1, \mathbf{u}_2$ , dovremmo iterare il procedimento come segue:

5. Scegliamo un vettore di  $\mathbb{R}^3$  indipendente da  $\mathbf{u}_1, \mathbf{u}_2$ , per esempio

$$\mathbf{v}_3 = (0, 0, 1)$$

6. Sottraiamo a  $\mathbf{v}_3$  la sua proiezione su  $V$ , ossia calcoliamo:

$$\begin{aligned} \mathbf{v}'_3 &= \mathbf{v}_3 - \{(\mathbf{v}_3 \cdot \mathbf{u}_1) \mathbf{u}_1 + (\mathbf{v}_3 \cdot \mathbf{u}_2) \mathbf{u}_2\} = \\ &= (0, 0, 1) - \left\{ -\frac{1}{\sqrt{2}} \left( \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right) - \frac{1}{\sqrt{6}} \left( -\frac{1}{\sqrt{6}}, \sqrt{\frac{2}{3}}, -\frac{1}{\sqrt{6}} \right) \right\} = \\ &= \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \end{aligned}$$

7. Normalizziamo  $\mathbf{v}'_3$ , ottenendo

$$\mathbf{u}_3 = \frac{\left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right)}{\sqrt{\frac{1}{3}}} = \left( \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)$$

Il lettore verifichi che  $\mathbf{u}_3$  è ortogonale a  $\mathbf{u}_1, \mathbf{u}_2$  ed ha modulo unitario, perciò  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$  sono una base ortonormale di  $\mathbb{R}^3$ , i cui primi due elementi sono una base ortonormale di  $V$ .

Più in generale, l'*algoritmo iterativo* di ortonormalizzazione di un sistema  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  di vettori linearmente indipendenti è il seguente:

$$\mathbf{u}_1 = \text{vers}(\mathbf{v}_1)$$

$$\mathbf{u}_k = \text{vers} \left( \mathbf{v}_k - \sum_{j=1}^{k-1} (\mathbf{u}_j \cdot \mathbf{v}_k) \mathbf{u}_j \right) \text{ per } k = 2, 3, \dots, n.$$

Lo studente è invitato a dimostrare che i vettori  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  così costruiti sono effettivamente ortonormali. Si noti che il sistema di vettori ortonormali  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  non dipende solo dal sistema di vettori  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  di partenza, ma anche dall'ordine in cui li consideriamo.

L'esistenza di una nozione di ortogonalità e la conoscenza di una base ortonormale gioca un ruolo importante in *problemi di approssimazione*. Si consideri il seguente problema geometrico:

assegnato, nello spazio tridimensionale, un piano  $\pi$  passante per l'origine e un punto  $P$  esterno ad esso, trovare il punto di  $\pi$  che ha distanza minima da  $P$ .

Chiaramente il punto di minima distanza è la *proiezione ortogonale* di  $P$  su  $\pi$ . (v. fig. 8.23).

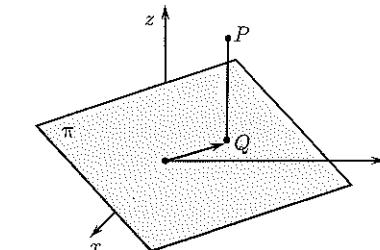


Figura 8.23. Punto di minima distanza da un piano.

Analogamente, supponiamo che  $V$  sia uno spazio vettoriale dotato di prodotto scalare,  $V_1$  un sottospazio di  $V$  e  $\mathbf{v}$  un elemento di  $V$  che non appartiene a  $V_1$ .

Qual è l'elemento di  $V_1$  che *approssima meglio*  $\mathbf{v}$ , ossia che ha distanza minima da  $\mathbf{v}$ ? Sarà, in analogia all'esempio tridimensionale, la proiezione ortogonale di  $\mathbf{v}$  su  $V_1$ , pur di definire opportunamente questo concetto.

**TEOREMA 8.3 (PROIEZIONE ED ELEMENTO DI MINIMA DISTANZA)** *Sia  $V$  uno spazio dotato di prodotto scalare,  $V_1$  un sottospazio di  $V$  di dimensione finita e  $\mathbf{v}$  un elemento di  $V$  che non appartiene a  $V_1$ . Allora esistono e sono univocamente determinati due vettori  $\mathbf{u}, \mathbf{w}$ , tali che:*

$$\mathbf{v} = \mathbf{u} + \mathbf{w} \quad \mathbf{u} \in V_1 \quad \mathbf{w} \in V_1^\perp$$

*Il vettore  $\mathbf{u}$  si dirà proiezione ortogonale di  $\mathbf{v}$  su  $V_1$  e, tra tutti i vettori di  $V_1$ , è quello di minima distanza da  $\mathbf{v}$ .*

Si noti che in questo enunciato lo spazio ambiente  $V$  può anche avere dimensione infinita, mentre  $V_1$  per ipotesi ha dimensione finita.

**DIMOSTRAZIONE.** Poiché  $V_1$  è uno spazio vettoriale di dimensione finita con prodotto interno, per il Teorema 3.6 esiste una base ortonormale  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  di  $V_1$ . Utilizzando questa base definiamo il vettore

$$\mathbf{u} = \sum_{i=1}^n (\mathbf{v} \cdot \mathbf{e}_i) \mathbf{e}_i$$

che rappresenta la proiezione di  $\mathbf{v}$  su  $V_1$ . Poniamo anche  $\mathbf{w} = \mathbf{v} - \mathbf{u}$ . Dalle definizioni segue subito che  $\mathbf{u} \in V_1$  e  $\mathbf{v} = \mathbf{u} + \mathbf{w}$ ; proviamo che  $\mathbf{w} \in V_1^\perp$ . È sufficiente mostrare l'ortogonalità di  $\mathbf{w}$  a ciascun elemento  $\mathbf{e}_j$  della base di  $V_1$ , in quanto l'ortogonalità a qualunque vettore di  $V_1$  seguirà allora per linearità. Si ha:

$$\begin{aligned} \mathbf{w} \cdot \mathbf{e}_j &= (\mathbf{v} - \mathbf{u}) \cdot \mathbf{e}_j = \mathbf{v} \cdot \mathbf{e}_j - \left( \sum_{i=1}^n (\mathbf{v} \cdot \mathbf{e}_i) \mathbf{e}_i \right) \cdot \mathbf{e}_j = \mathbf{v} \cdot \mathbf{e}_j - \sum_{i=1}^n (\mathbf{v} \cdot \mathbf{e}_i) (\mathbf{e}_i \cdot \mathbf{e}_j) \\ &= \mathbf{v} \cdot \mathbf{e}_j - (\mathbf{v} \cdot \mathbf{e}_j) = 0, \end{aligned}$$

essendo  $\mathbf{e}_i \cdot \mathbf{e}_j = 0$  per  $i \neq j$  e  $\mathbf{e}_i \cdot \mathbf{e}_i = 1$ ,

dunque  $\mathbf{w} \in V_1^\perp$ . Per mostrare che  $\mathbf{u}$  è l'elemento di  $V_1$  di minima distanza da  $\mathbf{v}$ , sia  $\mathbf{u}_1$  un altro generico elemento di  $V_1$ . Essendo

$$\mathbf{v} - \mathbf{u}_1 = \mathbf{w} + (\mathbf{u} - \mathbf{u}_1)$$

con  $(\mathbf{u} - \mathbf{u}_1) \in V_1$  e  $\mathbf{w} \in V_1^\perp$ , per il Teorema di Pitagora (3.4) si ha

$$|\mathbf{v} - \mathbf{u}_1|^2 = |\mathbf{w} + (\mathbf{u} - \mathbf{u}_1)|^2 = |\mathbf{w}|^2 + |\mathbf{u} - \mathbf{u}_1|^2 \geq |\mathbf{w}|^2 = |\mathbf{v} - \mathbf{u}|^2,$$

quindi la distanza di  $\mathbf{v}$  da  $\mathbf{u}$  è minore della distanza di  $\mathbf{v}$  da qualunque altro elemento  $\mathbf{u}_1$  di  $V_1$ .



**ESEMPIO 8.1** In  $\mathbb{R}^3$ , determinare il punto del piano  $x + z = 0$  che ha minima distanza da  $\mathbf{u} = (1, 2, 3)$ .

Poiché  $x + z = 0$  è un piano passante per l'origine, si può vedere come sottospazio vettoriale di  $\mathbb{R}^3$ , chiamiamolo  $V_1$ . Si verifica che una base ortonormale di  $V_1$  è:

$$\mathbf{v}_1 = \left( \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right) \quad \mathbf{v}_2 = (0, 1, 0)$$

Il punto di  $V_1$  che ha minima distanza da  $\mathbf{u}$  è la proiezione di  $\mathbf{u}$  su  $V_1$ , cioè:

$$(\mathbf{u} \cdot \mathbf{v}_1) \mathbf{v}_1 + (\mathbf{u} \cdot \mathbf{v}_2) \mathbf{v}_2 = -\frac{2}{\sqrt{2}} \mathbf{v}_1 + 2 \mathbf{v}_2 = (-1, 2, 1)$$

#### Somma diretta di sottospazi ortogonali

Sia  $V$  uno spazio vettoriale qualsiasi. Dati due sottospazi  $U$  e  $W$  qualunque, se ne possono formare altri due particolarmente significativi:  $V \cap W$ , il cui significato è noto, e

$$V + W = \{\mathbf{v} + \mathbf{w} : \mathbf{v} \in V, \mathbf{w} \in W\}$$

che coincide con il sottospazio generato da  $V \cup W$  (v. Esempio 3.8 per la definizione di sottospazio generato da un insieme di vettori). Vale il seguente risultato (di cui omettiamo la dimostrazione).

**PROPOSIZIONE 8.1** *Dati due sottospazi  $U$  e  $W$  di  $V$ , vale la formula*

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$$

Supponiamo ora che  $V$  sia uno spazio vettoriale dotato di prodotto scalare. Se i sottospazi  $U$  e  $W$  sono ortogonali, si ha evidentemente  $U \cap W = \{\mathbf{0}\}$  per cui

$$\dim(U + W) = \dim U + \dim W.$$

Naturalmente non è detto che risulti  $U + W = V$ , come dimostra l'Esempio 3.3, dove  $V = \mathbb{R}^4$  mentre  $U + W$  ha dimensione 3. Quando questo accade, la circostanza merita una definizione e una notazione speciale:

**DEFINIZIONE 8.6** Si dice che  $V$  è somma diretta dei sottospazi ortogonali  $U$  e  $W$  e si scrive

$$V = U \oplus W,$$

se ogni vettore  $\mathbf{v} \in V$  si può scrivere in uno e un solo modo nella forma

$$(3.5) \quad \mathbf{v} = \mathbf{u} + \mathbf{w} \text{ con } \mathbf{u} \in U, \mathbf{w} \in W$$

**PROPOSIZIONE 8.2** *Se  $V = U \oplus W$ , i due sottospazi  $U, W$  sono uno il complemento ortogonale dell'altro.*

**DIMOSTRAZIONE.** Facciamo vedere che, se  $V = U \oplus W$ , allora  $W = U^\perp$ . Ovviamente si ha  $W \subseteq U^\perp$ , perché  $W$  è ortogonale a  $U$ . Viceversa, sia  $\mathbf{v} \in U^\perp$  e, in base alla (3.5), scriviamolo nella forma

$$\mathbf{v} = \mathbf{u} + \mathbf{w}.$$

Facendo prodotto scalare con  $\mathbf{u}$  si ottiene

$$\mathbf{v} \cdot \mathbf{u} = \mathbf{u} \cdot \mathbf{u} + \mathbf{w} \cdot \mathbf{u}$$

Ma  $\mathbf{v} \cdot \mathbf{u} = 0$  essendo  $\mathbf{v}$  ortogonale a  $U$  e  $\mathbf{w} \cdot \mathbf{u} = 0$  perché  $W$  e  $U$  sono ortogonali. Si deduce che  $\mathbf{u} \cdot \mathbf{u} = |\mathbf{u}|^2 = 0$  che implica  $\mathbf{u} = \mathbf{0}$ . Ma allora  $\mathbf{v} = \mathbf{w} \in W$  e perciò  $U^\perp \subseteq W$ .

Nella (3.5)  $\mathbf{u}$  è la proiezione di  $\mathbf{v}$  su  $U$ , mentre  $\mathbf{w}$  è la proiezione di  $\mathbf{v}$  su  $W$  e vale il Teorema di Pitagora:

$$|\mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{w}|^2.$$

#### 3.4 Il concetto di linearità

Nel capitolo 2 abbiamo introdotto il concetto generale di funzione tra due insiemi, come legge che associa univocamente ad un dato “ingresso” una certa “uscita”. Ci occupiamo ora del caso in cui gli insiemi in questione siano 2 spazi vettoriali  $V_1, V_2$ ; una funzione  $f$  di dominio  $V_1$  e codominio  $V_2$ ,

$$f : V_1 \rightarrow V_2$$

sarà dunque una legge che ad ogni vettore di  $V_1$  associa uno e un sol vettore di  $V_2$ . Tra queste funzioni, particolare importanza hanno quelle che godono di una speciale proprietà, detta *linearità*:

**DEFINIZIONE 8.7** Siano  $V_1, V_2$  due spazi vettoriali su un campo  $\mathbb{K}$  (che per noi potrà essere  $\mathbb{R}$  o  $\mathbb{C}$ ) e sia  $f : V_1 \rightarrow V_2$ . Si dirà che  $f$  è una funzione lineare se per ogni  $\alpha \in \mathbb{K}$  e per ogni  $\mathbf{v}_1, \mathbf{v}_2 \in V_1$  si ha:

i.  $f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2)$  (additività);

ii.  $f(\alpha \mathbf{v}_1) = \alpha f(\mathbf{v}_1)$  (omogeneità).

Le i, ii si possono riassumere nell'unica formula

$$f(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2) = \alpha_1 f(\mathbf{v}_1) + \alpha_2 f(\mathbf{v}_2) \text{ per ogni } \alpha_1, \alpha_2 \in \mathbb{K} \quad \mathbf{v}_1, \mathbf{v}_2 \in V_1$$

Sinonimi spesso usati di funzione lineare sono *trasformazione lineare*, *operatore lineare*.



**ESEMPIO 8.2** Sia  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Se  $f$  è lineare si ha:  $f(x) = f(x \cdot 1) = xf(1) = \alpha x$  per  $\alpha = f(1)$ . Dunque le uniche funzioni lineari da  $\mathbb{R}$  a  $\mathbb{R}$  sono del tipo  $f(x) = \alpha x$ .

Sia  $V$  lo spazio dei vettori tridimensionali; consideriamo un vettore  $\mathbf{u} \in V$ , fissato una volta per tutte, e definiamo:

$$\begin{aligned} f : V &\rightarrow \mathbb{R} \\ f : \mathbf{v} &\mapsto \mathbf{u} \cdot \mathbf{v} \end{aligned}$$

(la funzione associa ad ogni vettore  $\mathbf{v}$  il suo prodotto scalare col vettore  $\mathbf{u}$  fissato). Dalle proprietà del prodotto scalare segue subito che  $f$  è lineare, infatti:

$$f(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) = \mathbf{u} \cdot (\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) = \alpha\mathbf{u} \cdot \mathbf{v}_1 + \beta\mathbf{u} \cdot \mathbf{v}_2 = \alpha f(\mathbf{v}_1) + \beta f(\mathbf{v}_2)$$

Il concetto di linearità è centrale in matematica e nelle sue applicazioni. Nel prossimo paragrafo 4 vedremo che il *calcolo matriciale* permette di descrivere completamente le *trasformazioni lineari tra spazi vettoriali di dimensione finita*. Molti concetti chiave del calcolo infinitesimale che lo studente ha già incontrato (come quelli di limite, derivata, integrale) corrispondono a opportuni operatori lineari che coinvolgono spazi di funzioni, di dimensione infinita.



L'integrale definito si può vedere come operatore lineare sullo spazio delle funzioni continue su un intervallo  $[a, b]$ , a valori reali:

$$\begin{aligned} L : C([a, b]) &\rightarrow \mathbb{R} \\ L : f &\mapsto \int_a^b f(x) dx. \end{aligned}$$

L'operatore  $L$  è lineare (linearità dell'integrale definito, v. Cap.6, Teorema 6.4); in questo caso lo spazio vettoriale di partenza ha dimensione infinita, mentre quello di arrivo è  $\mathbb{R}$ , di dimensione 1.

L'operatore che ad ogni funzione continua su  $[a, b]$  associa la sua funzione integrale, cioè:

$$L : f \mapsto \int_a^x f(t) dt$$

si può vedere come operatore lineare

$$L : C([a, b]) \rightarrow C([a, b]),$$

o meglio

$$L : C([a, b]) \rightarrow C^1([a, b]),$$

dove col simbolo  $C^1([a, b])$  si indica lo spazio delle funzioni derivabili, dotate di derivata continua su  $[a, b]$ . Infatti per il secondo Teorema fondamentale del calcolo integrale (Capitolo 6, Teorema 6.10) se la funzione integranda è continua, la funzione integrale è derivabile con continuità. In questo caso sia lo spazio vettoriale di partenza che quello di arrivo hanno dimensione infinita. Lo stesso accade nel prossimo esempio:

La derivata si può vedere come operatore lineare tra i seguenti spazi:

$$\begin{aligned} L : C^1([a, b]) &\rightarrow C([a, b]) \\ L : f &\mapsto f' \end{aligned}$$

(linearità della derivata, Capitolo 4, Teorema 4.2).

Molti problemi fisici si possono descrivere con modelli matematici di tipo lineare. Per dare un'idea elementare del significato di questa affermazione, facciamo un esempio specifico. La temperatura di un oggetto tridimensionale privo di sorgenti interne di calore, in condizioni di equilibrio termico, può essere determinata a partire dalla temperatura della sua superficie esterna, che è a contatto con l'ambiente. Temperatura all'interno e temperatura al bordo sono espresse da funzioni. La legge che associa al "dato" (temperatura al bordo) la "soluzione" (temperatura all'interno) è una trasformazione lineare tra due opportuni spazi di funzioni.

*La linearità ha allora il seguente significato fisico: la somma di due cause provoca la somma dei due effetti.*

Per esempio, se raddoppiamo il valore della temperatura al bordo in ogni punto, raddoppierà esattamente la temperatura all'interno, in ogni punto. Naturalmente non accade così per *ogni* fenomeno, ma solo, appunto, per quelli che sono ben descritti da modelli di tipo lineare.

Segnaliamo anche un tipo di trasformazione tra spazi vettoriali che è in stretta relazione con le trasformazioni lineari. Si dice *trasformazione affine*, o anche *lineare affine* tra due spazi vettoriali  $V_1, V_2$ , una trasformazione

$$f : V_1 \rightarrow V_2$$

che abbia la forma

$$f(\mathbf{v}) = l(\mathbf{v}) + \mathbf{b}$$

dove  $l : V_1 \rightarrow V_2$  è lineare e  $\mathbf{b} \in V_2$  è un elemento fissato. Ad esempio,

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R} \\ f(x) &= ax + b \end{aligned}$$

con  $a, b \in \mathbb{R}$  fissati, non è lineare ma è affine.



In  $\mathbb{R}^4$ , si considerino i seguenti gruppi di vettori, e per ognuno di essi si dica se sono linearmente indipendenti. In caso negativo, scrivere esplicitamente uno di essi come combinazione lineare degli altri.

- a.  $(1, 2, 0, 0) \quad (-1, 2, 0, 0) \quad (0, 0, 1, 1) \quad (0, 4, 2, 2)$
- b.  $(1, 1, 1, 1) \quad (1, 1, 1, 0) \quad (1, 1, 0, 0) \quad (1, 0, 0, 0)$
- c.  $(1, 2, 3, 0) \quad (2, 1, 0, 2) \quad (3, 1, 4, 5) \quad (1, 0, 0, 0) \quad (0, 1, 0, 0)$

Per quali valori del parametro reale  $t$  il vettore  $\mathbf{w} = (2, t, 0, 1)$  di  $\mathbb{R}^4$  appartiene al sottospazio generato da  $\mathbf{u} = (1, 0, 0, 1)$  e  $\mathbf{v} = (0, 1, 0, 1)$ ? Calcolare la dimensione dello spazio generato di  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  al variare di  $t$ .

Verificare che il seguente "prodotto scalare" in  $\mathbb{R}^2$  soddisfa effettivamente gli assiomi di prodotto scalare:

$$(x_1, y_1) \cdot (x_2, y_2) = 3x_1x_2 + 2x_1y_2 + 2y_1x_2 + 4y_1y_2$$

Rispetto a questa struttura di spazio vettoriale con prodotto interno, rispondere alle seguenti domande:

- I vettori  $(1, 0)$  e  $(0, 1)$  sono ortogonali?
- Quanto vale  $|(1, 1)|$ ?
- Determinare una base ortonormale in  $\mathbb{R}^2$ .

Si consideri lo spazio vettoriale  $V$  dei polinomi  $p(x)$  (in una variabile, a coefficienti reali), di grado  $\leq 3$ .

- Verificare che i seguenti vettori costituiscono una base di  $V$ :

$$\mathbf{e}_1 = 1 \quad \mathbf{e}_2 = x + 2 \quad \mathbf{e}_3 = 2x^2 + 1 \quad \mathbf{e}_4 = x^3 + x + 1$$

- Esprimere  $x$ ,  $x^2$ ,  $x^3$  come combinazione lineare dei vettori della base.
- Dire quali dei seguenti sottoinsiemi di  $V$  costituiscono un sottospazio vettoriale e, in caso affermativo, qual è la dimensione del sottospazio:
  - L'insieme dei polinomi di  $V$  con termine noto nullo.
  - L'insieme dei polinomi di  $V$  con termine noto positivo.
  - L'insieme dei polinomi di  $V$  di grado dispari.
  - L'insieme dei polinomi di  $V$  per i quali la somma dei coefficienti è nulla.

Stabilire quali delle seguenti funzioni  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  sono lineari:

- $f : (x, y) \mapsto (xy, x + y)$
- $f : (x, y) \mapsto (x - y, x + y + 1)$
- $f : (x, y) \mapsto (2x - y, x + 3y)$

## 4 MATRICI E TRASFORMAZIONI LINEARI

### 4.1 L'algebra delle matrici

Si può dire, informalmente, che una *matrice* è una *tabella a doppia entrata*. Ne sono esempi:

- la tavola pitagorica
- l'orario ferroviario
- le tabelle delle distanze chilometriche.

E ancora tabelle del tipo seguente:

$$a) \begin{pmatrix} i & i & -1 & -i \\ i & -1 & 1 & 1 \\ -i & -i & i & 1 \\ -i & 1 & 1 & i \end{pmatrix} \quad b) \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}$$

$$c) \begin{pmatrix} 7 & 0 & 0 & 5 \end{pmatrix} \quad d) \begin{pmatrix} \sqrt{2} \\ -\pi \\ -1 \end{pmatrix}$$

Possiamo dare la seguente definizione.

**DEFINIZIONE 8.8** Si dice *matrice* di tipo  $(m, n)$  su un insieme numerico  $\mathcal{A}$ , un insieme di  $m \cdot n$  numeri appartenenti ad  $\mathcal{A}$ , disposti in una tabella di  $m$  righe (linee orizzontali) ed  $n$  colonne (linee verticali).

La matrice  $a$ ) è di tipo  $(4, 4)$  su  $\mathbb{C}$ , la  $b$ ) di tipo  $(2, 4)$  su  $\mathbb{N}$ , la  $c$ ) di tipo  $(1, 4)$  su  $\mathbb{N}$ , la  $d$ ) di tipo  $(3, 1)$  su  $\mathbb{R}$ . Salvo avviso contrario, considereremo matrici su  $\mathbb{R}$ .

La generica matrice  $\mathbf{A}$  di tipo  $(m, n)$  si può scrivere nella forma

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Si noti il significato del doppio indice:  $a_{ij}$  si dice "elemento di posto  $i, j$ "; il primo dei due indici (qualunque sia la lettera con cui lo denotiamo!) rappresenta il numero della riga, il secondo indice il numero della colonna. I due indici costituiscono quindi una sorta di "coordinate" dell'elemento nella tabella.

Scriveremo sinteticamente  $\mathbf{A} = (a_{ij})$ ,  $\mathbf{B} = (b_{ij})$  indicando a parte il tipo di matrice o il campo di variabilità degli indici:  $i = 1, \dots, m$ ;  $j = 1, \dots, n$ .

Se  $m = n$ , come nel caso della  $a$ ), la matrice si dice *quadrata* di ordine  $m$ .

Due matrici dello stesso tipo  $(m, n)$  si dicono *uguali* se sono uguali i rispettivi elementi: in simboli, se

$$\mathbf{A} = (a_{ij}) \quad \mathbf{B} = (b_{ij})$$

allora

$$\mathbf{A} = \mathbf{B} \iff a_{ij} = b_{ij} \quad i = 1, \dots, m \\ j = 1, \dots, n$$

Due matrici dello stesso tipo si possono *sommare* ottenendo una nuova matrice ancora dello stesso tipo; se:

$$\mathbf{A} = (a_{ik}), \quad \mathbf{B} = (b_{ik}), \quad \mathbf{C} = (c_{ik}) \quad i = 1, \dots, m; \quad k = 1, \dots, n$$

si definisce

$$\mathbf{A} + \mathbf{B} = \mathbf{C} \iff a_{ik} + b_{ik} = c_{ik} \quad i = 1, \dots, m; \quad k = 1, \dots, n$$

Per esempio

$$\begin{pmatrix} 7 & 11 & 17 \\ 21 & -5 & -9 \end{pmatrix} + \begin{pmatrix} 4 & -10 & 3 \\ 7 & 81 & 32 \end{pmatrix} = \begin{pmatrix} 11 & 1 & 20 \\ 28 & 76 & 23 \end{pmatrix}$$

Con le matrici si possono effettuare altre operazioni. Per esempio si può *moltiplicare una matrice per un numero*:

$$3 \cdot \begin{pmatrix} 1 & 3 & 1 \\ 0 & -1 & 5 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 9 & 3 \\ 0 & -3 & 15 \\ 0 & 0 & 9 \end{pmatrix}$$

Cioè, se  $t$  è un numero e  $\mathbf{A} = (a_{ik})$ , allora

$$t \mathbf{A} = (ta_{ik})$$

Si verifica facilmente che, con queste definizioni, l'insieme  $M(m, n)$  di tutte le matrici su  $\mathbb{R}$  di tipo  $(m, n)$  è uno spazio vettoriale. In particolare, l'elemento  $\mathbf{0}$  è la matrice che ha tutti gli elementi nulli, e la matrice opposta di  $\mathbf{A}$ ,  $-\mathbf{A}$ , è la matrice che ha per elementi gli opposti degli elementi di  $\mathbf{A}$ .

La dimensione dello spazio vettoriale è  $n \cdot m$  e la base canonica è costituita dalle matrici  $n \times m$  aventi tutti gli elementi nulli, tranne uno uguale ad 1.

In particolare:

$M(1, n)$  si identifica con  $\mathbb{R}^n$ , pensato come spazio di vettori riga;

$M(n, 1)$  si identifica con  $\mathbb{R}^n$ , pensato come spazio di vettori colonna.

I vettori di  $\mathbb{R}^n$  sono quindi particolari matrici.

Le righe di una matrice  $\mathbf{A}$  di tipo  $(m, n)$  possono essere viste come *vettori riga* di  $\mathbb{R}^n$ , cioè

$$\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{in}) \quad i = 1, 2, \dots, m$$

e viceversa le colonne possono essere viste come *vettori colonna* di  $\mathbb{R}^m$ , cioè

$$\mathbf{a}^j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

quindi la matrice  $\mathbf{A}$  può anche essere scritta nei modi seguenti:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{pmatrix} = (\mathbf{a}^1 | \mathbf{a}^2 | \cdots | \mathbf{a}^m)$$

Possiamo definire anche un *prodotto tra matrici*; non per ogni coppia di matrici, però, come si vede nella seguente definizione.

**DEFINIZIONE 8.9 (PRODOTTO DI MATRICI)** Date due matrici  $\mathbf{A} = (a_{ik})$  e  $\mathbf{B} = (b_{ks})$  di tipo  $(m, n)$  e  $(n, p)$ , rispettivamente, indicheremo con  $\mathbf{A} \cdot \mathbf{B}$  o, più semplicemente, con  $\mathbf{AB}$ , la matrice  $\mathbf{C}$  di tipo  $(m, p)$  il cui elemento  $c_{ij}$  è il prodotto scalare della  $i$ -esima riga di  $\mathbf{A}$  e della  $j$ -esima colonna di  $\mathbf{B}$  (le quali sono entrambe vettori  $n$ -dimensionali). In simboli

$$c_{ij} = \mathbf{a}_i \cdot \mathbf{b}^j = \sum_{k=1}^n a_{ik} b_{kj}, \text{ dove } i = 1, 2, \dots, m \text{ e } j = 1, 2, \dots, p$$

Per esempio

$$(4.1) \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 \\ -7 & 8 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -13 & 22 \\ -31 & 58 \end{pmatrix}$$

$$(4.2) \quad \mathbf{A} \quad \mathbf{B} \quad \mathbf{C}$$

infatti

$$c_{11} = 1 \cdot 1 + 2 \cdot (-7) + 3 \cdot 0 = -13$$

$$c_{12} = 1 \cdot 3 + 2 \cdot 8 + 3 \cdot 1 = 22$$

$$c_{21} = 4 \cdot 1 + 5 \cdot (-7) + 6 \cdot 0 = -31$$

$$c_{22} = 4 \cdot 3 + 5 \cdot 8 + 6 \cdot 1 = 58$$

Due matrici  $\mathbf{A}$  e  $\mathbf{B}$  tali che il numero delle colonne della prima coincide col numero di righe della seconda si dicono *conformabili*. Quindi il prodotto definito sopra si può effettuare solo tra matrici conformabili; si chiama anche prodotto “righe per colonne” perché l'elemento  $c_{ij}$  è ottenuto facendo il prodotto scalare del vettore riga- $i$  per il vettore colonna- $j$ .

Se  $\mathbf{A}$  è una matrice  $(m, n)$  e  $\mathbf{B}$  una matrice  $(n, p)$ , con  $m \neq p$ , è definito il prodotto  $\mathbf{AB}$  ma non il prodotto  $\mathbf{BA}$ , per cui, in generale, non ha senso chiedersi se il prodotto tra due matrici sia commutativo.

Se però  $\mathbf{A}$  e  $\mathbf{B}$  sono quadrate dello stesso ordine, entrambi i prodotti

$$\mathbf{AB} \quad \text{e} \quad \mathbf{BA}$$

sono ben definiti, ma non sono in generale uguali, come mostra il seguente esempio:

$$(4.3) \quad \mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$$

$$(4.4) \quad \mathbf{AB} = \begin{pmatrix} 0 & 2 \\ 1 & 5 \end{pmatrix} \quad \mathbf{BA} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

quindi in generale

$$\mathbf{AB} \neq \mathbf{BA}$$

Si è quindi costretti, qualora si voglia moltiplicare una matrice quadrata  $\mathbf{A}$  per una dello stesso ordine  $\mathbf{B}$ , a specificare se la moltiplicazione avviene a destra (cioè  $\mathbf{AB}$ ) oppure a sinistra (cioè  $\mathbf{BA}$ ).

Il prodotto righe per colonne è però *associativo*. Più precisamente, date tre matrici

$$\mathbf{A} \text{ di tipo } (m, n)$$

$$\mathbf{B} \text{ di tipo } (n, p)$$

$$\mathbf{C} \text{ di tipo } (p, r)$$

i due prodotti

$$\mathbf{A}(\mathbf{BC}) \quad \text{e} \quad (\mathbf{AB})\mathbf{C}$$

sono ben definiti ed uguali, come si verifica dalla definizione. Potremo quindi scrivere  $\mathbf{ABC}$  senza alcuna parentesi. In particolare, se  $\mathbf{A}$  è una matrice quadrata ha senso definire

$$\mathbf{A}^n = \underbrace{\mathbf{A} \cdot \mathbf{A} \cdot \cdots \cdot \mathbf{A}}_{n \text{ volte}}$$

Per le matrici vale anche la *proprietà distributiva*:

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

quando ciò ha senso, ovvero se  $\mathbf{A}$  è di tipo  $(n, m)$  e  $\mathbf{B}, \mathbf{C}$  sono di tipo  $(m, r)$ . Analogamente  $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{B}\mathbf{A} + \mathbf{C}\mathbf{A}$  se  $\mathbf{B}, \mathbf{C}$  sono di tipo  $(n, m)$  e  $\mathbf{A}$  è di tipo  $(m, r)$ .

È importante notare che tra le matrici quadrate di ordine  $n$  ne esiste una, che indichiamo con  $\mathbf{I}_n$ , tale che per ogni altra matrice  $\mathbf{A}$  dello stesso ordine si ha:

$$\mathbf{A}\mathbf{I}_n = \mathbf{I}_n\mathbf{A}$$

Tale matrice si chiama matrice *identità* ed ha la forma

$$(4.5) \quad \mathbf{I}_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Si chiama matrice *trasposta* di una matrice  $\mathbf{A}$  di tipo  $(m, n)$  e la si indica con il simbolo  $\mathbf{A}^\top$ , la matrice di tipo  $(n, m)$  che si ottiene da  $\mathbf{A}$  scambiando le righe con le colonne. Per esempio

$$\mathbf{A} = \begin{pmatrix} 7 & 10 & 3 \\ -5 & 4 & 2 \end{pmatrix}, \quad \mathbf{A}^\top = \begin{pmatrix} 7 & -5 \\ 10 & 4 \\ 3 & 2 \end{pmatrix}$$

In simboli, se  $\mathbf{A} = (a_{ij})$ ,  $\mathbf{A}^\top$  avrà  $a_{ji}$  come elemento di posto  $ij$ . Si osservi la seguente proprietà:

$$(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$$

se  $\mathbf{A}$  è di tipo  $(n, m)$  e  $\mathbf{B}$  di tipo  $(m, r)$ .

Se  $\mathbf{A} = \mathbf{A}^\top$  (ed allora deve essere  $m = n$ ),  $\mathbf{A}$  si dice *simmetrica*.

Ciò significa che per ogni coppia di indici  $i, j$ , risulta  $a_{ij} = a_{ji}$ , ovvero la matrice è una tabella simmetrica rispetto alla *diagonale principale* (cioè la diagonale che attraversa la matrice quadrata dall'angolo in alto a sinistra a quello in basso a destra).



$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 0 & 3 \\ -1 & 3 & 4 \end{pmatrix}$$

è una matrice simmetrica.

Una matrice quadrata del tipo

$$U = \begin{pmatrix} d_1 & * & * & * & * \\ 0 & d_2 & * & * & * \\ 0 & 0 & d_3 & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{pmatrix}$$

dove gli asterischi indicano elementi qualsiasi, si chiama *triangolare alta*. Analogamente, una matrice quadrata si dice *triangolare bassa* se tutti gli elementi sopra

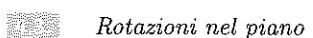
la diagonale principale sono nulli. Una matrice quadrata si dice *diagonale* se sono nulli tutti i suoi elementi fuori dalla diagonale principale. È facile controllare che: il prodotto di matrici quadrate dello stesso ordine triangolari alte (basse) è ancora triangolare alta (bassa). Evidentemente, se  $\mathbf{A}$  è triangolare e simmetrica allora deve essere diagonale; se  $\mathbf{A}$  è triangolare sia alta che bassa, allora è diagonale.



Le matrici

$$\begin{pmatrix} 2 & 3 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{pmatrix} \quad \begin{pmatrix} 2 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & -1 & 4 \end{pmatrix} \quad \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

sono, rispettivamente, triangolare alta, triangolare bassa, diagonale.



*Rotazioni nel piano*

- Consideriamo la seguente figura, nella quale un punto  $P$  di coordinate  $(x, y)$  è ruotato attorno all'origine di un angolo  $\alpha$  collocandosi nel punto  $P'$  di coordinate  $(x', y')$ . Vediamo quali sono le relazioni tra le coordinate di  $P$  e  $P'$

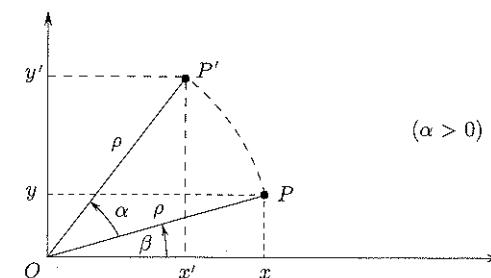


Figura 8.24.

Dalle note relazioni trigonometriche si ha:

$$x = \rho \cos \beta, \quad y = \rho \sin \beta$$

e

$$x' = \rho \cos(\alpha + \beta), \quad y' = \rho \sin(\alpha + \beta)$$

dove  $\rho$  indica la lunghezza di  $\overline{OP}$  e di  $\overline{OP'}$ . Poiché

$$\begin{aligned} \rho \cos(\alpha + \beta) &= \rho \cos \alpha \cos \beta - \rho \sin \alpha \sin \beta = x \cos \alpha - y \sin \alpha \\ \rho \sin(\alpha + \beta) &= \rho \sin \alpha \cos \beta + \rho \cos \alpha \sin \beta = x \sin \alpha + y \cos \alpha \end{aligned}$$

si ha:

$$(4.6) \quad \begin{cases} x' = x \cos \alpha - y \sin \alpha \\ y' = x \sin \alpha + y \cos \alpha \end{cases}$$

Introducendo i vettori colonna  $\begin{pmatrix} x \\ y \end{pmatrix}$  e  $\begin{pmatrix} x' \\ y' \end{pmatrix}$  e la matrice di tipo  $(2, 2)$ ,

$$(4.7) \quad \mathbf{A}_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$$

la (4.6) si può scrivere nel modo seguente:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Conclusione:

*Per trovare il punto corrispondente al punto di coordinate  $(x, y)$  dopo una rotazione (in senso antiorario) di un angolo  $\alpha$ , con centro nell'origine, è sufficiente moltiplicare a sinistra il vettore colonna  $\begin{pmatrix} x \\ y \end{pmatrix}$  per la matrice  $\mathbf{A}_\alpha$  data dalla (4.7).*

Possiamo vedere la rotazione di un angolo  $\alpha$  come una trasformazione del piano in sé, ossia come trasformazione  $L : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ . Si verifica facilmente (per esempio in base alle (4.2)) che tale trasformazione è lineare. Abbiamo visto inoltre che questa trasformazione lineare si realizza mediante moltiplicazione per un'opportuna matrice.

È altresì chiaro che ogni matrice della forma

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

con  $a^2 + b^2 = 1$  è univocamente associata ad una matrice di rotazione  $\mathbf{A}_\alpha$  che si ottiene risolvendo il sistema

$$\cos \alpha = a, \quad \sin \alpha = b$$

Quindi ogni matrice di questo tipo rappresenta una rotazione, nel senso appena spiegato. Ne concludiamo che è possibile identificare una certa classe di trasformazioni lineari del piano (le rotazioni) con una certa classe di matrici  $(2, 2)$ . Come vedremo nel prossimo paragrafo, questa affermazione può essere ampiamente generalizzata.

- Vediamo ora come si realizza l'applicazione successiva di 2 rotazioni. Supponiamo che il punto  $P$  di coordinate  $x$  e  $y$  venga trasformato da una rotazione rappresentata dalla matrice  $\mathbf{A}_\alpha$  data dalla (4.7) nel punto  $P'$  di coordinate  $x'$  e  $y'$ . Come abbiamo appena visto si ha:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \mathbf{A}_\alpha \begin{pmatrix} x \\ y \end{pmatrix}$$

Se ora al punto  $P'$  applichiamo una rotazione di angolo  $\beta$ , otterremo un punto  $P''$  le cui coordinate  $x'', y''$  si trovano con la formula

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix} \left( \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right) = \mathbf{A}_\beta \left( \mathbf{A}_\alpha \begin{pmatrix} x \\ y \end{pmatrix} \right)$$

Poiché il prodotto righe per colonne è associativo si può scrivere

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = (\mathbf{A}_\beta \mathbf{A}_\alpha) \begin{pmatrix} x \\ y \end{pmatrix}$$

cioè: se ad  $\mathbf{A}_\alpha$  è associata la rotazione di un angolo  $\alpha$ , e ad  $\mathbf{A}_\beta$  è associata la rotazione di un angolo  $\beta$ , alla matrice  $\mathbf{A}_\beta \mathbf{A}_\alpha$  risulta associata la rotazione di un angolo  $\alpha + \beta$ . In altre parole, la *composizione di due rotazioni* (operazione di carattere geometrico) si effettua con la *moltiplicazione per  $\mathbf{A}_\beta \mathbf{A}_\alpha$*  (operazione di carattere algebrico). Si osservi che, come è evidente dal significato geometrico dell'operazione, il prodotto  $\mathbf{A}_\beta \mathbf{A}_\alpha$ , per questa classe di matrici, è *commutativo*.



Dimostrare esplicitamente le seguenti proprietà delle operazioni su matrici:

$$\mathbf{A}(\mathbf{B}\mathbf{C}) = (\mathbf{AB})\mathbf{C}$$

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$$

*Suggerimento.* Ragionare pure per similità con matrici quadrate, di ordine  $n$ . Posto  $\mathbf{A} = (a_{ij})$ , ecc., occorre calcolare separatamente l'elemento di posto  $ij$  della matrice a primo e a secondo membro di ciascuna identità, e verificare che sono uguali.

Stabilire per ciascuna delle seguenti affermazioni se è vera (fornendone la dimostrazione) oppure falsa (fornendo un contesempio):

- Il prodotto di due matrici  $(n, n)$  diagonali è diagonale
- Il prodotto di due matrici  $(n, n)$  triangolari alte è triangolare alto
- Il prodotto di due matrici  $(n, n)$  triangolari è triangolare
- Il prodotto di due matrici  $(n, n)$  simmetriche è simmetrico

Scrivere la matrice che rappresenta, nel piano, la rotazione di:

- $\frac{\pi}{6}$
- $\pi$
- $-\frac{\pi}{2}$

## 4.2 Rappresentazione matriciale delle trasformazioni lineari

Abbiamo visto, nell'esempio 4.3, che la matrice quadrata  $\mathbf{A}_\alpha$  di ordine 2 realizza una rotazione del piano (attorno all'origine) di un angolo  $\alpha$ , nel senso che, preso il vettore colonna  $\begin{pmatrix} x \\ y \end{pmatrix}$ , il vettore trasformato

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \mathbf{A}_\alpha \begin{pmatrix} x \\ y \end{pmatrix}$$

si ottiene ruotando il primo di un angolo  $\alpha$  attorno all'origine (nel senso positivo).

In generale una matrice  $\mathbf{A}$ , quadrata di ordine  $n$ , trasforma un vettore di  $\mathbb{R}^n$  in un altro vettore di  $\mathbb{R}^n$  mediante il prodotto righe per colonne (se pensiamo sempre i vettori come colonne):  $\mathbf{x} \in \mathbb{R}^n$  è trasformato nel vettore  $\mathbf{Ax} \in \mathbb{R}^n$ .

Analogamente, una matrice  $\mathbf{A}$  di tipo  $(m, n)$  trasforma un vettore  $\mathbf{x} \in \mathbb{R}^n$  nel vettore  $\mathbf{Ax} \in \mathbb{R}^m$ .

Inoltre, questa trasformazione

$$L : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad L : \mathbf{x} \mapsto \mathbf{Ax}$$

è lineare. Si verifica subito, infatti, dalla definizione di prodotto righe per colonne, che per ogni  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $\lambda \in \mathbb{R}$ ,

$$\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{Ax} + \mathbf{Ay}$$

$$\mathbf{A}(\lambda \mathbf{x}) = \lambda \mathbf{Ax}$$

Infine, considerazioni analoghe valgono se al campo reale sostituiamo il campo complesso  $\mathbb{C}$ : una matrice di tipo  $(m, n)$  a elementi complessi realizza una trasformazione lineare di  $\mathbb{C}^n$  in  $\mathbb{C}^m$ .

Il prossimo importante teorema afferma, in sostanza, che *tutte le trasformazioni lineari* da  $\mathbb{R}^n$  a  $\mathbb{R}^m$  (o da  $\mathbb{C}^n$  a  $\mathbb{C}^m$ ) sono di questo tipo; anzi, tutte le trasformazioni lineari tra due spazi vettoriali qualsiasi, di dimensione finita, si possono rappresentare a questo modo, nel senso ora precisato:

**TEOREMA 8.4 (TEOREMA DI RAPPRESENTAZIONE)** *Siano  $V_n, V_m$  due spazi vettoriali sul campo  $\mathbb{K}$  (che può essere  $\mathbb{R}$  o  $\mathbb{C}$ ) di dimensione  $n, m$ , rispettivamente, e sia  $\mathcal{L} : V_n \rightarrow V_m$  una trasformazione lineare. Fissate due basi  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  e  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  in  $V_n$  e  $V_m$ , rispettivamente, esiste un'unica matrice  $\mathbf{A}$  di tipo  $(m, n)$  a elementi in  $\mathbb{K}$  che rappresenta  $\mathcal{L}$  nel senso che, se  $\mathbf{x} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \dots + x_n\mathbf{u}_n$  e  $\mathcal{L}(\mathbf{x}) = y_1\mathbf{v}_1 + \dots + y_m\mathbf{v}_m$ , allora*

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Si noti che, qualunque sia lo spazio vettoriale  $V_n$  di dimensione  $n$  sul campo  $\mathbb{K}$ , fissata una base nello spazio, ogni vettore è individuato dalla  $n$ -upla delle sue componenti rispetto a questa base, cioè da un elemento di  $\mathbb{K}^n$ . Il teorema afferma quindi l'esistenza di una matrice che rappresenta la trasformazione lineare, nel senso che, fissata una base nello spazio di partenza ed una nello spazio di arrivo, il vettore delle componenti di  $\mathcal{L}(\mathbf{x})$  si ottiene moltiplicando la matrice per il vettore delle componenti di  $\mathbf{x}$ .

Nel seguito, per non appesantire il linguaggio, parleremo di matrici a elementi reali e di trasformazioni lineari di  $\mathbb{R}^n$  in  $\mathbb{R}^m$ . Tutto quanto diremo, però, varrà pari pari sostituendo  $\mathbb{C}$  ad  $\mathbb{R}$ . Ambientare la teoria in  $\mathbb{C}^n$  anziché in  $\mathbb{R}^n$  diventerà importante quando parleremo di *diagonalizzazione* (par. 6).

**DIMOSTRAZIONE.** Essendo  $\mathcal{L}(\mathbf{u}_1) \in V_m$ , si può scrivere

$$\mathcal{L}(\mathbf{u}_1) = a_{11}\mathbf{v}_1 + \dots + a_{m1}\mathbf{v}_m$$

Analogamente, abbiamo

$$\mathcal{L}(\mathbf{u}_2) = a_{12}\mathbf{v}_1 + \dots + a_{m2}\mathbf{v}_m$$

⋮

$$\mathcal{L}(\mathbf{u}_n) = a_{1n}\mathbf{v}_1 + \dots + a_{mn}\mathbf{v}_m$$

per opportuni coefficienti  $a_{ij} \in \mathbb{R}$ .

Poniamo ora

$$\mathbf{A} = (a_{ij}) \quad i = 1, 2, \dots, m \quad j = 1, 2, \dots, n$$

e sia  $\mathbf{x} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \dots + x_n\mathbf{u}_n$ .

Si ha, per la linearità di  $\mathcal{L}$  e le equazioni precedenti:

$$\begin{aligned} \mathcal{L}(\mathbf{x}) &= \mathcal{L} \left( \sum_{j=1}^n x_j \mathbf{u}_j \right) = \sum_{j=1}^n x_j \mathcal{L}(\mathbf{u}_j) = \sum_{j=1}^n x_j \sum_{i=1}^m a_{ij} \mathbf{v}_i = \sum_{j=1}^n \sum_{i=1}^m a_{ij} x_j \mathbf{v}_i \\ &= \sum_{i=1}^m \left( \sum_{j=1}^n a_{ij} x_j \right) \mathbf{v}_i \end{aligned}$$

Si vede quindi che le componenti scalari di  $\mathcal{L}(\mathbf{x})$  rispetto alla base  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  sono date dalla formula

$$y_i = \sum_{j=1}^n a_{ij} x_j \quad i = 1, 2, \dots, m$$

che, raggruppate in forma matriciale, danno

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

La matrice  $\mathbf{A}$  rappresenta dunque  $\mathcal{L}$  nel senso precisato nell'enunciato del teorema. Tale matrice è unica, come conseguenza del fatto che ogni vettore  $\mathcal{L}(\mathbf{u}_i)$ ,  $i = 1, 2, \dots, n$  si può esprimere in un unico modo come combinazione lineare dei vettori  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ .



Le trasformazioni lineari  $\mathcal{L}$  da  $\mathbb{R}$  a  $\mathbb{R}$  si possono scrivere come

$$\mathcal{L}(x) = ax \quad a \in \mathbb{R}$$

(La matrice  $\mathbf{A} = (a)$  è di tipo  $(1, 1)$ .)

Le trasformazioni lineari da  $\mathbb{R}^2$  in  $\mathbb{R}$ , in riferimento alla base canonica, si possono scrivere nella forma

$$\mathcal{L}(x, y) = ax + by \quad a, b \in \mathbb{R}$$

Le trasformazioni lineari da  $\mathbb{R}^2$  a  $\mathbb{R}^2$  rappresentate da matrici del tipo

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

corrispondono, geometricamente, alle roto-omotetie nel piano, ossia alla composizione di una rotazione di angolo  $\theta$  e una dilatazione di coefficiente  $\lambda$ , con

$$\lambda = \sqrt{a^2 + b^2} \quad \cos \theta = \frac{a}{\lambda} \quad \sin \theta = \frac{b}{\lambda}$$

Infatti, con queste notazioni, la matrice si riscrive nella forma:

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \lambda \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

e, come visto nell'esempio 4.3, l'ultima matrice scritta rappresenta una rotazione di angolo  $\theta$ , mentre il prodotto per lo scalare  $\lambda$  rappresenta una dilatazione di coefficiente  $\lambda$ .

### Osservazioni sul teorema di rappresentazione

- Il teorema di rappresentazione chiarisce il motivo per cui il prodotto di matrici è stato definito proprio in quel modo (“righe per colonne”). Infatti, è quello il tipo di prodotto adatto a rappresentare le trasformazioni lineari.
- L’osservazione precedente è anche alla base dell’utilità del calcolo matriciale nella fisica. Nello studio della fisica elementare, si incontrano grandezze scalari (come la massa di un punto materiale) e grandezze vettoriali (come la velocità di un punto materiale). Ora, alcune importanti leggi fisiche affermano che una certa grandezza vettoriale dipende da un’altra grandezza vettoriale, secondo una legge lineare. In questo caso, la dipendenza è espressa, matematicamente, dal prodotto per una matrice. Un esempio importante di questo genere, preso dalla meccanica, è il seguente:



**Matrice d’inerzia.** Consideriamo un corpo rigido libero di ruotare attorno al proprio centro di massa, che supponiamo posto nell’origine. La grandezza cinematica rilevante è il vettore *velocità angolare*,  $\omega$ , che ha il significato cinematico seguente: in ogni istante il punto del corpo che si trova nella posizione  $\mathbf{r}$  ha velocità

$$\mathbf{v} = \omega \times \mathbf{r}.$$

La grandezza dinamica rilevante è invece il vettore *momento della quantità di moto*  $\mathbf{L}$ , che per esempio si mantiene costante se il corpo non è soggetto a forze (“moto per inerzia”). Tra queste due grandezze si dimostra esistere una relazione lineare. Ciò significa che esiste una matrice  $(3, 3)$ , che indichiamo con  $\mathbf{J}$ , tale che

$$\mathbf{L} = \mathbf{J}\omega.$$

Questa matrice, fisicamente, dipende da come la massa del corpo rigido è distribuita rispetto al centro di massa, ed è detta “matrice di inerzia”; si dimostra anche che  $\mathbf{J}$  è simmetrica. Più precisamente, l’oggetto rilevante dal punto di vista fisico è la *trasformazione lineare* che associa  $\mathbf{L}$  ad  $\omega$ . Questa si rappresenta con una matrice una volta fissato un sistema di riferimento in  $\mathbb{R}^3$ ; cambiando il sistema di riferimento, cambia la matrice ma non la trasformazione lineare che essa rappresenta. In fisica una trasformazione lineare tra due spazi vettoriali viene detta *tensore*. Perciò in questo caso si parla anche di *tensore d’inerzia*. Il tensore d’inerzia rappresenta una caratteristica intrinseca del corpo rigido, indipendente dal suo stato di moto, che determina la relazione esistente tra velocità angolare e momento della quantità di moto.

Vediamo quindi che accanto alle grandezze fisiche *scalari* o *vettoriali*, vi sono anche *grandezze fisiche matriciali* (o meglio *tensoriali*), ossia rappresentate da trasformazioni lineari tra spazi vettoriali).

Siano ora  $L_1 : \mathbb{R}^n \rightarrow \mathbb{R}^m$  e  $L_2 : \mathbb{R}^m \rightarrow \mathbb{R}^s$  due trasformazioni lineari, rappresentate rispettivamente da una matrice  $\mathbf{A}$ , di tipo  $(m, n)$ , e da una matrice  $\mathbf{B}$ , di tipo  $(s, m)$ , ottenute fissando una volta per tutte una base in  $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^s$  (è necessario che la

base considerata in  $\mathbb{R}^m$  sia la stessa per  $L_1$  e  $L_2$ ). Consideriamo la *trasformazione composta*,

$$\begin{aligned} L_2 \circ L_1 : \mathbb{R}^n &\rightarrow \mathbb{R}^s \\ (L_2 \circ L_1) : \mathbf{x} &\mapsto L_2(L_1\mathbf{x}). \end{aligned}$$

È immediato verificare che  $L_2 \circ L_1$  è anch’essa una trasformazione lineare; per il Teorema di rappresentazione, esisterà una matrice  $\mathbf{C}$ , di tipo  $(s, n)$  che rappresenta tale trasformazione rispetto alle stesse basi. Chi è questa matrice  $\mathbf{C}$ ? Semplicemente,

$$\mathbf{C} = \mathbf{BA}$$

ossia la matrice rappresentativa di  $L_2 \circ L_1$  è il prodotto righe per colonne delle matrici  $\mathbf{B}, \mathbf{A}$  che rappresentano  $L_2, L_1$ , rispettivamente. Questo risultato è anzi il motivo principale per cui è naturale definire il prodotto di matrici proprio in quel modo “complicato”. La verifica di questo fatto è immediata: scrivendo il vettore  $\mathbf{x}$  come  $n$ -upla rispetto alla base fissata, si ha che

$$L_2(L_1\mathbf{x}) = \mathbf{B} \cdot (\mathbf{A}\mathbf{x}) = (\mathbf{BA})\mathbf{x}$$

ossia  $\mathbf{BA}$  rappresenta  $L_2 \circ L_1$ .

Il discorso vale inalterato se anziché  $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^s$  si considerano tre spazi vettoriali qualsiasi di dimensioni, rispettivamente,  $n, m, s$ .

Notiamo infine che una *trasformazione affine* tra due spazi vettoriali (v. par. 3.2), cioè del tipo

$$\begin{aligned} f : V_n &\rightarrow V_m \\ f(\mathbf{v}) &= l(\mathbf{v}) + \mathbf{u} \end{aligned}$$

con  $l : V_n \rightarrow V_m$  lineare e  $\mathbf{u} \in V_m$  vettore fissato, si rappresenterà, rispetto a basi fissate in  $V_n, V_m$ , nella forma matriciale

$$L(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$$

dove  $\mathbf{A}$  è un’opportuna matrice  $(m, n)$  e  $\mathbf{b} \in \mathbb{R}^m$  il vettore che rappresenta  $\mathbf{u}$  rispetto alla base scelta in  $V_m$ .

### 4.3 Determinante

Uno dei concetti centrali associati ad una matrice quadrata è quello di *determinante*, la cui definizione generale, tuttavia, non è elementare.

D’altra parte, noi siamo interessati soprattutto al *calcolo* dei determinanti. Perciò, adottando un atteggiamento “ricorsivo”, definiremo esplicitamente il determinante per matrici di ordine  $n = 1$  e  $n = 2$ , indicando poi come si ottenga il determinante di una matrice di ordine  $n$  una volta spiegato come esso si ottenga per una matrice di ordine  $n - 1$ .

Per  $n = 1$ , cioè per una matrice costituita dal solo elemento  $a_{11}$ , il determinante è, per definizione, l'elemento stesso.

Per  $n = 2$ , cioè per una matrice quadrata di ordine 2, del tipo,

$$\mathbf{M} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

il determinante, indicato con  $\det \mathbf{M}$ , oppure con  $|\mathbf{M}|$ , è assegnato dalla formula

$$\boxed{\det \mathbf{M} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}}$$

Per esempio il determinante della matrice

$$\mathbf{M} = \begin{pmatrix} 1 & -3 \\ 2 & -4 \end{pmatrix}$$

è uguale a  $1 \cdot (-4) - 2(-3) = 2$ .

Sia data ora la matrice quadrata di ordine  $n$ .

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Conviene introdurre alcune definizioni.

- 1) Si chiama minore complementare di un elemento  $a_{ij}$  e lo si indica con  $M_{ij}$ , il determinante della matrice ottenuta cancellando la riga  $i$  e la colonna  $j$  dalla matrice  $\mathbf{A}$ .
- 2) Si chiama complemento algebrico di un elemento  $a_{ij}$  il numero

$$A_{ij} = (-1)^{i+j} M_{ij}$$

Notiamo che

$$\begin{aligned} (-1)^{i+j} &= 1 && \text{se } i + j \text{ è pari} \\ (-1)^{i+j} &= -1 && \text{se } i + j \text{ è dispari} \end{aligned}$$

Per esempio, sia:

$$(4.8) \quad \mathbf{A} = \begin{pmatrix} 8 & 2 & -1 \\ 4 & 3 & -5 \\ -6 & 0 & 4 \end{pmatrix}$$

Consideriamo l'elemento  $a_{13} = -1$ . Si ha:

$$M_{13} = \det \begin{pmatrix} 4 & 3 \\ -6 & 0 \end{pmatrix} = 18, \quad A_{13} = (-1)^{1+3} M_{13} = 18$$

- 3) Si chiama determinante della matrice quadrata  $\mathbf{A}$  la somma dei prodotti degli elementi di una qualunque linea (riga o colonna) per i loro complementi algebrici. In formule, per ogni  $k$ ,  $1 \leq k \leq n$ , si ha: fissando la riga  $k$ ,

$$(4.9) \quad \det \mathbf{A} = a_{k1} A_{k1} + a_{k2} A_{k2} + \dots + a_{kn} A_{kn}$$

oppure, fissando la colonna  $k$

$$\det \mathbf{A} = a_{1k} A_{1k} + a_{2k} A_{2k} + \dots + a_{nk} A_{nk}$$

Il determinante viene così definito in modo iterativo: infatti la (4.5) assegna il determinante di una matrice  $n \times n$ , posto che si sappia calcolare il determinante delle matrici  $(n-1) \times (n-1)$  (ricordare la definizione di complemento algebrico). La definizione precedente ha senso in virtù del seguente:

**TEOREMA 8.5 (TEOREMA DI LAPLACE)** Il risultato che si ottiene calcolando il determinante di una matrice come nella (4.9) non dipende dalla particolare riga o colonna scelta.

Per esempio, calcoliamo il determinante della matrice  $\mathbf{A}$  in (4.8) scegliendo la terza riga:

$$\det \mathbf{A} = -6 \cdot \begin{vmatrix} 2 & -1 \\ 3 & -5 \end{vmatrix} - 0 \cdot \begin{vmatrix} 8 & -1 \\ 4 & -5 \end{vmatrix} + 4 \cdot \begin{vmatrix} 8 & 2 \\ 4 & 3 \end{vmatrix} = 58$$

Il lettore è invitato a calcolare lo stesso determinante scegliendo, per esempio, la prima colonna: troverà lo stesso risultato.

Si voglia ora calcolare il determinante della matrice seguente:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 5 & 1 \\ 2 & 0 & 3 & 0 \\ 0 & -1 & -2 & 1 \end{pmatrix}$$

Applicando il teorema, considerando la prima colonna si ha:

$$\begin{aligned} \det \mathbf{A} &= 1 \cdot \begin{vmatrix} 2 & 5 & 1 \\ 0 & 3 & 0 \\ -1 & -2 & 1 \end{vmatrix} - 0 \cdot \begin{vmatrix} 2 & 3 & 4 \\ 0 & 3 & 0 \\ -1 & -2 & 1 \end{vmatrix} + 2 \cdot \begin{vmatrix} 2 & 3 & 4 \\ 2 & 5 & 1 \\ -1 & -2 & 1 \end{vmatrix} - 0 \cdot \begin{vmatrix} 2 & 3 & 4 \\ 0 & 3 & 0 \\ 0 & 3 & 0 \end{vmatrix} = \\ &= 2 \cdot \begin{vmatrix} 3 & 0 \\ -2 & 1 \end{vmatrix} - 1 \cdot \begin{vmatrix} 5 & 1 \\ 3 & 0 \end{vmatrix} + 2 \cdot \left\{ 2 \cdot \begin{vmatrix} 5 & 1 \\ -2 & 1 \end{vmatrix} - 2 \cdot \begin{vmatrix} 3 & 4 \\ -2 & 1 \end{vmatrix} - 1 \cdot \begin{vmatrix} 3 & 4 \\ 5 & 1 \end{vmatrix} \right\} = \\ &= 6 + 3 + 2(14 - 22 + 17) = 27 \end{aligned}$$

**TEOREMA 8.6 (PROPRIETÀ ELEMENTARI DEL DETERMINANTE)** Sia  $\mathbf{A}$  una matrice  $(n, n)$ .

- a. Se  $\mathbf{A}$  ha una riga o una colonna di zeri,  $\det \mathbf{A} = 0$ .

- b. Scambiando due righe o due colonne, il determinante cambia di segno.

c. Se  $\mathbf{A}$  ha due righe o due colonne uguali,  $\det \mathbf{A} = 0$ .

d. Il determinante è una funzione lineare di ciascuna sua riga (o colonna), ossia (esemplificando sulla prima riga)

$$d1 \quad \det \begin{pmatrix} \mathbf{a}_1 + \mathbf{b}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} + \det \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix}$$

$$d2 \quad \det \begin{pmatrix} \lambda \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} = \lambda \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix}$$

(gli  $\mathbf{a}_i$  sono i vettori riga che costituiscono la matrice)

e. Se ad una riga (colonna) si aggiunge una qualunque combinazione lineare delle altre righe (colonne), il determinante non cambia.

f. Se le righe (colonne) di  $\mathbf{A}$  sono vettori linearmente dipendenti,  $\det \mathbf{A} = 0$ .

g.  $\det(\lambda \mathbf{A}) = \lambda^n \det \mathbf{A}$ .

h. Se  $\mathbf{A}$  è triangolare (in particolare, se è diagonale)  $\det \mathbf{A} = a_{11} \cdot a_{22} \cdots \cdot a_{nn}$ .

#### DIMOSTRAZIONE.

a. Basta calcolare  $\det \mathbf{A}$  con la (4.9) scegliendo come riga  $k$  la riga di zeri.

b. Dimostriamolo ricorsivamente. Per le matrici di ordine 2 si verifica subito. Per le matrici di ordine 3: scambiamo tra loro 2 righe, e calcoliamo il determinante con la (4.9) scegliendo come riga  $k$  quella che non è stata mossa. Allora i 3 minori complementari risultano di segno cambiato, perché nelle 3 matrici  $(2,2)$  corrispondenti sono state scambiate le righe. Dunque il segno del determinante è cambiato. Sfruttando ora il risultato per le matrici  $(3,3)$  si dimostra analogamente quello per le  $(4,4)$ , e così via.

c. Scambiando tra loro le due righe uguali il segno del determinante deve cambiare, per b; d'altro canto la matrice non è cambiata, dunque il determinante è lo stesso, ossia  $\det \mathbf{A} = -\det \mathbf{A}$ , da cui  $\det \mathbf{A} = 0$ .

d1. Calcoliamo il determinante in base alla (4.9), mediante la prima riga, ponendo

$$\mathbf{a}_1 = (a_{11}, a_{12}, \dots, a_{1n}); \mathbf{b}_1 = (b_{11}, b_{12}, \dots, b_{1n}).$$

$$\begin{aligned} \det \begin{pmatrix} \mathbf{a}_1 + \mathbf{b}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} &= (a_{11} + b_{11}) A_{11} + \dots + (a_{1n} + b_{1n}) A_{1n} = \\ &= (a_{11} A_{11} + \dots + a_{1n} A_{1n}) + (b_{11} A_{11} + \dots + b_{1n} A_{1n}) = \\ &= \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} + \det \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} \end{aligned}$$

In modo analogo si verifica d2.

e. Indicando le righe della matrice con i vettori  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  si ha:

$$\det \begin{pmatrix} \mathbf{a}_1 + \lambda_1 \mathbf{a}_2 + \lambda_2 \mathbf{a}_3 + \dots + \lambda_n \mathbf{a}_n \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} =$$

per il punto d

$$= \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} + \lambda_1 \det \begin{pmatrix} \mathbf{a}_2 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} + \dots + \lambda_n \det \begin{pmatrix} \mathbf{a}_n \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix}$$

perché tutti gli addendi tranne il primo sono determinanti di matrici con due righe uguali, e quindi si annullano, per il punto c.

f. Se le righe sono vettori linearmente dipendenti, una di esse è combinazione lineare delle altre; supponiamo per esempio che la prima riga sia combinazione lineare delle altre, ossia:

$$\mathbf{A} = \begin{pmatrix} \lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + \dots + \lambda_n \mathbf{a}_n \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix}$$

Allora, per il punto d,

$$\det \mathbf{A} = \lambda_1 \det \begin{pmatrix} \mathbf{a}_2 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} + \dots + \lambda_n \det \begin{pmatrix} \mathbf{a}_n \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} = 0$$

per il punto c, perché ognuna delle matrici scritte ha due righe uguali.

g. Si applica  $n$  volte la proprietà d2:

$$\begin{aligned} \det(\lambda \mathbf{A}) &= \det \begin{pmatrix} \lambda \mathbf{a}_1 \\ \lambda \mathbf{a}_2 \\ \dots \\ \lambda \mathbf{a}_n \end{pmatrix} = \lambda \det \begin{pmatrix} \mathbf{a}_1 \\ \lambda \mathbf{a}_2 \\ \dots \\ \lambda \mathbf{a}_n \end{pmatrix} = \lambda^2 \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \lambda \mathbf{a}_n \end{pmatrix} = \\ &= \dots = \lambda^n \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \dots \\ \mathbf{a}_n \end{pmatrix} = \lambda^n \det \mathbf{A} \end{aligned}$$

h. Supponiamo per esempio che  $\mathbf{A}$  sia triangolare alta, e calcoliamo  $\det \mathbf{A}$  mediante la (4.9), sulla prima colonna. Allora  $\det \mathbf{A} = a_{11} A_{11}$ . Ma anche la matrice di cui  $A_{11}$  è il determinante è triangolare alta, perciò calcolando  $A_{11}$  allo stesso modo, ricorsivamente si giunge alla tesi.



Mostriamo come il Teorema precedente possa essere talvolta utilizzato per semplificare il calcolo di un determinante. Si vuole calcolare:

$$\begin{vmatrix} 2 & 1 & 2 \\ 1 & 3 & 1 \\ 4 & -1 & -2 \end{vmatrix}$$

Un momento di osservazione mostra che se si sottrae dalla prima riga il doppio della seconda, si ottiene una riga con due zeri, molto comoda per il calcolo del determinante. Per il punto e del Teorema, quest'operazione non altera il determinante. Perciò si ha:

$$\begin{vmatrix} 2 & 1 & 2 \\ 1 & 3 & 1 \\ 4 & -1 & -2 \end{vmatrix} = \begin{vmatrix} 0 & -5 & 0 \\ 1 & 3 & 1 \\ 4 & -1 & -2 \end{vmatrix} = 5 \begin{vmatrix} 1 & 1 \\ 4 & -2 \end{vmatrix} = 5(-2 - 4) = -30.$$

Un'altra importante proprietà del determinante, la cui dimostrazione è però meno elementare, è la seguente, che ci limitiamo ad enunciare:

**TEOREMA 8.7 (TEOREMA DI BINET)** *Se  $\mathbf{A}$  e  $\mathbf{B}$  sono matrici quadrate dello stesso ordine,*

$$\det \mathbf{AB} = \det \mathbf{A} \cdot \det \mathbf{B}.$$

- *Regola di Sarrus per i determinanti del terzo ordine.* Calcoliamo il determinante della matrice

$$\mathbf{M} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

mediante il teorema di Laplace, sviluppando secondo gli elementi della prima riga. Si trova

$$\det \mathbf{M} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} =$$

(raggruppando i termini con segno “+” e quelli con segno “-”)

$$= a_{11} a_{22} a_{33} + a_{12} a_{23} a_{31} + a_{13} a_{21} a_{32} - a_{13} a_{22} a_{31} - a_{11} a_{23} a_{32} - a_{12} a_{21} a_{33}$$

Questa somma può essere ottenuta col seguente procedimento, noto come *regola di Sarrus*.

Si accosta a destra di  $\mathbf{M}$  la prima e la seconda colonna:

$$\begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array} \quad \begin{array}{c} - \\ - \\ - \\ + \\ + \\ + \end{array}$$

Si sommano i prodotti degli elementi di ciascuna diagonale contrassegnata col segno “+” e si sottraggono i prodotti degli elementi di ciascuna diagonale contrassegnata col segno “-”.



Si voglia calcolare il determinante della matrice

$$\mathbf{M} = \begin{pmatrix} 1 & -1 & 3 \\ 2 & 4 & 0 \\ 1 & -2 & 4 \end{pmatrix}$$

Usando la regola di Sarrus si ha:

$$\begin{array}{ccccccc} 1 & -1 & 3 & 1 & -1 & 1 & \\ 2 & 4 & 0 & 2 & 4 & 0 & \\ 1 & -2 & 4 & 1 & -2 & 4 & \\ - & - & - & + & + & + & \end{array} \Rightarrow \det \mathbf{M} = 16 + 0 + (-12) - 12 - 0 - (-8) = 0.$$

#### Prodotto vettoriale e prodotto misto. Significato geometrico del determinante

Il prodotto vettoriale dei due vettori di  $\mathbb{R}^3$

$$\mathbf{v} = (x_1, x_2, x_3) \quad \mathbf{w} = (y_1, y_2, y_3)$$

assegnato dalla formula (1.13) si può riscrivere nel modo seguente:

$$(4.10) \quad \mathbf{v} \times \mathbf{w} = \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \mathbf{k}$$

La (4.10) si può ottenere sviluppando simbolicamente secondo la prima riga il determinante indicato sotto:

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

Se  $x_3 = y_3 = 0$ , la (4.10) si riduce a

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \mathbf{k}$$

e quindi

$$|\mathbf{v} \times \mathbf{w}| = \left| \det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \right|$$

Ricordando il significato del modulo del prodotto vettoriale (par. 1.2, fig. 8.16) si ricava che il *modulo del determinante*  $\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}$  coincide con l'area del parallelogramma costruito sui vettori piani  $(x_1, x_2), (y_1, y_2)$ .

Poiché l'area del parallelogramma è nulla se e solo se i due vettori che lo generano sono paralleli, si ha anche:

*il determinante  $2 \times 2$  è nullo se e solo se le sue righe (o colonne) sono due vettori paralleli, cioè linearmente dipendenti.*

Se ora  $\mathbf{u} = (z_1, z_2, z_3)$  è un terzo vettore, dalla (4.10) ricaviamo la seguente formula per il prodotto misto  $\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}$ :

$$(4.11) \quad \mathbf{u} \cdot \mathbf{v} \times \mathbf{w} = z_1 \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} - z_2 \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + z_3 \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = \begin{vmatrix} z_1 & z_2 & z_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

Ricordando il significato geometrico del prodotto misto (par. 1.2, fig. 8.15) si ricava che *il modulo del determinante (4.11) rappresenta il volume del parallelepipedo costruito sui vettori  $\mathbf{u}$ ,  $\mathbf{v}$  e  $\mathbf{w}$ .*

Poiché, a sua volta, il volume del parallelepipedo è nullo se e solo se i tre vettori che lo generano sono complanari, si ha anche:

*il determinante  $3 \times 3$  è nullo se e solo se le sue righe (o colonne) sono 3 vettori complanari, cioè linearmente dipendenti.*

Otteniamo quindi, nel caso particolare  $n = 2, 3$ , un comodo criterio di dipendenza o indipendenza lineare, mediante il calcolo del determinante.

La formula (4.11) permette anche di dimostrare facilmente la proprietà di *invarianza ciclica* del prodotto misto:

$$\mathbf{u} \cdot \mathbf{v} \times \mathbf{w} = \mathbf{w} \cdot \mathbf{u} \times \mathbf{v} = \mathbf{v} \cdot \mathbf{w} \times \mathbf{u}$$

Infatti, ricordando che il segno del determinante cambia scambiando 2 righe della matrice, si ha per esempio:

$$\begin{vmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{w} \end{vmatrix} = - \begin{vmatrix} \mathbf{v} \\ \mathbf{u} \\ \mathbf{w} \end{vmatrix} = \begin{vmatrix} \mathbf{w} \\ \mathbf{u} \\ \mathbf{v} \end{vmatrix}$$

che dimostra la prima uguaglianza (analoga si prova l'altra).

#### 4.4 Caratteristica di una matrice

Abbiamo visto che, se le righe (colonne) di  $\mathbf{A}$  sono vettori linearmente dipendenti, allora  $\det \mathbf{A} = 0$  (Teorema 8.6). Inoltre, abbiamo dimostrato che se  $n = 2$  o  $3$ , vale anche il viceversa (per il significato geometrico del determinante): il determinante si annulla se e solo se le righe della matrice sono linearmente dipendenti. Questo risultato è valido in realtà per  $n$  qualsiasi, e può essere anzi ulteriormente generalizzato allo studio di un qualsiasi numero  $r$  di vettori di  $\mathbb{R}^n$  (con  $r \leq n$ ):

**TEOREMA 8.8** *Siano  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r$   $r$  vettori riga di  $\mathbb{R}^n$  ( $r < n$ ), e sia  $\mathbf{A}$  la matrice  $(r, n)$  che ha per righe questi vettori. Allora i vettori  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r$  sono linearmente dipendenti se e solo se ogni matrice  $(r, r)$  estratta da  $\mathbf{A}$  ha determinante nullo; sono indipendenti se e solo se esiste almeno una matrice  $(r, r)$  estratta da  $\mathbf{A}$ , con determinante diverso da zero. Inoltre:  $n$  vettori di  $\mathbb{R}^n$  sono linearmente dipendenti (indipendenti) se e solo se la matrice  $(n, n)$  che si ottiene accostandoli ha determinante uguale a zero (diverso da zero).*

Omettiamo la dimostrazione di questo teorema, un po' laboriosa. Completiamo il quadro della situazione ricordando che  $r$  vettori di  $\mathbb{R}^n$  con  $r > n$  sono *sempre* linearmente dipendenti.

Introduciamo a questo punto il concetto di *rango* o *caratteristica* di una matrice.

**DEFINIZIONE 8.10 (CARATTERISTICA DI UNA MATRICE)** Dati una matrice  $\mathbf{A}$  di  $m$  righe e  $n$  colonne (con  $m$  non necessariamente uguale a  $n$ ) e un intero  $k \leq \min(m, n)$ , si dice *minore di ordine  $k$  estratto dalla matrice  $\mathbf{A}$* , il determinante di una qualsiasi matrice di ordine  $k$  ottenuta con gli elementi comuni a  $k$  righe e  $k$  colonne di  $\mathbf{A}$ .

Si definisce *caratteristica* o *rango* di  $\mathbf{A}$  l'intero  $r \geq 0$  tale che: esiste un minore estratto da  $\mathbf{A}$  di ordine  $r$  non nullo e ogni minore estratto da  $\mathbf{A}$  di ordine  $r+1$  è nullo.

Dal teorema 8.8 segue subito che

Il rango di una matrice rappresenta il massimo numero di righe o di colonne linearmente indipendenti.

Per la determinazione del rango risulta utile la seguente proposizione.

**PROPOSIZIONE 8.3 (DI KRONECKER)** Condizione necessaria e sufficiente affinché una matrice abbia rango  $k$  è che esista un minore di ordine  $k$  diverso da zero e siano nulli tutti i minori di ordine  $k+1$  ottenuti da quello orlandolo con una qualunque altra riga o colonna.



Applichiamo il metodo indicato nella proposizione 8.3 alla matrice:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 0 & 3 & 5 \\ 2 & 4 & 6 & 8 \end{pmatrix}$$

Consideriamo una matrice estratta da  $\mathbf{A}$  di ordine 1 con determinante diverso da zero, sia

$$\mathbf{A}^* = (1)$$

“Orliamo” tale matrice, in tutti i modi possibili, finché si ottenga una matrice del secondo ordine con determinante non nullo:

$$\mathbf{A}^{**} = \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix} \quad \det \mathbf{A}^{**} = -4.$$

Ripetendo il ragionamento, “orlando” cioè in tutti i modi possibili la  $\mathbf{A}^{**}$ :

$$\mathbf{A}_1^{**} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 3 \\ 2 & 4 & 6 \end{pmatrix} \quad \mathbf{A}_2^{**} = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 0 & 5 \\ 2 & 4 & 8 \end{pmatrix}$$

si ottengono due matrici i cui determinanti sono nulli. Il rango è dunque 2 in accordo con il teorema precedente.

Per la determinazione del rango di  $\mathbf{A}$  seguendo la definizione avremmo dovuto calcolare quattro determinanti del terzo ordine, mentre col metodo della proposizione 8.3 i determinanti  $3 \times 3$  da calcolare si sono ridotti a due.

Quanto più grandi sono le dimensioni della matrice tanto più evidenti risultano i vantaggi del metodo indicato nella proposizione.

#### 4.5 Matrice inversa

Se  $a$  è un numero non nullo, esiste un unico numero  $a^{-1} = \frac{1}{a}$  (il reciproco di  $a$ ) tale che

$$(4.12) \quad a \cdot a^{-1} = a^{-1} \cdot a = 1$$

Se  $\mathbf{A}$  è una matrice quadrata  $(n, n)$ , chiameremo *matrice inversa* di  $\mathbf{A}$  la matrice (se esiste)  $\mathbf{A}^{-1}$  tale che

$$(4.13) \quad \mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{I}_n$$

(dove  $\mathbf{I}_n$  è la matrice identità (4.5)), in perfetta analogia con la (4.12). La condizione che garantisce l'esistenza della matrice inversa, analoga alla condizione  $a \neq 0$  per l'esistenza del reciproco di un numero, è che sia  $\det \mathbf{A} \neq 0$ . Il seguente teorema precisa questa affermazione, e indica un modo per calcolare  $\mathbf{A}^{-1}$ :

**TEOREMA 8.9** Condizione necessaria e sufficiente affinché esista la matrice inversa  $\mathbf{A}^{-1}$  è che  $\mathbf{A}$  sia non singolare, cioè che  $\det \mathbf{A} \neq 0$ . In tal caso vale la formula:

$$(4.14) \quad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \cdot \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix}^{\top}$$

dove  $\top$  indica "trasposta" e gli  $A_{ij}$  sono i complementi algebrici degli elementi  $a_{ij}$  della matrice  $\mathbf{A}$ . Infine,  $\det(\mathbf{A}^{-1}) = \frac{1}{\det \mathbf{A}}$ .



Usiamo la formula (4.14) per calcolare l'inversa della matrice

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & 3 \\ 0 & 0 & 2 \end{pmatrix}$$

Essendo triangolare,  $\det \mathbf{A} = 1 \cdot (-1) \cdot 2 = -2 \neq 0$ ; quindi  $\mathbf{A}^{-1}$  esiste. Osserviamo ora che:

$$A_{11} = \begin{vmatrix} -1 & 3 \\ 0 & 2 \end{vmatrix} = -2, \quad A_{12} = -\begin{vmatrix} 0 & 3 \\ 0 & 2 \end{vmatrix} = 0, \quad A_{13} = \begin{vmatrix} 0 & -1 \\ 0 & 0 \end{vmatrix} = 0$$

$$A_{21} = -\begin{vmatrix} 2 & 3 \\ 0 & 2 \end{vmatrix} = -4, \quad A_{22} = \begin{vmatrix} 1 & 3 \\ 0 & 2 \end{vmatrix} = 2, \quad A_{23} = -\begin{vmatrix} 1 & 2 \\ 0 & 0 \end{vmatrix} = 0$$

$$A_{31} = \begin{vmatrix} 2 & 3 \\ -1 & 3 \end{vmatrix} = 9, \quad A_{32} = -\begin{vmatrix} 1 & 3 \\ 0 & 3 \end{vmatrix} = -3, \quad A_{33} = \begin{vmatrix} 1 & 2 \\ 0 & -1 \end{vmatrix} = -1$$

$$\mathbf{A}^{-1} = \frac{1}{-2} \begin{pmatrix} -2 & 0 & 0 \\ -4 & 2 & 0 \\ 9 & -3 & -1 \end{pmatrix}^{\top} = \begin{pmatrix} 1 & 2 & -9/2 \\ 0 & -1 & 3/2 \\ 0 & 0 & 1/2 \end{pmatrix}$$

Scriviamo la matrice inversa della generica matrice di ordine 2:

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

nell'ipotesi  $\det \mathbf{A} = ad - bc \neq 0$ . Si ha:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

e quindi

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

A titolo d'esercizio, si eseguano i prodotti  $\mathbf{A}\mathbf{A}^{-1}$  e  $\mathbf{A}^{-1}\mathbf{A}$ , controllando che si trova

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Sia

$$\mathbf{A}_{\alpha} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$$

la matrice che rappresenta la rotazione di angolo  $\alpha$  nel piano. Scriviamo la sua matrice inversa. In base al calcolo fatto nell'esempio precedente, si ha:

$$\mathbf{A}_{\alpha}^{-1} = \frac{1}{\cos^2 \alpha + \sin^2 \alpha} \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}$$

che possiamo riscrivere nella forma:

$$\mathbf{A}_{\alpha}^{-1} = \begin{pmatrix} \cos(-\alpha) & -\sin(-\alpha) \\ \sin(-\alpha) & \cos(-\alpha) \end{pmatrix}$$

In altre parole,  $\mathbf{A}_{\alpha}^{-1}$  rappresenta la rotazione di angolo  $-\alpha$  nel piano ossia, proprio la *trasformazione inversa* di quella di partenza.

**DIMOSTRAZIONE DEL TEOREMA 8.9** Proviamo prima che se esiste la matrice inversa di  $\mathbf{A}$ , allora necessariamente  $\det \mathbf{A} \neq 0$ . Infatti, poiché  $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}_n$ ,  $\det(\mathbf{A} \cdot \mathbf{A}^{-1}) = \det \mathbf{I}_n = 1$ . Per il Teorema di Binet si ha quindi:

$$(4.15) \quad 1 = \det \mathbf{A} \cdot \det(\mathbf{A}^{-1})$$

e questo implica  $\det \mathbf{A} \neq 0$ .

Viceversa, supponiamo che sia  $\det \mathbf{A} \neq 0$  e proviamo che la matrice definita dalla (4.14) è effettivamente la matrice inversa. Calcoliamo, per esempio, il prodotto  $\mathbf{A} \cdot \mathbf{A}^{-1}$  e verifichiamo che si ottiene la matrice identità. (Analogamente si calcola  $\mathbf{A}^{-1} \cdot \mathbf{A}$ .)

L'elemento di posto  $ki$  nella matrice  $\mathbf{A} \cdot \mathbf{A}^{-1}$  è, in base alla (4.14):

$$b_{ki} = \sum_{j=1}^n a_{kj} \cdot \frac{1}{\det \mathbf{A}} A_{ij} = \frac{1}{\det \mathbf{A}} \cdot \sum_{j=1}^n a_{kj} A_{ij}$$

Ora: se  $k = i$ , la sommatoria scritta vale  $\det \mathbf{A}$ , per la definizione di determinante (4.9), e quindi  $b_{ki} = 1$ ; se invece  $k \neq i$ , mostriamo che tale somma vale 0. Infatti: sia  $\mathbf{A}'$  la matrice ottenuta da  $\mathbf{A}$  rimpiazzando la riga  $i$  con una copia della riga  $k$ . Alterare la riga  $i$  non altera i numeri  $A_{i1}, \dots, A_{in}$ . Si vede allora che l'espressione

$$a_{k1}A_{i1} + a_{k2}A_{i2} + \dots + a_{kn}A_{in}$$

non è altro che  $\det \mathbf{A}'$ , per definizione di determinante. Ma  $\mathbf{A}'$  ha due righe uguali, quindi il suo determinante è nullo (Teorema 8.6 c). Abbiamo quindi provato che l'elemento di posto  $ki$  nella matrice  $\mathbf{A} \cdot \mathbf{A}^{-1}$  vale: 1 se  $k = i$ , 0 se  $k \neq i$ . Perciò  $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}_n$ . Infine, dalla (4.15) si ricava

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det \mathbf{A}}$$

e questo completa la dimostrazione. In base al teorema precedente, una matrice non singolare, cioè con determinante diverso da zero, viene anche detta *invertibile*. Vale anche la seguente:

**PROPOSIZIONE 8.4 (PRODOTTO DI MATRICI INVERTIBILI)** Se  $\mathbf{A}, \mathbf{B}$  sono due matrici quadrate non singolari, anche il prodotto  $\mathbf{AB}$  è non singolare, e vale:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \cdot \mathbf{A}^{-1}$$

**DIMOStrAZIONE.** Per il Teorema di Binet, se  $\det \mathbf{A} \neq 0$  e  $\det \mathbf{B} \neq 0$  si ha:

$$\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B} \neq 0$$

Quindi  $\mathbf{AB}$  è non singolare. Notiamo che:

$$(\mathbf{B}^{-1} \cdot \mathbf{A}^{-1}) \cdot (\mathbf{AB}) = \mathbf{B}^{-1} (\mathbf{A}^{-1} \mathbf{A}) \mathbf{B} = \mathbf{B}^{-1} \mathbf{I} \mathbf{B} = \mathbf{B}^{-1} \mathbf{B} = \mathbf{I}$$

Analogamente si verifica che  $(\mathbf{AB}) \cdot (\mathbf{B}^{-1} \mathbf{A}^{-1}) = \mathbf{I}$ , dunque effettivamente  $\mathbf{B}^{-1} \cdot \mathbf{A}^{-1}$  è la matrice inversa di  $\mathbf{AB}$ .



Siano:

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 3 & 1 \\ -1 & 0 \end{pmatrix}$$

Calcolare:  $\mathbf{A} + \mathbf{B}$ ,  $\mathbf{AB}$ ,  $\mathbf{BA}$ ,  $\mathbf{B}^T$ ,  $\mathbf{A}^{-1}$  (se esiste!).

Siano

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$$

Calcolare  $\mathbf{A}^n$  (cioè  $\mathbf{A} \cdot \mathbf{A} \cdot \dots \cdot \mathbf{A}$   $n$  volte) per ogni intero  $n$ . Lo stesso per  $\mathbf{B}^n$ .

Sia  $V = M(n, n)$  lo spazio vettoriale delle matrici quadrate di ordine  $n$ . Stabilire quali dei seguenti sottoinsiemi di  $V$  costituiscono un sottospazio vettoriale di  $V$ :

- a) L'insieme delle matrici di  $V$  simmetriche.
- b) L'insieme delle matrici di  $V$  triangolari alte.
- c) L'insieme delle matrici di  $V$  triangolari.
- d) L'insieme delle matrici di  $V$  diagonali.
- e) L'insieme delle matrici di  $V$  aventi tutti gli elementi sulla diagonale principale uguali a 0.
- f) L'insieme delle matrici di  $V$  aventi tutti gli elementi sulla diagonale principale uguali a 1.
- g) L'insieme delle matrici di  $V$  aventi determinante diverso da zero.

Nei casi in cui si è risposto affermativamente alla domanda, determinare la dimensione del sottospazio. (Se non si riesce a rispondere nel caso di  $n$  generico, supporre  $n = 3$  e ragionare in questo caso).

Date le matrici

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & -2 \\ 1 & 3 & 1 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 0 & -2 \end{pmatrix}$$

calcolare, dopo aver verificato che l'operazione è ben definita, (utilizzando eventualmente anche un software opportuno, per esempio, Matlab)

$$(\mathbf{AB})^2 \quad (\mathbf{AB})^{-1}$$

Dimostrare il teorema di Binet per le matrici  $2 \times 2$ , ossia: dette

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

verificare che  $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ , calcolando esplicitamente ambo i membri dell'ugualianza.

Scrivere la matrice inversa di

$$\mathbf{M} = \begin{pmatrix} 1 & -1 & 3 \\ 2 & 4 & 0 \\ 1 & -2 & 5 \end{pmatrix}$$

Sia

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}$$

Calcolare  $\det(\mathbf{A}^3)$ .

 Calcolare il determinante delle seguenti matrici. Quando questo semplifica i calcoli, utilizzare qualcuna delle proprietà espresse dal Teorema 4.1 invece di applicare direttamente la definizione.

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 3 \\ 2 & 1 & 4 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 3 & 3 & 2 \end{pmatrix}$$

 Utilizzando il significato geometrico del determinante di una matrice:

a) Calcolare  $\mathbf{u} \times \mathbf{v}$  con

$$\mathbf{u} = 2\mathbf{i} - 3\mathbf{j} + \mathbf{k} \quad \mathbf{v} = \mathbf{i} + 2\mathbf{j} - \mathbf{k}$$

b) Calcolare l'area del parallelogramma generato dai vettori  $\mathbf{u}, \mathbf{v}$ .

c) Calcolare il volume del parallelepipedo generato dai vettori  $\mathbf{u}, \mathbf{v}, \mathbf{w}$ , con

$$\mathbf{w} = \mathbf{i} - \mathbf{j}$$

e  $\mathbf{u}, \mathbf{v}$  come nel punto a.

 Scrivere la matrice inversa di ciascuna delle seguenti matrici:

$$\begin{pmatrix} 1 & 3 \\ 2 & -2 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

 Determinare il rango delle seguenti matrici:

$$\begin{pmatrix} 1 & 3 & 2 \\ 2 & 6 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & -2 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 \end{pmatrix}$$

 Siano  $\mathbf{u} = (1, 1, 0)$ ,  $\mathbf{v} = (0, 1, 0)$ ,  $\mathbf{w} = (0, 0, -1)$ . Calcolare il volume del parallelepipedo generato da  $\mathbf{u}, \mathbf{v}, \mathbf{w}$ .

## 5 SISTEMI LINEARI

### 5.1 Generalità. Metodo di Cramer

Applichiamo i concetti introdotti nel paragrafo 4 allo studio dei sistemi lineari.

Per sistema lineare s'intende un sistema di equazioni algebriche di primo grado. Il numero delle equazioni e quello delle incognite possono variare. Per esempio:

$$\begin{cases} 2x + y - 3z + 4w = -1 \\ -x - y + 7z - 6w = 0 \end{cases} \quad \text{2 equazioni, 4 incognite}$$

$$(5.1) \quad \begin{cases} 7x - 2y + z = 8 \\ 0,1x - 3y + 4z = \sqrt{2} \\ x + y + z = 0 \end{cases} \quad \text{3 equazioni, 3 incognite}$$

$$\begin{cases} x - y = 1 \\ 7x + 2y = 0 \\ 8x + y = 1 \end{cases} \quad \text{3 equazioni, 2 incognite}$$

Generalmente si scrivono al primo membro dell'equazione le incognite, moltiplicate per i loro coefficienti; a secondo membro i *termini noti*; se questi sono tutti nulli, il sistema si dice *omogeneo*.

In questo paragrafo ci occuperemo di sistemi nei quali il numero delle equazioni uguaglia quello delle incognite. Per questi sistemi i coefficienti delle incognite si possono ordinare in una matrice quadrata.

Per il sistema (5.1) la matrice dei coefficienti è la seguente:

$$\begin{pmatrix} 7 & -2 & 1 \\ 0,1 & -3 & 4 \\ 1 & 1 & 1 \end{pmatrix} \rightarrow \begin{array}{l} \text{coefficients delle incognite nella 1ª eq.} \\ \text{coefficients delle incognite nella 2ª eq.} \\ \text{coefficients delle incognite nella 3ª eq.} \end{array}$$

$\downarrow \quad \downarrow \quad \downarrow$

coefficients di  $x \quad y \quad z$

Dato un sistema di  $n$  equazioni in  $n$  incognite (che chiameremo anche "sistema  $n \times n$ "), il nostro scopo è quello di determinare le eventuali soluzioni.

Per soluzione si intende una  $n$ -upla di numeri che, sostituiti ordinatamente alle incognite, soddisfino simultaneamente tutte le equazioni del sistema.

Per esempio la coppia di numeri  $x = 1$  e  $y = -2$  è soluzione del sistema seguente:

$$(5.2) \quad \begin{cases} 2x - 3y = 8 \\ x + 7y = -13 \end{cases}$$

Una domanda naturale a questo punto è la seguente:

"Un sistema lineare di  $n$  equazioni ed  $n$  incognite ha sempre soluzioni? Ed eventualmente, quante?"

Qualche semplice esempio indica che vi sono tre possibilità.



 Equazione scalare:

$$(5.3) \quad ax = b$$

Se  $a \neq 0$  la (5.3) ammette un'unica soluzione data da  $x = \frac{b}{a}$ .

Se  $a = 0$  e  $b \neq 0$ , la (5.3) non ammette alcuna soluzione.

Se  $a = 0$  e  $b = 0$ , qualunque  $x$  è soluzione della (5.3).

 Consideriamo il sistema di 2 equazioni in 2 incognite:

$$\begin{cases} 2x + 3y = 1 \\ x - 2y = 2 \end{cases}$$

Dalla seconda equazione ricaviamo  $x$  in funzione di  $y$  e la sostituiamo nella prima. Otteniamo:

$$\begin{cases} x = 2y + 2 \\ 2(2y + 2) + 3y = 1 \end{cases}$$

Risolviamo la seconda rispetto ad  $y$  e sostituiamo il valore nella prima:

$$\begin{cases} y = -\frac{3}{7} \\ x = \frac{8}{7} \end{cases}$$

Abbiamo trovato un'unica soluzione: la coppia  $(\frac{8}{7}, -\frac{3}{7})$ . Diremo pertanto che il sistema è *determinato*.



$$(5.4) \quad \begin{cases} x + 2y = 3 \\ x + 2y = 5 \end{cases}$$

Evidentemente il sistema (5.4) è *impossibile*; non esiste cioè alcuna coppia di numeri che soddisfi simultaneamente le 2 equazioni in quanto la somma di  $x$  e  $2y$  non può essere simultaneamente 3 e 5.



$$(5.5) \quad \begin{cases} 3x + 2y = 1 \\ 6x + 4y = 2 \end{cases}$$

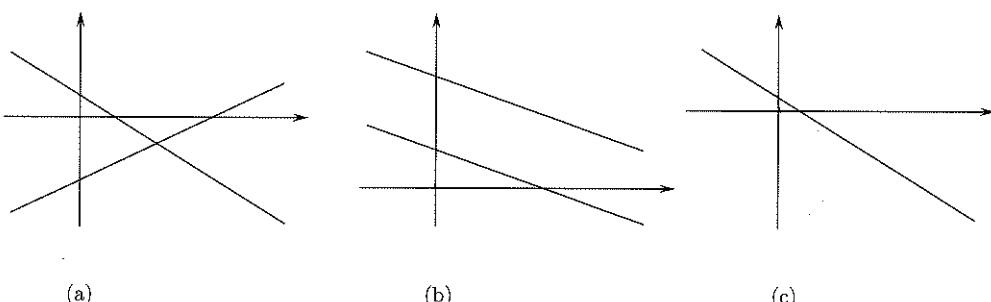
Questa volta si vede che la seconda equazione è stata ottenuta moltiplicando per 2 i membri della prima. Ne segue che una delle due equazioni è inutile, ovvero il sistema è equivalente ad una sola di esse, per esempio la prima. Poiché dalla prima ricaviamo  $y = \frac{1}{2} - \frac{3}{2}x$ , ne segue che il sistema ha le infinite soluzioni

$$\left( x, \frac{1}{2} - \frac{3}{2}x \right)$$

(per  $x$  reale qualunque). Diremo in questo caso che il sistema è *indeterminato* (anche se la locuzione è alquanto impropria, potendosi sempre dare una formula per le soluzioni).

Conclusione: *esistono sistemi determinati, indeterminati, impossibili*.

Gli esempi 5.2, 5.3, 5.4 hanno una semplice interpretazione geometrica: il sistema lineare di due equazioni in due incognite rappresenta la ricerca dell'intersezione di due rette nel piano:



Nell'esempio 5.2 (fig. a) le due equazioni rappresentano due rette incidenti in un punto: il sistema ha una e una sola soluzione; nell'esempio 5.3 (fig. b) le due equazioni rappresentano due rette parallele (prive di intersezioni): il sistema non ha soluzioni; nell'esempio 5.4 (fig. c), le due equazioni rappresentano la stessa retta: il sistema ha *infinte soluzioni* (tutti i punti della retta); notiamo che dire che il sistema ha infinite soluzioni è cosa ben diversa dal dire che *qualsiasi* coppia  $(x, y)$  è soluzione, come è evidente dal punto di vista geometrico: tra tutti i punti del piano, solo quelli che stanno su una certa retta, in questo caso, soddisfano l'equazione.

### Tecniche elementari di soluzione

Lo studente probabilmente conosce già le tecniche elementari con cui è possibile risolvere un sistema. Ricordiamo che due sistemi si dicono *equivalenti* quando hanno le *stesse soluzioni*. Le operazioni che permettono di passare da un sistema ad un altro equivalente al primo sono le seguenti:

1. Sostituire ad una equazione un'altra ad essa equivalente (ossia ottenuta dalla prima sommando membro a membro la stessa quantità, o moltiplicando ambo i membri per una stessa quantità non nulla);
2. Sommare o sottrarre membro a membro due equazioni del sistema e sostituire una delle due con quella così ottenuta.

L'applicazione congiunta (e ripetuta) delle due operazioni precedenti permette altre tipiche operazioni, come quella di *ricavare una variabile in funzione delle altre*, da un'equazione, e *sostituire questa in tutte le altre equazioni*, che così vengono a formare un sottosistema avente un'equazione in meno e un'incognita in meno, sul quale si può procedere ora allo stesso modo. Questo *metodo di sostituzione* è una tecnica ben nota che lo studente avrà certamente utilizzato molte volte nella risoluzione di sistemi di piccole dimensioni (tipicamente, 2 o 3 equazioni in altrettante incognite). Un esempio di risoluzione di sistema  $2 \times 2$  è stato dato nell'esempio 5.2; si consideri anche il prossimo:



Si vuole risolvere il sistema  $3 \times 3$ :

$$\begin{cases} x + y - z = 0 \\ y + z = 2 \\ 3x + z = -1 \end{cases}$$

Dalla seconda ricaviamo la  $z$  e sostituiamola nelle altre due:

$$(5.6) \quad \begin{cases} z = 2 - y \\ x + y - (2 - y) = 0 \\ 3x + (2 - y) = -1 \end{cases}$$

Le ultime due equazioni sono ora un sistema nelle due incognite  $(x, y)$ :

$$\begin{cases} x + 2y = 2 \\ 3x - y = -3 \end{cases}$$

Sommando alla prima equazione il doppio della seconda si ha:

$$\begin{cases} 7x = -4 \\ x + 2y = 2 \end{cases}$$

Dalla prima si ricava  $x = -\frac{4}{7}$ , che sostituita nella seconda dà  $y = \frac{9}{7}$ ; ricordando infine la prima delle (5.6) si ottiene  $z = \frac{5}{7}$ . Il sistema è determinato, la soluzione è il vettore  $(x, y, z) = (-\frac{4}{7}, \frac{9}{7}, \frac{5}{7})$ .

Nelle applicazioni dell'algebra lineare ad altre discipline, ed in particolare all'*analisi numerica*, si ha a che fare con sistemi di molte equazioni ed incognite, per i quali non è pensabile la soluzione manuale. Si capisce così l'importanza di:

1. Avere a disposizione una teoria che chiarisca il quadro concettuale, permettendo di rispondere a domande come: il sistema ha soluzione? La soluzione è unica? Come dipende dai termini noti? In particolare, interessa poter stabilire la solubilità o meno di un sistema *a priori*, senza effettuare tutti i calcoli necessari per determinare l'effettiva soluzione.
2. Conoscere algoritmi efficienti per il calcolo effettivo (ed automatico) delle soluzioni.

Lasciando all'*analisi numerica* il secondo problema, noi qui ci concentreremo sul primo, introducendo le idee fondamentali della teoria dei sistemi lineari.

#### Sistemi di $n$ equazioni in $n$ incognite, non omogenei

Consideriamo un sistema generale di  $n$  equazioni in  $n$  incognite.

$$(5.7) \quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n \end{cases}$$

La matrice  $\mathbf{A}$  dei coefficienti è

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

Introducendo i vettori colonna:

$$(5.8) \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ (delle incognite)} \text{ e } \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \text{ (dei termini noti),}$$

il sistema (5.7) si può porre nella forma matriciale seguente:

$$(5.9) \quad \mathbf{Ax} = \mathbf{b}$$

Per risolvere il sistema ricorriamo all'analogia con l'equazione scalare  $ax = b$  dell'esempio 5.1. Abbiamo visto che, se  $a \neq 0$ , l'unica soluzione è  $x = \frac{b}{a}$ .

*L'analogo della condizione  $a \neq 0$  è ora  $\det \mathbf{A} \neq 0$ .* Infatti, per il teorema sulla matrice inversa, se  $\det \mathbf{A} \neq 0$  esiste  $\mathbf{A}^{-1}$ ; moltiplicando a sinistra per  $\mathbf{A}^{-1}$  entrambi i membri di (5.9) si trova

$$\mathbf{A}^{-1}(\mathbf{Ax}) = \mathbf{A}^{-1} \cdot \mathbf{b}$$

Utilizzando la proprietà associativa per il prodotto tra matrici, si ricava:

$$(\mathbf{A} \cdot \mathbf{A}^{-1})\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{b}$$

ed infine, poiché  $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}_n$  e  $\mathbf{I}_n \cdot \mathbf{x} = \mathbf{x}$ , si ottiene la formula seguente, per il vettore soluzione del sistema:

$$(5.10) \quad \boxed{\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{b}}$$

Si osserva che, sviluppando i calcoli indicati dalla (5.10), le componenti  $x_i$  ( $i = 1, 2, \dots, n$ ) di  $\mathbf{x}$  si ottengono dalla formula seguente:

$$(5.11) \quad x_i = \frac{|\mathbf{B}_i|}{|\mathbf{A}|}$$

dove a numeratore compare il determinante della matrice  $\mathbf{B}_i$  che si ottiene dalla matrice  $\mathbf{A}$  sostituendo, alla colonna  $i$ -esima, la colonna dei termini noti  $\mathbf{b}$ :

$$\mathbf{B}_i = \begin{pmatrix} a_{11} & a_{12} & \dots & b_1 & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & b_2 & \dots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & b_n & \dots & a_{nn} \end{pmatrix}$$

$\uparrow$   
 $i$ -esima colonna

Infatti: ricordando l'espressione (4.14) della matrice inversa, si ha:

$$x_i = (\mathbf{A}^{-1}\mathbf{b})_i = \frac{1}{\det \mathbf{A}} \sum_{j=1}^n A_{ji} b_j = \frac{1}{\det \mathbf{A}} \cdot \det \mathbf{B}_i$$

come si vede calcolando il determinante di  $\mathbf{B}_i$  rispetto alla  $i$ -esima colonna.

Abbiamo così dimostrato un primo importante risultato:

**TEOREMA 8.10 (TEOREMA DI CRAMER)** Consideriamo il sistema di  $n$  equazioni in  $n$  incognite

$$\mathbf{Ax} = \mathbf{b}$$

(con  $\mathbf{A}$  matrice  $(n,n)$ ,  $\mathbf{b} \in \mathbb{R}^n$  assegnato e  $\mathbf{x} \in \mathbb{R}^n$  incognito). Se  $\det \mathbf{A} \neq 0$ , il sistema è determinato, ossia ha una e una sola soluzione, assegnata dalla (5.11) (o equivalentemente dalla (5.10)).



Consideriamo il sistema:

$$\begin{cases} 2x + y + 3z = 12 \\ 4y - z = -7 \\ 5x + 8z = 34 \end{cases} \quad \mathbf{A} = \begin{pmatrix} 2 & 1 & 3 \\ 0 & 4 & -1 \\ 5 & 0 & 8 \end{pmatrix}$$

Poiché  $|\mathbf{A}| = -1$ , il sistema ammette soluzione unica; dalla (5.11) abbiamo:

$$x = -\begin{vmatrix} 12 & 1 & 3 \\ -7 & 4 & -1 \\ 34 & 0 & 8 \end{vmatrix} = 2$$

$$y = -\begin{vmatrix} 2 & 12 & 3 \\ 0 & -7 & -1 \\ 5 & 34 & 8 \end{vmatrix} = -1$$

$$z = -\begin{vmatrix} 2 & 1 & 12 \\ 0 & 4 & -7 \\ 5 & 0 & 34 \end{vmatrix} = 3$$

### Sistemi omogenei di $n$ equazioni in $n$ incognite

Interpretiamo ora il Teorema di Cramer per i sistemi omogenei:

$$\mathbf{Ax} = \mathbf{0}.$$

In questo caso, se  $\det \mathbf{A} \neq 0$ , l'unica soluzione è  $\mathbf{x} = \mathbf{0}$ . (Infatti,  $\mathbf{A}^{-1} \cdot \mathbf{0} = \mathbf{0}$ ).

Viceversa, supponiamo che sia  $\det \mathbf{A} = 0$ . Ciò significa, per il Teorema 8.8, che le colonne di  $\mathbf{A}, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  sono vettori linearmente dipendenti, ossia esistono numeri  $x_1, x_2, \dots, x_n$ , non tutti nulli, tali che

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_n = \mathbf{0}.$$

Dunque, detto  $\mathbf{x}$  il vettore  $(x_1, x_2, \dots, x_n)$ , si ha

$$\mathbf{Ax} = (\mathbf{a}_1 | \mathbf{a}_2 | \dots | \mathbf{a}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_n = \mathbf{0}$$

ossia: il sistema  $\mathbf{Ax} = \mathbf{0}$  in questo caso ammette almeno una soluzione  $\mathbf{x}$  non banale (ossia una soluzione  $\mathbf{x} \neq \mathbf{0}$ ). Si noti, d'altro canto, che se  $\mathbf{Ax} = \mathbf{0}$ , si ha anche, per ogni  $\lambda \in \mathbb{R}$ ,

$$\mathbf{A}(\lambda \mathbf{x}) = \lambda \mathbf{Ax} = \mathbf{0}$$

perciò in questo caso le soluzioni del sistema omogeno sono infinite. Riassumendo, abbiamo dimostrato il:

**TEOREMA 8.11** *Il sistema omogeno di  $n$  equazioni in  $n$  incognite*

$$\mathbf{Ax} = \mathbf{0}$$

ha solo la soluzione banale ( $\mathbf{x} = \mathbf{0}$ ) se e solo se  $\det \mathbf{A} \neq 0$ ; ha almeno una soluzione non banale (e in questo caso ne ha infinite) se e solo se  $\det \mathbf{A} = 0$ .

Cosa succede invece al sistema *non omogeneo* quando  $\det \mathbf{A} = 0$ ? A questo risponderemo completamente nel paragrafo 5.3 col Teorema di Rouché-Capelli. Possiamo fin d'ora fare però la seguente osservazione.

Supponiamo che  $\mathbf{x}_0$  sia una soluzione del sistema omogeneo, ossia  $\mathbf{Ax}_0 = \mathbf{0}$ , e  $\mathbf{x}_1$  sia una soluzione del sistema non omogeneo,  $\mathbf{Ax}_1 = \mathbf{b}$ . Allora anche  $\mathbf{x}_0 + \mathbf{x}_1$  è soluzione del sistema non omogeneo, infatti:

$$\mathbf{A}(\mathbf{x}_1 + \mathbf{x}_0) = \mathbf{Ax}_1 + \mathbf{Ax}_0 = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

Se  $\det \mathbf{A} = 0$  sappiamo che il sistema omogeneo ha sempre infinite soluzioni. Ne concludiamo che:

Se  $\det \mathbf{A} = 0$  e il sistema non omogeneo  $\mathbf{Ax} = \mathbf{b}$  ha una soluzione, allora ne ha infinite. In altre parole, quando  $\det \mathbf{A} = 0$ , per il sistema non omogeneo viene necessariamente a cadere o l'esistenza o l'unicità della soluzione: il sistema è impossibile o indeterminato. Vedremo in seguito come si può prevedere se accade l'una o l'altra cosa.

## 5.2 Immagine e nucleo di una trasformazione lineare da $\mathbb{R}^n$ a $\mathbb{R}^m$

Per affrontare lo studio dei sistemi lineari di  $n$  equazioni in  $m$  incognite è utile prima studiare qualche proprietà delle trasformazioni lineari.

**DEFINIZIONE 8.11 (IMMAGINE)** Se  $\mathcal{L}$  è lineare da  $\mathbb{R}^n$  a  $\mathbb{R}^m$ , si chiama *immagine*<sup>2</sup> di  $\mathcal{L}$  l'insieme dei vettori di  $\mathbb{R}^m$  che sono i trasformati di qualche vettore di  $\mathbb{R}^n$ . Tale insieme si indica con

$$\text{Im}(\mathcal{L})$$

Se  $\mathbf{y} \in \text{Im}(\mathcal{L})$  deve essere  $\mathbf{y} = \mathcal{L}(\mathbf{x})$  per qualche  $\mathbf{x} \in \mathbb{R}^n$ .

Le proprietà principali dell'immagine sono espresse dal seguente:

**TEOREMA 8.12**

- 1) L'insieme  $\text{Im}(\mathcal{L})$  è un sottospazio vettoriale di  $\mathbb{R}^m$ .
- 2) Fissate le basi in  $\mathbb{R}^n$  e  $\mathbb{R}^m$ , sia  $\mathbf{A}$  una matrice di tipo  $(m, n)$  che rappresenta  $\mathcal{L}$ . Allora

$$\dim \text{Im}(\mathcal{L}) = \text{rango di } \mathbf{A}$$

**DIMOSTRAZIONE.** La proprietà 1) segue subito dalla linearità di  $\mathcal{L}$ : se  $\mathbf{y}_1 = \mathcal{L}(\mathbf{x}_1)$ ,  $\mathbf{y}_2 = \mathcal{L}(\mathbf{x}_2)$  e  $\alpha, \beta \in \mathbb{R}$ , si ha

$$\alpha \mathbf{y}_1 + \beta \mathbf{y}_2 = \alpha \mathcal{L}(\mathbf{x}_1) + \beta \mathcal{L}(\mathbf{x}_2) = \mathcal{L}(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2)$$

e perciò anche  $\alpha \mathbf{y}_1 + \beta \mathbf{y}_2 \in \text{Im}(\mathcal{L})$  essendo il trasformato del vettore  $\alpha \mathbf{x}_1 + \beta \mathbf{x}_2$ . Il criterio di riconoscimento dei sottospazi vettoriali implica ora che  $\text{Im}(\mathcal{L})$  è un sottospazio di  $\mathbb{R}^m$ .

Per la 2), osserviamo che, dalla dimostrazione del teorema di rappresentazione, si vede che  $\text{Im}(\mathcal{L})$  è generato dai vettori  $\mathcal{L}(\mathbf{u}_1), \dots, \mathcal{L}(\mathbf{u}_n)$ . Le colonne della matrice  $\mathbf{A}$  sono le componenti scalari degli  $\mathcal{L}(\mathbf{u}_i)$ , rispetto alla base  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  in  $\mathbb{R}^m$ . Se  $r$  tra queste colonne sono indipendenti i corrispondenti vettori  $\mathcal{L}(\mathbf{u}_j)$  sono indipendenti e viceversa. Da ciò segue subito che  $r = \dim \text{Im}(\mathcal{L})$ .

<sup>2</sup>È un caso particolare della definizione di immagine di una funzione qualsiasi, data nel capitolo 2, paragrafo 1.

Da questo teorema segue subito che le matrici che rappresentano (rispetto a basi diverse) la stessa trasformazione lineare hanno lo stesso rango.

**DEFINIZIONE 8.12 (NUCLEO)** Il nucleo di un'applicazione lineare  $\mathcal{L}$  si indica con il simbolo  $\text{Ker}(\mathcal{L})$ <sup>3</sup> ed è l'insieme dei vettori di  $\mathbb{R}^n$  che hanno come immagine il vettore nullo di  $\mathbb{R}^m$ . In simboli

$$\text{Ker}(\mathcal{L}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathcal{L}(\mathbf{x}) = \mathbf{0} \in \mathbb{R}^m\}$$

Anche  $\text{Ker}(\mathcal{L})$  è uno spazio vettoriale (lasciamo la facile verifica al lettore). Vale l'importante formula che lega le dimensioni di immagine e nucleo per una generica trasformazione lineare  $\mathcal{L}$ :

**TEOREMA 8.13** Se  $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  è una trasformazione lineare, allora

$$(5.12) \quad \dim \text{Ker}(\mathcal{L}) + \dim \text{Im}(\mathcal{L}) = n$$

**DIMOSTRAZIONE.** Poniamo  $k = \dim \text{Ker}(\mathcal{L})$  e scegliamo una base  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k$  per il nucleo. Aggiungiamo a questa base altri  $n - k$  vettori linearmente indipendenti  $\mathbf{u}_{k+1}, \dots, \mathbf{u}_n$  in modo che

$$\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n$$

formino una base per  $\mathbb{R}^n$ . Se  $\mathbf{x}$  è un generico vettore di  $\mathbb{R}^n$  si può perciò scrivere

$$\mathbf{x} = \alpha_1 \mathbf{w}_1 + \dots + \alpha_k \mathbf{w}_k + \alpha_{k+1} \mathbf{u}_{k+1} + \dots + \alpha_n \mathbf{u}_n$$

Essendo  $\mathcal{L}(\mathbf{w}_j) = \mathbf{0}$ ,  $\forall j = 1, 2, \dots, k$ , si ha:

$$\mathcal{L}(\mathbf{x}) = \alpha_{k+1} \mathcal{L}(\mathbf{u}_{k+1}) + \dots + \alpha_n \mathcal{L}(\mathbf{u}_n)$$

per cui l'insieme dei vettori  $\mathbf{v}_{k+1} = \mathcal{L}(\mathbf{u}_{k+1}), \dots, \mathbf{v}_n = \mathcal{L}(\mathbf{u}_n)$  risulta un sistema di generatori di  $\text{Im}(\mathcal{L})$  e quindi  $\dim \text{Im}(\mathcal{L}) \leq n - k$ . Facciamo vedere che, in realtà, i vettori  $\mathbf{v}_{k+1}, \dots, \mathbf{v}_n$  sono indipendenti, per cui  $\dim \text{Im}(\mathcal{L}) = n - k$  e la formula (5.12) è dimostrata. Se fossero linearmente dipendenti si avrebbe

$$0 = \beta_{k+1} \mathbf{v}_{k+1} + \dots + \beta_n \mathbf{v}_n = \beta_{k+1} \mathcal{L}(\mathbf{u}_{k+1}) + \dots + \beta_n \mathcal{L}(\mathbf{u}_n) = \mathcal{L}(\beta_{k+1} \mathbf{u}_{k+1} + \dots + \beta_n \mathbf{u}_n)$$

con qualche  $\beta_j$  diverso da zero. Ciò implicherebbe che il vettore non nullo  $\beta_{k+1} \mathbf{u}_{k+1} + \dots + \beta_n \mathbf{u}_n \in \text{Ker}(\mathcal{L})$ ; ma allora questo vettore si potrebbe scrivere come combinazione lineare dei  $\mathbf{w}_i$ , assurdo perché i  $\mathbf{w}_i$  e gli  $\mathbf{u}_i$  sono tra loro indipendenti.

### Esempio

Sia  $\mathcal{L}$  la trasformazione lineare da  $\mathbb{R}^3$  in  $\mathbb{R}^3$  che, rispetto alle basi canoniche, è rappresentata dalla matrice

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ -2 & 3 & 1 \\ -1 & 2 & 1 \end{pmatrix}$$

Il rango di  $\mathbf{A}$  è  $r = 2$ , essendo la terza colonna (o riga) somma delle due precedenti e

$$\begin{vmatrix} 1 & -1 \\ -2 & 3 \end{vmatrix} = 1 \neq 0$$

<sup>3</sup>Dall'inglese *kernel* o dal tedesco *Kern* = nucleo, nocciola.

Dunque

$$\dim \text{Im}(\mathcal{L}) = 2 \quad \dim \text{Ker}(\mathcal{L}) = n - r = 1$$

Per determinare l'immagine di  $\mathcal{L}$ , calcoliamo

$$\mathbf{A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ -2 & 3 & 1 \\ -1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ -2x_1 + 3x_2 + x_3 \\ -x_1 + 2x_2 + x_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Poiché  $y_3 = y_2 + y_1$ ,  $\text{Im}(\mathcal{L})$  coincide col piano di equazione

$$y_1 + y_2 - y_3 = 0$$

Detto altrimenti: il generico elemento di  $\text{Im}(\mathcal{L})$  è del tipo:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_1 + y_2 \end{pmatrix} \quad \text{ossia} \quad y_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + y_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

pertanto una base di  $\text{Im}(\mathcal{L})$  è costituita dai 2 vettori

$$\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Per determinare il nucleo occorre risolvere il sistema omogeneo

$$\begin{cases} x_1 - x_2 = 0 \\ -2x_1 + 3x_2 + x_3 = 0 \end{cases}$$

dal quale abbiamo omesso la terza equazione, in quanto dipendente dalle prime due. Si trova subito

$$x_2 = x_1 \quad x_3 = -x_1$$

per cui il generico vettore del nucleo è della forma  $\lambda \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ ,  $\lambda$  reale.

### Trasformazioni iniettive e suriettive

**DEFINIZIONE 8.13** Se  $\text{Im}(\mathcal{L}) = \mathbb{R}^m$ , si dice che  $\mathcal{L}$  è *suriettiva*.

Ciò significa che l'equazione  $\mathcal{L}(\mathbf{x}) = \mathbf{b}$  ha almeno una soluzione  $\mathbf{x}$ , comunque si assegni il termine noto  $\mathbf{b}$ .

**DEFINIZIONE 8.14** Se  $\text{Ker}(\mathcal{L}) = \{\mathbf{0}\}$ , si dice che  $\mathcal{L}$  è *iniettiva*.

Ciò significa che l'equazione  $\mathcal{L}(\mathbf{x}) = \mathbf{b}$  ha solo la soluzione nulla. Per la linearità di  $\mathcal{L}$ , questo è equivalente ad affermare che l'equazione  $\mathcal{L}(\mathbf{x}) = \mathbf{b}$  ha al più una soluzione  $\mathbf{x}$ , comunque si assegni il termine noto  $\mathbf{b}$ . (Infatti, se  $\mathbf{x}_1, \mathbf{x}_2$  sono due soluzioni di  $\mathcal{L}(\mathbf{x}) = \mathbf{b}$ , per linearità  $\mathcal{L}(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{b} - \mathbf{b} = \mathbf{0}$ , dunque  $\mathbf{x}_1 = \mathbf{x}_2$ ).

**DEFINIZIONE 8.15** Se  $\mathcal{L}$  è iniettiva e suriettiva, si dice che  $\mathcal{L}$  è *bijettiva*; la corrispondenza stabilita da  $\mathcal{L}$  è *biunivoca*.

In tal caso l'equazione  $\mathcal{L}(\mathbf{x}) = \mathbf{b}$  ha una e una sola soluzione  $\mathbf{x}$ , comunque si assegni il termine noto  $\mathbf{b}$ .

Interessante è il caso in cui  $n = m$ . In tal caso la (5.12) implica

$$\dim \text{Ker}(\mathcal{L}) = 0 \Leftrightarrow \dim \text{Im}(\mathcal{L}) = n$$

e perciò, per trasformazioni lineari  $\mathcal{L}$  da  $\mathbb{R}^n$  a  $\mathbb{R}^n$ , *iniettività equivale a suriettività* (e quindi biunivocità).

Detto altrimenti:

**PROPOSIZIONE 8.5** *Se  $\mathcal{L}$  è una trasformazione da  $\mathbb{R}^n$  a  $\mathbb{R}^n$ , per l'equazione  $\mathcal{L}(\mathbf{x}) = \mathbf{b}$  si può affermare che c'è esistenza di soluzione (per ogni  $\mathbf{b}$ ) se e solo se c'è unicità di soluzione (per ogni  $\mathbf{b}$ ).*

Sempre dalla (5.12), infine, si legge che se  $n \neq m$  la trasformazione lineare  $\mathcal{L}$  non può essere biunivoca.



• Sia  $\mathcal{L} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  la trasformazione lineare rappresentata nella base canonica da

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 1 \\ 3 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$$

Determinare  $\text{Ker}(\mathcal{L})$ ,  $\text{Im}(\mathcal{L})$ , le loro dimensioni e una loro base.

• Sia  $\mathcal{L} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  la trasformazione lineare rappresentata nella base canonica da

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \end{pmatrix}$$

Determinare  $\text{Ker}(\mathcal{L})$ ,  $\text{Im}(\mathcal{L})$ , le loro dimensioni e una loro base.

• Si consideri la base di  $\mathbb{R}^2$ :

$$\mathbf{e}_1 = (1, 2); \mathbf{e}_2 = (2, 0)$$

e sia  $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  così definita:

$$\mathcal{L}(\alpha \mathbf{e}_1 + \beta \mathbf{e}_2) = (2\alpha + \beta)\mathbf{e}_1 - \beta \mathbf{e}_2$$

- Scrivere la matrice che rappresenta  $\mathcal{L}$  rispetto alla base  $\mathbf{e}_1, \mathbf{e}_2$ .
- Scrivere la matrice che rappresenta  $\mathcal{L}$  rispetto alla base canonica.
- Determinare  $\text{Ker}(\mathcal{L})$ ,  $\text{Im}(\mathcal{L})$ , le loro dimensioni e una loro base.

• Sia  $\mathcal{L} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  la trasformazione lineare rappresentata nella base canonica dalla matrice  $\mathbf{A}$ :

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & -2 \\ -1 & 3 & 1 \end{pmatrix}$$

Determinare  $\text{Ker}(\mathcal{L})$ ,  $\text{Im}(\mathcal{L})$ , le loro dimensioni e una loro base.

• Si consideri la trasformazione lineare  $L$  rappresentata dalla matrice

$$\mathbf{A} = \begin{pmatrix} \alpha & 0 & 1 & 1 \\ \alpha - 1 & -1 & 0 & 1 \\ 2\alpha & 0 & 2 & 1 \end{pmatrix}$$

Per quali valori del parametro  $\alpha$ ,  $L$  è iniettiva? Per quali è suriettiva? Per quali è biunivoca? Per quali  $\dim \text{Ker}(L) = 1$ ?

• Sia  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  la trasformazione lineare definita dalle relazioni:

$$\begin{aligned} f(\mathbf{e}_1) &= \mathbf{e}_1 - \mathbf{e}_2 \\ f(\mathbf{e}_2) &= \mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_3 \\ f(\mathbf{e}_3) &= \mathbf{e}_2 - \mathbf{e}_3 \end{aligned}$$

dove  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  sono la base canonica in  $\mathbb{R}^3$ . Dire se  $f$  è iniettiva, se è suriettiva, se è biunivoca. Determinare  $\dim \text{Im}(f)$  e il rango della matrice che rappresenta  $f$ .

### 5.3 Sistemi generali. Teorema di Rouché-Capelli

Consideriamo ora il sistema

$$(5.13) \quad \left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{array} \right.$$

costituito da  $m$  equazioni in  $n$  incognite, ( $m, n$  qualsiasi). Sia  $\mathbf{A}$  la matrice  $(m, n)$  dei coefficienti del sistema. Pensando di fissare in  $\mathbb{R}^n$  e  $\mathbb{R}^m$  le rispettive basi canoniche, sia  $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$  la trasformazione lineare rappresentata da  $\mathbf{A}$ . Il sistema lineare può dunque risciversi nella forma

$$L(\mathbf{x}) = \mathbf{b}$$

o anche

$$\mathbf{Ax} = \mathbf{b}$$

con  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{b} \in \mathbb{R}^m$ . Risolvere il sistema significa trovare quali vettori  $\mathbf{x} \in \mathbb{R}^n$  hanno come immagine attraverso  $L$  il vettore  $\mathbf{b} \in \mathbb{R}^m$ . Vale in proposito il seguente risultato:

**TEOREMA 8.14 (TEOREMA DI ROUCHÉ-CAPELLI)** *Sia  $\mathbf{A}$  la matrice dei coefficienti del sistema (5.13) e  $\mathbf{B} = (\mathbf{A}|\mathbf{b})$  la matrice completa, ottenuta orlando la matrice  $\mathbf{A}$  con la colonna dei termini noti. Allora il sistema ha soluzioni se e solo se la matrice  $\mathbf{A}$  e la matrice  $\mathbf{B}$  hanno lo stesso rango.*

**DIMOSTRAZIONE.** Per quanto premesso all'enunciato del teorema, il sistema è risolubile se e solo se  $\mathbf{b} \in \text{Im}(L)$ . Indicando con  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  le colonne di  $\mathbf{A}$  e con  $\mathbf{e}_j$  i vettori della base canonica di  $\mathbb{R}^n$ , si ha

$$\mathbf{A}\mathbf{e}_j = \mathbf{a}_j$$

cioè le colonne della matrice  $\mathbf{A}$  sono i trasformati dei vettori della base canonica in  $\mathbb{R}^n$ . Di conseguenza, ogni elemento di  $\text{Im}(L)$  è combinazione lineare delle colonne di  $\mathbf{A}$ , perché per ogni vettore  $\mathbf{x} = \sum x_j \mathbf{e}_j \in \mathbb{R}^n$  si ha

$$\mathbf{Ax} = \sum x_j \mathbf{Ae}_j = \sum x_j \mathbf{a}_j.$$

Dunque il sistema è solubile se e solo se  $\mathbf{b}$  è combinazione lineare delle colonne di  $\mathbf{A}$ , ossia, ricordando che  $\dim \text{Im}(L) = \text{rango di } \mathbf{A}$ , se e solo se

$$\text{rango di } (\mathbf{A}|\mathbf{b}) = \text{rango di } \mathbf{A}.$$



Osserviamo che, se il sistema è omogeneo ( $\mathbf{b} = \mathbf{0}$ ), la condizione  $\text{rango di } \mathbf{A} = \text{rango di } \mathbf{B}$  è sempre soddisfatta. In questo caso l'insieme delle soluzioni del sistema è lo spazio  $\text{Ker}(L)$ . Per la (5.12),

$$\dim \text{Ker}(L) + \dim \text{Im}(L) = n.$$

Inoltre  $\dim \text{Im}(L) = \text{rango di } \mathbf{A}$ . Perciò se  $\text{rango di } \mathbf{A} = r$

$$\dim \text{Ker}(L) = n - r.$$

Abbiamo così dimostrato il seguente:

**TEOREMA 8.15** *Le soluzioni del sistema omogeneo formano uno spazio vettoriale di dimensione  $n - r$ . Si dice anche che il sistema ha  $\infty^{n-r}$  soluzioni, per indicare che le soluzioni sono infinite e dipendono da  $n - r$  parametri arbitrari.*

Supponiamo ora che il sistema non omogeneo sia risolubile, ovvero supponiamo sia soddisfatta l'ipotesi del Teorema di Rouché-Capelli, e chiediamoci: quante sono le sue soluzioni?

**TEOREMA 8.16** *Se il sistema (5.13) ammette una soluzione  $\mathbf{x}_1$ , le soluzioni del sistema sono tutte e sole le  $n$ -uple del tipo*

$$\mathbf{x}_1 + \mathbf{x}_0$$

*al variare di  $\mathbf{x}_0$  in  $\text{Ker}(L)$  (ossia al variare di  $\mathbf{x}_0$  tra le soluzioni del sistema omogeneo). Pertanto il sistema non omogeneo ha  $\infty^{n-r}$  soluzioni, dove  $r = \text{rango di } \mathbf{A}$ .*

**DIMOSTRAZIONE.** Se  $L(\mathbf{x}_1) = \mathbf{b}$  e  $\mathbf{x}_0 \in \text{Ker}(L)$ ,

$$L(\mathbf{x}_1 + \mathbf{x}_0) = L(\mathbf{x}_1) + L(\mathbf{x}_0) = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

perciò ogni  $n$ -upla del tipo  $\mathbf{x}_1 + \mathbf{x}_0$  è soluzione del sistema.

Viceversa, sia  $\mathbf{x}_2$  una soluzione del sistema, e definiamo  $\mathbf{x}_0 = \mathbf{x}_2 - \mathbf{x}_1$ . Risulta

$$L(\mathbf{x}_0) = L(\mathbf{x}_2 - \mathbf{x}_1) = L(\mathbf{x}_2) - L(\mathbf{x}_1) = \mathbf{b} - \mathbf{b} = \mathbf{0},$$

perciò  $\mathbf{x}_0 \in \text{Ker}(L)$ . Dunque  $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{x}_0$  con  $\mathbf{x}_0 \in \text{Ker}(L)$ , ossia ogni soluzione del sistema si può scrivere a questo modo. Poiché  $\mathbf{x}_0$  dipende da  $n - r$  parametri arbitrari e  $\mathbf{x}_1$  è fissato, anche la generica soluzione del sistema non omogeneo dipende da  $n - r$  parametri arbitrari.

Abbiamo fin qui chiarito le condizioni che permettono di decidere se il sistema è risolubile o no, e in caso affermativo abbiamo “contato” le soluzioni. Illustriamo ora il metodo con cui si possono effettivamente determinare.

Supponendo verificata l'ipotesi

$$\text{rango di } (\mathbf{A}|\mathbf{b}) = \text{rango di } \mathbf{A} = r$$

si può seguire il seguente *schema risolutivo*:

1. Si isola un minore di ordine  $r$  estratto dalla matrice  $\mathbf{A}$ , non singolare.
2. Del sistema si considerano solo le  $r$  equazioni corrispondenti alle  $r$  righe del minore; le altre  $m - r$  equazioni vengono eliminate: infatti, queste sono automaticamente soddisfatte quando lo sono le prime  $r$ .
3. A primo membro si mantengono le  $r$  incognite i cui coefficienti costituiscono le  $r$  colonne del minore; i termini contenenti le altre  $n - r$  incognite si trasportano al secondo membro.
4. Si ottiene un sistema di  $r$  equazioni ed  $r$  incognite, al quale si può applicare il metodo di Cramer (o qualunque altro metodo si conosca).
5. Le soluzioni trovate dipendono dalle  $n - r$  incognite portate a secondo membro, che possono assumere valori arbitrari.

Si noti che l'unico caso in cui il sistema è determinato è quando  $r = n$  (si ricordi che  $n$  è il numero di incognite che compaiono nel sistema, mentre  $r$ , rango della matrice, uguaglia il numero di incognite che si riescono a ricavare). Se  $r < n$ , se il sistema è risolubile è sempre indeterminato, con  $\infty^{n-r}$  soluzioni. Notiamo anche che nella teoria esposta in questo paragrafo rientra in particolare il caso dei sistemi non omogenei di  $n$  equazioni in  $n$  incognite, con determinante nullo, caso che era stato lasciato in sospeso nel paragrafo 5.1.

I risultati di questo paragrafo riguardanti le soluzioni dei sistemi lineari si possono riformulare in termini di relazioni tra particolari sottospazi associati alle righe o colonne della matrice del sistema. Possiamo associare ad una matrice  $\mathbf{A}$  di ordine  $m \times n$  (per fissare le idee, ad elementi reali) i seguenti sottospazi:

- Lo spazio generato dalle colonne di  $\mathbf{A}$ , indicato con  $R(\mathbf{A})$ , sottospazio di  $\mathbb{R}^m$ .
- Il nucleo di  $\mathbf{A}$ ,  $\text{Ker}(\mathbf{A})$ , sottospazio di  $\mathbb{R}^n$ .
- Lo spazio generato dalle righe di  $\mathbf{A}$ , ossia  $R(\mathbf{A}^\top)$ , sottospazio di  $\mathbb{R}^n$ .
- Il nucleo di  $\mathbf{A}^\top$ ,  $\text{Ker}(\mathbf{A}^\top)$ , sottospazio di  $\mathbb{R}^m$ .

Sappiamo che

$$\dim R(\mathbf{A}) = \dim R(\mathbf{A}^\top) = r$$

e che

$$\dim \text{Ker}(\mathbf{A}) = n - r, \quad \dim \text{Ker}(\mathbf{A}^\top) = m - r.$$

Il seguente teorema permette di rileggere in termini di relazioni tra tali sottospazi tutta la teoria svolta.

**TEOREMA 8.17** Si ha:

$$1. R(\mathbf{A}) = [\text{Ker}(\mathbf{A}^T)]^\perp \text{ ovvero}$$

$$(5.14) \quad R(\mathbf{A}) \oplus \text{Ker}(\mathbf{A}^T) = \mathbb{R}^m.$$

$$2. R(\mathbf{A}^T) = [\text{Ker}(\mathbf{A})]^\perp \text{ ovvero}$$

$$(5.15) \quad R(\mathbf{A}^T) \oplus \text{Ker}(\mathbf{A}) = \mathbb{R}^n.$$

**DIMOSTRAZIONE.** Siano  $\mathbf{b} \in R(\mathbf{A})$  e  $\mathbf{y} \in \text{Ker}(\mathbf{A}^T)$ . Allora  $\mathbf{b} = \mathbf{Ax}$ , per qualche  $\mathbf{x} \in \mathbb{R}^n$ . Si ha:

$$\mathbf{y}^T \mathbf{b} = \mathbf{y}^T (\mathbf{Ax}) = (\mathbf{A}^T \mathbf{y})^T \mathbf{x} = \mathbf{0}^T \mathbf{x} = 0.$$

Pertanto  $R(\mathbf{A})$  e  $\text{Ker}(\mathbf{A}^T)$  sono ortogonali. Per mostrare che  $R(\mathbf{A}) \oplus \text{Ker}(\mathbf{A}^T) = \mathbb{R}^m$  basta far vedere che qualunque vettore ortogonale a  $R(\mathbf{A})$  appartiene a  $\text{ker}(\mathbf{A}^T)$ .

Sia dunque  $\mathbf{y} \in R(\mathbf{A})^T$ . Allora, qualunque sia  $\mathbf{x} \in \mathbb{R}^n$ ,

$$(\mathbf{A}^T \mathbf{y})^T \mathbf{x} = \mathbf{y}^T (\mathbf{Ax}) = \mathbf{0}^T \mathbf{x} = 0$$

che implica  $\mathbf{A}^T \mathbf{y} = \mathbf{0}$  e cioè  $\mathbf{y} \in \text{Ker}(\mathbf{A}^T)$ .

La (5.15) si dimostra in modo perfettamente analogo, scambiando  $\mathbf{A}$  con  $\mathbf{A}^T$ .  $\diamond$

Interpretiamo il teorema in termini di soluzione del sistema  $\mathbf{Ax} = \mathbf{b}$ . La (5.14) equivale ad affermare che il sistema è risolubile se e solo se  $\mathbf{b}$  è ortogonale al nucleo di  $\mathbf{A}^T$  (versione del teorema di Rouché-Capelli). La (5.15) equivale ad affermare che ogni soluzione si ottiene aggiungendo ad una soluzione particolare un elemento del nucleo di  $\mathbf{A}$ .

Una conseguenza di queste considerazioni è:

**COROLLARIO 8.18** La matrice  $\mathbf{A}$  (come trasformazione lineare) realizza una corrispondenza biunivoca tra  $R(\mathbf{A}^T)$  e  $R(\mathbf{A})$  ossia tra lo spazio delle sue righe e quello delle sue colonne.

In altri termini, il corollario afferma che, per ogni elemento  $\mathbf{b}$  nello spazio delle colonne di  $\mathbf{A}$ , esiste un'unica soluzione  $\mathbf{x}$  appartenente allo spazio delle righe di  $\mathbf{A}$  del sistema  $\mathbf{Ax} = \mathbf{b}$ . È poi facile convincersi, dalla (5.15), che tale soluzione è quella di minimo modulo tra tutte le possibili soluzioni del sistema.



Consideriamo il sistema di quattro equazioni in tre incognite:

$$(5.16) \quad \begin{cases} x + 2y + 3z = 1 \\ x - 2y - z = 0 \\ 2x + 8y + 10z = 3 \\ 2x + + 2z = 1 \end{cases}$$

Si ha

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & -2 & -1 \\ 2 & 8 & 10 \\ 2 & 0 & 2 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & -2 & -1 & 0 \\ 2 & 8 & 10 & 3 \\ 2 & 0 & 2 & 1 \end{pmatrix}$$

In entrambe le matrici la terza riga si ottiene moltiplicando la prima per 3 e sottraendo la seconda mentre la quarta riga è la somma delle prime due. Quindi rango di  $\mathbf{A} \leq 2$ , rango  $\mathbf{B} \leq 2$ .

Essendo

$$\det \begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix} = -4 \neq 0$$

il rango di  $\mathbf{A}$  e di  $\mathbf{B}$  è 2. In base al teorema di Rouché-Capelli il sistema è risolubile. Per determinare le soluzioni usiamo la procedura indicata sopra.

1. Isoliamo il minore  $\begin{vmatrix} 1 & 2 \\ 1 & -2 \end{vmatrix}$  corrispondente alle prime due righe e due colonne di  $\mathbf{A}$ .
2. Eliminiamo la terza e la quarta equazione del sistema.
3. Portiamo a secondo membro i termini contenenti  $z$ .
4. Si ottiene il sistema di due equazioni in due incognite:

$$\begin{cases} x + 2y = 1 - 3z \\ x - 2y = z \end{cases}$$

Sottraendo la seconda equazione dalla prima, si ottiene:

$$\begin{cases} x + 2y = 1 - 3z \\ 4y = 1 - 4z \end{cases}$$

da cui  $y = \frac{1-4z}{4}$ ,  $x = 1 - 3z - 2(\frac{1-4z}{4}) = 1/2 - z$

In conclusione le soluzioni sono assegnate dalla formula

$$x = 1/2 - z \quad y = \frac{1-4z}{4}$$

con  $z$  arbitrario. Detto altrimenti, le soluzioni sono tutti i vettori del tipo

$$\left( 1/2 - z, \frac{1-4z}{4}, z \right)$$

con  $z$  arbitrario numero reale. Le soluzioni sono  $\infty^1$ , ovvero dipendono da un parametro arbitrario:  $z$ .

Se al vettore colonna

$$\mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 3 \\ 1 \end{pmatrix}$$

del sistema (5.16) sostituiamo il vettore  $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$  il sistema diventa impossibile. Infatti, mentre

il rango di  $\mathbf{A}$  è 2 il rango della nuova matrice completa

$$\mathbf{B} = \left( \mathbf{A} \left| \begin{matrix} 0 \\ 0 \\ 1 \\ 0 \end{matrix} \right. \right) = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 1 & -2 & -1 & 0 \\ 2 & 8 & 10 & 1 \\ 2 & 0 & 2 & 0 \end{pmatrix}$$

è 3, essendo il minore

$$\begin{vmatrix} 1 & 2 & 0 \\ 1 & -2 & 0 \\ 2 & 8 & 1 \end{vmatrix} = -4$$

diverso da zero.



Risolvere i sistemi

$$\text{i) } \begin{cases} 6x - 2y - 2z - 8t = 7 \\ -9x + 3y + 3z + 12t = 13 \end{cases}$$

$$\text{ii) } \begin{cases} x - 3y + 5z = -3 \\ 3x + 2y + 4z = -9 \\ -x - 3y + z = 3 \\ 2x + y + 3z = -6 \end{cases}$$

Al variare dei parametri reali  $a, b$  calcolare il rango della seguente matrice:

$$\begin{pmatrix} a & b & 0 \\ a & a & a \\ 0 & b & a \end{pmatrix}$$

Trovare poi tutte le soluzioni del sistema:

$$\begin{cases} x + \frac{1}{2}y = 0 \\ x + y + z = 1 \\ y + 2z = 2 \end{cases}$$

Discutere, al variare del parametro reale  $a$ , la risolubilità o meno del seguente sistema e, quando è possibile, risolverlo:

$$\begin{cases} ax + 2y = 2 \\ 2x + \frac{a}{4}y = 1 \end{cases}$$

Stabilire per quali valori di  $k$  il sistema

$$\begin{cases} x + y + z = k \\ x + y - z = 1 \\ kz + y + z = -1 \end{cases}$$

ammette un'unica soluzione, e determinarla.

Discutere l'esistenza di soluzioni del sistema lineare omogeneo di 3 equazioni in 4 incognite  $x, y, z, t$

$$\begin{cases} hx - hy + t = 0 \\ x - 2y - z = 0 \\ y + hz + t = 0 \end{cases}$$

al variare del parametro  $h$  e, quando esistono, trovare le soluzioni.

Per quali valori del parametro  $h$  è invertibile la matrice

$$\mathbf{A} = \begin{pmatrix} 0 & h & 0 \\ 1 & 1 & h \\ 2 & h & 1 \end{pmatrix} ?$$

Calcolare  $\mathbf{A}^{-1}$ , quando esiste. Quante soluzioni ha il sistema  $\mathbf{Ax} = \mathbf{b}$ , al variare del termine noto  $\mathbf{b}$ ?

Nei prossimi due esercizi si considerano sistemi lineari nel campo complesso, contenenti parametri complessi. La teoria fin qui sviluppata si applica inalterata.

Trovare, per ogni  $\lambda$  complesso, tutte le soluzioni dei seguenti sistemi (omogenei):

$$\text{i) } \begin{cases} \lambda x - y = 0 \\ \lambda y - z = 0 \\ x - \lambda z = 0 \end{cases} \quad \text{ii) } \begin{cases} x + y + z = 0 \\ x + y + \lambda^3 z = 0 \\ 2y - \lambda^3 z = 0 \end{cases} \quad \text{iii) } \begin{cases} 2x + \lambda^3 y - z = 0 \\ x + y - 2z = 0 \\ x - y = 0 \end{cases}$$

Discutere, al variare del parametro complesso  $\lambda$ , la risolubilità dei sistemi seguenti e calcolare esplicitamente le soluzioni, quando esistono.

$$\begin{cases} x + y + z = 0 \\ \lambda x + 2z = \lambda \\ x - \lambda z = 0 \end{cases} \quad \begin{cases} \lambda x + z = 0 \\ 2y + z = 1 \\ \frac{1}{2}x + \lambda^3(y + z) = -1 \end{cases} \quad \begin{cases} 2x + \lambda^3 y - z = 0 \\ x + y - 2z = 0 \\ x - y = 0 \end{cases}$$

#### 5.4 Metodo di eliminazione di Gauss

Presentiamo ora un metodo per risolvere i sistemi lineari, che è alternativo rispetto a quello basato sul teorema di Cramer incontrato in precedenza e si presta meglio ad essere tradotto in un effettivo algoritmo risolutivo.

Cominciamo con un sistema di  $n$  equazioni in  $n$  incognite

$$\mathbf{Ax} = \mathbf{b}$$

L'idea, molto semplice, è quella di procedere con una serie di operazioni elementari sulle equazioni del sistema, ossia sulle righe della matrice completa, in modo da ottenere ogni volta un sistema equivalente (cioè con le stesse soluzioni) cercando di ridursi alla fine ad un sistema

$$\mathbf{A}^* \mathbf{x} = \mathbf{b}^*$$

del tipo *triangolare* seguente:

$$(5.17) \quad \left\{ \begin{array}{l} d_1 x_1 + a_{12}^* x_2 + a_{13}^* x_3 + \cdots + a_{1n}^* x_n = b_1^* \\ d_2 x_2 + a_{23}^* x_3 + \cdots + a_{2n}^* x_n = b_2^* \\ \vdots \\ d_{n-1} x_{n-1} + a_{n-1,n}^* x_n = b_{n-1}^* \\ d_n x_n = b_n^* \end{array} \right.$$

I coefficienti sulla diagonale principale  $d_1, d_2, \dots, d_n$  alla fine del processo prendono il nome di *elementi pivotali* o *pivot*. Si noti che  $\det \mathbf{A}^* = d_1 d_2 \cdots d_n$ .

Se i pivot sono tutti diversi da zero, il sistema (5.17) ha una ed una sola soluzione, facilmente calcolabile con un procedimento di sostituzione *all'indietro*:

dall'ultima equazione si ricava  $x_n = b_n^* d_n^{-1}$ ; sostituendo nella penultima si trova

$$x_{n-1} = (-a_{n-1,n}^* b_n^* + b_{n-1}^*) d_{n-1}$$

e così via fino a determinare  $x_1$ .

Se qualche pivot è nullo la matrice  $\mathbf{A}^*$  è singolare (di conseguenza lo è anche la matrice  $\mathbf{A}$  del sistema originale). Abbiamo già visto che in tal caso il sistema può non essere risolubile oppure ammettere infinite soluzioni. Esamineremo più avanti come si sviluppa il metodo in questi casi ritrovando i risultati teorici delle sezioni precedenti.

Illustriamo il metodo di eliminazione anzitutto su un esempio.



**ESEMPIO** Sia da risolvere il sistema

$$(5.18) \quad \begin{cases} 2x + y + z = 1 \\ 4x + y = -2 \\ -2x + 2y + z = -5. \end{cases}$$

Si comincia col determinare un sistema equivalente (cioè con le stesse soluzioni) eliminando l'incognita  $x$  nella seconda e terza equazione, mediante le operazioni seguenti.

1. Si sottrae dalla seconda equazione la prima equazione moltiplicata per 2. In simboli:

$$2^{\text{a}} \text{ equaz.} - (2 \times 1^{\text{a}} \text{ equaz.}) .$$

2. Si somma alla terza equazione la prima equazione:

$$3^{\text{a}} \text{ equaz.} + 1^{\text{a}} \text{ equaz.}$$

Si ottiene il sistema

$$(5.19) \quad \begin{cases} 2x + y + z = 1 \\ -y - 2z = -4 \\ 3y + 2z = -4. \end{cases}$$

Evidentemente il primo pivot è  $d_1 = 2$ .

3. Si ignora la prima equazione e si elimina l'incognita  $y$  dalla terza equazione con l'operazione

$$3^{\text{a}} \text{ equaz.} + (3 \times 1^{\text{a}} \text{ equaz.}).$$

Si trova

$$(5.20) \quad \begin{cases} 2x + y + z = 1 \\ -y - 2z = -4 \\ -4z = -16. \end{cases}$$

Gli altri pivot sono  $d_2 = -1, d_3 = -4$ . La soluzione è ora immediata:

$$z = 4, y = -4, x = 1/2.$$

Si capisce facilmente come il metodo sia generalizzabile ad un sistema di  $n$  equazioni in  $n$  incognite, dove possiamo sempre supporre  $d_1 = a_{11} \neq 0$ . Il primo passo consiste nell'aggiungere alle altre equazioni opportuni multipli della prima, in modo da azzerare tutti i coefficienti sotto il primo pivot  $d_1$ . Nel secondo si ignora la prima equazione e si aggiungono alle altre equazioni opportuni multipli della seconda, in modo da azzerare tutti i coefficienti sotto il secondo pivot  $d_2$ .

Se ad ogni passo si produce un pivot non nullo, l'azzeramento dei coefficienti sotto ogni pivot è sempre possibile e si procede così fino alla forma triangolare alta finale. Una sostituzione all'indietro fornisce immediatamente la soluzione. Vediamo ora come in generale si traduce l'idea precedente in un algoritmo.

#### Sistemi $n \times n$ . Il caso dei pivot tutti diversi da zero. Fattorizzazione LU di una matrice quadrata

Introduciamo anzitutto qualche notazione matriciale che ci servirà a rappresentare le operazioni elementari che trasformano il sistema in uno ad esso equivalente.

- *Matrici di permutazione.* matrici!di permutazione Prendiamo la matrice identità di ordine  $n$  e scambiamo le due righe  $j$  e  $k$ . Indichiamo la nuova matrice con  $\mathbf{P}_{jk}$ . Sia ora  $\mathbf{A}$  una matrice di ordine  $n \times n$ . Allora la matrice  $\mathbf{P}_{jk}\mathbf{A}$  si ottiene da  $\mathbf{A}$  scambiando le righe  $j$  e  $k$ . Per esempio, se

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & 5 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix}, \quad \mathbf{P}_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{P}_{13}\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 1 & 5 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 2 & 1 \\ 4 & 1 & 0 \\ 3 & 1 & 5 \end{pmatrix}$$

La prima e la terza riga sono scambiate. Se postmoltiplichiamo per  $\mathbf{P}_{jk}$ , cioè calcoliamo  $\mathbf{AP}_{jk}$ , sono le colonne di  $\mathbf{A}$  che vengono scambiate:

$$\mathbf{AP}_{13} = \begin{pmatrix} 3 & 1 & 5 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 1 & 3 \\ 0 & 1 & 4 \\ 1 & 2 & -2 \end{pmatrix}$$

- *Matrici elementari.* matrici!elementari Introduciamo ora delle matrici che servono invece a rappresentare l'operazione di sommare o sottrarre a un'equazione un multiplo di un'altra equazione. Indichiamo con  $\mathbf{E}_{ij}(q)$  la matrice costruita sostituendo nella matrice identità lo zero in posizione  $ij$ ,  $i > j$ , con il numero  $q$ . Pre (post) moltiplicando una matrice  $\mathbf{A}$  per  $\mathbf{E}_{ij}(q)$  si aggiunge alla riga (colonna)  $i$  il multiplo  $q$  della riga  $j$ . Per esempio:

$$\mathbf{E}_{31}(-2)\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 1 \\ 4 & 1 & 0 \\ 1 & -3 & 5 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 1 \\ 4 & 1 & 0 \\ -3 & -1 & 3 \end{pmatrix}$$

e si vede che la terza riga della matrice prodotto è ottenuta sommando alla terza riga di  $\mathbf{A}$  la prima riga moltiplicata per  $-2$ . Queste matrici si chiamano matrici *elementari*, sono triangolari basse e hanno 1 sulla diagonale principale.

Rileggiamo ora il metodo di eliminazione in termini matriciali. Usiamo ancora il sistema (5.18):

$$\begin{pmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ -5 \end{pmatrix}.$$

La prima operazione

$$(5.21) \quad 2^{\text{a}} \text{ equaz.} - (2 \times 1^{\text{a}} \text{ equaz.})$$

può essere realizzata mediante premoltiplicazione della matrice dei coefficienti e del vettore dei termini noti per la matrice elementare

$$\mathbf{E}_{21}(-2) = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Analogamente, la seconda e terza operazione, date da

$$(5.22) \quad 3^{\text{a}} \text{ equaz.} + 1^{\text{a}} \text{ equaz.} \text{ e } 3^{\text{a}} \text{ equaz.} + (3 \times 2^{\text{a}} \text{ equaz.}),$$

si realizzano premoltiplicando rispettivamente per le matrici

$$\mathbf{E}_{31}(1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \text{ e } \mathbf{E}_{32}(3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}.$$

Si ha:

$$\mathbf{E}_{32}(3) \mathbf{E}_{31}(1) \mathbf{E}_{21}(-2) = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -5 & 3 & 1 \end{pmatrix}$$

che è ovviamente triangolare bassa. Poniamo poi

$$\mathbf{U} = \mathbf{E}_{32}(3) \mathbf{E}_{31}(1) \mathbf{E}_{21}(-2) \mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{pmatrix}$$

che è *triangolare alta* ed ha sulla diagonale principale i pivot. Abbiamo dunque

$$\mathbf{A} = \mathbf{LU}$$

dove

$$\mathbf{L} = \mathbf{E}_{21}(-2)^{-1} \mathbf{E}_{31}(1)^{-1} \mathbf{E}_{32}(3)^{-1}.$$

Si può controllare facilmente che le inverse delle  $\mathbf{E}_{ij}(q)$  si ottengono semplicemente sostituendo  $-q$  a  $q$ ; cioè

$$\mathbf{E}_{ij}(q)^{-1} = \mathbf{E}_{ij}(-q).$$

Perciò

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{pmatrix}$$

che è triangolare bassa, ha gli elementi della diagonale principale tutti uguali ad 1, gli altri elementi  $q_{ij}$  sono gli opposti degli elementi  $i, j$  che compaiono nelle matrici  $\mathbf{E}_{ij}(q)$  utilizzate per costruire  $\mathbf{L}$ .

Possiamo quindi scrivere

$$\underbrace{\begin{pmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{pmatrix}}_{\mathbf{L}} \underbrace{\begin{pmatrix} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{pmatrix}}_{\mathbf{U}}.$$

La matrice  $\mathbf{A}$  è stata quindi fattorizzata nel prodotto di due matrici triangolari. Le due matrici  $\mathbf{L}$  ed  $\mathbf{U}$  sono univocamente determinate. Infatti se

$$\mathbf{A} = \mathbf{LU} = \mathbf{L}'\mathbf{U}'$$

con  $\mathbf{L}'$  ed  $\mathbf{U}'$  rispettivamente triangolare bassa ed alta<sup>4</sup>, si ha anche

$$(5.23) \quad (\mathbf{L}')^{-1}\mathbf{L} = \mathbf{U}'\mathbf{U}^{-1}.$$

Ora  $(\mathbf{L}')^{-1}$  ed  $(\mathbf{L}')^{-1}\mathbf{L}$  sono triangolari basse mentre  $\mathbf{U}^{-1}$  ed  $\mathbf{U}'\mathbf{U}^{-1}$  sono triangolari alte, con elementi uguali ad 1 sulla diagonale principale. Ne segue che entrambi i membri di (5.23) sono matrici diagonalari. Ma allora deve essere

$$(\mathbf{L}')^{-1}\mathbf{L} = \mathbf{U}'\mathbf{U}^{-1} = \mathbf{I}$$

da cui  $\mathbf{L} = \mathbf{L}'$ ,  $\mathbf{U} = \mathbf{U}'$ .

Riassumiamo generalizzando ad una matrice quadrata di ordine  $n$ .

**PROPOSIZIONE 8.6** *Se nel processo di eliminazione di Gauss tutti i pivot sono diversi da zero, la matrice  $\mathbf{A}$  ammette la fattorizzazione*

$$\mathbf{A} = \mathbf{LU}$$

dove  $\mathbf{L}$ ,  $\mathbf{U}$  sono univocamente determinate dalle seguenti proprietà:

- a)  $\mathbf{L}$  è triangolare bassa con elementi sulla diagonale principale tutti uguali a 1; gli altri elementi non nulli sono i  $q_{ij}$  descritti in precedenza.
- b)  $\mathbf{U}$  è la matrice che si ottiene al termine del processo di eliminazione, è triangolare alta e gli elementi sulla diagonale principale sono i pivot.

Di conseguenza:

**COROLLARIO 8.19** *Nelle ipotesi della Proposizione precedente il sistema (di  $n$  equazioni in  $n$  incognite)*

$$\mathbf{Ax} = \mathbf{b}$$

è equivalente al sistema triangolare alto

$$\mathbf{Ux} = \mathbf{L}^{-1}\mathbf{b}$$

ed ha esattamente una soluzione.

Infatti, essendo il sistema  $\mathbf{Ux} = \mathbf{L}^{-1}\mathbf{b}$  triangolare e con elementi 1 sulla diagonale, può essere univocamente risolto per sostituzione all'indietro, un'equazione dopo l'altra.

<sup>4</sup> $L$  sta per *lower*,  $U$  sta per *upper*.

**Sistemi  $n \times n$ . Pivot non tutti diversi da zero**

Esaminiamo ora il caso in cui, durante il processo di eliminazione, i primi  $k - 1$  pivot siano non nulli, mentre  $d_k = 0$ . Controlliamo gli elementi della colonna  $k$ -esima, sotto  $d_k$ .

Vi sono due possibilità: (a) o tutti questi elementi sono nulli oppure (b) ne esiste uno non nullo, diciamo l'elemento  $a_{lk}$ .

Nel primo caso la difficoltà è più seria: la matrice  $\mathbf{A}$  è *singolare*. Le incognite rimaste nelle ultime  $n - k + 1$  equazioni sono meno di  $n - k + 1$  e quindi o il sistema ha infinite soluzioni oppure non ne ha alcuna. Completeremo l'analisi di questo caso considerando direttamente sistemi generali di  $m$  equazioni in  $n$  incognite.

Nel secondo caso la difficoltà si supera facilmente scambiando le righe  $k$  ed  $l$  ed il procedimento di eliminazione può continuare. Se proseguendo non ci si trova mai nel caso (a) l'eliminazione si può portare a termine. Per quanto riguarda la fattorizzazione, lo scambio di due righe  $k$  ed  $l$  si effettua premoltiplicando per la matrice di permutazione  $\mathbf{P}_{kl}$  ottenuta dalla matrice identità scambiando quelle due righe. Nella fattorizzazione intervengono allora le matrici di permutazione  $\mathbf{P}_{kl}^{-1} = \mathbf{P}_{lk}$  che non mantengono la forma triangolare delle  $\mathbf{E}_{ij}(q)$ . Il prodotto di tali matrici non è triangolare bassa. La fattorizzazione può tuttavia essere quasi del tutto recuperata utilizzando la seguente osservazione. Esaminiamo la lista di scambi tra righe necessari per portare a termine l'eliminazione ed eseguiamole *prima* di incominciarla. In altri termini, rimpiazziamo la matrice originale  $\mathbf{A}$  con una matrice  $\mathbf{PA}$ , dove  $\mathbf{P}$  realizza tutte le permutazioni richieste di righe.  $\mathbf{P}$  è il prodotto di matrici  $\mathbf{P}_{kl}$  e contiene esattamente un 1 in ogni riga. La matrice  $\mathbf{PA}$  ammette la fattorizzazione  $\mathbf{LU}$  con pivot tutti diversi da zero.

**PROPOSIZIONE 8.7** *Se  $\mathbf{A}$  è non singolare esiste una matrice di permutazione  $\mathbf{P}$  tale che la matrice  $\mathbf{PA}$  ammette la fattorizzazione*

$$\mathbf{PA} = \mathbf{LU}$$

con tutti i pivot diversi da zero.

**COROLLARIO 8.20** *Nelle ipotesi della Proposizione precedente il sistema (di  $n$  equazioni in  $n$  incognite)*

$$\mathbf{Ax} = \mathbf{b}$$

è equivalente al sistema triangolare alto

$$\mathbf{Ux} = \mathbf{L}^{-1}\mathbf{Pb}$$

ed ha esattamente una soluzione.

**Sistemi lineari di  $m$  equazioni in  $n$  incognite. Fattorizzazione di una matrice rettangolare**

Esaminiamo ora il caso di un sistema generale di  $m$  equazioni in  $n$  incognite

$$\mathbf{Ax} = \mathbf{b}$$

con  $m \neq n$ . Il metodo di eliminazione di Gauss si può applicare anche in questo caso. Quando però tutto viene riletto in termini di fattorizzazione della matrice  $\mathbf{A}$ , si presenta qualche differenza.

Ci serviamo di un esempio di 4 equazioni in 5 incognite, ma lavorando direttamente sulla matrice dei coefficienti del sistema, che supponiamo sia

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 3 & 2 & 0 \\ 2 & 6 & 9 & 1 & -6 \\ -1 & -3 & 3 & 0 & -2 \\ -1 & -3 & 1 & -6 & -8 \end{pmatrix}$$

Il primo pivot è  $a_{11} = 1 \neq 0$ , per cui le consuete operazioni elementari permettono di produrre zeri nella prima colonna sotto quel pivot; si ottiene:

$$\begin{pmatrix} 1 & 3 & 3 & 2 & 0 \\ 0 & 0 & 3 & -3 & -6 \\ 0 & 0 & 6 & 2 & -2 \\ 0 & 0 & 4 & -4 & -8 \end{pmatrix}$$

Il secondo pivot è zero, gli elementi sottostanti sono nulli e lo scambio di righe non può produrre un pivot diverso da zero.

Per procedere oltre con l'eliminazione si va al primo elemento non nullo in seconda riga, nelle colonne successive; nell'esempio è 3 in terza colonna. Annulliamo gli elementi sotto il 3 con le operazioni

$$3^{\text{a}} \text{ riga} - (2 \times 2^{\text{a}} \text{ riga}) \quad \text{e} \quad 4^{\text{a}} \text{ riga} - \left(\frac{4}{3} \times 2^{\text{a}} \text{ riga}\right).$$

Si perviene alla matrice

$$\mathbf{U} = \begin{pmatrix} 1 & 3 & 3 & 2 & 0 \\ 0 & 0 & 3 & -3 & -6 \\ 0 & 0 & 0 & 8 & 10 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Si considera ora la terza colonna dove si riconosce che il nuovo pivot è 8. Essendoci solo zeri nell'ultima riga, il processo di eliminazione è finito. La forma finale di  $\mathbf{U}$  è triangolare alta, o meglio, *trapezoidale alta* con pivot uguali a 1 (in prima riga), 3 (in seconda riga) e 8 (in terza riga) non nulli.

In generale si arriva a matrici dalla forma tipica seguente:

$$\mathbf{U} = \begin{pmatrix} \clubsuit & * & * & * & * & * & * & * \\ 0 & \clubsuit & * & * & * & * & * & * \\ 0 & 0 & \clubsuit & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \clubsuit & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

dove i fiori indicano i pivot non nulli e gli asterischi indicano elementi generici.

Le caratteristiche della forma trapezoidale alta sono:

- Le righe non nulle vengono prima (dall'alto) e i pivot sono i primi elementi (da sinistra) non nulli in tali righe.

- Il numero di pivot non nulli coincide con il numero di colonne indipendenti di  $\mathbf{U}$  e con il numero di righe non nulle di  $\mathbf{U}$ : coincide perciò con il rango di  $\mathbf{U}$ .
- Sotto ogni pivot c'è una colonna di zeri, ottenuti con l'eliminazione di Gauss.
- Ogni pivot si trova alla destra del pivot nella riga superiore.

Come prima, a partire da  $\mathbf{U}$ , possiamo ricostruire la matrice  $\mathbf{A}$  premoltiplicando  $\mathbf{U}$  per le inverse delle matrici elementari  $E_{jk}(q)$ . Il prodotto  $\mathbf{L}$  di queste ultime è una matrice quadrata e *triangolare bassa* di ordine  $m$ , il numero di righe di  $\mathbf{A}$ . Essendo  $\mathbf{L}$  invertibile,  $\mathbf{A} = \mathbf{LU}$  ha lo stesso rango di  $\mathbf{U}$ .

Nel caso siano necessari scambi di righe nulla cambia rispetto al caso di una matrice quadrata, con l'introduzione della matrice di permutazione  $\mathbf{P}$ . Riassumiamo le conclusioni.

**PROPOSIZIONE 8.8** *Sia  $\mathbf{A}$  una matrice  $m \times n$ . Esistono una matrice di permutazione  $\mathbf{P}$ , una matrice quadrata  $\mathbf{L}$  di ordine  $m$ , triangolare bassa, ed una matrice  $\mathbf{U}$  di ordine  $m \times n$ , trapezoidale alta, tali che*

$$\mathbf{PA} = \mathbf{LU}.$$

Inoltre, rango di  $\mathbf{U}$  = rango di  $\mathbf{A}$ .

Il sistema  $\mathbf{Ax} = \mathbf{b}$  è equivalente al sistema trapezoidale

$$(5.24) \quad \mathbf{Ux} = \mathbf{L}^{-1}\mathbf{Pb}.$$

Una volta arrivati alla forma trapezoidale (5.24) è immediato trarre conclusioni sulla risolubilità o meno del sistema. Tornando all'esempio, poiché vi sono tre pivot non nulli, il rango della matrice  $\mathbf{A}$  è tre.

Una prima conclusione è che il sistema omogeneo finale

$$\begin{cases} x_1 + 3x_2 + 3x_3 + 2x_4 = 0 \\ 3x_3 - 3x_4 - 6x_5 = 0 \\ 8x_4 + 10x_5 = 0 \\ 0 = 0 \end{cases}$$

equivale al sistema triangolare

$$\begin{cases} x_1 + 3x_3 + 2x_4 = -3x_2 \\ 3x_3 - 3x_4 = 6x_5 \\ 8x_4 = -10x_5 \end{cases}$$

ottenuto lasciando a primo membro le variabili corrispondenti ai pivot non nulli (cosiddette variabili *di base*). Si vede che esistono  $\infty^2$  soluzioni, essendo  $x_2, x_5$  variabili libere.

La risolubilità del sistema non omogeneo si legge dalle righe di zeri. Per esempio, se

$$\mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ -1 \\ 2 \end{pmatrix}$$

il sistema trapezoidale finale è

$$\begin{cases} x_1 + 3x_2 + 3x_3 + 2x_4 = 2 \\ 3x_3 - 3x_4 - 6x_5 = -4 \\ 8x_4 + 10x_5 = 9 \\ 0 = 28/3 \end{cases}$$

che evidentemente non ha soluzione. Ciò riflette il fatto che il vettore  $\mathbf{L}^{-1}\mathbf{Pb}$  non appartiene allo spazio generato dalle colonne di  $\mathbf{U}$  (Teorema di Rouché-Capelli).

## 6 AUTOVETTORI ED AUTOVALORI. DIAGONALIZZAZIONE

### 6.1 Matrici diagonalizzabili

Consideriamo una trasformazione  $\mathcal{L}$  da  $\mathbb{R}^n$  in sé. La rappresentazione di  $\mathcal{L}$  per mezzo di una matrice  $\mathbf{A}$  (in questo caso quadrata di ordine  $n$ ) dipende dalla scelta della base in  $\mathbb{R}^n$ . È chiaro che se si potesse scegliere una base rispetto alla quale  $\mathcal{L}$  fosse rappresentata da una matrice di forma speciale, per esempio diagonale, si avrebbe un indubbio vantaggio, almeno sul piano dell'esecuzione dei calcoli.

Supponiamo, per esempio, che una trasformazione  $\mathcal{L} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ , rispetto ad una opportuna base, si possa rappresentare mediante la matrice diagonale

$$\mathbf{A} = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{pmatrix}$$

ciò significa che la trasformazione  $\mathbf{y} = \mathcal{L}(\mathbf{x})$  rispetto a questa base si scrive così:

$$\begin{cases} y_1 = c_1 x_1 \\ y_2 = c_2 x_2 \\ y_3 = c_3 x_3 \end{cases}$$

Si tratta di un'espressione particolarmente semplice, in cui ogni variabile  $y_i$  dipende solo dalla variabile  $x_i$ , anziché da tutte e tre ("separazione di variabili"). Geometricamente, significa che abbiamo scelto un opportuno sistema di riferimento in  $\mathbb{R}^3$  (non necessariamente ortogonale), rispetto al quale la trasformazione agisce come una dilatazione di un certo coefficiente  $c_i$  su ciascun asse. Rispetto a questa base i calcoli risultano più semplici, e le proprietà della trasformazione appaiono in modo più trasparente. Diremo in questo caso che *la trasformazione ha forma diagonale*.

Cominciamo a esaminare come cambia la rappresentazione di  $\mathcal{L}$  cambiando la base in  $\mathbb{R}^n$ .

Supponiamo che  $\mathbb{R}^n$  sia riferito alla base canonica  $\mathbf{e}^1, \mathbf{e}^2, \dots, \mathbf{e}^n$ . Scegliamo ora  $n$  vettori linearmente indipendenti  $\tilde{\mathbf{e}}^1, \tilde{\mathbf{e}}^2, \dots, \tilde{\mathbf{e}}^n$  ed assumiamoli come nuova base. La matrice

$$\mathbf{S} = (\tilde{\mathbf{e}}^1 | \tilde{\mathbf{e}}^2 | \dots | \tilde{\mathbf{e}}^n)$$

ottenuta accostando i vettori (colonna)  $\tilde{\mathbf{e}}^1, \tilde{\mathbf{e}}^2, \dots, \tilde{\mathbf{e}}^n$  rappresenta, nella base canonica, la trasformazione che fa passare da una base all'altra. Infatti, come si verifica subito:

$$\mathbf{Se}^i = \tilde{\mathbf{e}}^i \quad i = 1, 2, \dots, n$$

Poiché i vettori  $\tilde{\mathbf{e}}^i$  sono indipendenti,  $\det \mathbf{S} \neq 0$ , quindi  $\mathbf{S}$  è invertibile, e possiamo anche scrivere:

$$\mathbf{e}^i = \mathbf{S}^{-1} \tilde{\mathbf{e}}^i.$$

Se quindi un vettore di  $\mathbb{R}^n$  ha, rispetto alla base canonica, le coordinate  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , nella nuova base avrà coordinate  $\tilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$  date da

$$\tilde{\mathbf{x}} = \mathbf{S}^{-1} \mathbf{x}$$

Infatti: se la matrice  $\mathbf{S} = (s_{ij})$  ha per colonne le rappresentazioni dei vettori  $\tilde{\mathbf{e}}_i$  nella base di  $\mathbf{e}_i$ , significa che  $s_{ij} = (\tilde{\mathbf{e}}_j)_i$ , quindi da

$$\mathbf{x} = \sum_{j=1}^n \tilde{x}_j \tilde{\mathbf{e}}_j$$

ricaviamo

$$x_i = \sum_{j=1}^n \tilde{x}_j (\tilde{\mathbf{e}}_j)_i = \sum_{j=1}^n \tilde{x}_j s_{ij}$$

ossia

$$\mathbf{x} = \mathbf{S} \tilde{\mathbf{x}}.$$

Allora, se  $\mathbf{y} = \mathbf{Ax}$  nella base canonica, risulta

$$\tilde{\mathbf{y}} = \mathbf{S}^{-1} \mathbf{y} = \mathbf{S}^{-1} \mathbf{Ax} = \mathbf{S}^{-1} \mathbf{AS} \tilde{\mathbf{x}}$$

e perciò la trasformazione  $\mathcal{L}$ , nella nuova base è rappresentata dalla matrice

$$(6.1) \quad \tilde{\mathbf{A}} = \mathbf{S}^{-1} \mathbf{AS}$$

Si dice che due matrici  $\mathbf{A}, \tilde{\mathbf{A}}$  legate dalla relazione (6.1) (con  $\mathbf{S}$  ovviamente matrice invertibile) sono *simili*, o equivalenti.

Si noti che dal Teorema di Binet (v. par. 4.3) segue subito che *due matrici simili hanno lo stesso determinante*. Infatti se  $\tilde{\mathbf{A}} = \mathbf{S}^{-1} \mathbf{AS}$  si ha:

$$|\tilde{\mathbf{A}}| = |\mathbf{S}^{-1} \mathbf{AS}| = |\mathbf{S}^{-1}| |\mathbf{A}| |\mathbf{S}| = \frac{1}{|\mathbf{S}|} |\mathbf{A}| |\mathbf{S}| = |\mathbf{A}|.$$

È utile evidenziare anche un'altra proprietà, legata a questa:

**PROPOSIZIONE 8.9** *Due matrici simili hanno la stessa caratteristica.*

**DIMOSTRAZIONE.** Iniziamo con la seguente osservazione. Se  $\mathbf{S}$  è una matrice  $n \times n$  non singolare,  $k$  vettori riga  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbb{R}^n$  sono linearmente indipendenti se e solo se lo sono i  $k$  vettori  $\mathbf{v}_1 \mathbf{S}, \mathbf{v}_2 \mathbf{S}, \dots, \mathbf{v}_k \mathbf{S}$ . Infatti:

$$\sum_{j=1}^k c_j (\mathbf{v}_j \mathbf{S}) = 0 \iff \left( \sum_{j=1}^k c_j \mathbf{v}_j \right) \mathbf{S} = 0 \iff \sum_{j=1}^k c_j \mathbf{v}_j = 0$$

(essendo  $\mathbf{S}$  non singolare). Analogamente  $k$  vettori colonne  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \in \mathbb{R}^n$  sono linearmente indipendenti se e solo se lo sono i  $k$  vettori  $\mathbf{S} \mathbf{u}_1, \mathbf{S} \mathbf{u}_2, \dots, \mathbf{S} \mathbf{u}_k$ . Ricordando ora che la caratteristica di una matrice  $\mathbf{A}$ ,  $n \times n$ , è il massimo numero di righe (o colonne) di  $\mathbf{A}$

linearmente indipendenti, ne segue subito che la caratteristica di una matrice quadrata  $\mathbf{A}$  non cambia se la moltiplichiamo a destra o a sinistra per una matrice non singolare. Pertanto  $\mathbf{S}^{-1} \mathbf{AS}$  ha la stessa caratteristica di  $\mathbf{A}$ .

Quanto detto, inoltre, si può ripetere pari pari per matrici a elementi complessi, viste come trasformazioni di  $\mathbb{C}^n$  in sé. Diamo ora una definizione generale:

**DEFINIZIONE 8.16** Diremo che una matrice quadrata  $\mathbf{A}$  a elementi in  $\mathbb{K}$  (dove  $\mathbb{K}$  è  $\mathbb{R}$  o  $\mathbb{C}$ ) è diagonalizzabile (su  $\mathbb{K}$ ) se esistono due matrici  $\mathbf{A}$  e  $\mathbf{S}$  a elementi in  $\mathbb{K}$ ,  $\mathbf{A}$  diagonale e  $\mathbf{S}$  non singolare, tali che

$$\mathbf{A} = \mathbf{S} \mathbf{AS}^{-1}.$$

La matrice  $\mathbf{S}$  si dice matrice di passaggio.

In questo caso esiste una base di  $\mathbb{K}^n$  in cui la trasformazione che rispetto alla base canonica si rappresentava con la matrice  $\mathbf{A}$ , si rappresenta invece con una matrice diagonale. Nella nuova base, quindi, la trasformazione è più semplice da studiare, in quanto ha forma diagonale.

Il problema che ci poniamo è quindi: data una matrice, decidere se essa è diagonalizzabile, e in caso affermativo determinare la matrice diagonale e la matrice di passaggio. Questo problema è legato ad un altro, che ora introdurremo, quello della ricerca di *autovalori* e *autovettori* della matrice.

Notiamo esplicitamente quanto segue: poiché i numeri reali sono particolari numeri complessi, una matrice  $(n, n)$  a elementi reali può vedersi sia come una trasformazione lineare di  $\mathbb{R}^n$  in sé che come trasformazione lineare di  $\mathbb{C}^n$  in sé; in entrambi i casi ci si può chiedere se la matrice  $\mathbf{A}$  è diagonalizzabile. Ora, può accadere, come vedremo, che  $\mathbf{A}$  sia diagonalizzabile su  $\mathbb{C}$  ma non su  $\mathbb{R}$ . Per una matrice a elementi reali, dunque, le due nozioni di diagonalizzabilità vanno distinte.

## 6.2 Autovalori ed autovettori di una matrice

Chiediamoci se esistono vettori (non nulli) di  $\mathbb{K}^n$  che vengono trasformati da  $\mathbf{A}$  in vettori paralleli. Se  $\mathbf{v} \in \mathbb{K}^n$  è un tale vettore, dovrà risultare

$$(6.2) \quad \mathbf{Av} = \lambda \mathbf{v}$$

per qualche  $\lambda \in \mathbb{K}$ . In altre parole, ci stiamo chiedendo se esistono *particolari vettori* sui quali la trasformazione agisce come se fosse di tipo diagonale. In tal caso, preso un qualsiasi vettore  $t\mathbf{v}$  parallelo a  $\mathbf{v}$  ( $t \in \mathbb{K}$ ) si avrà anche, per linearità:

$$\mathbf{A}(t\mathbf{v}) = t\mathbf{Av} = (t\lambda)\mathbf{v}$$

e quindi, in questo caso, possiamo affermare che la direzione di  $\mathbb{K}^n$  individuata dal vettore  $\mathbf{v}$  è trasformata in se stessa.

Scrivendo  $\mathbf{I}_n \mathbf{v}$  al posto di  $\mathbf{v}$  possiamo riscrivere l'equazione (6.2) nella forma

$$(\mathbf{A} - \lambda \mathbf{I}_n) \mathbf{v} = \mathbf{0}$$

In questa equazione il vettore  $\mathbf{v}$  (come lo scalare  $\lambda$ ) è incognito: le componenti del vettore soddisfano perciò un sistema lineare omogeneo  $n \times n$ . Sappiamo che, perché esistano vettori  $\mathbf{v}$  (non nulli) che risolvono la (6.2), dovrà essere

$$(6.3) \quad D(\lambda) = |\mathbf{A} - \lambda \mathbf{I}_n| = 0$$

Per esempio, per  $n = 2$ , se  $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = 0$$

L'equazione (6.3) è un'equazione algebrica in  $\lambda$ , di grado  $n$ , detta *equazione caratteristica* della matrice  $\mathbf{A}$ ; il polinomio  $D(\lambda)$  è detto *polinomio caratteristico* della matrice  $\mathbf{A}$ . Questo polinomio dipende solo dalla trasformazione lineare, e non dal riferimento in cui questa è descritta; in altri termini, non varia se si sostituisce alla matrice  $\mathbf{A}$  un'altra matrice ad essa equivalente. Infatti si può scrivere

$$\mathbf{S}^{-1}\mathbf{AS} - \lambda\mathbf{I} = \mathbf{S}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{S}$$

e quindi, usando le proprietà del determinante:

$$|\mathbf{S}^{-1}\mathbf{AS} - \lambda\mathbf{I}| = |\mathbf{S}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{S}| = \frac{1}{|\mathbf{S}|}|\mathbf{A} - \lambda\mathbf{I}||\mathbf{S}| = D(\lambda)$$

È utile fissare il risultato appena ottenuto, che ci servirà più volte:

**PROPOSIZIONE 8.10** Due matrici  $n \times n$  simili hanno lo stesso polinomio caratteristico.

Una prima conseguenza di questa proprietà è la seguente. Si può osservare facilmente che nella scrittura esplicita del polinomio  $\det(\mathbf{A} - \lambda\mathbf{I})$ , il termine contenente la potenza  $\lambda^{n-1}$  è

$$(-\lambda)^{n-1}(a_{11} + a_{22} + \dots + a_{nn}).$$

La somma degli elementi sulla diagonale principale di una matrice quadrata è detta *traccia* della matrice:

$$Tr(\mathbf{A}) = a_{11} + a_{22} + \dots + a_{nn}.$$

Possiamo quindi affermare che *due matrici simili hanno la stessa traccia, oltre che lo stesso determinante e la stessa caratteristica*.

**DEFINIZIONE 8.17 (AUTOVALORI E AUTOVETTORI)** Chiameremo *autovalore* della matrice  $\mathbf{A}$  (o della trasformazione che la matrice  $\mathbf{A}$  rappresenta) qualsiasi numero  $\lambda \in \mathbb{C}$  soddisfi l'equazione caratteristica; chiameremo *autovettore* (corrispondente all'autovalore  $\lambda$ ) ogni vettore  $\mathbf{v} \neq 0$  ( $\mathbf{v} \in \mathbb{K}^n$ ) che risolva l'equazione (6.2). 

Alcune osservazioni sono opportune. Si noti anzitutto che, poiché matrici simili hanno lo stesso polinomio caratteristico, avranno pure gli stessi autovalori. Dunque gli autovalori dipendono solo dalla trasformazione lineare, e non dalla particolare matrice rappresentativa (ossia dalla base scelta).

Nella definizione di autovalore, anche se la matrice  $\mathbf{A}$  è reale e la pensiamo come rappresentativa di una trasformazione di  $\mathbb{R}^n$  in sé, gli autovalori possono essere complessi. In altre parole, i problemi della diagonalizzazione e della ricerca di autovalori e autovettori hanno la loro impostazione naturale nel campo complesso.

Notiamo che, mentre un autovettore corrisponde sempre ad un unico autovalore (perché l'identità  $\mathbf{Av} = \lambda\mathbf{v}$  determina univocamente  $\lambda$ ), ad un autovalore corrispondono sempre infiniti autovettori, in quanto se  $\mathbf{Av} = \lambda\mathbf{v}$ , è anche  $\mathbf{A}(c\mathbf{v}) = \lambda(c\mathbf{v})$ , come già osservato, e quindi tutti i vettori  $c\mathbf{v}$  paralleli a  $\mathbf{v}$  sono pure autovettori relativi a  $\lambda$ .

Se la matrice  $\mathbf{A}$  è reale, il polinomio caratteristico è a coefficienti reali per cui gli autovalori sono reali oppure a coppie complessi coniugati. Gli autovettori possono essere scelti in modo che agli autovalori reali corrispondano autovettori reali e ad autovalori complessi coniugati corrispondano autovettori complessi coniugati.



Trovare autovalori e autovettori della matrice

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Calcoliamo

$$\det(\mathbf{A} - \lambda\mathbf{I}_3) = \begin{vmatrix} -\lambda & 1 & 1 \\ 1 & 1-\lambda & 0 \\ 1 & 0 & 1-\lambda \end{vmatrix} =$$

calcolando il determinante dalla 3<sup>a</sup> riga

$$= (\lambda - 1) + (1 - \lambda)(-\lambda + \lambda^2 - 1) = (1 - \lambda)(\lambda^2 - \lambda - 2) = (1 - \lambda)(\lambda + 1)(\lambda - 2)$$

Gli autovalori (valori di  $\lambda$  che annullano il determinante di  $\mathbf{A} - \lambda\mathbf{I}_3$ ) sono

$$\lambda_1 = 1 \quad \lambda_2 = -1 \quad \lambda_3 = 2$$

Gli autovettori corrispondenti si trovano risolvendo il sistema omogeneo  $(\mathbf{A} - \lambda\mathbf{I}_3)\mathbf{x} = 0$ . Se  $\lambda = 1$  tale sistema diventa

$$\begin{cases} -x + y + z = 0 \\ x = 0 \\ x = 0 \end{cases} \quad \text{ossia} \quad \begin{cases} y + z = 0 \\ x = 0 \end{cases}$$

e perciò abbiamo  $\infty^1$  soluzioni date da

$$\begin{cases} x = 0 \\ y = -c \\ z = c \end{cases}$$

con  $c$  parametro arbitrario. Queste soluzioni rappresentano  $\infty^1$  vettori, tutti paralleli tra di loro. Uno qualsiasi di essi si otterrà dando a  $c$  un particolare valore, per esempio  $c = 1$ . Il

vettore  $\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$  è perciò un autovettore della matrice  $\mathbf{A}$  corrispondente all'autovalore 1.

Se  $\lambda = -1$ , il sistema diventa

$$\begin{cases} x + y + z = 0 \\ x + 2y = 0 \\ x + 2z = 0 \end{cases} \quad \text{cioè} \quad \begin{cases} y = -\frac{x}{2} \\ z = -\frac{x}{2} \end{cases}$$

quindi le soluzioni sono del tipo

$$\begin{cases} x = c \\ y = -\frac{c}{2} \\ z = -\frac{c}{2} \end{cases}$$

e si hanno ancora  $\infty^1$  autovettori paralleli tra loro; uno di essi è (per  $c = 2$ )  $\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}$ .

Se  $\lambda = 2$ , il sistema diventa

$$\begin{cases} -2x + y + z = 0 \\ x - y = 0 \\ x - z = 0 \end{cases} \quad \text{ossia} \quad \begin{cases} y = x \\ z = x \end{cases}$$

quindi le soluzioni sono del tipo

$$\begin{cases} x = c \\ y = c \\ z = c \end{cases}$$

e si hanno ancora  $\infty^1$  autovettori paralleli tra loro; uno di essi è (per  $c = 1$ )  $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ .

Trovare autovalori e autovettori della matrice

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

Si ha

$$\det(\mathbf{A} - \lambda \mathbf{I}_2) = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0 \Rightarrow \lambda = \pm i$$

(autovalori complessi coniugati). Questo spinge a vedere  $\mathbf{A}$  come trasformazione lineare di  $\mathbb{C}^2$  in sé. Gli autovettori corrispondenti a  $\lambda = i$  si trovano risolvendo il sistema:

$$(\mathbf{A} - i\mathbf{I}_2)\mathbf{x} = 0$$

cioè

$$\begin{cases} -ix + y = 0 \\ -x - iy = 0 \end{cases} \quad \text{ossia} \quad y = ix$$

Si hanno dunque gli  $\infty^1$  autovettori del tipo  $\begin{pmatrix} c \\ ic \end{pmatrix}$ , al variare di  $c \in \mathbb{C}$ . In particolare, uno di essi è  $\begin{pmatrix} 1 \\ i \end{pmatrix}$ . Con calcoli analoghi si vede che gli autovettori relativi a  $\lambda = -i$  sono del tipo  $\begin{pmatrix} c \\ -ic \end{pmatrix}$ , al variare di  $c \in \mathbb{C}$ . In particolare, uno di essi è  $\begin{pmatrix} 1 \\ -i \end{pmatrix}$ . In questo caso non esistono vettori  $\mathbf{v} \in \mathbb{R}^2$  tali che  $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ .

### 6.3 Condizioni di diagonalizzabilità

La relazione tra il problema della diagonalizzabilità di una matrice e quello della ricerca dei suoi autovalori e autovettori è espressa dal seguente:

**TEOREMA 8.21** La matrice  $\mathbf{A}$  è diagonalizzabile su  $\mathbb{K}$  se e solo se  $\mathbb{K}^n$  possiede una base di autovettori di  $\mathbf{A}$ . In tal caso, detti  $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n$  gli autovettori di tale base, e  $\lambda_1, \lambda_2, \dots, \lambda_n$  gli autovalori corrispondenti ( $\lambda_i \in \mathbb{K}$ , non necessariamente distinti), risulta

$$\mathbf{A} = \mathbf{S}\Lambda\mathbf{S}^{-1}$$

dove  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  e  $\mathbf{S} = (\mathbf{h}_1 | \mathbf{h}_2 | \dots | \mathbf{h}_n)$ .

**DIMOSTRAZIONE.** Supponiamo che  $\mathbf{A}$  sia diagonalizzabile su  $\mathbb{K}$ , ossia valga  $\mathbf{A} = \mathbf{S}\Lambda\mathbf{S}^{-1}$ , con  $\mathbf{S}, \mathbf{A}$  matrici a elementi in  $\mathbb{K}$ ,  $\mathbf{S}$  non singolare e  $\Lambda$  diagonale. Perciò  $\mathbf{AS} = \mathbf{S}\Lambda$ . Chiamando  $\lambda_1, \lambda_2, \dots, \lambda_n$  gli elementi della diagonale di  $\Lambda$ , e  $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n$  le colonne di  $\mathbf{S}$ , l'equazione  $\mathbf{AS} = \mathbf{S}\Lambda$  si riscrive  $\mathbf{Ah}_i = \lambda_i \mathbf{h}_i$  ( $i = 1, 2, \dots, n$ ). Questo prova che gli  $\mathbf{h}_i$  sono autovettori e i  $\lambda_i$  sono gli autovalori corrispondenti. Poiché  $\mathbf{S}$  è una matrice non singolare, i vettori  $\mathbf{h}_i$  sono indipendenti, e quindi sono effettivamente una base di  $\mathbb{K}^n$  di autovettori.

Viceversa, supponiamo ora che esista una base di  $\mathbb{K}^n$  di autovettori di  $\mathbf{A}$ ,  $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n$ , e siano  $\lambda_1, \lambda_2, \dots, \lambda_n$  gli autovalori corrispondenti. Definendo le matrici  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  e  $\mathbf{S} = (\mathbf{h}_1 | \mathbf{h}_2 | \dots | \mathbf{h}_n)$ , le relazioni  $\mathbf{Ah}_i = \lambda_i \mathbf{h}_i$  ( $i = 1, 2, \dots, n$ ) si riscrivono  $\mathbf{AS} = \mathbf{S}\Lambda$ , cioè  $\mathbf{A} = \mathbf{S}\Lambda\mathbf{S}^{-1}$  (l'invertibilità di  $\mathbf{S}$  segue dall'ipotesi di indipendenza dei vettori  $\mathbf{h}_i$ ), perciò  $\mathbf{A}$  è diagonalizzabile.

Si noti in particolare che una condizione *necessaria* (ma non sufficiente) affinché una matrice *reale* sia diagonalizzabile su  $\mathbb{R}$  è che i suoi autovalori siano reali.



Studiamo la diagonalizzabilità o meno della matrice:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$$

Per quanto visto, i passi da seguire sono i seguenti:

1. Determiniamo gli autovalori risolvendo l'equazione caratteristica:

$$\begin{vmatrix} 1 - \lambda & 1 \\ 0 & 2 - \lambda \end{vmatrix} = (1 - \lambda)(2 - \lambda) = 0 \Rightarrow \lambda = 1, \lambda = 2.$$

2. Determiniamo gli autovettori relativi agli autovalori trovati. Ciò significa risolvere, per ciascun autovalore, il sistema omogeneo:

$$\begin{pmatrix} 1 - \lambda & 1 \\ 0 & 2 - \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

Per  $\lambda = 1$ :

$$\begin{cases} y = 0 \\ y = 0 \end{cases}$$

perciò gli autovettori sono  $(x, 0)$ ; uno di essi è  $(1, 0)$ .

Per  $\lambda = 2$ :

$$\begin{cases} -x + y = 0 \\ 0 = 0 \end{cases}$$

perciò gli autovettori sono  $(x, x)$ ; uno di essi è  $(1, 1)$ .

3. Osserviamo ora che i due autovettori  $(1, 0), (1, 1)$  sono indipendenti, perciò sono una base di  $\mathbb{R}^2$ . Pertanto  $\mathbf{A}$  è diagonalizzabile, con

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

calcolando anche

$$\mathbf{S}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

si può scrivere la relazione  $\mathbf{A} = \mathbf{SAS}^{-1}$ :

$$\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

Abbiamo così diagonalizzato la matrice  $\mathbf{A}$ .

L'ipotesi del teorema precedente non è sempre di verifica agevole, specialmente se le dimensioni della matrice non sono piccole. Vediamo quindi di comprendere più in dettaglio quali sono le situazioni in cui una matrice è o non è diagonalizzabile. Diamo anzitutto qualche definizione.

**DEFINIZIONE 8.18 (MOLTEPLICITÀ ALGEBRICA)** Si dice che un autovalore  $\lambda_0$  ha *molteplicità algebrica*  $m$  se  $\lambda_0$  è una radice del polinomio caratteristico  $D(\lambda)$  di molteplicità  $m$ , cioè se  $D(\lambda)$  è divisibile per  $(\lambda - \lambda_0)^m$  ma non per  $(\lambda - \lambda_0)^{m+1}$ .

Siano  $\lambda_1, \lambda_2, \dots, \lambda_k$  ( $k \leq n$ ) gli autovalori distinti di  $\mathbf{A}$  (reali o complessi), di molteplicità  $m_1, m_2, \dots, m_k$ , rispettivamente. Per il Teorema fondamentale dell'algebra,  $\sum_{i=1}^k m_i = n$ .

L'insieme degli autovettori corrispondenti all'autovalore  $\lambda_j$  non è altro che il nucleo della trasformazione  $\mathbf{A} - \lambda_j \mathbf{I}$ . Più precisamente, poiché per definizione gli autovettori devono essere non nulli, diciamo che l'insieme degli autovettori corrispondenti all'autovalore  $\lambda_j$ , con l'aggiunta del vettore  $\mathbf{0}$ , coincide con il nucleo della trasformazione  $\mathbf{A} - \lambda_j \mathbf{I}$ .

**DEFINIZIONE 8.19 (AUTOSPAZIO, MOLTEPLICITÀ GEOMETRICA)** Lo spazio vettoriale  $\text{Ker}(\mathbf{A} - \lambda_j \mathbf{I})$  viene detto *autospazio* associato all'autovalore  $\lambda_j$  e la sua dimensione  $d_j$  è detta *molteplicità geometrica* dell'autovalore  $\lambda_j$ .

Valgono i seguenti fatti, che non dimostriamo:

**TEOREMA 8.22**

- Autovettori relativi ad autovalori distinti sono linearmente indipendenti.
- La molteplicità geometrica di un autovalore non supera la sua molteplicità algebrica:  $d_j \leq m_j$ .

**DEFINIZIONE 8.20** Un autovalore  $\lambda_j$  si dice *regolare* se  $d_j = m_j$ .

Se un autovalore è semplice, cioè ha molteplicità algebrica 1, anche  $d_j$  dovrà essere uguale a 1, perciò un autovalore semplice è sempre regolare. La distinzione tra autovalori regolari o meno ha quindi senso per gli autovalori di molteplicità algebrica  $> 1$ .

Arriviamo così al seguente:

**TEOREMA 8.23** La matrice  $\mathbf{A}$  a elementi in  $\mathbb{K}$  è diagonalizzabile su  $\mathbb{K}$  se e solo se i suoi autovalori sono tutti regolari. (Se  $\mathbb{K} = \mathbb{R}$ : “se e solo se i suoi autovalori sono tutti reali e regolari”).

**DIMOSTRAZIONE.** Infatti se la matrice è diagonalizzabile,  $\mathbb{K}^n$  ha una base di autovettori di  $\mathbf{A}$ . Ora, se  $\lambda_1, \lambda_2, \dots, \lambda_k$  sono gli autovalori distinti di  $\mathbf{A}$  (se  $\mathbb{K} = \mathbb{R}$  reali perché  $\mathbf{A}$  è diagonalizzabile), ciascun autospazio fornisce  $d_j$  autovettori indipendenti tra loro (per definizione di molteplicità geometrica), e indipendenti da quelli degli altri autopazi (per il teorema precedente, punto a). Il numero totale di autovettori indipendenti è quindi

$$\sum_{j=1}^k d_j.$$

D'altro canto  $d_j \leq m_j$  e  $\sum_{j=1}^k m_j = n$ ; perciò l'unica possibilità per avere  $n$  autovettori reali e indipendenti è che sia  $d_j = m_j$  per ogni  $j$ , quindi gli autovalori sono regolari.

Viceversa, supponiamo ora che gli autovalori  $\lambda_1, \lambda_2, \dots, \lambda_k$  siano tutti regolari, quindi ogni autospazio fornisce  $m_j$  autovettori indipendenti tra loro (e indipendenti da quelli degli altri autopazi). In conclusione, il numero totale di autovettori indipendenti è  $\sum_{j=1}^k m_j = n$ , dunque  $\mathbb{K}^n$  ha una base di autovettori di  $\mathbf{A}$ .  $\diamond$

Dal teorema precedente segue in particolare che *se una matrice reale ha tutti gli autovalori reali e semplici, è diagonalizzabile su  $\mathbb{R}$ ; se ha tutti gli autovalori semplici (ma non tutti reali) è diagonalizzabile su  $\mathbb{C}$* . Infatti gli autovalori semplici sono automaticamente regolari.



Sia

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

e chiediamoci se la matrice è diagonalizzabile.

L'equazione caratteristica è:

$$\lambda^2 + 1 = 0, \lambda = \pm i$$

quindi non esistono autovalori reali. Pertanto  $\mathbf{A}$  non è diagonalizzabile su  $\mathbb{R}$ . D'altro canto gli autovalori, in  $\mathbb{C}$ , sono semplici, quindi la matrice è diagonalizzabile su  $\mathbb{C}$ .



Sia

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Equazione caratteristica:

$$(\lambda - 1)^2 = 0 \Rightarrow \lambda = 1 \text{ autovalore doppio.}$$

Cercando gli autovettori si trova:

$$\begin{pmatrix} x \\ 0 \end{pmatrix}$$

pertanto l'autospazio relativo a  $\lambda = 1$  ha dimensione 1, l'autovalore non è regolare, e la matrice non è diagonalizzabile.

Sia

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Equazione caratteristica:

$$(2 - \lambda)^2 (1 - \lambda) = 0,$$

perciò  $\lambda = 1$  autovalore semplice (quindi regolare),  $\lambda = 2$  autovalore doppio. L'autospazio relativo a  $\lambda = 2$  si determina risolvendo il sistema:

$$\begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

che dà  $z = 0$ ; perciò l'autospazio consiste dei vettori  $(x, y, 0)$ , e ha quindi dimensione 2.  $\lambda = 2$  è regolare, e la matrice è diagonalizzabile.

Per diagonalizzarla effettivamente, occorre calcolare anche un autovettore relativo a  $\lambda = 1$ . Si trova, per esempio,  $(-3, 0, 1)$ . Quindi si può porre  $\mathbf{A} = \mathbf{SAS}^{-1}$  con:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} -3 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \mathbf{S}^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 3 \\ 0 & 1 & 0 \end{pmatrix}$$

Si noti che le ultime due colonne della matrice  $\mathbf{S}$  sono una base dell'autospazio relativo all'autovalore  $\lambda = 2$ .

• **Osservazione.** Nel caso di autovalori non regolari non è possibile trovare una matrice equivalente a quella data, che sia diagonale. È tuttavia possibile ridurre la matrice ad una forma “canonica” più semplice, detta *forma di Jordan*, ma la discussione di questo fatto ci porterebbe troppo lontano.

#### Matrici reali simmetriche. Matrici ortogonali

Una classe particolare di matrici reali che, come vedremo, risultano sempre diagonalizzabili, è quella delle *matrici simmetriche*. Per discutere questo risultato dobbiamo però prima introdurre un'altra classe di matrici che giocano un ruolo importante in questo problema, quella delle *matrici ortogonali*.

**DEFINIZIONE 8.21** Una matrice  $\mathbf{A}$  reale di tipo  $(n, n)$  si dice *ortogonale* se:

$$\mathbf{A} \cdot \mathbf{A}^T = \mathbf{A}^T \cdot \mathbf{A} = \mathbf{I}$$

ossia se  $\mathbf{A}$  è una matrice non singolare la cui inversa coincide con la trasposta.

Ricordiamo che la trasposta  $\mathbf{A}^T$  di una matrice  $\mathbf{A}$  è quella che si ottiene scambiando le righe con le colonne. Perciò se  $\mathbf{A} = (a_{ij})$ ,  $\mathbf{A}^T = (a_{ji})$ . Inoltre, se  $\mathbf{A}, \mathbf{B}$  sono due matrici  $(n, n)$ , vale la proprietà  $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$ .

#### TEOREMA 8.24 (PROPRIETÀ DELLE MATRICI ORTOGONALI)

a.  $\mathbf{A}$  è una matrice ortogonale di tipo  $(n, n)$  se e solo se le sue righe sono una base ortonormale di  $\mathbb{R}^n$ ; lo stesso vale per le colonne.

- b. Il determinante di una matrice ortogonale vale 1 o -1.
- c. Il prodotto di due matrici ortogonali è ortogonale.
- d. La trasformazione lineare rappresentata da una matrice ortogonale conserva il modulo dei vettori:  $|\mathbf{Ax}| = |\mathbf{x}|$  per ogni  $\mathbf{x} \in \mathbb{R}^n$ , e conserva il loro prodotto scalare:  $\mathbf{Ax} \cdot \mathbf{Ay} = \mathbf{x} \cdot \mathbf{y}$ .

#### DIMOSTRAZIONE.

a. Poniamo

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{pmatrix}$$

La relazione  $\mathbf{A} \cdot \mathbf{A}^T = \mathbf{I}_n$  dà, ragionando sugli elementi:

$$\sum_{k=1}^n a_{ik} a_{jk} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

ossia

$$\mathbf{a}_i \cdot \mathbf{a}_j = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

il che significa che i vettori riga di  $\mathbf{A}$  sono ortonormali (e quindi sono una base, essendo in numero  $n$ ). Analogamente, ponendo

$$\mathbf{A} = (\mathbf{a}'_1 | \mathbf{a}'_2 | \dots | \mathbf{a}'_n)$$

la relazione  $\mathbf{A}^T \cdot \mathbf{A} = \mathbf{I}_n$  dà

$$\mathbf{a}'_i \cdot \mathbf{a}'_j = \sum_{k=1}^n a_{ki} a_{kj} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

e quindi anche le colonne di  $\mathbf{A}$  sono ortonormali. Ripercorrendo i passaggi all'inverso, si vede che se, viceversa, le righe (rispettivamente, le colonne) di  $\mathbf{A}$  sono una base ortonormale di  $\mathbb{R}^n$ , allora  $\mathbf{A} \cdot \mathbf{A}^T = \mathbf{I}_n$  (rispettivamente,  $\mathbf{A}^T \cdot \mathbf{A} = \mathbf{I}_n$ ).

b. Dalla relazione  $\mathbf{A}^T \cdot \mathbf{A} = \mathbf{I}_n$ , per il teorema di Binet si ha

$$\det(\mathbf{A}^T) \cdot \det(\mathbf{A}) = 1.$$

Ricordando che  $\det(\mathbf{A}^T) = \det(\mathbf{A})$  si ha  $\det^2(\mathbf{A}) = 1$ ,  $\det(\mathbf{A}) = \pm 1$ .

c. Siano  $\mathbf{A}, \mathbf{B}$  ortogonali.

$$(\mathbf{AB}) \cdot (\mathbf{AB})^T = (\mathbf{AB}) \cdot (\mathbf{B}^T \mathbf{A}^T) = \mathbf{A} (\mathbf{B} \mathbf{B}^T) \mathbf{A}^T = \mathbf{A} \mathbf{I}_n \mathbf{A}^T = \mathbf{AA}^T = \mathbf{I}_n.$$

Analogamente si prova che  $(\mathbf{AB})^T \cdot (\mathbf{AB}) = \mathbf{I}_n$ , perciò  $\mathbf{AB}$  è ortogonale.

d. Con le notazioni del prodotto righe per colonne, il prodotto scalare  $\mathbf{x} \cdot \mathbf{y}$  è  $\mathbf{x}^T \mathbf{y}$ , quello dei trasformati è

$$(\mathbf{Ax})^T \cdot (\mathbf{Ay}) = \mathbf{x}^T \mathbf{A}^T \mathbf{Ay} = \mathbf{x}^T \mathbf{I}_n \mathbf{y} = \mathbf{x}^T \mathbf{y}$$

quindi il prodotto scalare si conserva; in particolare allora si conserva il modulo dei vettori:

$$|\mathbf{Ax}|^2 = (\mathbf{Ax})^T (\mathbf{Ax}) = \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} = |\mathbf{x}|^2.$$

Se  $n = 2$  o  $3$ , la conservazione del prodotto scalare ha il significato geometrico di conservazione degli angoli e delle lunghezze. Geometricamente, le trasformazioni lineari che conservano angoli e lunghezze sono le rotazioni rispetto all'origine, eventualmente accompagnate da una riflessione (rispetto a una retta, nel piano; rispetto a un piano, nello spazio)<sup>5</sup>. Precisamente, si osserva che le trasformazioni lineari che rappresentano una rotazione rispetto all'origine senza alcuna riflessione sono rappresentate da matrici ortogonali aventi determinante  $1$ ; queste matrici si chiamano anche *ortogonalmente speciali*; noi le chiameremo semplicemente *matrici di rotazione*. Invece, le trasformazioni lineari che rappresentano una rotazione rispetto all'origine accompagnata da un riflessione sono rappresentate da matrici ortogonali aventi determinante  $-1$ .

Per le matrici ortogonali  $n \times n$  adotteremo la stessa terminologia, continuando a pensare come “rotazioni in  $\mathbb{R}^n$ ” le trasformazioni ortogonali con determinante  $1$ , e come “rotazioni in  $\mathbb{R}^n$  accompagnate da una riflessione rispetto a un iperpiano” quelle di determinante  $-1$ .

Si noti che, in base al punto  $a$  del teorema precedente, scambiando tra loro due colonne (o due righe) di una matrice ortogonale  $n \times n$  si ottiene una matrice ancora ortogonale, che però avrà determinante di segno opposto. In particolare, data una matrice ortogonale avente determinante  $-1$ , è sufficiente scambiare tra loro due righe o due colonne per ottenere una matrice di rotazione. Quest'ultimo fatto ci sarà utile nelle applicazioni geometriche della diagonalizzazione, che vedremo nel prossimo capitolo.

Possiamo ora enunciare l'importante

**TEOREMA 8.25** *Sia  $\mathbf{A}$  una matrice reale e simmetrica. Allora  $\mathbf{A}$  è diagonalizzabile, con una matrice di passaggio ortogonale, ossia:*

$$\mathbf{A} = \mathbf{M} \Delta \mathbf{M}^T$$

con  $\Delta$  diagonale e  $\mathbf{M}$  ortogonale. In particolare, ciò significa che  $\mathbb{R}^n$  possiede una base ortonormale di autovettori di  $\mathbf{A}$ .

Ricordiamo esplicitamente come si costruiscono le matrici  $\mathbf{M}$  e  $\Delta$ :

1. Si calcolano gli autovalori di  $\mathbf{A}$ ,  $\lambda_1, \lambda_2, \dots, \lambda_n$  e  $n$  autovettori corrispondenti; questi risultano a due a due ortogonali.
2. Si normalizzano gli  $n$  autovettori, ottenendo  $\mathbf{h}_1, \dots, \mathbf{h}_n$  vettori colonna ortonormali.
3. Si ha allora

$$\mathbf{M} = (\mathbf{h}_1 | \mathbf{h}_2 | \mathbf{h}_3 | \dots | \mathbf{h}_n) \text{ e } \Delta = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n).$$

<sup>5</sup>Anche le *traslazioni* conservano angoli e lunghezze, ma non sono trasformazioni *lineari*, quindi non sono rappresentate dal prodotto per una matrice.

4. Nel caso risultasse  $\det \mathbf{M} = -1$ , per ottenere una matrice  $\mathbf{M}$  di rotazione è sufficiente scambiare tra loro (ad esempio) i primi due autovettori e i primi due autovalori, ossia porre

$$\mathbf{M} = (\mathbf{h}_2 | \mathbf{h}_1 | \mathbf{h}_3 | \dots | \mathbf{h}_n) \text{ e } \Delta = \text{diag}(\lambda_2, \lambda_1, \lambda_3, \dots, \lambda_n).$$



A titolo d'esempio, dimostriamo il teorema per  $n = 2$ . La generica matrice simmetrica di tipo  $(2, 2)$  è

$$\mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

I suoi autovalori sono dati da:

$$\begin{aligned} \begin{vmatrix} a - \lambda & b \\ b & c - \lambda \end{vmatrix} &= \lambda^2 - (a + c)\lambda + ac - b^2 = 0 \\ \lambda &= \frac{(a + c) \pm \sqrt{(a + c)^2 - 4(ac - b^2)}}{2} = \frac{(a + c) \pm \sqrt{(a - c)^2 + 4b^2}}{2} \end{aligned}$$

Come si vede, il discriminante è sempre  $\geq 0$ , il che implica autovalori reali. In particolare, il discriminante si annulla se e solo se

$$a = c \text{ e } b = 0$$

ossia se e solo se  $\mathbf{A}$  è già un matrice diagonale. Altrimenti si hanno 2 autovalori reali e distinti, quindi semplici, quindi regolari, e perciò la matrice è diagonalizzabile.

Mostriamo che è possibile diagonalizzarla con una matrice ortogonale. Siano  $\lambda_1, \lambda_2$  i due autovalori, e  $\mathbf{v}_1, \mathbf{v}_2$  due autovettori corrispondenti, rispettivamente. Mostriamo che  $\mathbf{v}_1$  e  $\mathbf{v}_2$  sono ortogonali. Infatti:

$$\mathbf{v}_1 \cdot (\mathbf{A} \mathbf{v}_2) = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2 \quad \mathbf{v}_2 \cdot (\mathbf{A} \mathbf{v}_1) = \lambda_1 \mathbf{v}_2 \cdot \mathbf{v}_1$$

D'altro canto la simmetria della matrice  $\mathbf{A}$  implica che  $\mathbf{v}_1 \cdot (\mathbf{A} \mathbf{v}_2) = \mathbf{v}_2 \cdot (\mathbf{A} \mathbf{v}_1)$ , infatti:

$$\mathbf{v}_1 \cdot (\mathbf{A} \mathbf{v}_2) = \sum_{i=1}^2 v_1^i \sum_{j=1}^2 a_{ij} v_2^j = \sum_{i=1}^2 v_1^i \sum_{j=1}^2 a_{ji} v_2^j = \sum_{j=1}^2 v_2^j \sum_{i=1}^2 a_{ji} v_1^i = \mathbf{v}_2 \cdot (\mathbf{A} \mathbf{v}_1)$$

Ne segue che  $\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2$ , ed essendo  $\lambda_1 \neq \lambda_2$  dev'essere  $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ . Basta ora normalizzare i due autovettori per averne due ortonormali; la matrice che ha questi come colonne è quindi una matrice ortogonale, che diagonalizza  $\mathbf{A}$ .



Diagonalizziamo la matrice

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Gli autovalori sono dati da:

$$\lambda^2 - 2\lambda = 0$$

cioè  $\lambda = 0, \lambda = 2$ . Gli autovettori relativi a  $\lambda = 0$  sono dati da  $x + y = 0$ , cioè sono del tipo  $(x, -x)$ , per esempio  $(1, -1)$ . Gli autovettori relativi a  $\lambda = 2$  sono dati da  $-x + y = 0$ , per esempio  $(1, 1)$ . Possiamo quindi scrivere:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}^{-1}$$

Si noti che i due autovettori scelti sono ortogonali ma non sono normalizzati, quindi la matrice  $S$  qui utilizzata non è ortogonale. Normalizzando gli autovettori, otteniamo una matrice di passaggio ortogonale:

$$S = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

che in questo caso rappresenta una rotazione di  $-\frac{\pi}{4}$  nel piano. La sua inversa è la matrice trasposta, perciò possiamo scrivere:

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

**ESEMPIO** Matrice d'inerzia e assi principali d'inerzia. Consideriamo la matrice d'inerzia di un corpo rigido. (v. esempio 4.7). Essendo una matrice reale e simmetrica, è senz'altro diagonalizzabile, con matrice di passaggio ortogonale. Ciò significa che esiste una terna di assi cartesiani (ruotata rispetto agli assi  $x, y, z$  originali) rispetto ai quali la matrice d'inerzia  $J$  è diagonale. In questo sistema di riferimento, la relazione tra velocità angolare  $\omega'$  e momento della quantità di moto  $L'$  è semplicemente:

$$L'_i = J_{ii}\omega'_i \quad \text{con } i = 1, 2, 3$$

Le direzioni di questi nuovi assi, ossia le direzioni degli autovettori della matrice, si dicono *assi principali d'inerzia*; gli autovalori corrispondenti sono i *momenti principali d'inerzia*. La descrizione dinamica del sistema è più semplice se come assi coordinati si scelgono gli assi principali d'inerzia. Se il corpo ha particolari simmetrie, questi assi si possono individuare con considerazioni geometriche, senza bisogno di diagonalizzare la matrice.

### Rotazioni dello spazio

Abbiamo visto che una matrice ortogonale  $(2, 2)$  o  $(3, 3)$  rappresenta geometricamente una rotazione nel piano o nello spazio, eventualmente accompagnata da una riflessione. Delle rotazioni nel piano abbiamo già detto nell'esempio 4.3. Soffermiamoci ora sul caso tridimensionale. Dal punto di vista geometrico, una rotazione nello spazio (senza riflessione) è determinata da un *asse di rotazione*, e da un *angolo* di rotazione attorno a questo asse. Possiamo per esempio individuare una rotazione con un vettore  $\alpha n$ , dove  $n$  è un versore che individua l'asse di rotazione e  $\alpha$  l'angolo di rotazione antioraria rispetto all'asse orientato come  $n$ . Quindi una rotazione è individuata da 3 parametri indipendenti. D'altro canto, dal punto di vista algebrico, una rotazione è individuata da una matrice  $(3, 3)$  ortogonale. Le condizioni di ortogonalità (v. teorema 6.5) implicano 6 relazioni indipendenti tra i 9 coefficienti della matrice, e quindi i parametri indipendenti che individuano la matrice sono 3. Chiediamoci come si legge, dalla matrice, l'informazione geometrica su quale sia l'asse e l'angolo di rotazione. A questo risponde il prossimo:

**TEOREMA 8.26 (TEOREMA DI EULERO)** Una matrice  $M$  ortogonale  $(3, 3)$  avente determinante 1 ha sempre un autovalore uguale ad 1 con un autovettore reale, e due autovalori complessi coniugati di modulo unitario,  $e^{i\alpha}, e^{-i\alpha}$ , per qualche  $\alpha$ . Il versore

**n** dell'autovettore reale indica l'asse di rotazione, mentre il numero  $\alpha$  indica l'angolo. L'angolo  $\alpha$  si può anche calcolare direttamente dalla relazione:

$$\text{Tr } M = 1 + 2 \cos \alpha$$

dove  $\text{Tr } M$  (traccia di  $M$ ) indica la somma degli elementi di  $M$  posti sulla diagonale principale

### ESEMPIO

**ESEMPIO** Le rotazioni di angolo  $\alpha$  attorno agli assi  $x, y, z$  sono date rispettivamente dalle matrici:

$$R_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{pmatrix} \quad R_y(\alpha) = \begin{pmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{pmatrix}$$

$$R_z(\alpha) = \begin{pmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Il lettore verifichi su queste matrici le affermazioni del Teorema di Eulero.

### ESEMPIO

**ESEMPIO** Calcolare gli autovalori (reali o complessi) e gli autovettori corrispondenti per le seguenti matrici:

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 2 \\ -4 & 2 \end{pmatrix} \quad C = \begin{pmatrix} 2 & -3 \\ 3 & 1 \end{pmatrix}$$

Diagonalizzare poi la matrice  $A$ .

### ESEMPIO

Trovare autovalori e autovettori delle seguenti matrici:

$$\text{i)} \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{ii)} \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Lo studente può immaginare (e poi cercare di dimostrare) quali siano gli autovalori e i corrispondenti autovettori della matrice  $n \times n$  con tutti gli elementi uguali a 1.

### ESEMPIO

Determinare autovalori e autovettori della seguente matrice; se è possibile, diagonalizzarla:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

### ESEMPIO

Dire se le seguenti matrici hanno autovalori regolari o no:

$$A = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

 Determinare i valori del parametro  $k$  per i quali la matrice:

$$\mathbf{A} = \begin{pmatrix} 1-k & 0 & 0 \\ -1 & 2 & -1 \\ -k & 2 & -1 \end{pmatrix}$$

ammette come autovettore  $x = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ . Scelto uno di questi valori di  $k$ , dire se la matrice corrispondente è diagonalizzabile.

 Determinare una base ortonormale di autovettori della matrice

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

# 9

## Coniche e quadriche

Nel capitolo precedente abbiamo introdotto alcuni concetti e strumenti fondamentali dell'algebra lineare e abbiamo sviluppato i primi elementi della geometria analitica dello spazio limitandoci alla geometria degli enti lineari (rette e piani). Naturalmente, molti degli aspetti più interessanti e utili della geometria riguardano gli oggetti geometrici curvi: linee nel piano e superfici nello spazio. Dal punto di vista algebrico, i più semplici oggetti geometrici curvi che è naturale considerare sono quelli descritti da equazioni algebriche di secondo grado in due variabili (che rappresentano particolari curve nel piano) o in tre variabili (che rappresentano superfici nello spazio). Queste curve e superfici prendono il nome di *coniche* e *quadriche*, rispettivamente, e costituiscono l'oggetto di questo capitolo. Come vedremo, nello studio delle proprietà geometriche di questi oggetti, l'algebra lineare studiata nel capitolo precedente risulta essere uno strumento molto potente. Nel volume di Analisi Matematica 2 lo studente vedrà invece come gli strumenti del calcolo differenziale si possano impiegare per uno studio di curve e superfici più generali di queste.

### 1 LE CONICHE

Le coniche sono, dopo la circonferenza, le prime curve piane che storicamente sono state studiate, non solo per la loro semplicità, ma perché fanno naturalmente la loro comparsa in moltissime applicazioni geometriche, meccaniche, ecc. Nella matematica greca erano studiate coi metodi della geometria euclidea (sintetica), del piano e dello spazio. Già note a Euclide e Archimede, intorno al 225 a.C. Apollonio di Perga scrisse un trattato sulle coniche che rimase famoso per secoli; nel '600, Cartesio e Fermat iniziarono invece lo studio di queste curve con i metodi della geometria analitica. Noi discuteremo brevemente le coniche dal punto di vista della geometria elementare, per poi concentrarci sul punto di vista della geometria analitica.

#### 1.1 Le coniche dal punto di vista della geometria elementare e le loro equazioni canoniche

##### Le coniche come sezioni di un cono

Un primo modo per definire le coniche utilizza il punto di vista della geometria elementare dello spazio. Consideriamo anzitutto un *cono a due falde* (fig. 9.1), ovvero

la superficie, nello spazio, così definita: si considera una retta  $a$ , detta asse del cono (che pensiamo verticale), e un punto  $V$  su  $a$ , detto vertice del cono; su un piano orizzontale e non passante per  $V$ , si considera una circonferenza avente il centro su  $a$ ; il cono a due falde è la superficie formata da tutte le rette (dette generatrici del cono) che passano per  $V$  e per un punto qualsiasi della circonferenza.

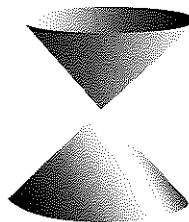


Figura 9.1. Cono a due falde.

Si dice *sezione conica* (o semplicemente *conica*) qualsiasi curva si ottenga intersecando un cono a due falde con un piano qualsiasi dello spazio, non passante per il vertice  $V$ . Precisamente:

Se il piano è meno inclinato delle generatrici (rispetto all'orizzontale), allora interseca una sola delle due falde del cono, e taglia su di esse una curva (limitata) detta *ellisse*. Se il piano è orizzontale, l'ellisse è una circonferenza.

Se il piano è parallelo a una generatrice, interseca una sola delle due falde del cono, e taglia su di esse una curva (illimitata) detta *parabola*.

Se il piano è più inclinato delle generatrici, interseca entrambe le falde del cono, e taglia su di esse una curva, illimitata e spezzata in due rami, detta *iperbole*.

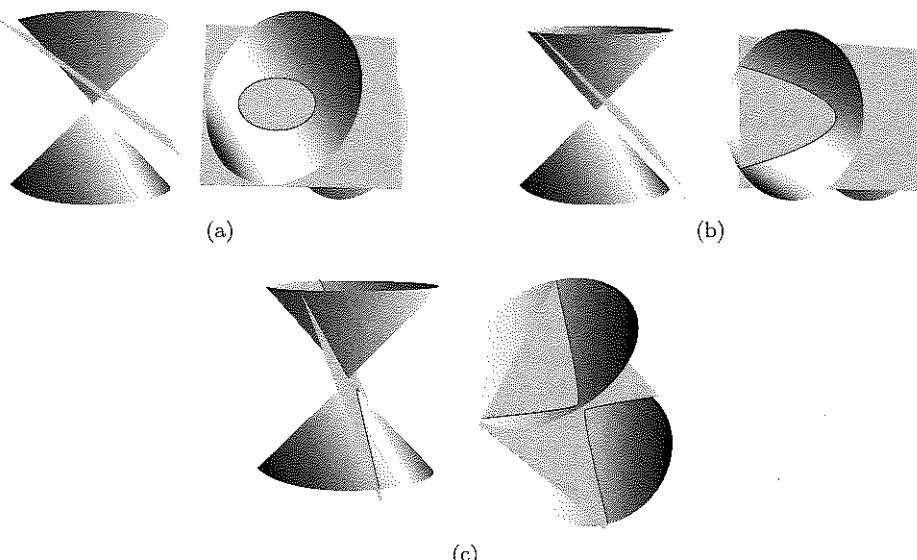


Figura 9.2. a) L'ellisse come sezione di un cono; b) la parabola come sezione di un cono; c) l'iperbole come sezione di un cono.

### Le coniche come luoghi geometrici piani e le loro equazioni canoniche

Senza fare uso di costruzioni tridimensionali, le coniche possono anche essere definite direttamente come opportuni luoghi geometrici, nel piano, nei modi seguenti.

- *Ellisse*. ellisse Siano  $F_1, F_2$  due punti nel piano, eventualmente coincidenti, e  $d$  un numero reale maggiore della distanza  $\overline{F_1 F_2}$ . Si chiama *ellisse di fuochi*  $F_1, F_2$  l'insieme dei punti  $P$  del piano tali che  $\overline{P F_1} + \overline{P F_2} = d$ , ossia tali che la somma delle loro distanze dai 2 fuochi è costante. (v. fig. 9.3)

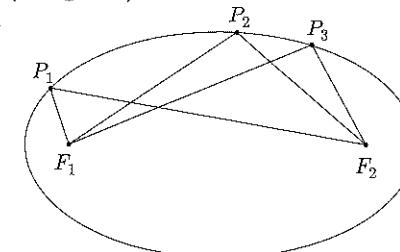


Figura 9.3. Proprietà dell'ellisse:  $\overline{F_1 P_1} + \overline{F_2 P_1} = \overline{F_1 P_2} + \overline{F_2 P_2} = \overline{F_1 P_3} + \overline{F_2 P_3}$ .

Se in particolare  $F_1 \equiv F_2$ , l'ellisse è la circonferenza di diametro  $d$  e centro  $F_1$ .

Scegliamo un sistema di riferimento cartesiano in cui sia  $F_1 \equiv (c, 0)$ ;  $F_2 \equiv (-c, 0)$ , per un certo  $c > 0$ . Se  $P \equiv (x, y)$ , imponendo la relazione  $\overline{P F_1} + \overline{P F_2} = d$  (con  $2c < d$ ) si ottiene (lasciamo i calcoli al lettore) un'equazione del tipo

$$(1.1) \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad \text{con } a, b > 0$$

con  $a > b$  detta *equazione canonica dell'ellisse*. I numeri  $a, b$  sono detti *semiassi* dell'ellisse e sono legati ai fuochi dalla relazione  $c = \sqrt{a^2 - b^2}$ , mentre  $d = 2a$ .

Una curva di equazione (1.1) con  $b > a$  rappresenta invece un'ellisse avente i fuochi sull'asse  $y$ :

$$F_1 \equiv (0, c); F_2 \equiv (0, -c), \text{ con } c = \sqrt{b^2 - a^2}, d = 2b.$$

I fuochi dell'ellisse stanno quindi sempre sull'asse maggiore. I punti  $(\pm a, 0), (0, \pm b)$  si dicono *vertici* dell'ellisse. Se  $a = b$  l'ellisse è la circonferenza di centro l'origine e raggio  $R = a = b$ .

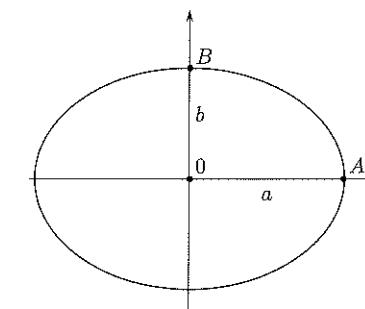


Figura 9.4. L'ellisse di equazione  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ .

- *Parabola.* parabola Sia  $F$  un punto del piano e  $r$  una retta non passante per  $F$ . Si dice *parabola di fuoco  $F$  e direttrice  $r$*  il luogo dei punti equidistanti da  $F$  e  $r$ . In altre parole: per ogni punto  $P$  della parabola, detta  $H$  la sua proiezione su  $r$ , risulta  $\overline{PH} = \overline{PF}$ . (v. fig. 9.5).

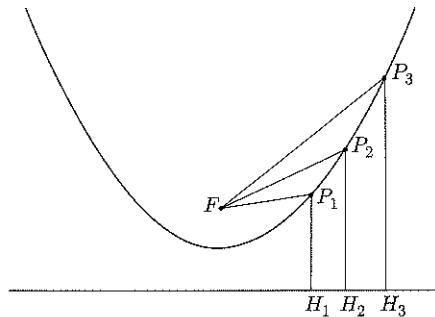


Figura 9.5. Proprietà della parabola:  $\overline{FP_1} = \overline{P_1H_1}$ ;  $\overline{FP_2} = \overline{P_2H_2}$ ;  $\overline{FP_3} = \overline{P_3H_3}$ .

Se sceglio un riferimento in cui il fuoco della parabola è  $F \equiv (a, 0)$  e la direttrice è la retta:  $x = -a$ , uguagliando le distanze di  $P(x, y)$  dal fuoco e dalla direttrice si ottiene l'equazione:

$$(1.2) \quad y^2 = 4ax,$$

*equazione canonica della parabola.* L'equazione

$$(1.3) \quad x^2 = 4ay$$

rappresenta invece la parabola di fuoco  $F \equiv (0, a)$  e direttrice la retta:  $y = -a$  (v. fig. 9.6).

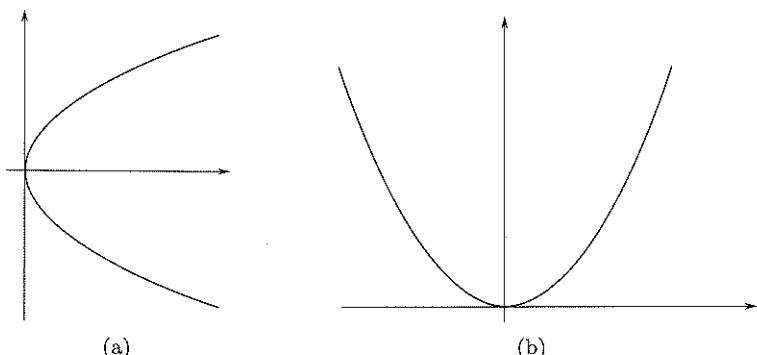


Figura 9.6. a) La parabola  $y^2 = 4ax$  con  $a > 0$ ; b) la parabola  $x^2 = 4ay$  con  $a > 0$ .

- *Iperbole.* iperbole Siano  $F_1, F_2$  due punti nel piano e  $d$  un numero reale positivo minore della distanza  $\overline{F_1F_2}$ . Si chiama *iperbole di fuochi  $F_1, F_2$*  il luogo dei punti  $P$  del piano tali che  $|\overline{PF_1} - \overline{PF_2}| = d$ , ossia tali che il modulo della differenza delle loro distanze dai 2 fuochi sia costante. (v. fig. 9.7). La relazione precedente equivale a  $\overline{PF_1} - \overline{PF_2} = \pm d$ ; ciascuna delle due equazioni (con il segno + oppure - davanti a  $d$ ) individua uno dei due *rami dell'iperbole*.

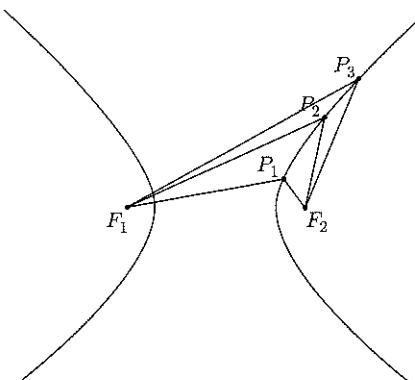


Figura 9.7. Proprietà dell'iperbole:  $\overline{F_1P_1} - \overline{F_2P_1} = \overline{F_1P_2} - \overline{F_2P_2} = \overline{F_1P_3} - \overline{F_2P_3}$ .

Se sceglio un riferimento in cui i fuochi sono i punti  $F_1 \equiv (c, 0); F_2 \equiv (-c, 0)$ , per un certo  $c > 0$ , imponendo la relazione  $|\overline{PF_1} - \overline{PF_2}| = d$  (con  $2c > d$ ) si ottiene un'equazione del tipo:

$$(1.4) \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

detta *equazione canonica dell'iperbole*, dove i coefficienti  $a, b$  sono legati ai fuochi dalla relazione

$$c = \sqrt{a^2 + b^2}.$$

I punti  $(\pm a, 0)$  sono detti *vertici* dell'iperbole; le rette  $y = \pm \frac{b}{a}x$  sono dette *asintoti* dell'iperbole.

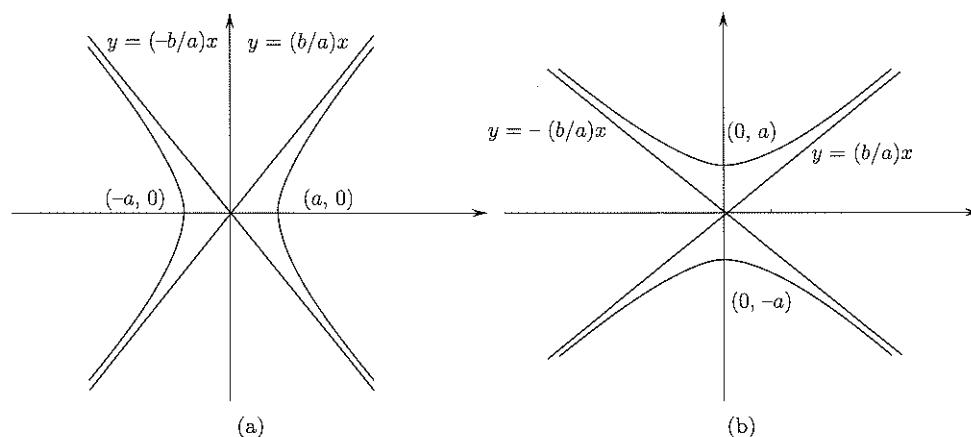
L'analogia equazione

$$(1.5) \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = -1 \quad \text{con } a, b > 0$$

rappresenta un'iperbole avente fuochi sull'asse  $y$ ,  $F_1 \equiv (0, c), F_2 \equiv (0, -c)$ , vertici  $(0, \pm b)$  e gli stessi asintoti.

Se in particolare  $a = b$ , l'iperbole si dice *equilatera*. In questo caso gli asintoti sono le rette  $y = \pm x$ , ortogonali tra loro.

Concludiamo questa prima panoramica sulle coniche ricordando che si può dimostrare che una curva è un'ellisse (rispettivamente: una parabola, un'iperbole) rispetto alla prima definizione (ossia come opportuna sezione di un cono) se e solo se lo è rispetto alla seconda definizione (ossia come luogo geometrico definito mediante opportune distanze). Le due definizioni date sono dunque equivalenti, anche se a prima vista molto diverse.

Figura 9.3. a) L'iperbole di equazione  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ ; b) L'iperbole di equazione  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1$ .

## 1.2 L'equazione generale delle coniche. Riduzione a forma canonica

La forma canonica delle coniche è quella che l'equazione di una curva di questo tipo assume in un sistema di riferimento *opportuno*, ossia scelto appositamente affinché l'equazione della conica sia la più semplice possibile. Chiediamoci ora: se cambiamo sistema di riferimento, come cambia l'equazione della conica? Prima di rispondere alla domanda è opportuno aprire una parentesi da cui apparirà meglio l'interesse della domanda stessa.

### Movimenti rigidi, trasformazioni di coordinate, proprietà geometriche

Cominciamo a ricordare che, nel piano, il più generale *movimento rigido* (cioè che non altera le mutue distanze tra i punti) consiste in una *rototraslazione* (eventualmente combinata con una riflessione rispetto a una retta): eseguire una rototraslazione di assi equivale quindi a cambiare il sistema di riferimento (cartesiano ortogonale), senza cambiare unità di misura sugli assi (ed eventualmente cambiando l'orientazione di uno dei due assi, se abbiamo eseguito una riflessione). A sua volta, sappiamo che (v. cap. 8, par. 4.1) una rotazione di angolo  $\alpha$  rispetto all'origine è espressa dalle equazioni

$$\begin{cases} x = x' \cos \alpha - y' \sin \alpha \\ y = x' \sin \alpha + y' \cos \alpha \end{cases}$$

o anche, in forma compatta matriciale, dalla moltiplicazione per una *matrice di rotazione* (v. cap. 8, par. 6.3):

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{R} \begin{pmatrix} x' \\ y' \end{pmatrix}$$

con  $\mathbf{R}$  matrice  $(2, 2)$  non singolare tale che  $\mathbf{R}\mathbf{R}^\top = \mathbf{R}^\top\mathbf{R} = \mathbf{I}$  (ossia  $\mathbf{R}$  è ortogonale) e  $\det \mathbf{R} = 1$ . Una rotazione eventualmente combinata con una riflessione rispetto a una retta è espressa da una matrice ortogonale qualsiasi (cioè avente determinante  $1$  o  $-1$ ); chiameremo *trasformazione ortogonale* una trasformazione di questo tipo.

Una *traslazione*, secondo un vettore  $(u, v)^\top$ , è espressa dalle equazioni

$$\begin{cases} x = x' + u \\ y = y' + v \end{cases}$$

Due trasformazioni ortogonali, nel piano, commutano tra loro e la loro composizione è un'altra trasformazione ortogonale; analogamente, due traslazioni commutano tra loro e la loro composizione è un'altra traslazione. Invece una traslazione e una trasformazione ortogonale non commutano tra loro. Ad esempio: ruotare di un angolo  $\alpha$  e poi traslare di un vettore  $[u, v]^\top$  si esprime mediante le equazioni

$$(1.6) \quad \begin{cases} x = x' \cos \alpha - y' \sin \alpha + u \\ y = x' \sin \alpha + y' \cos \alpha + v \end{cases}$$

mentre traslare di un vettore  $[u, v]^\top$  e poi ruotare di un angolo  $\alpha$  si esprime mediante le equazioni

$$\begin{cases} x = (x' + u) \cos \alpha - (y' + v) \sin \alpha \\ y = (x' + u) \sin \alpha + (y' + v) \cos \alpha \end{cases}$$

che comunque si possono riscrivere nella forma (1.6)

$$\begin{cases} x = x' \cos \alpha - y' \sin \alpha + u' \\ y = x' \sin \alpha + y' \cos \alpha + v' \end{cases} \text{ con } \begin{cases} u' = u \cos \alpha - v \sin \alpha \\ v' = u \sin \alpha + v \cos \alpha \end{cases}$$

Dalla discussione precedente discende la prossima

**PROPOSIZIONE 9.1** *Componendo, in qualsiasi ordine, un numero finito di traslazioni e rotazioni rispetto all'origine, nel piano, si ottiene sempre una trasformazione di equazioni (1.6), per qualche  $\alpha, u, v \in \mathbb{R}$ , detta appunto rototraslazione. In forma compatta matriciale, e ammettendo anche riflessioni rispetto a una retta per l'origine, questa assume la forma*

$$(1.7) \quad \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{R} \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} u \\ v \end{pmatrix}$$

con  $\mathbf{R}$  matrice ortogonale  $(2, 2)$  e  $[u, v]^\top \in \mathbb{R}^2$  assegnato.

Se una curva è rappresentata, rispetto ad un certo sistema di riferimento cartesiano ortogonale, da una certa equazione  $f(x, y) = 0$ , cambiando riferimento sarà rappresentata da una nuova equazione  $g(x', y') = 0$ , che si ottiene dalla precedente eseguendo sulle variabili  $(x, y)$  un'opportuna trasformazione (1.7). L'equazione è cambiata ma, geometricamente, la curva è la stessa. È naturale chiedersi se nell'equazione sia rimasta traccia del fatto che la curva è la stessa. In altre parole: c'è qualche proprietà algebrica dell'equazione  $f(x, y) = 0$  (ad esempio, qualche quantità calcolata a partire dai coefficienti che compaiono nell'equazione) che non cambia se eseguiamo una trasformazione di variabili del tipo (1.7)? Diremo che queste (eventuali) proprietà algebriche dell'equazione esprimono una *proprietà geometrica* della curva, per sottolineare il fatto che dipendono dalla curva e non dal particolare sistema di riferimento scelto per descriverla.

Anche l'operazione di moltiplicare ambo i membri dell'equazione per una costante  $\lambda \neq 0$  porta formalmente a un'equazione diversa,  $\lambda f(x, y) = 0$ , eppure non cambia ovviamente il luogo geometrico descritto. Affinché una proprietà algebrica dell'equazione esprima una proprietà geometrica della curva occorre quindi che la proprietà si conservi anche moltiplicando ambo i membri dell'equazione per una costante.

Si può dimostrare ad esempio la seguente

**PROPOSIZIONE 9.2** *Se  $f(x, y) = 0$  è un'equazione algebrica di grado  $n$  in  $x, y$  (ossia  $f$  è un polinomio di grado  $n$ ), eseguendo sulle variabili  $(x, y)$  una rototraslazione (1.6) si ottiene un'equazione  $g(x', y') = 0$  che è ancora algebrica di grado  $n$  in  $x', y'$ .*

In altre parole: *il grado  $n$  di un'equazione algebrica  $f(x, y) = 0$  esprime una proprietà geometrica della curva da essa rappresentata*. Si dice anche che il grado  $n$  è un *invariante* (cioè una quantità che non cambia eseguendo una rototraslazione sulle variabili)<sup>1</sup>.

Ha senso quindi parlare, ad esempio di “curve (e non solo di equazioni) del second'ordine”, “curve del terz'ordine”, e così via, per indicare curve piane la cui equazione è espressa da un'equazione algebrica di grado 2, 3, e così via: infatti, se questo è vero in un particolare sistema di riferimento cartesiano, sarà vero rispetto ad *ogni* sistema di riferimento cartesiano: è una proprietà *della curva*. Dimostreremo questa proposizione solo nel caso particolare  $n = 2$ , quello che qui ci interessa<sup>2</sup>.

#### Le coniche come curve algebriche del second'ordine

Consideriamo ora l'equazione canonica di un'ellisse, un'iperbole o una parabola e chiediamoci come cambia eseguendo una rototraslazione. Se nelle (1.1), (1.2), (1.3), (1.4) o (1.5) sostituiamo le (1.6), si vede subito che in generale otterremo un'equazione di secondo grado in  $x, y$  completa di tutti i termini, ossia del tipo:

$$(1.8) \quad ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

Questo significa che, in un sistema di riferimento generico (ossia: non scelto con criteri particolari) l'equazione di una conica assume la forma (1.8), ossia è *una curva del second'ordine*. La discussione precedente suggerisce allora la seguente

**DEFINIZIONE 9.1 (CONICHE COME CURVE DEL SECOND'ORDINE)** Si dice *conica* una qualunque curva piana del second'ordine, ossia espressa da un'equazione del tipo (1.8) in cui i coefficienti  $a, b, c$  non siano tutti nulli (altrimenti la curva è del prim'ordine, cioè una retta).

Per ora sappiamo che le coniche che già conosciamo (ossia di cui abbiamo scritto le equazioni canoniche) si rappresentano in qualsiasi sistema di riferimento come curve del second'ordine, quindi sono coniche anche nel senso di quest'ultima definizione. Non è chiaro se valga il viceversa: qualsiasi curva del second'ordine è un'ellisse, iperbole o

<sup>1</sup>Nel caso particolare  $n = 2$  la scoperta di questa proprietà è sostanzialmente contenuta nell'opera di Fermat del 1637 che, insieme alla contemporanea *Geometria* di Cartesio, segnò l'inizio della geometria analitica.

<sup>2</sup>Si veda la dimostrazione del Teorema 9.1.

parabola nel senso elementare del termine? I prossimi esempi mostrano che non è così, e quindi l'ultima definizione data appare più ampia delle definizioni del paragrafo 1.1:



Consideriamo le seguenti equazioni, tutte di tipo (1.8):

- |                     |                           |
|---------------------|---------------------------|
| (a) $x^2 - y^2 = 0$ | (b) $x^2 = 0$             |
| (c) $x^2 + y^2 = 0$ | (d) $x^2 + y^2 + 1 = 0$ . |

L'equazione (a) è equivalente a  $(x+y)(x-y) = 0$ , perciò rappresenta le due rette  $y = x$ ,  $y = -x$ ; la (b) rappresenta una sola retta:  $x = 0$ ; la (c) individua un solo punto, l'origine; infine, l'equazione (d) non è mai verificata, perciò individua l'insieme vuoto.

Inoltre, per quanto ne sappiamo fin qui, potrebbe restare il dubbio che esistano curve del second'ordine che hanno come immagine geometrica una curva effettiva ma di un tipo diverso dalle ellissi, parabole e iperboli che già conosciamo. Data un'equazione di tipo (1.8), occorrono dei criteri per decidere anzitutto se essa rappresenti effettivamente una curva nel senso intuitivo del termine (si dirà: *una conica reale non degenere*), e in secondo luogo di che tipo di curva si tratti. I problemi sono quindi: *riconoscere il tipo di curva* a partire dall'equazione, e saperla *ridurre a forma canonica*, cioè trovare un sistema di riferimento in cui abbia un'equazione semplice, che ci permetta di tracciarla.

Entrambi i problemi si affrontano utilizzando il *calcolo matriciale*. Cominciamo col chiederci come si trasforma un'equazione (1.8) sotto l'azione di una rototraslazione e che cosa eventualmente resta invariato.

In vista dell'utilizzo del calcolo matriciale, è più comodo riscrivere l'equazione (1.8) nella forma

$$(1.9) \quad a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0.$$

Introducendo la matrice  $(3, 3)$  simmetrica

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

l'equazione (1.9) si riscrive nella forma matriciale seguente:

$$(1.10) \quad (x, y, 1) \mathbf{A} (x, y, 1)^T = 0$$

o anche

$$\mathbf{x} \mathbf{A} \mathbf{x}^T = 0$$

avendo posto

$$\mathbf{x} = (x, y, 1).$$

Poniamo anche

$$\mathbf{B} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}.$$

Le matrici  $\mathbf{A}, \mathbf{B}$  si dicono *matrice dei coefficienti della conica* e *matrice dei termini di secondo grado della conica*, rispettivamente.

Supponiamo ora di eseguire una rototraslazione d'assi, di tipo (1.6); la nuova equazione, nelle variabili  $(x', y')$ , può essere riscritta nella forma (1.9), con nuovi coefficienti  $a'_{ij}$ , che raccogliamo in una nuova matrice  $\mathbf{A}'$ . Si dimostra che, anche se  $\mathbf{A}'$  in generale può avere tutti gli elementi diversi dai corrispondenti elementi di  $\mathbf{A}$ , alcune quantità calcolate su questa matrice rimangono invariate (si dicono perciò *invarianti*):

**TEOREMA 9.1 (INVARIANTI DI UNA CONICA)** *A seguito di una rototraslazione di assi, la conica avente matrice dei coefficienti  $\mathbf{A}$  e matrice dei termini di secondo grado  $\mathbf{B}$  si trasforma in un'altra conica e le seguenti quantità rimangono invariate:*

$$\begin{aligned} I_3 &= \det \mathbf{A} \text{ (invariante cubico);} \\ I_2 &= \det \mathbf{B} \text{ (invariante quadratico);} \\ I_1 &= \text{Tr } \mathbf{B} \text{ (invariante lineare).}^3 \end{aligned}$$

In particolare, l'annullarsi o meno di un invariante  $I_1, I_2$  o  $I_3$  esprime una proprietà geometrica della curva.

**OSSERVAZIONE** Notiamo che se moltiplichiamo ambo i membri dell'equazione (1.9) della conica per una costante  $\lambda \neq 0$ , tutti i coefficienti  $a_{ij}$  risultano moltiplicati per  $\lambda$  e quindi le quantità  $I_1, I_2, I_3$  si trasformano rispettivamente in  $\lambda I_1, \lambda^2 I_2, \lambda^3 I_3$ , da cui il nome di invariante lineare, quadratico, cubico, che si dà rispettivamente a  $I_1, I_2, I_3$ . Da questo fatto segue che anche il segno dell'invariante  $I_2$  (ma non quello di  $I_1$  o  $I_3$ ) esprime una proprietà geometrica della curva.

Per la dimostrazione ci servirà anche la seguente

**OSSERVAZIONE (MATRICI A BLOCCHI)** A volte è comodo rappresentare una matrice dalla struttura complessa in termini di blocchi più piccoli. Ad esempio, se

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

la scrittura

$$\mathbf{B} = \begin{pmatrix} \mathbf{A} & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix}$$

indica la matrice  $(3, 3)$  definita da

$$\mathbf{B} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(La dimensione dei vettori nulli indicati coi simboli  $\mathbf{0}, \mathbf{0}^\top$  nell'espressione di  $\mathbf{B}$  si deduce in questo caso dal fatto che nell'angolo in basso a destra c'è un elemento singolo – il numero 1 – e in quello in alto a sinistra c'è una matrice  $(2, 2)$ ). Dovendo eseguire

<sup>3</sup>Si ricordi che  $\text{Tr } \mathbf{B}$  indica la *traccia* della matrice  $\mathbf{B}$ , cioè la somma degli elementi sulla sua diagonale principale.

il prodotto di due “matrici a blocchi” espresse in termini di blocchi di dimensioni opportune, si può operare come se i blocchi fossero numeri. Ad esempio,

$$\begin{pmatrix} \mathbf{A} & \mathbf{0}^\top \\ \mathbf{0} & a \end{pmatrix} \begin{pmatrix} \mathbf{A}' & \mathbf{0}^\top \\ \mathbf{0} & a' \end{pmatrix} = \begin{pmatrix} \mathbf{A}\mathbf{A}' & \mathbf{0}^\top \\ \mathbf{0} & aa' \end{pmatrix},$$

come il lettore può verificare riscrivendo il calcolo senza far uso della notazione a blocchi. Useremo talvolta questo genere di scrittura abbreviata.

**DIMOSTRAZIONE DEL TEOREMA 9.1.** La dimostrazione per verifica diretta (cioè scrivendo esplicitamente la matrice  $\mathbf{A}'$  e i suoi invarianti) porta a calcoli molto laboriosi. Il calcolo matriciale consente di alleggerire l'argomento e renderlo più trasparente. Conviene inoltre separare l'azione della rotazione da quella della traslazione. Consideriamo quindi prima una rotazione

$$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{R} \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

Ponendo  $\mathbf{x} = (x, y, 1)$ ,  $\mathbf{x}' = (x', y', 1)$ , possiamo riscrivere (utilizzando la scrittura delle matrici a blocchi e indicando  $(0, 0)$  con  $\mathbf{0}$ ):

$$\mathbf{x}^\top = \begin{pmatrix} \mathbf{R} & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{x}'^\top, \text{ quindi } \mathbf{x} = \mathbf{x}' \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix}$$

(si ricordi che  $(\mathbf{A}\mathbf{b}^\top)^\top = \mathbf{b}\mathbf{A}^\top$ ) che, sostituendo nell'equazione matriciale della conica  $\mathbf{x}\mathbf{A}\mathbf{x}^\top = 0$  danno:

$$\mathbf{x}' \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{A} \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{x}'^\top = 0$$

da cui leggiamo che

$$(1.11) \quad \mathbf{A}' = \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{A} \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \equiv \mathcal{R}^\top \mathbf{A} \mathcal{R}$$

e in particolare

$$(1.12) \quad \mathbf{B}' = \mathbf{R}^\top \mathbf{B} \mathbf{R}.$$

Notiamo che la matrice  $(3, 3)$

$$\mathcal{R} = \begin{pmatrix} \mathbf{R} & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix}$$

è anch'essa ortogonale, perciò la (1.11) dice che  $\mathbf{A}$  e  $\mathbf{A}'$  sono *matrici simili* e in particolare hanno lo stesso determinante (v. cap. 8, par. 6.1). Anche  $\mathbf{B}$  e  $\mathbf{B}'$  sono simili per (1.12), perciò hanno lo stesso determinante e la stessa traccia. Dunque una rotazione conserva effettivamente gli invarianti  $I_1, I_2$  e  $I_3$ .

Verifichiamo ora che anche una traslazione conserva gli invarianti. Si vede facilmente che una traslazione non altera la matrice  $\mathbf{B}$  dei coefficienti dei termini di secondo grado, perciò conserva  $I_1$  e  $I_2$  (che dipendono solo dai coefficienti di  $\mathbf{B}$ ). Occorre quindi verificare che la traslazione

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} u \\ v \end{pmatrix}$$

non cambia il determinante di  $\mathbf{A}$ . Il calcolo diretto mostra che

$$(1.13) \quad \mathbf{A}' = \begin{pmatrix} a_{11} & a_{12} & a_{13} + a_{11}u + a_{12}v \\ \dots & a_{22} & a_{23} + a_{12}u + a_{22}v \\ \dots & \dots & a_{33} + a_{11}u^2 + a_{22}v^2 + 2a_{12}uv + 2a_{13}u + 2a_{13}v \end{pmatrix}$$

(gli elementi indicati con ... si deducono per la simmetria della matrice). Osserviamo ora che la matrice  $\mathbf{A}'$  si ottiene dalla matrice  $\mathbf{A}$  applicando le seguenti due operazioni, ognuna delle quali non altera il determinante (in base al Teorema 6.6 del cap. 8):

- sommare alla 3<sup>a</sup> colonna  $u$  volte la 1<sup>a</sup> più  $v$  volte la 2<sup>a</sup>;
- sommare alla 3<sup>a</sup> riga  $u$  volte la 1<sup>a</sup> più  $v$  volte la 2<sup>a</sup>.

Di conseguenza  $\det \mathbf{A} = \det \mathbf{A}'$ , e anche  $I_3$  si conserva.

A titolo di esempio, calcoliamo gli invarianti  $I_2, I_3$  per l'ellisse, la parabola e l'iperbole in forma canonica.

• *Ellisse:*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1; \mathbf{A} = \begin{pmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$I_3 = \det \mathbf{A} = -\frac{1}{a^2 b^2}; I_2 = \det \mathbf{B} = \frac{1}{a^2 b^2}.$$

• *Parabola:*

$$y = ax^2; \mathbf{A} = \begin{pmatrix} a & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix};$$

$$I_3 = \det \mathbf{A} = -a; I_2 = \det \mathbf{B} = 0.$$

• *Iperbole:*

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \pm 1; \mathbf{A} = \begin{pmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & -\frac{1}{b^2} & 0 \\ 0 & 0 & \mp 1 \end{pmatrix};$$

$$I_3 = \det \mathbf{A} = \pm \frac{1}{a^2 b^2}; I_2 = \det \mathbf{B} = -\frac{1}{a^2 b^2}.$$

Notiamo che in tutti e tre i casi risulta  $I_3 \neq 0$ ; inoltre si ha

$$I_2 \begin{cases} > 0 & \text{per l'ellisse} \\ = 0 & \text{per la parabola} \\ < 0 & \text{per l'iperbole.} \end{cases}$$

D'altro canto sappiamo che una rototraslazione e una moltiplicazione dell'equazione per una costante non nulla non alterano il segno di  $I_2$ , quindi se una curva è un'ellisse, una parabola o un'iperbole (nel senso della geometria elementare, v. par. 1.1), *in qualunque sistema di riferimento* la sua equazione ci darà  $I_2$  rispettivamente positivo, nullo o negativo. Il calcolo algebrico di questa quantità permette quindi immediatamente di riconoscere il tipo di conica. Quest'argomentazione lascia però in ombra un

punto delicato: *presuppone infatti a priori che la curva sia una di queste tre*. Ciò che ci aspetta ora è quest'ultima parte dell'indagine: a partire da una qualsiasi equazione algebrica di secondo grado, riconoscere se e quando questa rappresenta effettivamente una conica nel senso elementare del termine. Cominciamo dalla prossima

**DEFINIZIONE 9.2** Si dice che una conica (cioè una curva del second'ordine) è *degenera*, o *riducibile*, quando la sua equazione si può riscrivere nella forma

$$(ax + by + c)(a'x + b'y + c') = 0$$

per certi  $a, b, c, a', b', c'$  reali o complessi.

Vale il seguente teorema che ci limitiamo ad enunciare:

**TEOREMA 9.2** Una conica è degenera se e solo se si annulla il suo invarianto  $I_3$ , cioè il determinante della matrice dei coefficienti. In questo caso il luogo geometrico descritto nel piano dall'equazione può essere una coppia di rette, oppure una retta, oppure un punto.

Ci concentriamo ora sulle coniche non degeneri, per le quali vogliamo ottenere una caratterizzazione completa in base ai loro invarianti.

**TEOREMA 9.3 (RICONOSCIMENTO DI UNA CONICA NON DEGENERE)** Consideriamo una conica non degenera, cioè con  $I_3 \neq 0$ . Allora:

Se  $I_2 > 0$  esiste una rototraslazione che la riduce alla forma canonica di un'ellisse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \text{ oppure immaginaria } \frac{x^2}{a^2} + \frac{y^2}{b^2} = -1.$$

Se  $I_2 < 0$  esiste una rototraslazione che la riduce alla forma canonica di una iperbole

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \text{ oppure } \frac{x^2}{a^2} - \frac{y^2}{b^2} = -1.$$

Se  $I_2 = 0$  esiste una rototraslazione che la riduce alla forma canonica di una parabola

$$y = ax^2 \text{ oppure } x = by^2.$$

Notiamo che l'enunciato del teorema insegnava a riconoscere una conica non degenera per mezzo dei suoi invarianti. La dimostrazione del teorema mostrerà come questo sia possibile costruendo effettivamente una rototraslazione che riduca la conica a forma canonica, permettendo così di riconoscerne il tipo, ma anche di tracciare effettivamente la curva. Il teorema mostra anche che le coniche non degeneri coincidono effettivamente con le coniche nel senso della geometria elementare, con un'unica eccezione: l'ellisse immaginaria, curva priva di punti reali, che tuttavia in base alla definizione algebrica di conica come curva del second'ordine risulta una conica non degenera.

**DIMOSTRAZIONE.** Per prima cosa cerchiamo una traslazione che ottenga come risultato di annullare i termini di primo grado (cioè annullare i coefficienti  $a_{13}, a_{23}$ ). Dalla (1.13) leggiamo che questo equivale a cercare  $u, v$  per cui si abbia

$$\begin{cases} a_{13} + a_{11}u + a_{12}v = 0 \\ a_{23} + a_{12}u + a_{22}v = 0. \end{cases}$$

Questo è un sistema lineare di due equazioni nelle due incognite  $u, v$ , risolubile purché sia

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \neq 0, \text{ ossia } I_2 \neq 0.$$

Cominciamo perciò a trattare questo caso. Se  $I_2 \neq 0$  esiste una traslazione che porta la conica alla forma

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + a_{33} = 0.$$

(Per semplicità continuiamo a chiamare i coefficienti  $a_{ij}$  anziché  $a'_{ij}$ ). Si osservi che questa conica ha l'origine come centro di simmetria, ossia: se il punto  $(x, y)$  soddisfa l'equazione, anche  $(-x, -y)$  (il suo simmetrico rispetto all'origine) la soddisfa. Le coniche con  $I_2 \neq 0$  si dicono perciò *coniche a centro*. Ora cerchiamo una rotazione che abbia come effetto di annullare il coefficiente  $a_{12}$ . Dalle (1.11)-(1.12) leggiamo che qualsiasi rotazione trasformerà la conica in un'altra che ha ancora i termini  $a_{13}, a_{23}$  mancanti, e la matrice  $\mathbf{B}$  si trasformerà secondo la legge

$$\mathbf{B}' = \mathbf{R}^\top \mathbf{B} \mathbf{R}.$$

Noi desideriamo che  $\mathbf{B}'$  risulti una matrice diagonale; poiché  $\mathbf{B}$  è reale simmetrica, sappiamo (Teorema 6.6) che è sempre possibile diagonalizzarla mediante una matrice di rotazione, quindi: esiste una rotazione  $\mathbf{R}$  che trasforma la conica in una di equazione<sup>4</sup> (indicando i nuovi coefficienti ancora con i simboli  $a_{ij}$  senza apici)

$$a_{11}x^2 + a_{22}y^2 + a_{33} = 0, \text{ ossia } \mathbf{A} = \text{diag}(a_{11}, a_{22}, a_{33}).$$

Poiché siamo nell'ipotesi di conica non degenere,  $I_3 \neq 0$  perciò i tre coefficienti  $a_{11}, a_{22}, a_{33}$  sono tutti diversi da zero. Distinguiamo ora i vari casi possibili in base al segno dei tre coefficienti rimasti. Possiamo sempre supporre  $a_{11} > 0$  (eventualmente cambiando tutti i segni nell'equazione), perciò i casi possibili sono:

1.  $a_{22} > 0, a_{33} < 0$ . Dividendo per  $a_{33}$  si può riscrivere l'equazione nella forma

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

che è un'*ellisse reale* in forma canonica.

2.  $a_{22} > 0, a_{33} > 0$ . Dividendo per  $a_{33}$  si può riscrivere l'equazione nella forma

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$$

che dà una curva priva di punti reali, detta *ellisse immaginaria*. Si osservi che  $I_2 > 0$  per l'ellisse sia reale che immaginaria.

3.  $a_{22} < 0, a_{33} < 0$ . Dividendo per  $a_{33}$  si può riscrivere l'equazione nella forma

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

che è un'*iperbole* in forma canonica.

<sup>4</sup>Nel seguito ci servirà saper calcolare esplicitamente la matrice di rotazione  $\mathbf{R}$ . A questo scopo è utile ricordare (v. Teorema 8.19 del cap. 8) che le sue colonne sono gli autovettori, normalizzati, relativi agli autovalori di  $\mathbf{B}$ .

4.  $a_{22} < 0, a_{33} > 0$ . Dividendo per  $a_{33}$  si può riscrivere l'equazione nella forma

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1$$

che è ancora un'*iperbole* in forma canonica. Si osservi che per entrambe le iperboli è  $I_2 < 0$ .

Consideriamo ora il caso lasciato in sospeso, in cui  $I_2 = 0$ . In questo caso si dice che la *conica non è a centro*. Cerchiamo direttamente una rotazione che annulli il coefficiente  $a_{12}$ . Ancora dalle (1.11)-(1.12) leggiamo che esiste sempre una rotazione  $\mathbf{R}$  che trasforma la matrice  $\mathbf{A}$  in modo che  $\mathbf{B}$  risulti diagonalizzata (e in questo caso non diciamo nulla su come si trasformano i termini di ordine inferiore). Dopo la rotazione la conica ha quindi una nuova matrice

$$\mathbf{B} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}.$$

Ricordiamo però che siamo nelle ipotesi  $I_2 = 0$ , perciò si ha  $a_{11}a_{22} = 0$ , quindi si annulla anche uno di questi coefficienti. Supponiamo ad esempio che sia  $a_{11} = 0$  (l'altro caso è analogo). La nuova equazione è quindi:

$$a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0, \quad \mathbf{A} = \begin{pmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

con  $a_{22} \neq 0$ , altrimenti la curva non sarebbe del second'ordine, e  $a_{13} \neq 0$ , altrimenti si avrebbe  $I_3 = 0$ , ossia la conica sarebbe degenere. Cerchiamo una traslazione che annulli gli altri due coefficienti,  $a_{23}, a_{33}$ . Con la *tecnica di completamento del quadrato* possiamo eliminare il coefficiente della  $y$ :

$$\begin{aligned} a_{22}y^2 + 2a_{23}y &= a_{22} \left( y^2 + 2\frac{a_{23}}{a_{22}}y \right) = a_{22} \left( y^2 + 2\frac{a_{23}}{a_{22}}y + \left( \frac{a_{23}}{a_{22}} \right)^2 \right) - \frac{a_{23}^2}{a_{22}} \\ &= a_{22} \left( y + \frac{a_{23}}{a_{22}} \right)^2 - \frac{a_{23}^2}{a_{22}}, \end{aligned}$$

perciò ponendo  $y' = y + \frac{a_{23}}{a_{22}}$  l'equazione diventa

$$a_{22}y'^2 + 2a_{13}x + a_{33} - \frac{a_{23}^2}{a_{22}} = 0.$$

Ora dalle identità

$$2a_{13}x + a_{33} - \frac{a_{23}^2}{a_{22}} = 2a_{13} \left( x + \frac{a_{22}a_{33} - a_{23}^2}{2a_{13}a_{22}} \right)$$

leggiamo che ponendo  $x' = x + \frac{a_{22}a_{33} - a_{23}^2}{2a_{13}a_{22}}$  l'equazione diventa

$$\begin{aligned} a_{22}y'^2 + 2a_{13}x' &= 0 \\ x' &= -\frac{a_{22}}{2a_{13}}y'^2 \end{aligned}$$

che è una parabola in forma canonica.

Si osservi che abbiamo considerato tutti i casi possibili nell'ipotesi  $I_3 \neq 0$  e abbiamo riscontrato che la conica è effettivamente un'iperbole se  $I_2 < 0$ , un'ellisse reale o immaginaria se  $I_2 > 0$ , una parabola se  $I_2 = 0$ . Il teorema è quindi completamente dimostrato. 



Riconoscere le seguenti coniche e ricondurle a forma canonica:

$$(a) \quad x^2 + 2y^2 - 4x + 3y + 1 = 0 \quad (b) \quad x^2 - 4y^2 + x - 1 = 0 \quad (c) \quad 2xy + x + 1 = 0$$

(a)  $I_3 \neq 0, I_2 > 0$ , perciò è una conica non degenera e precisamente un'ellisse. Non è immediatamente evidente se si tratti di una ellisse reale o immaginaria, ma con la tecnica di completamento dei quadrati, scriviamo

$$x^2 - 4x = (x - 2)^2 - 4$$

$$2y^2 + 3y = 2\left(y + \frac{3}{4}\right)^2 - \frac{9}{8}$$

e quindi l'equazione diventa

$$\begin{aligned} (x - 2)^2 - 4 + 2\left(y + \frac{3}{4}\right)^2 - \frac{9}{8} + 1 &= 0 \\ \frac{(x - 2)^2}{\frac{33}{8}} + \frac{\left(y + \frac{3}{4}\right)^2}{\frac{33}{16}} &= 1. \end{aligned}$$

Perciò si tratta dell'ellisse reale di centro  $(2, -\frac{3}{4})$  e semiassi  $\sqrt{\frac{33}{8}}, \frac{\sqrt{33}}{4}$ .

(b)  $I_3 \neq 0, I_2 < 0$ , perciò è una conica non degenera e precisamente un'iperbole. Riscrivendo

$$x^2 + x - 1 = \left(x + \frac{1}{2}\right)^2 - \frac{1}{4} - 1$$

l'equazione prende la forma

$$\begin{aligned} \left(x + \frac{1}{2}\right)^2 - 4y^2 &= \frac{5}{4} \\ \frac{\left(x + \frac{1}{2}\right)^2}{\left(\frac{\sqrt{5}}{2}\right)^2} - \frac{y^2}{\left(\frac{\sqrt{5}}{4}\right)^2} &= 1 \end{aligned}$$

da cui leggiamo che è un'iperbole di centro  $(-\frac{1}{2}, 0)$ , asintoti  $y = \pm \frac{\left(\frac{\sqrt{5}}{4}\right)}{\left(\frac{\sqrt{5}}{2}\right)}(x + \frac{1}{2})$ , cioè  $y = \pm (\frac{1}{2}x + \frac{1}{4})$ .

(c) Si ha  $I_3 \neq 0, I_2 < 0$ , perciò si tratta di un'iperbole, non degenera. Posto

$$\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

cerchiamo anzitutto autovalori e autovettori di  $\mathbf{B}$ . Un facile calcolo dà:

$$\lambda_1 = 1, \lambda_2 = -1$$

con autovettori relativi (ad esempio):

$$\mathbf{v}_1 = (1, 1); \mathbf{v}_2 = (1, -1).$$

Si noti che  $\mathbf{v}_1, \mathbf{v}_2$  sono ortogonali, ma non hanno modulo unitario. Normalizzandoli, cioè calcolando

$$\mathbf{v}'_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right); \mathbf{v}'_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right),$$

otteniamo la matrice di rotazione richiesta:

$$\mathbf{R} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

(si riconosce che è una rotazione di  $\frac{\pi}{4}$ ). Eseguendo dunque la rotazione:

$$\begin{cases} x = \frac{1}{\sqrt{2}}x' + \frac{1}{\sqrt{2}}y' \\ y = \frac{1}{\sqrt{2}}x' - \frac{1}{\sqrt{2}}y' \end{cases}$$

L'equazione della conica diventa:

$$x'^2 - y'^2 + \frac{1}{\sqrt{2}}x' + \frac{1}{\sqrt{2}}y' + 1 = 0.$$

Il completamento dei quadrati ci suggerisce ora la traslazione corretta da eseguire:

$$\left(x' - \frac{1}{2\sqrt{2}}\right)^2 - \frac{1}{8} - \left(y' + \frac{1}{2\sqrt{2}}\right)^2 + \frac{1}{8} + 1 = 0.$$

Ponendo quindi:

$$\begin{cases} x'' = x' - \frac{1}{2\sqrt{2}} \\ y'' = y' + \frac{1}{2\sqrt{2}} \end{cases}$$

l'equazione dell'iperbole si può riscrivere in forma canonica:

$$x''^2 - y''^2 = -1.$$

Nel Teorema 9.3 abbiamo considerato solo coniche non degeneri. Per analogia a quanto stabilito in quel teorema possiamo ora dare la seguente

**DEFINIZIONE 9.3** Consideriamo una conica degenera. Diremo ancora che la conica è una

$$\begin{cases} \text{ellisse (degenera)} \\ \text{parabola (degenera)} \\ \text{iperbole (degenera)} \end{cases} \quad \text{se } I_2 \begin{cases} > 0 \\ = 0 \\ < 0. \end{cases}$$

Si può dimostrare la seguente:

**PROPOSIZIONE 9.3 (TIPI DI CONICHE DEGENERI)** Le coniche degeneri possono essere solo dei seguenti tipi:

- un'ellisse degenera è costituita da un solo punto reale (ad es.  $x^2 + y^2 = 0$ );

- una parabola degenere è costituita da una sola retta (reale) o da due rette parallele (reali) o dall'insieme vuoto, (ad es., rispettivamente,  $(y+1)^2 = 0$ ,  $y(y+1) = 0$ ,  $y^2 + 1 = 0$ );
- un'iperbole degenere è costituita da due rette (reali) distinte e incidenti (es.  $x^2 - y^2 = 0$ ).

Per analogia con le coniche spezzate in due rette reali, nel caso dell'ellisse degenere (un punto) si dice anche che la conica si spezza in due rette complesse coniugate che si intesecano in un punto reale (ad es.,  $y = \pm ix$ , che si intersecano nell'origine) e nel caso della parabola che degenera all'insieme vuoto si dice anche che la conica si spezza in due rette parallele immaginarie (ad es.  $y = \pm i$ , che non hanno alcun punto reale). I prossimi esempi illustrano alcune di queste situazioni.

### Esempio

Consideriamo la conica

$$x^2 - 2y^2 + xy - x + y = 0.$$

La matrice dei coefficienti è:

$$\mathbf{A} = \begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -2 & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

e il calcolo mostra che  $I_3 = \det \mathbf{A} = 0$  e  $I_2 = \det \mathbf{B} < 0$ . Si tratta quindi di un'iperbole degenere che si spezza in due rette distinte. Per scomporre effettivamente il primo membro possiamo procedere così: risolviamo l'equazione come equazione di secondo grado in  $x$ , al modo seguente<sup>5</sup>:

$$\begin{aligned} x^2 - x(1-y) + (y-2y^2) &= 0 \\ x &= \frac{(1-y) \pm \sqrt{(1-y)^2 - 4(y-2y^2)}}{2} = \\ &= \frac{(1-y) \pm \sqrt{9y^2 - 6y + 1}}{2} = \frac{(1-y) \pm (3y-1)}{2} \end{aligned}$$

da cui otteniamo le due rette

$$x = y; x = 1 - 2y$$

ossia l'equazione di partenza si può riscrivere così:

$$(x-y)(x+2y-1) = 0.$$

La conica degenere in questo caso si spezza in due rette reali e distinte. Il fatto che nella risoluzione dell'equazione di secondo grado il radicando risulti un quadrato perfetto (e quindi  $x$  si esprima come funzione lineare di  $y$ ) dipende proprio dall'ipotesi  $I_3 = 0$ .

<sup>5</sup> Allo stesso risultato finale si arriverebbe risolvendo l'equazione rispetto ad  $y$ .

### La conica

$$x^2 - 4xy + 4y^2 + 2x - 4y + 1 = 0$$

ha matrice dei coefficienti

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

da cui  $I_3 = 0$  e  $I_2 = 0$ . Si tratta quindi di una parabola degenere. Risolvendo ad esempio rispetto a  $x$  si trova:

$$\begin{aligned} x^2 + 2x(1-2y) + (4y^2 - 4y + 1) &= 0 \\ x = 2y - 1 \pm \sqrt{(2y-1)^2 - (4y^2 - 4y + 1)} &= 2y - 1 \end{aligned}$$

ossia la parabola si riduce ad un'unica retta reale,  $x - 2y + 1 = 0$ . L'equazione della conica si può riscrivere come

$$(x - 2y + 1)^2 = 0.$$

### La conica

$$x^2 + 2xy + 4y^2 = 0$$

ha matrice dei coefficienti

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

da cui  $I_3 = 0$  e  $I_2 > 0$ . Si tratta quindi di un'ellisse degenere che si riduce ad un solo punto. Infatti risolvendo rispetto a  $x$  si ha

$$x = -y \pm \sqrt{y^2 - 4y^2} = y(-1 \pm i\sqrt{3})$$

dunque la conica si spezza in due rette complesse coniugate

$$(x + y + i\sqrt{3}y)(x + y - i\sqrt{3}y) = 0$$

che si intersecano nell'unico punto reale  $(0, 0)$ .

### 1.3 Coniche ed eccentricità

È possibile dare un'ulteriore caratterizzazione di ellisse, parabola e iperbole come opportuni luoghi geometrici, che rispetto alle definizioni in termini di fuochi o direttrice ricordate nel paragrafo 1.1 ha il pregio di trattare i tre tipi di coniche in modo più uniforme:

**TEOREMA 9.4 (CARATTERIZZAZIONE DELLE CONICHE MEDIANTE L'ECCENTRICITÀ)** Consideriamo, nel piano, un punto, detto fuoco, ed una retta non passante per il fuoco, detta direttrice. Il luogo dei punti  $P$  per cui il rapporto tra la distanza di  $P$  dal fuoco e la distanza di  $P$  dalla direttrice è costante è una conica; questa costante si dice eccentricità ( $e$ ) della conica. Precisamente, la conica risulta essere

un'iperbole se  $e > 1$

una parabola se  $e = 1$

un'ellisse se  $0 < e < 1$ .

Segnaliamo che questa caratterizzazione delle coniche mediante l'eccentricità presenta un'eccezione: la *circonferenza* non può essere ottenuta con questo procedimento (la circonferenza rappresenta in un certo senso il caso limite di conica con eccentricità nulla).

Come lo studente vedrà quando studierà più in generale le curve nel piano, questo teorema è alla base della rappresentazione delle coniche come *curve in forma polare*, rappresentazione molto utile, ad esempio, nella descrizione dei moti dei pianeti<sup>6</sup>.

**DIMOSTRAZIONE.** Scegliamo un riferimento cartesiano in cui, detto  $F$  il fuoco e  $d$  la direttrice, risulti, per qualche  $a > 0$ ,

$$F(a, 0) \text{ e } d : x = -a.$$

La definizione richiede che, detto  $P(x, y)$  il generico punto della conica, il rapporto tra la distanza di  $P$  da  $F$  e la distanza di  $P$  da  $d$  sia uguale a una costante  $e > 0$ , detta eccentricità della conica. Questa condizione si traduce nell'equazione:

$$\frac{\sqrt{(x-a)^2 + y^2}}{|x+a|} = e$$

da cui ricaviamo successivamente

$$\begin{aligned}\sqrt{(x-a)^2 + y^2} &= e|x+a| \\ (x-a)^2 + y^2 &= e^2(x+a)^2\end{aligned}$$

e, sviluppando e semplificando,

$$x^2(1-e^2) + y^2 - 2ax(1+e^2) + a^2(1-e^2) = 0,$$

che è la conica avente matrice dei coefficienti

$$\mathbf{A} = \begin{pmatrix} 1-e^2 & 0 & -a(1+e^2) \\ 0 & 1 & 0 \\ -a(1+e^2) & 0 & a^2(1-e^2) \end{pmatrix}$$

da cui calcoliamo

$$\begin{aligned}I_3 &= -4a^2e^2 \\ I_2 &= 1 - e^2.\end{aligned}$$

La conica ottenuta è quindi non degenere per  $a > 0, e > 0$ , e risulta un'iperbole per  $e > 1$ , un'ellisse per  $0 < e < 1$ , una parabola per  $e = 1$ .

Notiamo ancora che nel caso  $0 < e < 1$  l'ellisse non è mai una circonferenza. Per diventare una circonferenza dovrebbe infatti essere  $e = 0$ , il che però corrisponde a una conica degenere.

<sup>6</sup>Questi fatti saranno approfonditi in Analisi Matematica 2, v. cap. 2, paragrafo 3.5.



Riconoscere, ridurre a forma canonica e disegnare le seguenti coniche:

1.  $x^2 + xy + 2y^2 = 1$

2.  $x^2 + 4xy + 4y^2 + 2x = 0$

3.  $x^2 - 2y^2 + x + 1 = 0$

4.  $x^2 - 3y^2 = 0$

5.  $x^2 + 2y^2 + 4x + 4 = 0$

6.  $x^2 + 3y^2 - 2x + 3 = 0$

7.  $x^2 + y^2 + 2x - 4y = 0$

8.  $y^2 + 3x - 2y + 1 = 0$

9.  $x^2 + 4xy + 4y^2 - 3x - 6y + 2 = 0$

10.  $7x^2 + 13y^2 + 6\sqrt{3}xy - 16 = 0$

11. Scrivere l'equazione dell'ellisse di centro  $(1, -1)$ , assi paralleli agli assi coordinati, semiassi  $(2, 3)$ .

12. Scrivere l'equazione dell'ellisse di fuochi  $(0, 2), (0, -2)$  e passante per il punto  $(1, 3)$ .

13. Scrivere l'equazione dell'iperbole avente centro  $(0, 0)$ , asintoti di coefficienti angolari  $\pm \frac{1}{2}$ , un vertice in  $(2, 0)$ .

14. Scrivere l'equazione dell'iperbole avente centro  $(2, 1)$ , asintoti di coefficienti angolari  $\pm 2$ , un vertice in  $(2, 2)$ .

15. Scrivere l'equazione della parabola di vertice  $(1, 2)$ , asse parallelo all'asse  $y$ , e passante per il punto  $(3, 4)$ .

16. Scrivere l'equazione della parabola di vertice  $(1, -1)$ , asse parallelo all'asse  $x$ , e passante per l'origine.

17. Scrivere l'equazione della conica di fuoco l'origine, direttrice la retta  $x = 2$  e eccentricità 3. Determinare gli asintoti dell'iperbole.

18. Scrivere l'equazione della conica di fuoco l'origine, direttrice la retta  $y = 1$  e eccentricità 1/2. Determinare i semiassi dell'ellisse.

## 2 LE SUPERFICI QUADRICHE

Se facciamo ruotare una parabola attorno al suo asse, oppure un'ellisse o un'iperbole attorno a uno dei suoi due assi di simmetria, otteniamo superfici dall'aspetto abbastanza semplice che, prendendo il nome dalla conica generatrice, si dicono paraboloidi, ellissoide o iperboloidi di rotazione (v. fig. 9.9): questi sono gli esempi più semplici di superfici quadriche (come vedremo, ce ne sono anche altre) e sono note fin dall'antichità<sup>7</sup>.

Per dare una definizione generale di superficie quadrica e poter studiare queste superfici è naturale utilizzare fin da subito il linguaggio della geometria analitica e del calcolo matriciale. Diamo quindi la seguente

<sup>7</sup>Ad esempio Archimede dedicò a queste superfici il suo trattato *Sui conoidi e gli sferoidi*.

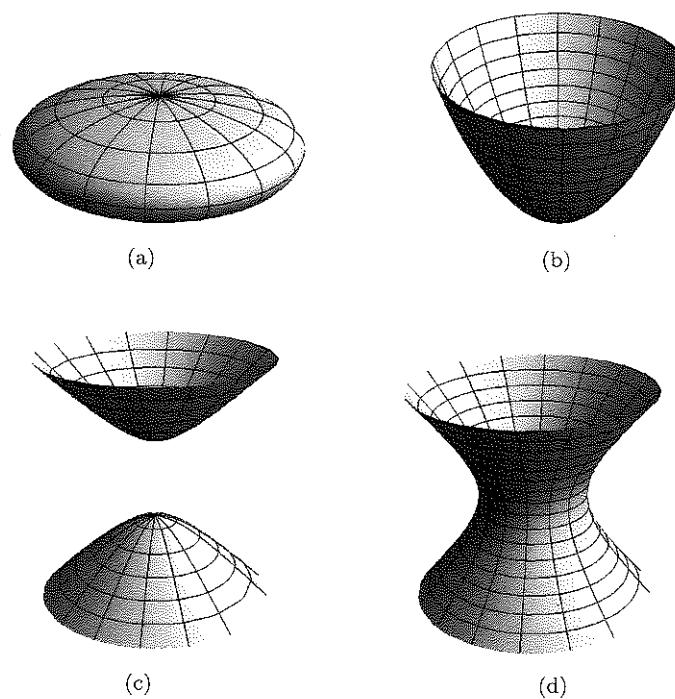


Figura 9.9. a) Ellisse di rotazione; b) paraboloide di rotazione; c) iperboloidi di rotazione; l'iperbole ruota attorno all'asse di simmetria che contiene i vertici; d) iperboloidi di rotazione; l'iperbole ruota attorno all'asse di simmetria che non contiene i vertici.

**DEFINIZIONE 9.4 (QUADRICE COME SUPERFICI DEL SECOND'ORDINE)** Si dice *superficie quadrica* (o semplicemente *quadrica*) il luogo dei punti dello spazio le cui coordinate  $(x, y, z)$  soddisfano un'equazione algebrica di secondo grado in  $x, y, z$ , del tipo:

$$(2.1) \quad a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz + 2a_{14}x + 2a_{24}y + 2a_{34}z + a_{44} = 0$$

ovvero, in forma matriciale:

$$(x, y, z, 1) \mathbf{A} (x, y, z, 1)^T = 0$$

dove

$$(2.2) \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{23} & a_{33} & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{pmatrix}$$

è la matrice reale simmetrica  $4 \times 4$  di coefficienti  $a_{ij}$ . Supponiamo inoltre che la *matrice dei termini di secondo grado*

$$(2.3) \quad \mathbf{B} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

non sia la matrice nulla (altrimenti non si tratta di una superficie del second'ordine ma del primo, cioè di un piano).

Anche se la definizione data è puramente algebrica, ci interessa studiare queste superfici dal punto di vista geometrico. Come per le coniche, le proprietà geometriche sono quelle che non dipendono dal particolare sistema di riferimento scelto, ossia sono *invarianti per rototraslazioni*. Il filo del discorso sarà quindi il seguente: dopo aver richiamato qualche proprietà delle rototraslazioni nello spazio tridimensionale e aver individuato alcuni *invarianti* delle quadriche, cercheremo una rototraslazione che porti una quadrica ad assumere un'equazione la più semplice possibile. Individueremo così facendo le *equazioni in forma canonica* dei vari tipi di quadriche, arrivando ad una *classificazione delle quadriche* stesse e al riconoscimento del loro aspetto geometrico.

## 2.1 Rototraslazioni nello spazio e invarianti di una quadrica

Abbiamo visto nel capitolo 8, paragrafo 6.3 che, nello spazio tridimensionale, una rotazione rispetto all'origine, eventualmente accompagnata da una riflessione rispetto a un piano passante per l'origine, si può rappresentare mediante una matrice  $\mathbf{R}$ ,  $(3, 3)$  e ortogonale<sup>8</sup>:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{R} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

A differenza di quanto accade nel piano, il prodotto di due trasformazioni ortogonali nello spazio *non è commutativo*. Ad ogni modo la composizione di due trasformazioni ortogonali è ancora una trasformazione ortogonale.

Una traslazione è rappresentata da

$$(2.4) \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} u \\ v \\ w \end{pmatrix}.$$

La composizione di un numero finito di traslazioni e di trasformazioni ortogonali, in qualunque ordine, assume sempre la forma

$$(2.5) \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{R} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

per qualche matrice ortogonale  $\mathbf{R}$  e qualche vettore  $(u, v, w)^T \in \mathbb{R}^3$ ; la (2.5) rappresenta quindi la generica rototraslazione (con eventuale riflessione) nello spazio tridimensionale. Sotto l'azione di una rototraslazione qualsiasi, l'equazione di una generica quadrica (2.1) si trasformerà ancora in un'equazione del second'ordine.

<sup>8</sup>Più precisamente, come abbiamo visto nel capitolo 8, paragrafo 6.3, una generica matrice ortogonale  $(3, 3)$  rappresenta una rotazione nello spazio eventualmente accompagnata da una riflessione rispetto ad un piano passante per l'origine. Le matrici ortogonali che rappresentano una rotazione pura si dicono *ortogonalî speciali* e hanno l'ulteriore proprietà di avere determinante uguale a 1. Nel seguito non sarà importante per noi distinguere le due cose.

Vediamo più da vicino come cambia l'equazione di una quadrica sotto l'azione di una trasformazione ortogonale oppure di una traslazione. Il calcolo, in forma matriciale, è analogo a quello già fatto per le coniche.

Se  $\mathbf{R}$  è una matrice ortogonale  $(3, 3)$ , ponendo  $\mathbf{x} = (x, y, z, 1)$ ,  $\mathbf{x}' = (x', y', z', 1)$ , possiamo riscrivere (utilizzando la scrittura delle matrici a blocchi e ponendo  $(0, 0, 0) = \mathbf{0}$ ):

$$\mathbf{x}^\top = \begin{pmatrix} \mathbf{R} & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{x}'^\top, \text{ quindi } \mathbf{x} = \mathbf{x}' \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix}$$

che, sostituendo nell'equazione matriciale della quadrica  $\mathbf{x}^\top \mathbf{A} \mathbf{x} = 0$  danno:

$$\mathbf{x}'^\top \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{A} \begin{pmatrix} \mathbf{R} & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{x}'^\top = 0$$

da cui leggiamo che la matrice  $\mathbf{A}$  si trasforma nella matrice

$$(2.6) \quad \mathbf{A}' = \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{A} \begin{pmatrix} \mathbf{R} & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \equiv \mathcal{R}^\top \mathbf{A} \mathcal{R}$$

e in particolare  $\mathbf{B}$  si trasforma in

$$(2.7) \quad \mathbf{B}' = \mathbf{R}^\top \mathbf{B} \mathbf{R}.$$

Analogamente a quanto abbiamo visto per le coniche, l'azione di una rotazione  $\mathbf{R}$  su una quadrica trasforma quindi la matrice  $\mathbf{B}$  dei coefficienti dei termini di 2° grado della quadrica in una matrice simile  $\mathbf{B}'$ . In particolare,  $\mathbf{B}$  e  $\mathbf{B}'$  hanno lo stesso determinante. Inoltre, scegliendo opportunamente la matrice  $\mathbf{R}$  ortogonale nella (2.7),  $\mathbf{B}$  può essere diagonalizzata, ossia portata alla forma

$$\mathbf{B}' = \text{diag}(a'_{11}, a'_{22}, a'_{33}).$$

Questo fatto sarà utile nella riduzione a forma canonica della quadrica.

Notiamo anche che la matrice  $\mathcal{R}$  che compare nella (2.6) è ancora una matrice ortogonale,  $(4, 4)$ , perciò la (2.6) dice che anche  $\mathbf{A}$  e  $\mathbf{A}'$  sono matrici simili e in particolare hanno lo stesso determinante.

Eseguendo invece una traslazione (2.4), l'equazione della quadrica si trasforma in

$$\mathbf{x}'^\top \mathbf{A}' \mathbf{x}'^\top = 0$$

dove la matrice  $\mathbf{A}'$  ha gli stessi coefficienti di  $\mathbf{A}$  nella parte di secondo grado ( $\mathbf{B}' = \mathbf{B}$ ), mentre gli altri coefficienti cambiano secondo la legge:

$$(2.8) \quad \left\{ \begin{array}{l} a'_{14} = a_{11}u + a_{12}v + a_{13}w + a_{14} \\ a'_{24} = a_{12}u + a_{22}v + a_{23}w + a_{24} \\ a'_{34} = a_{13}u + a_{23}v + a_{33}w + a_{34} \\ a'_{44} = a_{11}u^2 + a_{22}v^2 + a_{33}w^2 + 2a_{12}uv + 2a_{13}uw + \\ \quad + 2a_{23}vw + 2a_{14} + 2a_{24} + 2a_{34} + a_{44} \end{array} \right.$$

come si verifica con un po' di pazienza. Anche queste relazioni ci saranno utili nella riduzione a forma canonica delle quadriche.

Possiamo ora enunciare il seguente

**TEOREMA 9.5 (INVARIANTI DI UNA QUADRICA)** *Con le notazioni introdotte nella Definizione 9.4, le seguenti quantità risultano invarianti sotto l'effetto di una rototraslazione:*

$$\mathcal{I}_1 = \text{Tr} \mathbf{B} = a_{11} + a_{22} + a_{33}$$

$$\mathcal{I}_2 = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} + \det \begin{pmatrix} a_{11} & a_{13} \\ a_{13} & a_{33} \end{pmatrix} + \det \begin{pmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{pmatrix};$$

$$\mathcal{I}_3 = \det \mathbf{B};$$

$$\mathcal{I}_4 = \det \mathbf{A}.$$

Sono inoltre invarianti la caratteristica della matrice  $\mathbf{A}$  e gli autovalori di  $\mathbf{B}$ .

**DIMOSTRAZIONE.** Ricordiamo che le matrici  $\mathbf{A}$ ,  $\mathbf{B}$  di una conica si trasformano per effetto di una rototraslazione in matrici  $\mathbf{A}'$ ,  $\mathbf{B}'$  simili rispettivamente ad  $\mathbf{A}$ ,  $\mathbf{B}$  e che, per quanto visto nel capitolo 8, paragrafo 6.1, matrici simili hanno lo stesso determinante, la stessa caratteristica e lo stesso polinomio caratteristico, quindi gli stessi autovalori. Ciò implica subito che  $\mathcal{I}_4$ ,  $\mathcal{I}_3$ , la caratteristica di  $\mathbf{A}$  e gli autovalori di  $\mathbf{B}$  siano invarianti. Inoltre, scrivendo esplicitamente il polinomio caratteristico di  $\mathbf{B}$  si ottiene (lasciamo i calcoli al lettore):

$$\det(\mathbf{B} - \lambda \mathbf{I}) = -\lambda^3 + \lambda^2 \mathcal{I}_1 - \lambda \mathcal{I}_2 + \mathcal{I}_3$$

da cui leggiamo anche l'invarianza di  $\mathcal{I}_1$  e  $\mathcal{I}_2$ . ◆

Osserviamo anche che, moltiplicando l'equazione della quadrica per un coefficiente  $\lambda > 0$ , gli invarianti  $\mathcal{I}_1$ ,  $\mathcal{I}_2$ ,  $\mathcal{I}_3$ ,  $\mathcal{I}_4$  risultano moltiplicati rispettivamente per  $\lambda$ ,  $\lambda^2$ ,  $\lambda^3$ ,  $\lambda^4$ , mentre gli autovalori di  $\mathbf{B}$  sono moltiplicati per  $\lambda$ . Tenendo conto anche di quest'osservazione possiamo enunciare la seguente proposizione, che raccoglie i fatti che operativamente utilizzeremo nella classificazione delle quadriche:

**PROPOSIZIONE 9.4** *I seguenti fatti esprimono una proprietà geometrica della superficie quadrica:*

- l'annullarsi o meno di ciascuno degli invarianti  $\mathcal{I}_1$ ,  $\mathcal{I}_2$ ,  $\mathcal{I}_3$ ,  $\mathcal{I}_4$  (utilizzeremo solo gli ultimi due);
- il segno degli invarianti  $\mathcal{I}_2$  e  $\mathcal{I}_4$  (utilizzeremo solo il segno di  $\mathcal{I}_4$ );
- il numero di autovalori di  $\mathbf{B}$  che si annullano e il fatto che gli autovalori non nulli siano concordi tra loro oppure no;
- la caratteristica di  $\mathbf{A}$ .

Diamo anche la seguente

**DEFINIZIONE 9.5** Si dice che una quadrica è *degenera* se  $\mathcal{I}_4 = 0$ . ◆

Vedremo un po' alla volta che forme possono assumere le quadriche degeneri.

## 2.2 Riduzione a forma canonica delle quadriche: quadriche a centro

Consideriamo ora una generica quadrica (2.1). Ci proponiamo di determinare una rototraslazione che ne trasformi l'equazione in una forma abbastanza semplice da permetterci di riconoscere le proprietà importanti della superficie. Così facendo arriveremo ad una classificazione delle quadriche.

Chiediamoci anzitutto se esiste una traslazione che abbia come effetto quello di annullare i coefficienti dei termini di primo grado, cioè  $a_{14}, a_{24}, a_{34}$ . Dalle (2.8) leggiamo che questo equivale a cercare  $u, v, w$  che soddisfano il sistema:

$$\begin{cases} a_{11}u + a_{12}v + a_{13}w + a_{14} = 0 \\ a_{12}u + a_{22}v + a_{23}w + a_{24} = 0 \\ a_{13}u + a_{23}v + a_{33}w + a_{34} = 0 \end{cases}$$

che è un sistema lineare di 3 equazioni in 3 incognite, avente per matrice dei coefficienti la matrice  $\mathbf{B}$ . La traslazione cercata quindi esiste certamente se  $\mathcal{I}_3 \neq 0$ . Mettiamoci in quest'ipotesi. Dopo la traslazione quindi l'equazione della quadrica sarà (chiamando ancora, per semplicità, i coefficienti con i simboli  $a_{ij}$  anziché  $a'_{ij}$ )

$$(2.9) \quad a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz + a_{44} = 0.$$

Si osserva che questa superficie è *simmetrica rispetto all'origine*: se la terna  $x, y, z$  soddisfa l'equazione, anche  $-x, -y, -z$  la soddisfa; questo si può esprimere dicendo che l'origine è *centro di simmetria* per la quadrica, o anche che la superficie è una *quadrica a centro*. Abbiamo quindi stabilito la seguente

**PROPOSIZIONE 9.5** *Se  $\mathcal{I}_3 \neq 0$  la quadrica è a centro; un'opportuna traslazione la riconduce alla forma (2.9), simmetrica rispetto all'origine.*

Esaminiamo ora l'effetto di una rotazione su una quadrica a centro, per la quale cioè risulta

$$\mathbf{A} = \begin{pmatrix} \mathbf{B} & \mathbf{0}^\top \\ \mathbf{0} & a_{44} \end{pmatrix}$$

(avendo posto  $\mathbf{0} = (0, 0, 0)$ ). Dalle (2.6) leggiamo che la matrice  $\mathbf{A}$  si trasforma in

$$\mathbf{A}' = \begin{pmatrix} \mathbf{R}^\top & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{B} & \mathbf{0}^\top \\ \mathbf{0} & a_{44} \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{0}^\top \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}^\top \mathbf{B} \mathbf{R} & \mathbf{0}^\top \\ \mathbf{0} & a_{44} \end{pmatrix}$$

ossia la quadrica rimane a centro, e la matrice  $\mathbf{B}$  si trasforma secondo la legge  $\mathbf{B}' = \mathbf{R}^\top \mathbf{B} \mathbf{R}$ . Con una scelta opportuna della rotazione possiamo quindi diagonalizzare la matrice  $\mathbf{B}$ , ossia ridurre l'equazione della quadrica alla forma

$$(2.10) \quad a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + a_{44} = 0$$

con  $\mathcal{I}_3 = a_{11}a_{22}a_{33} \neq 0$  per ipotesi. Si noti che in questo caso risulta

$$\mathcal{I}_4 = a_{44}\mathcal{I}_3.$$

Pertanto  $\mathcal{I}_4 = 0$ , ossia la quadrica è degenere, se e solo se  $a_{44} = 0$ . Discutiamo subito questa situazione, ossia consideriamo una *quadrica a centro, degenere*:

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 = 0.$$

L'unica distinzione di rilievo, per quanto riguarda la forma della superficie descritta da questa equazione, sta nei segni dei tre coefficienti  $a_{11}, a_{22}, a_{33}$  (tutti diversi da zero per l'ipotesi  $\mathcal{I}_3 \neq 0$ ).

Se i 3 coefficienti non hanno tutti lo stesso segno, ad esempio  $a_{11}$  e  $a_{22}$  hanno lo stesso segno e  $a_{33}$  ha segno diverso (gli altri casi si ottengono da questo semplicemente cambiando nomi alle variabili), l'equazione si può riscrivere nella forma

$$a^2x^2 + b^2y^2 = c^2z^2$$

per qualche  $a, b, c > 0$ , e rappresenta un *cono a due falde* che ha vertice nell'origine e ha per generatrice ad esempio l'ellisse di equazione

$$(2.11) \quad a^2x^2 + b^2y^2 = c^2 \text{ nel piano } z = 1.$$

Infatti, preso un punto  $(x_0, y_0, 1)$  su tale curva, ossia tale che  $a^2x_0^2 + b^2y_0^2 = c^2$ , la retta passante per l'origine e  $(x_0, y_0, 1)$  ha equazioni parametriche

$$\begin{cases} x = tx_0 \\ y = ty_0 \\ z = t \end{cases} \quad t \in \mathbb{R}$$

e giace interamente sulla quadrica, in quanto

$$a^2x^2 + b^2y^2 = a^2(tx_0)^2 + b^2(ty_0)^2 = t^2(a^2x_0^2 + b^2y_0^2) = t^2c^2 = c^2z^2.$$

Pertanto la superficie coincide con l'unione delle rette passanti per l'origine e i punti dell'ellisse (2.11), ossia è un cono avente tale ellisse come generatrice.

Se invece i 3 coefficienti  $a_{11}, a_{22}, a_{33}$  hanno tutti lo stesso segno, l'equazione si può riscrivere nella forma

$$a^2x^2 + b^2y^2 + c^2z^2 = 0$$

per qualche  $a, b, c > 0$ , ed è evidentemente soddisfatta solo dal punto  $(0, 0, 0)$ . In questo caso quindi la quadrica degenera si riduce a un punto. Per analogia col caso precedente (cono a due falde) si parla anche di *cono immaginario*.

Occupiamoci ora delle *quadriche a centro non degeneri*, ossia del tipo (2.10) con tutti i coefficienti non nulli, che quindi dividendo per  $a_{44}$  si possono riscrivere nella forma

$$\alpha x^2 + \beta y^2 + \gamma z^2 = 1$$

Le uniche differenze importanti dipendono dai segni dei coefficienti  $\alpha, \beta, \gamma$ : ciò che conta è semplicemente quanti di questi coefficienti sono positivi (*quali* siano positivi è solo una questione di *nomi* dati alle variabili). Le possibilità sono quindi 4, a seconda che il numero di coefficienti positivi tra  $\alpha, \beta, \gamma$  sia 3, 2, 1, 0. In corrispondenza di questi casi si ottengono rispettivamente le seguenti *equazioni canoniche*:

$$(2.12) \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (\text{ellissoide reale})$$

$$(2.13) \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \quad (\text{iperboloide a una falda})$$

$$(2.14) \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \quad (\text{iperboloide a due falde})$$

$$(2.15) \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1 \quad (\text{ellissoide immaginario})$$

per opportuni  $a, b, c > 0$ . Descriviamo brevemente queste 4 superfici.

La (2.12) è l'equazione canonica di un **ellissoide (reale)** (v. fig. 9.10).

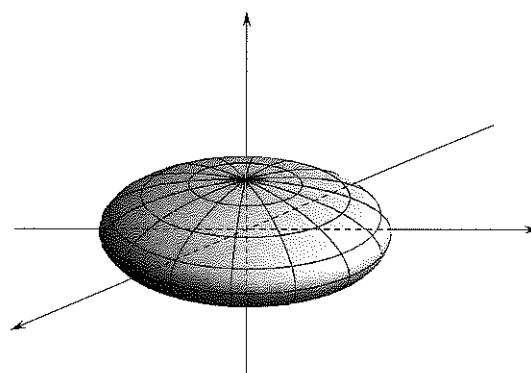


Figura 9.10. L'ellissoide di equazione  $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ .

Il modo più semplice di capire la forma di questa superficie è pensare che si può ottenere da una superficie sferica eseguendo una dilatazione di fattore diverso su ogni asse:  $x = ax'$ ,  $y = by'$ ,  $z = cz'$ , che dà all'ellissoide la forma di una "sfera allungata". I numeri  $a, b, c$  sono i *semiassi* dell'ellissoide, che ha per *vertici* i punti  $(\pm a, 0, 0)$ ,  $(0, \pm b, 0)$ ,  $(0, 0, \pm c)$ . Se due dei tre semiassi sono uguali tra loro l'ellissoide è di rotazione: ad esempio per  $a = b$ , la superficie taglia i piani  $z = z_0$  (quando  $|z_0| < c$ ) lungo le circonferenze

$$x^2 + y^2 = a^2 \left(1 - \frac{z_0^2}{c^2}\right),$$

pertanto la superficie si può vedere ottenuta facendo ruotare attorno all'asse  $z$  l'ellisse (nel piano  $xz$ )  $\frac{x^2}{a^2} + \frac{z^2}{c^2} = 1$ . Se tutti e tre i semiassi sono uguali, la superficie è una sfera di raggio  $a$ .

La (2.13) è l'equazione canonica di un **iperboloido a una falda** (v. fig. 9.11).

Si osserva che la superficie taglia qualsiasi piano  $z = z_0$  lungo l'ellisse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 + \frac{z_0^2}{c^2}.$$

I semiassi delle sezioni ellittiche crescono al crescere di  $|z_0|$ . La superficie è illimitata e connessa (cioè "formata da un pezzo solo", per cui si dice *a una falda*). In particolare, per  $a = b$  l'ellisse è una circonferenza e la quadrica è di rotazione, ottenuta facendo ruotare attorno all'asse  $z$  l'iperbole  $\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1$  nel piano  $xz$  (da cui il nome di iperboloido). Sezionando invece la superficie con un piano contenente l'asse  $z$ , ossia del tipo  $\alpha x + \beta y = 0$ , si ottiene un'iperbole, come si verifica facilmente.

Una particolarità di questa superficie è di *contenere due famiglie di rette*. Infatti, riscrivendo l'equazione nella forma

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1 - \frac{y^2}{b^2}$$

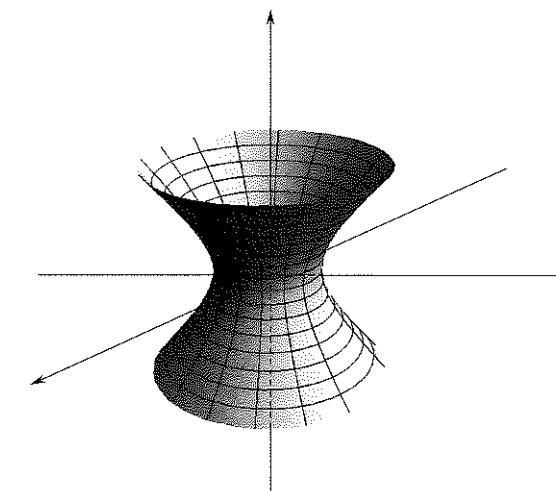


Figura 9.11. L'ellissoide di equazione  $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$ .

si osserva che le due famiglie di rette:

$$(2.16) \quad \begin{cases} \frac{x}{a} + \frac{z}{c} = h \left(1 + \frac{y}{b}\right) \\ \frac{x}{a} - \frac{z}{c} = \frac{1}{h} \left(1 - \frac{y}{b}\right) \end{cases} \quad \text{e} \quad \begin{cases} \frac{x}{a} + \frac{z}{c} = h \left(1 - \frac{y}{b}\right) \\ \frac{x}{a} - \frac{z}{c} = \frac{1}{h} \left(1 + \frac{y}{b}\right) \end{cases}$$

(al variare del parametro reale  $h$ ) sono contenute nella superficie stessa, come si verifica moltiplicando membro a membro le due equazioni di ciascun sistema. Per questa ragione si dice che l'iperboloido a una falda è una *superficie rigata*. Precisamente:

**PROPOSIZIONE 9.6** *Le due famiglie di rette (2.16) che giacciono sull'iperboloido a una falda hanno le seguenti proprietà:*

- a) *per ogni punto della superficie passano esattamente una retta della prima famiglia e una retta della seconda famiglia;*
- b) *due rette qualsiasi della stessa famiglia sono sghembe;*
- c) *ogni retta della prima famiglia interseca ogni retta della seconda famiglia.*

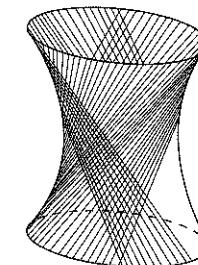


Figura 9.12. Rigatura dell'iperboloido a una falda.

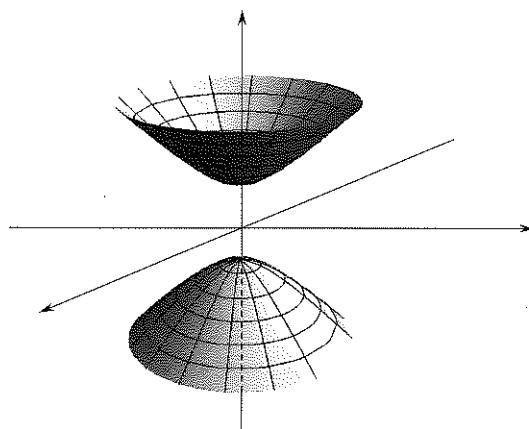


Figura 9.13. L'iperboloido a due falde di equazione  $\frac{x^2}{a^2} - 1 = \frac{y^2}{b^2} + \frac{z^2}{c^2}$ .

La (2.14) è l'equazione canonica di un **iperboloido a due falde** (v. fig. 9.13). Riscrivendo l'equazione nella forma

$$\frac{x^2}{a^2} - 1 = \frac{y^2}{b^2} + \frac{z^2}{c^2}$$

si osserva che la superficie è contenuta nelle regioni in cui  $\frac{x^2}{a^2} - 1 \geq 0$ , ossia nelle due regioni, separate,  $x \geq a$  e  $x \leq -a$ . Perciò la superficie consiste di *due falde* distinte. Più precisamente, la superficie è contenuta nella regione

$$\frac{y^2}{b^2} + \frac{z^2}{c^2} < \frac{x^2}{a^2}$$

(come segue dal fatto che  $\frac{x^2}{a^2} - 1 < \frac{x^2}{a^2}$ ), che rappresenta una delle due regioni in cui il cono a due falde

$$\frac{y^2}{b^2} + \frac{z^2}{c^2} = \frac{x^2}{a^2}$$

divide lo spazio (precisamente, la regione contenente l'asse del cono). Questo cono si dice *cono asintotico*, perché la superficie si avvicina indefinitamente ad esso così come un'iperbole si avvicina ai suoi asintoti. Tagliando la superficie con un piano  $x = x_0$  (per  $|x_0| > a$ ) si trova l'ellisse

$$\frac{y^2}{b^2} + \frac{z^2}{c^2} = \frac{x_0^2}{a^2} - 1.$$

In particolare se  $b = c$  quest'ellisse è una circonferenza e la quadrica è di rotazione, ottenuta facendo ruotare attorno all'asse  $x$  l'iperbole

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

(nel piano  $xy$ ). Tagliando invece la superficie con un piano contenente l'asse  $x$ , cioè del tipo  $\alpha y + \beta z = 0$ , si ottiene sempre un'iperbole, i cui asintoti giacciono sul cono asintotico, come si verifica facilmente. A differenza dell'iperboloido a una falda, l'iperboloido a due falde non è una superficie regata.

La (2.15) è l'equazione canonica di un **ellissoide immaginario**. Si vede subito che l'equazione

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = -1$$

non è soddisfatta da alcuna terna  $(x, y, z)$  di numeri reali, quindi nello spazio  $\mathbb{R}^3$  rappresenta l'insieme vuoto. Per analogia con l'ellissoide (2.12), la superficie è detta ellissoide immaginario.

### 2.3 Riduzione a forma canonica delle quadriche: quadriche non a centro

Consideriamo ora le quadriche non a centro, ossia quelle per cui risulta  $I_3 = 0$  (cfr. con la Proposizione 9.5). In tal caso non sempre esiste una traslazione che elimini i termini di primo grado. Possiamo in compenso diagonalizzare subito la matrice  $B$ , ossia effettuare un'opportuna rotazione di assi che elimini i termini  $a_{12}, a_{23}, a_{13}$ . Dopo questa rotazione l'equazione della quadrica non a centro risulta quindi del tipo:

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{14}x + 2a_{24}y + 2a_{34}z + a_{44} = 0,$$

inoltre

$$I_3 = a_{11}a_{22}a_{33} = 0,$$

quindi almeno uno dei coefficienti  $a_{11}, a_{22}, a_{33}$  si annulla. Supponiamo ad esempio  $a_{33} = 0$ . La matrice dei coefficienti è quindi:

$$(2.17) \quad \mathbf{A} = \begin{pmatrix} a_{11} & 0 & 0 & a_{14} \\ 0 & a_{22} & 0 & a_{24} \\ 0 & 0 & 0 & a_{34} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{pmatrix}.$$

Notiamo che

$$I_4 = -a_{11}a_{22}a_{34}^2.$$

Studiamo prima il caso della quadrica non a centro e non degenere; quindi  $I_4 \neq 0$  e i coefficienti  $a_{11}, a_{22}, a_{34}$  sono effettivamente diversi da zero.

Cerchiamo una traslazione che annulli i coefficienti  $a_{14}, a_{24}$ . Dalle (2.8) leggiamo che questo significa risolvere il sistema in  $u, v$  (in altre parole, la traslazione non agisce sulla terza variabile):

$$\begin{cases} a_{11}u + a_{14} = 0 \\ a_{22}v + a_{24} = 0, \end{cases}$$

il che è immediato perché stiamo supponendo  $a_{11}, a_{22} \neq 0$ . Arriviamo quindi alla quadrica di equazione:

$$a_{11}x^2 + a_{22}y^2 + 2a_{34}z + a_{44} = 0.$$

Infine la traslazione

$$z = z' + w$$

elimina il termine noto  $a_{44}$  scegliendo

$$w = -\frac{a_{44}}{2a_{34}}.$$

in quanto contiene le due famiglie di rette:

$$\begin{cases} \frac{x}{a} + \frac{y}{b} = 2hpz \\ \frac{x}{a} - \frac{y}{b} = \frac{1}{h} \end{cases} \quad \text{e} \quad \begin{cases} \frac{x}{a} + \frac{y}{b} = -2hpz \\ \frac{x}{a} - \frac{y}{b} = -\frac{1}{h} \end{cases}$$

(al variare del parametro reale  $h$ ), come si verifica moltiplicando membro a membro le due equazioni di ciascun sistema.

Precisamente, valgono ancora le proprietà espresse nella Proposizione 9.6.

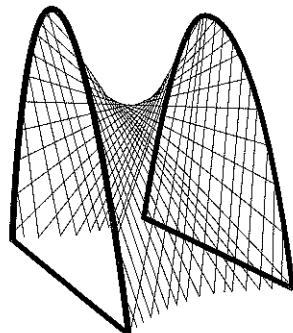


Figura 9.16. Rigatura del paraboloido iperbolico.

Dobbiamo ancora considerare le quadriche *non a centro e degeneri*, ossia quelle aventi matrice dei coefficienti (2.17) ma  $\mathcal{I}_4 = -a_{11}a_{22}a_{34}^2 = 0$ .

Senza stare a illustrare in dettaglio i vari passi della riduzione a forma canonica, limitiamoci a descrivere in questo caso le situazioni essenzialmente distinte che si possono presentare. La superficie può essere un *cilindro* o una *quadrica spezzata*. Per cilindro (in senso lato) si intende una superficie formata da tutte le rette parallele a un certo *asse* fissato e passanti ciascuna per un punto di una curva fissata, che definisce la *sezione* del cilindro:

Cilindri (con asse parallelo all'asse  $z$ )

$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	cilindro ellittico (cioè a sezione ellittica);
$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$	cilindro iperbolico;
$x^2 = 2py$	cilindro parabolico
$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$	cilindro immaginario (insieme vuoto)

Le quadriche non a centro e spezzate sono invece le seguenti:

Quadriche spezzate

$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$	due piani (reali, $y = \pm \frac{b}{a}x$ ), incidenti su una retta (asse $z$ )
$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$	due piani complessi coniugati ( $y = \pm i\frac{b}{a}x$ ), incidenti su una retta reale (asse $z$ )
$x^2 - a^2 = 0$	due piani paralleli reali ( $x = \pm a$ )
$x^2 + a^2 = 0$	due piani immaginari ( $x = \pm ia$ ) (insieme vuoto)
$x^2 = 0$	un piano ( $x = 0$ )

Si osservi che anche le quadriche spezzate, a rigore, soddisfano la definizione che abbiamo dato di cilindro (consistono tutte di rette parallele a un asse); solitamente si riserva il nome di cilindro alle superfici che non si spezzano in piani.

#### 2.4 Riconoscimento di una quadrica mediante gli invarianti

Nel paragrafo precedente abbiamo passato in rassegna tutti i casi possibili, ottenendo quindi una classificazione completa delle quadriche. È utile a questo punto riepilogare i casi ottenuti, calcolando in ciascuno di essi gli invarianti della quadrica, al fine di poter poi procedere in senso inverso, ossia: data l'equazione di una quadrica qualsiasi, dal calcolo degli invarianti risalire immediatamente al tipo di quadrica. Il prossimo schema riassume gli elementi utili al riconoscimento delle quadriche in base agli invarianti. Lasciamo al lettore la verifica delle informazioni qui contenute, che seguono dalla casistica stabilita nel paragrafo precedente.

La prima distinzione da fare è:

A.  $\mathcal{I}_4 \neq 0 \Rightarrow$  Quadrica non degenere

B.  $\mathcal{I}_4 = 0 \Rightarrow$  Quadrica degenere

- A. *Riconoscimento delle quadriche non degeneri ( $\mathcal{I}_4 \neq 0$ )*  
 $\mathcal{I}_3 \neq 0 \Rightarrow$  Quadrica (non degenere) a centro.

Esistono 4 tipi, che si riconoscono in base al segno dell'invariante  $\mathcal{I}_4$  e alle relazioni tra i segni degli autovalori<sup>9</sup> di B:

	$\mathcal{I}_4 < 0$	$\mathcal{I}_4 > 0$
Autovalori di B concordi	ellissoide reale	ellissoide immaginario
Autovalori di B discordi	iperboloide a due falda	iperboloide a una falda

<sup>9</sup>Si ricordi che se B è diagonale, come accade quando l'equazione della quadrica è in forma canonica, i suoi autovalori sono semplicemente gli elementi sulla diagonale.

$\mathcal{I}_3 = 0 \implies$  Quadrica (non degenere) non a centro  
Esistono 2 tipi:

$\mathcal{I}_4 < 0$	$\mathcal{I}_4 > 0$
paraboloide ellittico	paraboloide iperbolico

• B. Riconoscimento delle quadriche degeneri ( $\mathcal{I}_4 = 0$ )

Se  $\mathcal{I}_3 \neq 0$  si ha un cono (reale o immaginario) e precisamente:

Autovalori di $\mathbf{B}$ concordi	cono immaginario (un punto reale)
Autovalori di $\mathbf{B}$ discordi	cono reale (a sezione ellittica)

Se  $\mathcal{I}_3 = 0$  si ha un cilindro (reale o immaginario) oppure una quadrica spezzata e precisamente:

Gli autovalori di $\mathbf{B}$ sono:			
uno nullo e due concordi	uno nullo e due discordi	due nulli	
Car $\mathbf{A} = 3$	cilindro ellittico oppure cilindro immaginario ( $\emptyset$ )	cilindro iperbolico	cilindro parabolico
Car $\mathbf{A} = 2$	2 piani immaginari incidenti (una retta reale)	2 piani (reali) incidenti	2 piani paralleli oppure 2 piani immaginari ( $\emptyset$ )
Car $\mathbf{A} = 1$			un piano (reale)

Per distinguere tra loro il caso del cilindro reale o immaginario (o tra due piani reali o immaginari) la cosa più semplice solitamente è cercare almeno un punto reale che soddisfi l'equazione.



Riconoscere le seguenti quadriche:

- a)  $xy + z^2 + 1 = 0$
- b)  $x^2 - y^2 + 4z^2 - 4xz - z + 3 = 0$
- c)  $x^2 + y^2 + z^2 + 2xy - 3 = 0$

a) La matrice dei coefficienti è

$$\mathbf{A} = \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

da cui:

$$\mathcal{I}_4 = -1/4 < 0, \mathcal{I}_3 = -1/4 \neq 0,$$

quadrica non degenere a centro. Poiché  $\det \mathbf{A} < 0$  si tratta di un ellissoide reale o di un iperboloido a due falde. Per decidere calcoliamo gli autovalori di  $\mathbf{B}$ :

$$\det(\mathbf{B} - \lambda \mathbf{I}) = (1 - \lambda) \left( \lambda^2 - \frac{1}{4} \right) = 0 \Rightarrow \lambda_1 = 1, \lambda_2 = \frac{1}{2}, \lambda_3 = -\frac{1}{2}.$$

Poiché gli autovalori sono discordi, si tratta di un iperboloido a due falde.

b) La matrice dei coefficienti è

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & -1 & 0 & 0 \\ -2 & 0 & 4 & -1/2 \\ 0 & 0 & -1/2 & 3 \end{pmatrix},$$

da cui:

$$\mathcal{I}_4 > 0, \mathcal{I}_3 = 0,$$

quadrica non degenere non a centro. Poiché  $\det \mathbf{A} > 0$  si tratta di un paraboloide iperbolico.

c) La matrice dei coefficienti è

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix},$$

da cui:

$$\mathcal{I}_4 = 0, \mathcal{I}_3 = 0,$$

quadrica degenere, cilindro o spezzata. Per decidere, osserviamo che

$$\text{Car } \mathbf{A} = 3 \text{ in quanto } \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{vmatrix} \neq 0,$$

quindi si tratta di un cilindro. Per determinarne il tipo, calcoliamo gli autovalori di  $\mathbf{B}$ :

$$\det(\mathbf{B} - \lambda \mathbf{I}) = (1 - \lambda) [(1 - \lambda)^2 - 1] = 0 \Rightarrow \lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 0.$$

Essendo gli autovalori uno nullo e due concordi, il cilindro ha sezione ellittica oppure è immaginario. Ma si vede subito che il cilindro possiede punti reali, ad es.  $(0, 0, \sqrt{3})$ , perciò si tratta di un cilindro (reale) ellittico.



Riconoscere le seguenti quadriche:

- |  |                                      |
|--|--------------------------------------|
| 1) $x^2 + 2y^2 + z^2 + 2yz + 2x + 1 = 0$ | 2) $xy + z^2 = 0$                    |
| 3) $x^2 + z^2 + 2y + 2x = 0$             | 4) $y^2 + z^2 + 2yz + 3 = 0$         |
| 5) $x^2 + y^2 + z^2 - 4yz - 2 = 0$       | 6) $x^2 + 2y^2 + 3z^2 - 2yz + 5 = 0$ |

Scrivere l'equazione delle seguenti quadriche degeneri:

- 7) Generico cono ellittico di vertice l'origine avente per asse l'asse  $x$ .

- 1 Cilindro parabolico avente asse parallelo all'asse  $y$ .
- 2 Cilindro iperbolico avente asse parallelo all'asse  $x$ .
- 3 Dimostrare la Proposizione 9.6 a partire dalle equazioni parametriche delle rette (2.16).

### 3 ESEMPI DI APPLICAZIONI DI CONICHE E QUADRICHÉ

Segnaliamo alcune questioni più o meno elementari in cui sono coinvolte le coniche e le quadriche.

1. L'ombra proiettata da una circonferenza su un piano è un'ellisse.
2. La sezione di un cilindro con un piano (non parallelo all'asse del cilindro) è un'ellisse.
3. L'orbita di un pianeta intorno al sole è un'ellisse, di cui il sole occupa uno dei fuochi. Questa è la famosa "prima legge di Keplero" sul moto dei pianeti, formulata nel 1609.
4. Più in generale, applicando le leggi della meccanica, Newton dimostrò che qualsiasi corpo celeste sia attratto dal sole si muove lungo un'orbita descritta da una conica, quindi non necessariamente un'ellisse. Si noti, però, che tra ellisse, parabola e iperbole, l'ellisse è l'unica curva chiusa. Perciò se il corpo celeste passa vicino al sole lungo un'orbita parabolica o iperbolica, significa che ci passa una volta sola e poi si allontana per sempre. Nel 1705 Sir Edmond Halley (1656-1742) mostrò che la cometa che oggi porta il suo nome si muove intorno al sole su un'orbita ellittica, ritornando ogni 76 anni nella stessa posizione. Fu la prima verifica sperimentale di questo fatto per un corpo celeste diverso da un pianeta del nostro sistema solare.
5. In base alle leggi della cinematica classica, il moto di un proiettile (nel vuoto) avviene lungo una parabola. Per "proiettile" si intende qualsiasi corpo il cui moto sia dato dalla composizione di un'accelerazione costante in verticale (accelerazione di gravità) e una velocità costante in una qualsiasi direzione diversa dalla verticale.
6. La terra ha, in prima approssimazione, la forma di un ellissoide di rotazione.
7. La parabola ha la proprietà che ogni raggio che giunge sul suo interno parallelamente all'asse, viene riflesso<sup>10</sup> nel fuoco (fig. 9.11) (da cui l'origine del nome "fuoco"). Su questo principio si basano molti dispositivi, tra cui:
  - a. L'*antenna parabolica*. Questa antenna ha la forma della superficie generata dalla rotazione di una parabola intorno al suo asse (paraboloide di rotazione); perciò tutti i raggi (onde elettromagnetiche) che arrivano parallelamente al suo asse vengono riflesse in un unico punto, il fuoco, in cui è posto il ricevitore. Le onde captate dall'antenna parabolica si possono supporre viaggiare su linee parallele in quanto

<sup>10</sup>Supponendo che valga la *legge della riflessione*: l'angolo formato dal raggio incidente con la tangente alla parabola è uguale all'angolo formato dal raggio riflesso con la stessa tangente.

- provengono da una sorgente molto lontana. Dunque la forma dell'antenna parabolica è quella adatta a convogliare in un unico punto le onde elettromagnetiche provenienti da molto lontano.
- b. Per lo stesso motivo, i telescopi sono dotati di *specchi parabolici*, costruiti in modo da convogliare in un punto i raggi luminosi provenienti da molto lontano.
  - c. Nel *microfono parabolico* le onde sonore provenienti dall'infinito vengono a concentrarsi nel fuoco del paraboloide in cui si trova un microfono. Le cuffie collegate al microfono permettono di sentire queste onde sonore. Questo oggetto è comunemente utilizzato dagli ornitologi e permette di sentire il canto degli uccelli ad un chilometro di distanza.
  - d. Reciprocamente, i fari delle automobili hanno la forma di un paraboloide affinché la luce emessa nel fuoco venga riflessa e inviata in raggi luminosi paralleli.
  8. L'ellisse ha la proprietà che un raggio uscente da uno dei due fuochi viene riflesso dalla curva in modo da passare per l'altro fuoco. Un'analogia proprietà è goduta di conseguenza dall'ellissoide di rotazione: un raggio uscente da uno dei fuochi dell'ellisse generatore si riflette sulla superficie interna dell'ellissoide passando per l'altro fuoco.

Supponiamo ad esempio di essere in una sala di forma ellittica, oppure sormontata da una volta ellissoidale. Il suono emesso in uno dei due fuochi, anche se molto debole, si sente molto bene nell'altro fuoco.

La proprietà dei punti dell'ellisse, di avere somma delle distanze dai fuochi costante, significa che le onde luminose o sonore che sono emesse in un fuoco e, riflesse dalla parete, raggiungono l'altro fuoco, percorrono tutte la stessa distanza e perciò arrivano contemporaneamente (si dice che sono *in fase*).

  9. La forma assunta dalla superficie libera del liquido contenuto in un cilindro che ruota attorno al proprio asse è un paraboloide di rotazione.
  10. Quando una grandezza  $y$  dipende da un'altra grandezza  $x$  in modo inversamente proporzionale, significa che  $y = \frac{k}{x}$  (con  $k$  costante). La curva  $y = \frac{k}{x}$  è un'iperbole equilatera. Si può anche dire così: se il prodotto di due variabili  $(x, y)$  si mantiene costante, il punto  $(x, y)$  si muove su un'iperbole.
  11. L'area racchiusa dall'ellisse vale  $\pi ab$  (v. cap. 6, Esempio 5.27). Il risultato è "ragionevole" se si pensa che il cerchio ha area  $\pi r^2$ , ed è un'ellisse con  $a = b = r$ . Invece, non esiste una formula semplice che assegna la lunghezza dell'ellisse in funzione dei semiassi  $a, b$ . Anzi, lo studio della relazione che lega la lunghezza dell'ellisse ai semiassi ha condotto a un intero settore di ricerche in matematica, la *teoria delle funzioni ellittiche*.
  12. L'iperboideo a una falda e il paraboloide iperbolico condividono, oltre alla proprietà di essere superfici rigate, quella di essere *superfici a curvatura negativa*: sezionando in qualsiasi punto queste superfici con una coppia di piani ortogonali tra loro e ortogonali al piano tangente alla superficie stessa, si trovano due curve aventi concavità opposte. Queste due proprietà fanno sì che le superfici materiali aventi queste forme abbiano particolari proprietà di stabilità strutturale, che le rendono interessanti dal punto di vista costruttivo: torri a forma di iperboideo a una falda hanno un

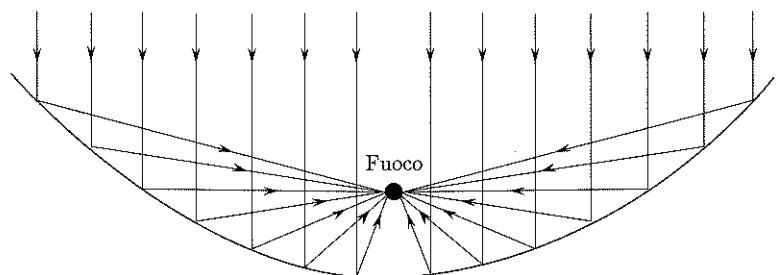


Figura 9.17. Riflessione dei raggi in una parabola.

certo impiego industriale (ad es. nelle torri di raffreddamento delle acque industriali), mentre la forma di paraboloidi iperbolico è utile a realizzare coperture di edifici a luce molto ampia con spessori ridotti.

**13.** Abbiamo visto che, data l'equazione di una conica

$$ax^2 + bxy + cy^2 + dx + ey + f = 0,$$

si può decidere se si tratti di ellisse, parabola o iperbole esaminando il segno del discriminante ( $b^2 - 4ac$ ). Questo può vedersi come il discriminante di un trinomio di secondo grado che avrà, rispettivamente 2, 1, 0 radici reali nel caso iperbolico, parabolico, ellittico. In matematica ci sono molte altre questioni in cui si presenta un'analogia classificazione: un certo oggetto matematico può essere di 3 tipi diversi, a seconda del segno del discriminante di un certo trinomio di secondo grado. Per analogia con la classificazione delle coniche, questi 3 "casi" sono detti allora caso ellittico, caso parabolico, caso iperbolico. Questo succede ad esempio nello studio della geometria delle superfici (punti ellittici, parabolici, iperbolici), nello studio delle equazioni a derivate parziali (equazioni ellittiche, paraboliche, iperboliche), ecc.

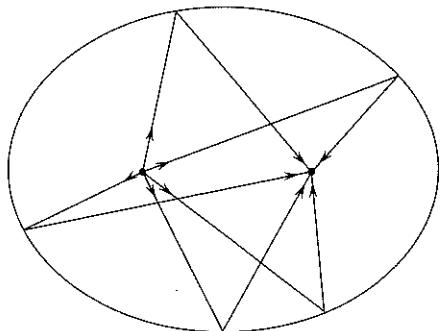


Figura 9.18. Riflessione dei raggi in un'ellisse.



## Formule utili

### 1 COSTANTI MATEMATICHE

$e$	2,7182818285...
$\pi$	3,1415926536...
$\log_{10} 2$	0,3010299957...
$\log_{10} e$	0,4342944819...
$\log_{10} \pi$	0,4971498727...
$\log_e 2$	0,6931471806...
$\log_e \pi$	1,1447298858...
$\log_e 10$	2,3025850930...
$\sqrt{2}$	1,4142135624...
$\sqrt{e}$	1,6487212707...
$\sqrt{3}$	1,7320508076...
$\sqrt{\pi}$	1,7724538509...
$\sqrt{5}$	2,2360679775...
$\sqrt{10}$	3,1622776602...
$1^\circ$	0,0174532925... radianti
1 radiante	57°17'44",806...

### 2 FUNZIONI TRIGONOMETRICHE

$$\begin{aligned} \sin x & \quad \cos x & \quad \operatorname{tg} x = \frac{\sin x}{\cos x} & \quad \operatorname{cotg} x = \frac{\cos x}{\sin x} \\ & & (\sin x)^2 + (\cos x)^2 = 1 & \end{aligned}$$

**Angoli notevoli**

$x$	$\cos x$	$\sin x$	$\operatorname{tg} x$	$\operatorname{cotg} x$
0	1	0	0	$\pm\infty$
$\frac{\pi}{10} = 18^\circ$	$\frac{1}{4}\sqrt{10+2\sqrt{5}}$	$\frac{\sqrt{5}-1}{4}$	$\sqrt{\frac{5-2\sqrt{5}}{5}}$	$\sqrt{5+2\sqrt{5}}$
$\frac{\pi}{6} = 30^\circ$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{3}$	$\sqrt{3}$
$\frac{\pi}{5} = 36^\circ$	$\frac{\sqrt{5}+1}{4}$	$\frac{1}{4}\sqrt{10-2\sqrt{5}}$	$\sqrt{5-2\sqrt{5}}$	$\sqrt{\frac{5+2\sqrt{5}}{5}}$
$\frac{\pi}{4} = 45^\circ$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	1	1
$\frac{3\pi}{10} = 54^\circ$	$\frac{1}{4}\sqrt{10-2\sqrt{5}}$	$\frac{\sqrt{5}+1}{4}$	$\sqrt{\frac{5+2\sqrt{5}}{5}}$	$\sqrt{5-2\sqrt{5}}$
$\frac{\pi}{3} = 60^\circ$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\sqrt{3}$	$\frac{\sqrt{3}}{3}$
$\frac{2\pi}{5} = 72^\circ$	$\frac{\sqrt{5}-1}{4}$	$\frac{1}{4}\sqrt{10+2\sqrt{5}}$	$\sqrt{5+2\sqrt{5}}$	$\sqrt{\frac{5-2\sqrt{5}}{5}}$
$\frac{\pi}{2} = 90^\circ$	0	1	$\pm\infty$	0

**Simmetrie, archi complementari e supplementari**

$$\sin(-x) = -\sin x \quad \cos(-x) = \cos x \quad \operatorname{tg}(-x) = -\operatorname{tg} x$$

$$\sin\left(\frac{\pi}{2} \pm x\right) = \cos x \quad \cos\left(\frac{\pi}{2} \pm x\right) = \mp \sin x \quad \operatorname{tg}\left(\frac{\pi}{2} \pm x\right) = \mp \operatorname{cotg} x$$

$$\sin(\pi \pm x) = \mp \sin x \quad \cos(\pi \pm x) = -\cos x \quad \operatorname{tg}(\pi \pm x) = \pm \operatorname{tg} x$$

**Formule di addizione**

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y$$

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y$$

$$\operatorname{tg}(x \pm y) = \frac{\operatorname{tg} x \pm \operatorname{tg} y}{1 \mp \operatorname{tg} x \operatorname{tg} y}$$

Per le formule di addizione si ricorda che  $\operatorname{tg} x = \frac{\sin x}{\cos x}$

**Formule di duplicazione**

$$\sin 2x = 2 \sin x \cos x$$

$$\cos 2x = (\cos x)^2 - (\sin x)^2 = 2(\cos x)^2 - 1 = 1 - 2(\sin x)^2$$

$$\operatorname{tg} 2x = \frac{2 \operatorname{tg} x}{1 - (\operatorname{tg} x)^2}$$

**Formule di bisezione** (scegliere il segno corretto)

$$\sin \frac{x}{2} = \pm \sqrt{\frac{1 - \cos x}{2}}$$

$$\cos \frac{x}{2} = \pm \sqrt{\frac{1 + \cos x}{2}}$$

$$\operatorname{tg} \frac{x}{2} = \frac{1 - \cos x}{\sin x} = \frac{\sin x}{1 + \cos x}$$

**Formule di prostaferesi**

$$\sin u + \sin v = 2 \sin \frac{u+v}{2} \cos \frac{u-v}{2}$$

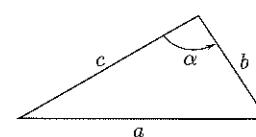
$$\sin u - \sin v = 2 \cos \frac{u+v}{2} \sin \frac{u-v}{2}$$

$$\cos u + \cos v = 2 \cos \frac{u+v}{2} \cos \frac{u-v}{2}$$

$$\cos u - \cos v = -2 \sin \frac{u+v}{2} \sin \frac{u-v}{2}$$

**Formule parametriche** Posto  $t = \operatorname{tg}(x/2)$ :

$$\sin x = \frac{2t}{1+t^2} \quad \cos x = \frac{1-t^2}{1+t^2} \quad \operatorname{tg} x = \frac{2t}{1-t^2}$$

**Teorema di Carnot**

$$a^2 = b^2 + c^2 - 2bc \cdot \cos \alpha$$

**3 FUNZIONI IPERBOLICHE**

$$\begin{aligned}\operatorname{Sh} x &= \operatorname{Ch} x - \operatorname{Th} x = \frac{\operatorname{Sh} x}{\operatorname{Ch} x} \\ (\operatorname{Ch} x)^2 - (\operatorname{Sh} x)^2 &= 1\end{aligned}$$

$$\operatorname{Sh}(-x) = -\operatorname{Sh} x \quad \operatorname{Ch}(-x) = \operatorname{Ch}(x) \quad \operatorname{Th}(-x) = -\operatorname{Th} x$$

**Formule di addizione**

$$\operatorname{Sh}(x \pm y) = \operatorname{Sh} x \operatorname{Ch} y \pm \operatorname{Ch} x \operatorname{Sh} y$$

$$\operatorname{Ch}(x \pm y) = \operatorname{Ch} x \operatorname{Ch} y \pm \operatorname{Sh} x \operatorname{Sh} y$$

$$\operatorname{Th}(x \pm y) = \frac{\operatorname{Th} x \pm \operatorname{Th} y}{1 \pm \operatorname{Th} x \operatorname{Th} y}$$

**Formule di duplicazione**

$$\operatorname{Sh} 2x = 2 \operatorname{Sh} x \operatorname{Ch} x$$

$$\operatorname{Ch} 2x = (\operatorname{Ch} x)^2 + (\operatorname{Sh} x)^2 = 2(\operatorname{Ch} x)^2 - 1 = 1 + 2(\operatorname{Sh} x)^2$$

$$\operatorname{Th} 2x = \frac{2 \operatorname{Th} x}{1 + (\operatorname{Th} x)^2}$$

**Formule di bisezione** (scegliere il segno corretto)

$$\operatorname{Sh} \frac{x}{2} = \pm \sqrt{\frac{\operatorname{Ch} x - 1}{2}}$$

$$\operatorname{Ch} \frac{x}{2} = \sqrt{\frac{\operatorname{Ch} x + 1}{2}}$$

$$\operatorname{Th} \frac{x}{2} = \frac{\operatorname{Ch} x - 1}{\operatorname{Sh} x} = \frac{\operatorname{Sh} x}{\operatorname{Ch} x + 1}$$

**Formule di prostaferesi**

$$\operatorname{Sh} u + \operatorname{Sh} v = 2 \operatorname{Sh} \frac{u+v}{2} \operatorname{Ch} \frac{u-v}{2}$$

$$\operatorname{Sh} u - \operatorname{Sh} v = 2 \operatorname{Ch} \frac{u+v}{2} \operatorname{Sh} \frac{u-v}{2}$$

$$\operatorname{Ch} u + \operatorname{Ch} v = 2 \operatorname{Ch} \frac{u+v}{2} \operatorname{Ch} \frac{u-v}{2}$$

$$\operatorname{Ch} u - \operatorname{Ch} v = 2 \operatorname{Sh} \frac{u+v}{2} \operatorname{Sh} \frac{u-v}{2}$$

**Formule parametriche** Posto  $t = \operatorname{Th}(x/2)$ :

$$\begin{aligned}\operatorname{Sh} x &= \frac{2t}{1-t^2} & \operatorname{Ch} x &= \frac{1+t^2}{1-t^2} & \operatorname{Th} x &= \frac{2t}{1+t^2}\end{aligned}$$

**4 DERIVATE ELEMENTARI**

$f(x)$	$f'(x)$
$x^\alpha$	$\alpha x^{\alpha-1}$
$ x $	$\operatorname{sgn} x$
$\log  x $	$1/x$
$\log_a  x $	$1/(x \log a)$
$e^x$	$e^x$
$a^x$	$a^x \log a$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\operatorname{tg} x$	$1 + (\operatorname{tg} x)^2 = 1/(\cos x)^2$
$\operatorname{ctg} x$	$-1 - (\operatorname{ctg} x)^2 = -1/(\sin x)^2$
$\operatorname{Sh} x$	$\operatorname{Ch} x$
$\operatorname{Ch} x$	$\operatorname{Sh} x$
$\operatorname{Th} x$	$1 - (\operatorname{Th} x)^2 = 1/(\operatorname{Ch} x)^2$
$\operatorname{Cth} x$	$1 - (\operatorname{Cth} x)^2 = -1/(\operatorname{Sh} x)^2$
$\log  \sin x $	$-\operatorname{ctg} x$
$\log  \cos x $	$\operatorname{tg} x$
$\log  \operatorname{Sh} x $	$\operatorname{Cth} x$
$\log \operatorname{Ch} x$	$\operatorname{Th} x$
$\arcsin x$	$1/\sqrt{1-x^2}$
$\arccos x$	$-1/\sqrt{1-x^2}$
$\arctg x$	$1/(1+x^2)$
$\operatorname{arccot} x$	$-1/(1+x^2)$

**5 REGOLE DI DERIVAZIONE**

$$\begin{aligned}D(\lambda f(x) + \mu g(x)) &= \lambda f'(x) + \mu g'(x) \\ D(f(x)g(x)) &= f'(x)g(x) + f(x)g'(x) \\ D \frac{f(x)}{g(x)} &= \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2} \\ D f(g(x)) &= f'(g(x))g'(x) \\ D f(g(h(x))) &= f'(g(h(x)))g'(h(x))h'(x) \\ D e^{f(x)} &= e^{f(x)} f'(x)\end{aligned}$$

$$D \log |f(x)| = \frac{f'(x)}{f(x)}$$

$$D[f(x)]^{g(x)} = [f(x)]^{g(x)} \left\{ g'(x) \log f(x) + \frac{g(x)f'(x)}{f(x)} \right\}$$

## 6 SVILUPPI DI MAC LAURIN DELLE PRINCIPALI FUNZIONI

$f(x)$	Sviluppo
$e^x$	$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots$
$\operatorname{Sh} x$	$x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{x^{2n+1}}{(2n+1)!} + \cdots$
$\operatorname{Ch} x$	$1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + \frac{x^{2n}}{(2n)!} + \cdots$
$\sin x$	$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \cdots$
$\cos x$	$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots$
$\log(1+x)$	$x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n-1} \frac{x^n}{n} + \cdots$
$\operatorname{arctg} x$	$x - \frac{x^3}{3} + \cdots + (-1)^n \frac{x^{2n+1}}{2n+1} + \cdots$
$(1+x)^\alpha$	$1 + \alpha x + \binom{\alpha}{2} x^2 + \cdots + \binom{\alpha}{n} x^n + \cdots$

In particolare:

$\frac{1}{1+x}$	$1 - x + x^2 - x^3 + \cdots + (-1)^n x^n + \cdots$	$(\alpha = -1)$
$\sqrt{1+x}$	$1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - \cdots + \frac{(-1)^{n+1}(2n-3)!!x^n}{(2n)!!}$	$(\alpha = \frac{1}{2})$
$\frac{1}{\sqrt{1+x}}$	$1 - \frac{x}{2} + \frac{3x^2}{8} - \frac{5x^3}{16} + \cdots + \frac{(-1)^n(2n-1)!!x^n}{(2n)!!}$	$(\alpha = -\frac{1}{2})$

ove

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)}{n!} \quad (\text{coefficiente binomiale generalizzato})$$

$$k!! = k(k-2)(k-4)\cdots 2 \quad (k \text{ semifattoriale})$$

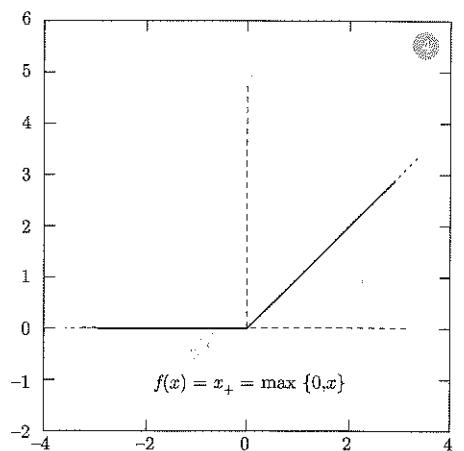
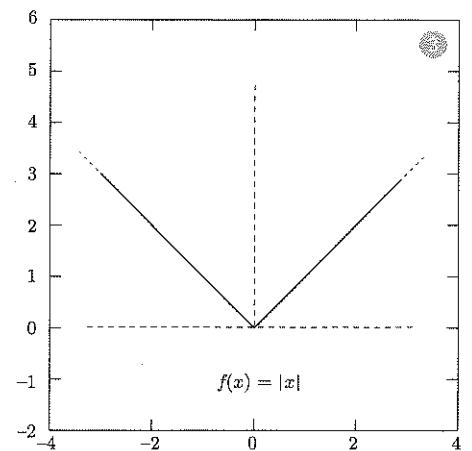
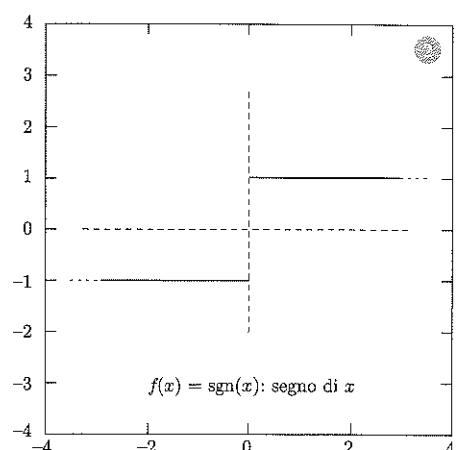
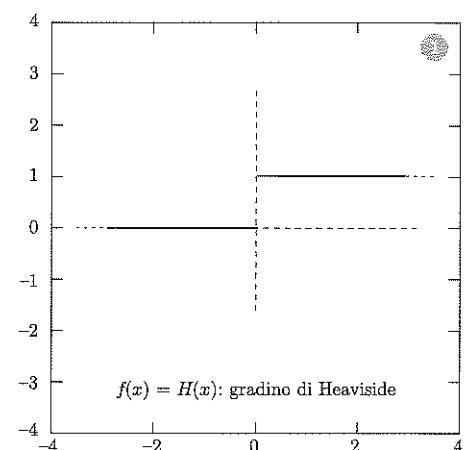
## 7 TABELLA DI PRIMITIVE

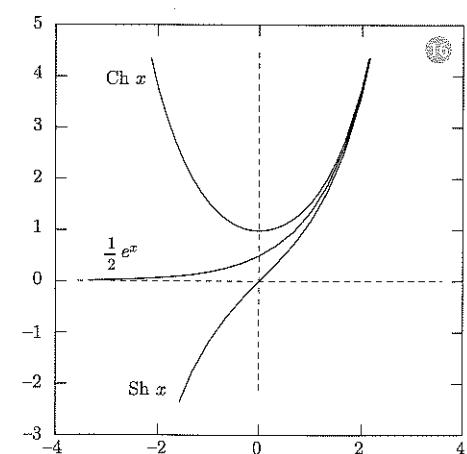
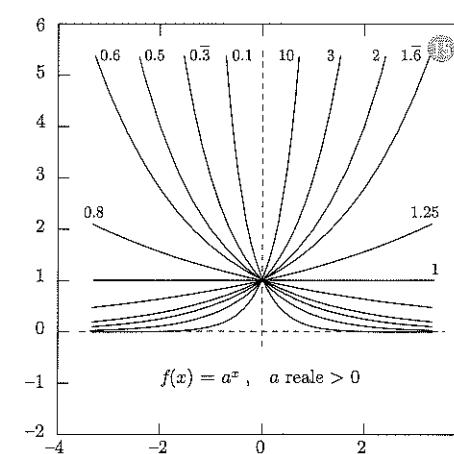
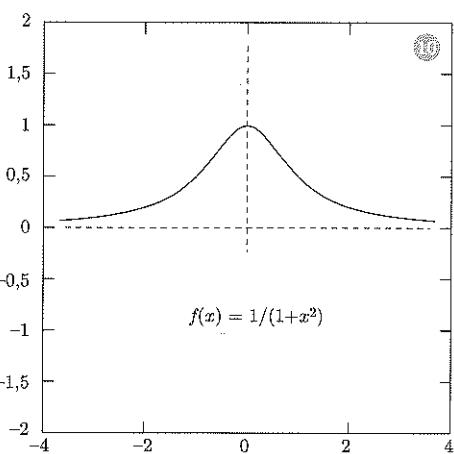
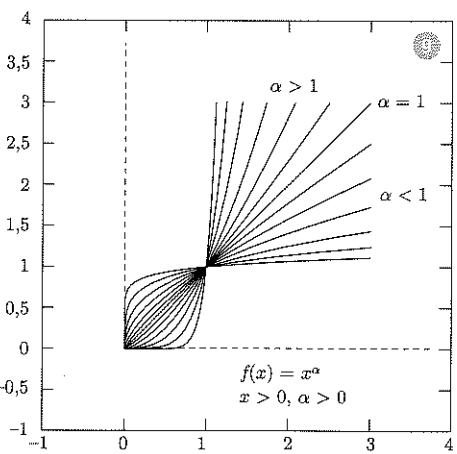
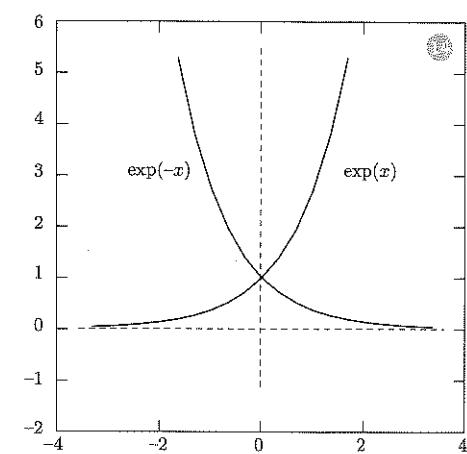
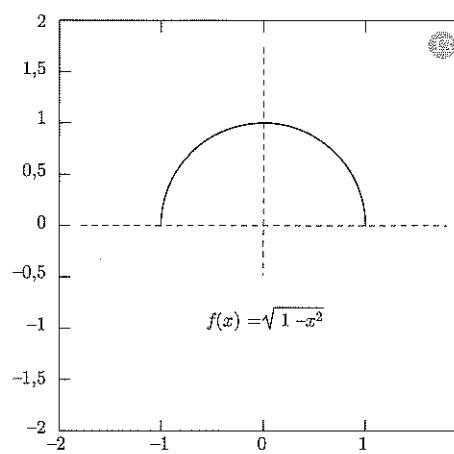
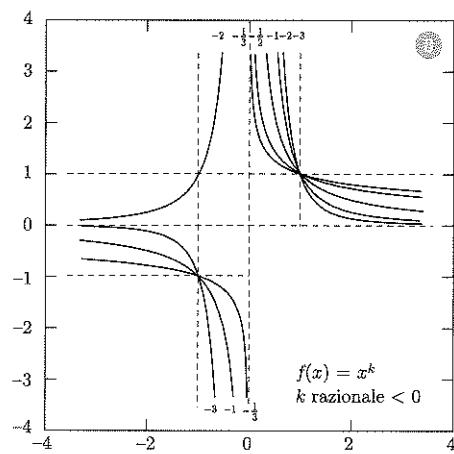
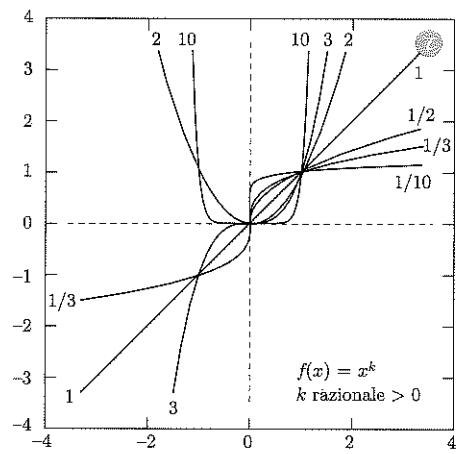
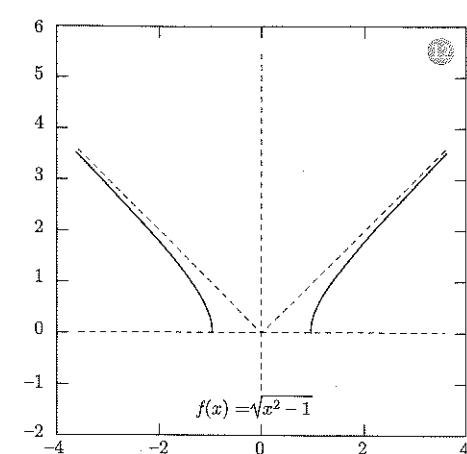
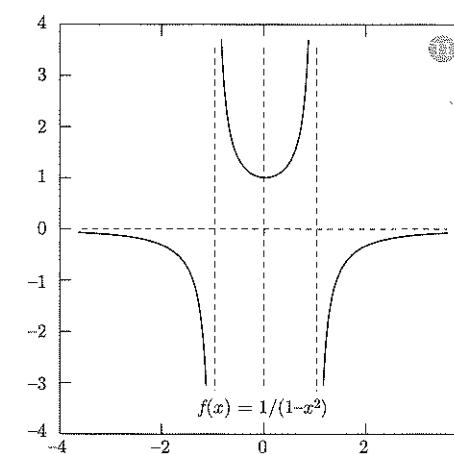
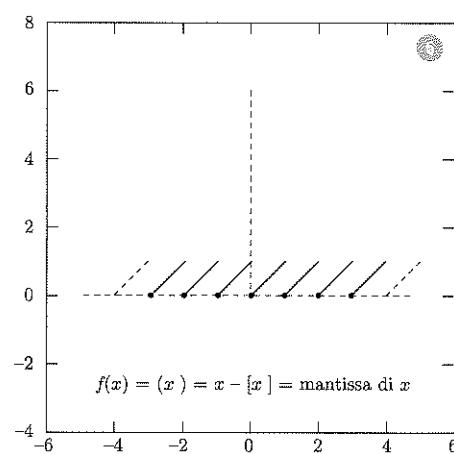
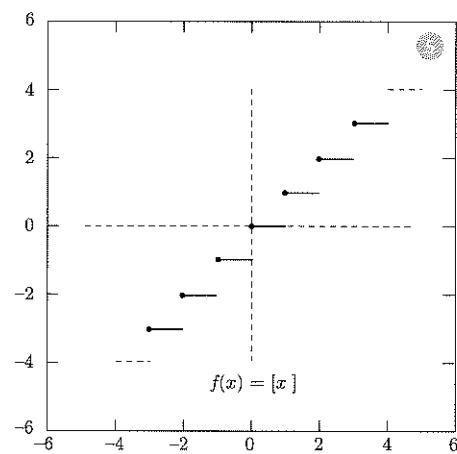
$f(x)$	$F(x)$
$x^\alpha$	$\frac{x^{\alpha+1}}{(\alpha+1)}$ se $\alpha \neq -1$
$\frac{1}{x}$	$\log x $
$e^x$	$e^x$
$a^x$	$\frac{a^x}{\log a}$ se $a > 0, a \neq 1$
$\sin x$	$-\cos x$
$\cos x$	$\sin x$
$\operatorname{tg} x$	$-\log \cos x $
$\operatorname{cotg} x$	$-\log \sin x $
$\operatorname{Sh} x$	$\operatorname{Ch} x$
$\operatorname{Ch} x$	$\operatorname{Sh} x$
$\operatorname{Th} x$	$\log \operatorname{Ch} x$
$\operatorname{Cth} x$	$\log \operatorname{Sh} x $
$\frac{1}{(\cos x)^2}$	$\operatorname{tg} x$
$\frac{1}{(\sin x)^2}$	$-\operatorname{cotg} x$
$\frac{1}{(\operatorname{Ch} x)^2}$	$\operatorname{Th} x$
$\frac{1}{(\operatorname{Sh} x)^2}$	$-\operatorname{Cth} x$
$\frac{1}{1+x^2}$	$\operatorname{arctg} x$
$\frac{1}{1-x^2}$	$\frac{1}{2} \log \frac{1+x}{1-x}$ se $ x  < 1$
$\frac{1}{1-x^2}$	$\frac{1}{2} \log \frac{x+1}{x-1}$ se $ x  > 1$
$\frac{1}{\sqrt{1-x^2}}$	$\operatorname{arcsin} x$
$\frac{1}{\sqrt{1+x^2}}$	$\log x+\sqrt{1+x^2} $
$\frac{1}{\sqrt{x^2-1}}$	$\log x+\sqrt{x^2-1} $ se $ x  > 1$

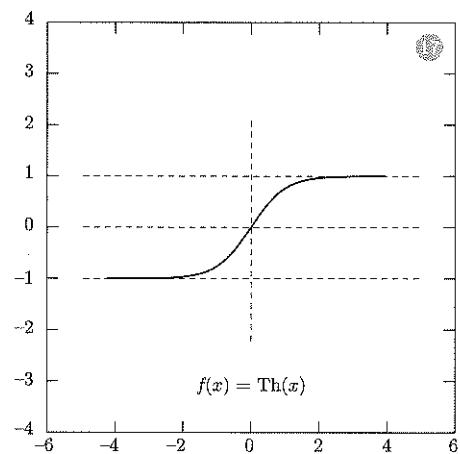
$f(x)$	$F(x)$
$f^\alpha f'$	$\frac{f^{\alpha+1}}{(\alpha+1)}$ se $\alpha \neq -1$
$\frac{f'}{f}$	$\log  f $
$\frac{f'}{1+f^2}$	$\operatorname{arctg} f$
$\frac{f'}{\sqrt{1-f^2}}$	$\arcsin f$

## B

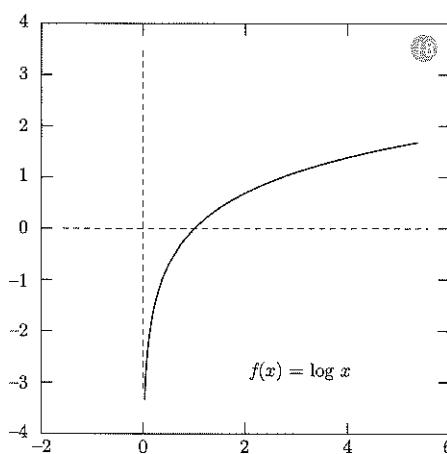
## Grafici



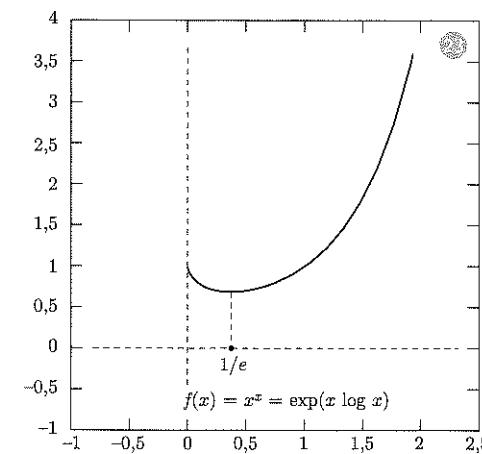




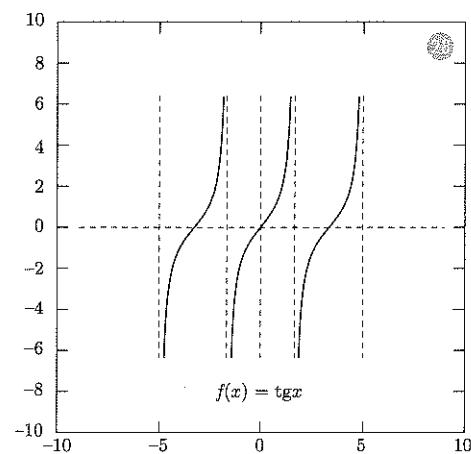
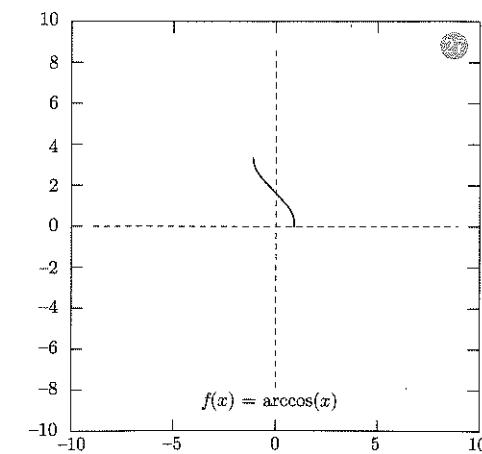
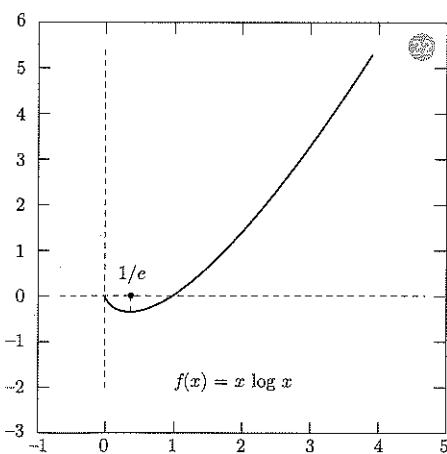
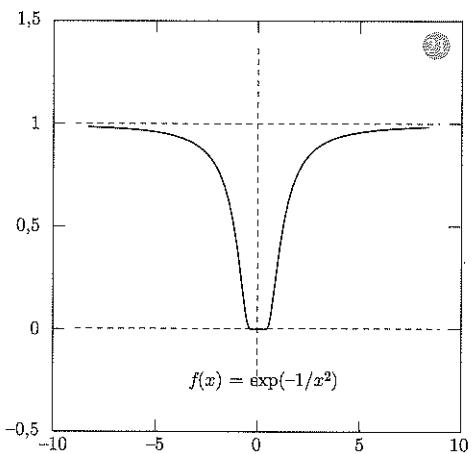
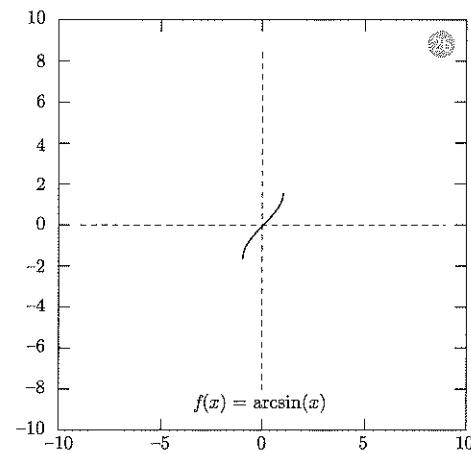
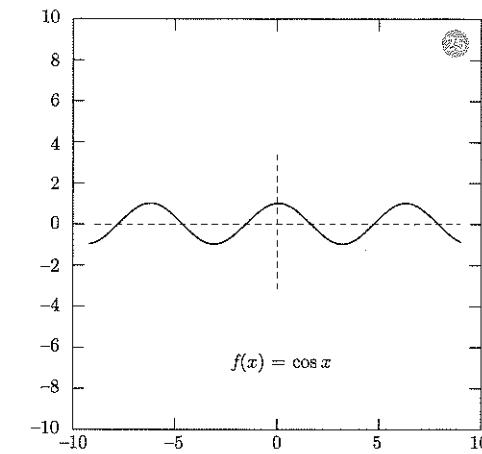
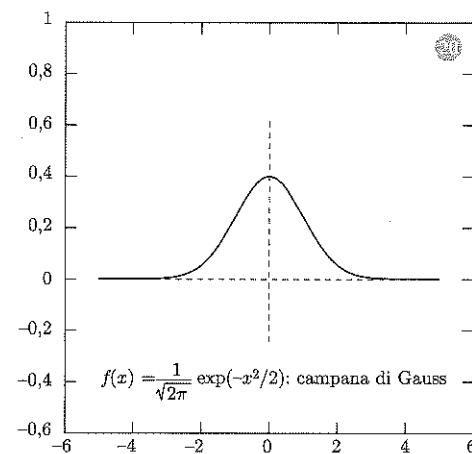
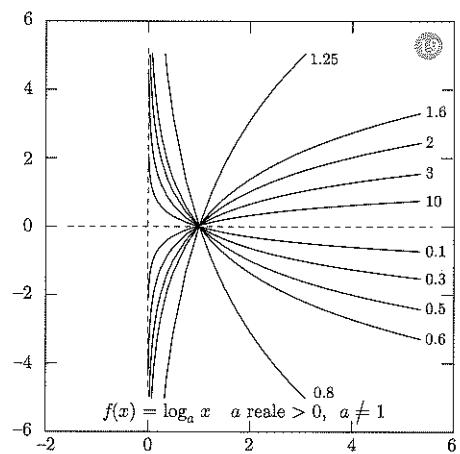
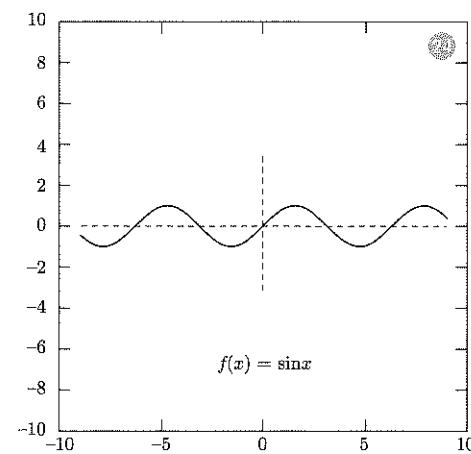
© 978-88-08-25421-4

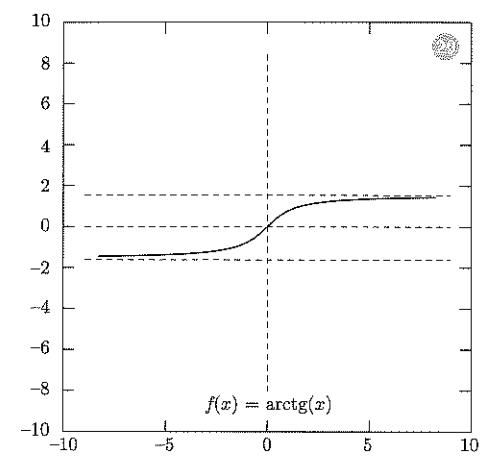


© 978-88-08-25421-4

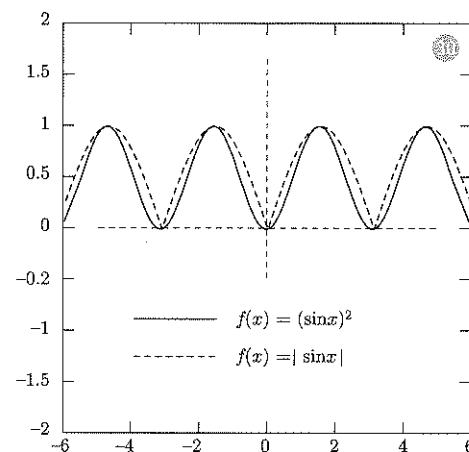


Appendice B. Grafici 519

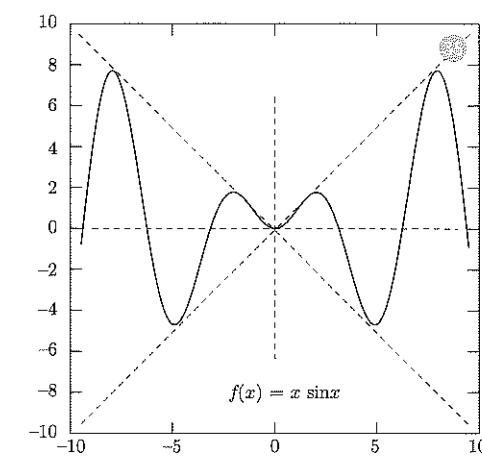




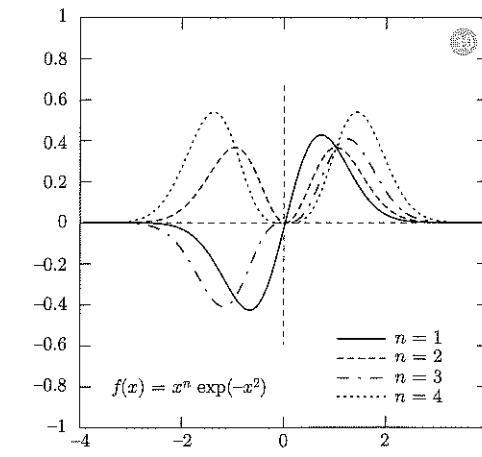
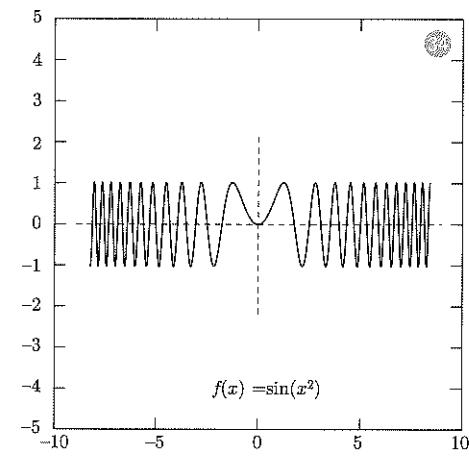
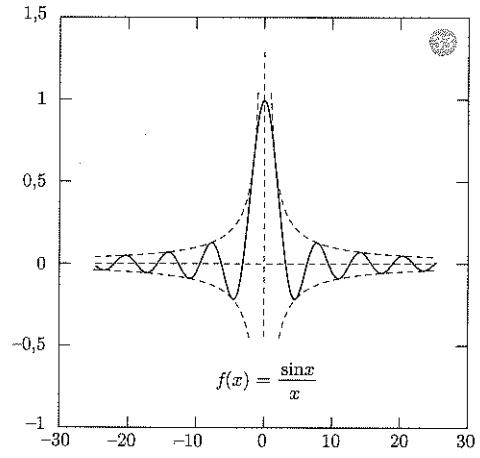
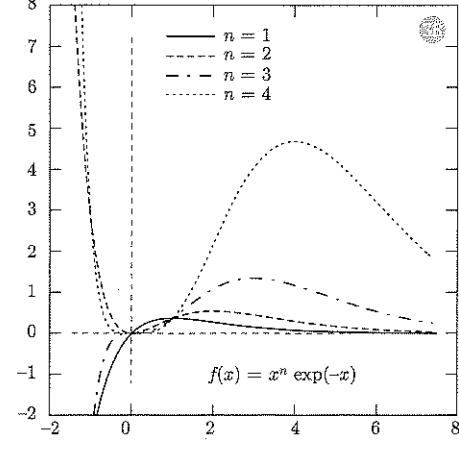
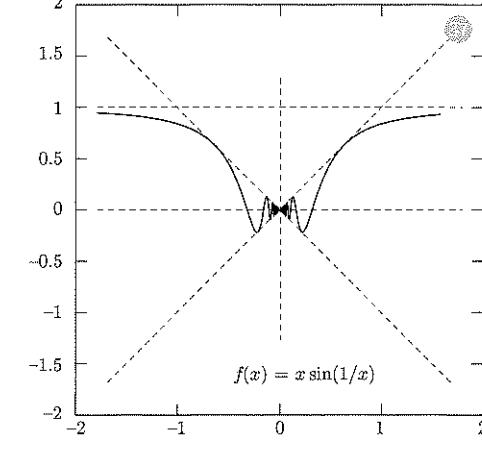
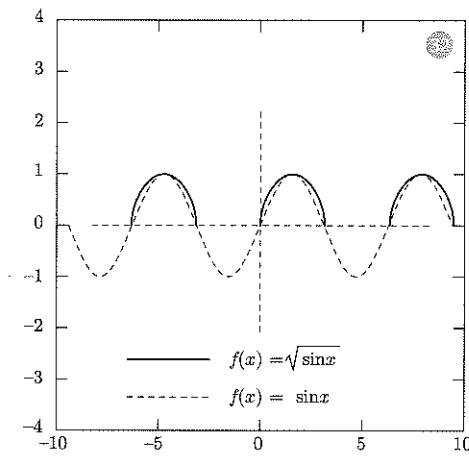
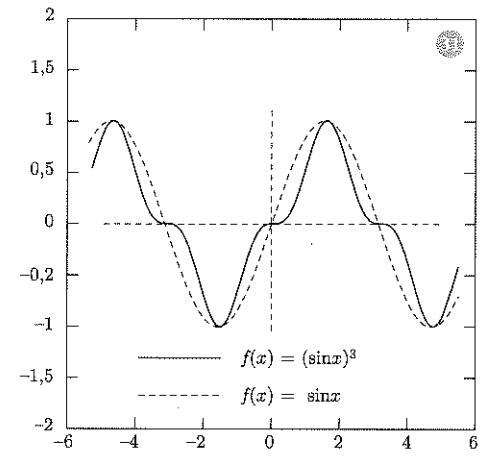
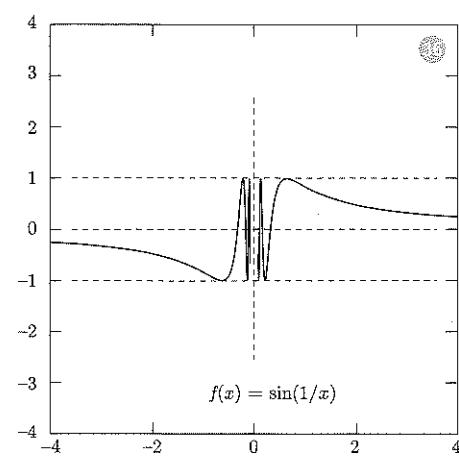
© 978-88-08-25421-4



© 978-88-08-25421-4



Appendice B. Grafici 521



# Indice analitico

## A

accelerazione, 153  
ampiezza, 64  
angolo polare, 40  
antiderivata, 267  
antisimmetrica, proprietà, 19  
appartenenza, 1  
approssimazione, 28  
– lineare, 208  
– polinomiale, 213  
arcocoseno, 80  
arcoseno, 80  
arcotangente, 81  
area di una figura piana, 260  
argomento  
– di un numero complesso, 40  
– principale, 41  
aritmetizzazione parziale del simbolo di infinito, 100, 122  
arrotondamento, 28  
asintotico  
– per funzioni, 130, 212  
– per successioni, 104  
asintoto  
– obliquo, 114  
– orizzontale, 112  
– verticale, 116  
asse  
– immaginario, 37  
– reale, 37  
assioma  
– di continuità, 22  
– di Dedekind, 22  
autospazio, 458  
autovalore, 454  
– regolare, 458  
– semplice, 458  
autovettore, 454

## B

bacino di attrazione, 336  
base  
– canonica, 389  
– di uno spazio vettoriale, 388  
– naturale, 62, 102  
– ortonormale, 394  
Bernoulli, diseguaglianza di, 33  
biforcazione, 346

## C

calcolo numerico di un integrale, 294  
campo, 20  
– di forze, 373  
– ordinato, 20  
Cantor procedimento diagonale di, 30  
caos deterministico, 343  
carattere della serie, 231  
caratteristica di una matrice, 420  
cardinalità, 29  
catenaria, 70  
Cauchy  
– condizione di, 315  
– successione di, 315  
cilindro, 500  
codominio, 51  
coefficiente  
– angolare, 152  
– binomiale, 16  
coefficienti binomiali, 16  
combinazione lineare, 388  
– di vettori, 367  
complementare, 7  
complemento algebrico, 414  
complesso coniugato, 37  
componente vettoriale, 370  
componenti scalari, 364, 388  
condizione di Cauchy, 315  
condizioni iniziali, 329  
congiunzione logica, 8

conica, 468, 474  
 – degenere, o riducibile, 479  
 – direttrice di una, 485  
 – eccentricità di una, 485  
 – fuoco di una, 485  
 coniugio, 38  
 connessione, 25  
 cono a due falde, 467  
 continuità uniforme, 316  
 controesempio, 10  
 controinversa, implicazione, 12  
 convergenza  
 – assoluta, 239  
 – semplice, 239  
 convoluzione, 310  
 coordinate polari, 40  
 corrispondenza biunivoca, 29  
 coseni  
 – direttori del piano, 378  
 – direttori della retta, 375  
 coseno, 63  
 – iperbolico, 69  
 costante di Feigenbaum, 347  
 cotangente, 63  
 crescita  
 – lineare, 134  
 – sopralineare, 134  
 – sottolineare, 134  
 curvatura, 195  
 cuspidi, 158

**D**

De Moivre, formule di, 42  
 De Morgan, leggi di, 9  
 Dedekind, assioma di, 22  
 definitivamente, 89, 119  
 derivata, 152  
 – destra, 158  
 – di Schwarz, 342  
 – logaritmica, 166  
 – prima, 152  
 – sinistra, 158  
 determinante, 413, 415  
 – di una matrice, 413  
 diagonale principale, 406  
 diagramma  
 – a gradino, 334  
 – di fase, 334  
 differenza, tra due insiemi, 6  
 differenziale, 208  
 dimensione

– di uno spazio vettoriale, 388  
 dimostrazione  
 – indiretta, 11  
 – per assurdo, 12  
 – per induzione, 32  
 dinamica, caotica, 343  
 dipolo, elettrico, 210  
 Dirac, impulso di, 311  
 direttrice  
 – della parabola, 470  
 – di una conica, 485  
 direzione di un vettore, 361  
 discontinuità a salto, 117  
 disgiunzione logica, 8  
 disuguaglianza  
 – di Bernoulli, 33  
 – di Cauchy-Schwartz, 391  
 – triangolare, 391  
 – in  $\mathbb{C}$ , 38  
 – in  $\mathbb{R}$ , 24  
 dominio, 50

**E**

eccentricità di una conica, 485  
 elasticità, 166  
 elementi pivotali o pivot, 443  
 elemento, 1  
 ellisse, 468  
 ellissoide  
 – immaginario, 497  
 – reale, 494  
 energia cinetica, 211  
 enunciato, 9  
 epigrafico, 196  
 equazione  
 – alle differenze  
 – autonoma, 333  
 – del prim'ordine, 329  
 – del second'ordine, 351  
 – di ordine  $k$ , 328  
 – autonoma, 329, 333  
 – caratteristica, 352, 454  
 – del piano, 377  
 – della retta, 374  
 – differenziale, 156  
 – logistica discreta, 326  
 – non omogenea, 356  
 – omogenea, 329, 351  
 esponenziale complesso, 250  
 estremo, 172  
 – inferiore, 22

– superiore, 22  
 – – proprietà dell', 22  
 Eulero, formule di, 250

**F**

fattoriale, 15  
 fattorizzazione LU di una matrice, 445  
 Feigenbaum, costante di, 347  
 Fibonacci, numeri di, 354  
 flesso a tangente verticale, 158  
 forma  
 – di indecisione, 100  
 – esponenziale dei numeri complessi, 251  
 – trigonometrica dei numeri complessi, 41  
 forma algebrica dei numeri complessi, 36  
 formula  
 – di De Moivre, 42  
 – di MacLaurin, 214  
 – di Newton, 16, 33  
 – di Taylor, 216  
 formule di Eulero, 250  
 frequenza, 64  
 Fresnel  
 – funzioni integrali di, 306  
 – integrale di, 304  
 funzione, 49  
 – composta, 75  
 – concava, 196  
 – – strettamente, 198  
 – continua, 117  
 – convessa, 196  
 – – strettamente, 198  
 – crescente, 55  
 – – strettamente, 55  
 – decrescente, 55  
 – – strettamente, 55  
 – definita a tratti, 74  
 – degli errori, 305  
 – derivabile, 152  
 – discontinua, 117  
 – dispari, 54  
 – elementare, 56  
 – esponenziale, 61  
 – generatrice, 334  
 – infinitesima, 111  
 – integrabile, 259, 296, 300  
 – – assolutamente, 299, 304  
 – integrale, 305  
 – – di Fresnel, 306  
 – inversa, 77, 143  
 – invertibile, 77, 143

– limitata, 53  
 – lineare, 399  
 – logaritmo, 61  
 – monotona, 55  
 – non crescente, 55  
 – non decrescente, 55  
 – pari, 54  
 – periodica, 55  
 – potenza, 56  
 – prolungamento per continuità di una, 129  
 – simmetrica, 54  
 – uniformemente continua, 316

funzione  
 – crescente, 176  
 – decrescente, 176

funzioni  
 – iperboliche, 69  
 – – inverse, 82  
 – trigonometriche, 63  
 – – inverse, 80  
 – asintotiche, 130  
 fuochi  
 – dell'ellisse, 469  
 – dell'iperbole, 471  
 fuoco  
 – della parabola, 470  
 – di una conica, 485

**G**

Gauss  
 – metodo di eliminazione di, 443  
 – piano di, 37  
 gerarchia degli infiniti  
 – per funzioni, 132  
 – per successioni, 105

**I**

immagine, 51  
 – di una trasformazione lineare, 433  
 implicazione, 3  
 – controinversa, 12  
 – universale, 9, 11  
 impulso, di Dirac, 311  
 inclusione, 3, 11  
 – stretta, 3  
 indice  
 – di sommatoria, 13  
 – muto, 14  
 indipendenza lineare, 367, 388  
 infinitesimi dello stesso ordine, 104

infinitesimo di ordine  
 – inferiore, 104  
 – superiore, 104

infiniti dello stesso ordine, 104

infinito di ordine

– inferiore, 104

– superiore, 104

insieme, 1

– convesso, 196

– delle parti, 5

– infinito, 28

– limitato, 21

– – inferiormente, 21

– – superiormente, 21

– ordinato, 19

– totalmente ordinato, 19

– vuoto, 4

insiemi operazioni su, 6

instabilità

– di un punto d'equilibrio, 336

– di un'orbita periodica, 341

integrabilità delle funzioni continue, 318

integrale

– convergente, 296, 300

– definito, 258, 305

– di Fresnel, 304

– divergente, 296, 300

– generalizzato, 296

– indefinito, 269, 305

integrazione

– delle funzioni irrazionali, 285

– delle funzioni razionali, 273

– delle funzioni trigonometriche, 281

– per parti, 277

– per scomposizione, 270

– per sostituzione, 270

intersezione di due insiemi, 6

intervallo, 24

– di periodicità, 55

intorno, 119

invarianti

– di una conica, 476

– di una quadrica, 491

iperbole, 468

– equilatera, 471

– fuochi dell', 471

iperboloida

– a due falde, 496

– a una falda, 494

## L

Lagrange, resto secondo, 219

lavoro di una forza, 289, 371

legge delle controinverse, 12

leggi di De Morgan, 9

limite, 90

– destro, 115

– di funzione

– – definizione successionale, 111

– – definizione topologica, 119

– – per difetto, 112

– – per eccesso, 112

– di una successione, 90

– – per difetto, 93

– – per eccesso, 93

– sinistro, 115

limiti notevoli, 128

linearizzazione, 208, 337

logaritmo, 27

– nel campo complesso, 251

lunghezza di un grafico, 290

## M

MacLaurin

– formula di, 214

– polinomio di, 214

maggiorante, 22

Malthus, modello di, 324

mantissa, 68

massimo

– di funzione, 138

– di un insieme, 21

– di una funzione, 172

– locale, 172

matrice, 402, 403

– caratteristica di una, 420

– d'inerzia, 464

– d'inerzia (o tensore d'inerzia), 412

– dei coefficienti della conica, 475

– di passaggio, 453

– diagonale, 407

– diagonalizzabile, 453

– fattorizzazione LU di una, 445

– identità, 406

– inversa, 422

– quadrata, 403

– rango di una, 421

– simmetrica, 406

– traccia della, 454

– trasposta, 406

– triangolare

– alta, 406

– bassa, 406

matrici

– a blocchi, 476

– conformabili, 405

– di rotazione, 462

– equivalenti, 452

– ortogonali, 460

– – speciali, 462

– simili, 452

media

– aritmetica, 176

– geometrica, 176

Mengoli, serie di, 232

metodo

– del punto medio, 294

– di eliminazione di Gauss, 443

– di Newton, 221

– di somiglianza, 357

minimo

– di funzione, 138

– di un insieme, 21

– di una funzione, 172

– locale, 172

minorante, 22

minore complementare, 414

modello

– dell'acceleratore, 327

– della ragnatela, 332

– di Malthus, 324

– logistico, 326

– matematico, 323

modulo, 23

– di un numero complesso, 38

– di un vettore, 361, 391

molteplicità di una radice di un'equazione, 46

molteplicità algebrica di un autovalore, 458

molteplicità geometrica di un autovalore, 458

moltiplicazione di un vettore per uno scalare, 363

momento di una forza, 372

movimento rigido, 472

## N

negazione, 11

– logica, 8

Nepero, numero e di, 62, 101

Newton

– formula di, 16, 33

– metodo di, 221

norma di un vettore, 391

normalizzazione, 364, 391

nucleo di una trasformazione lineare, 434

numerabile, 29

numeri complessi, 6, 35

– forma algebrica, 36

– – di Fibonacci, 354

– immaginari puri, 37

– interi, 5

– naturali, 5

– razionali, 5, 18

– reali, 5, 18

– – definizione assiomatica, 22

numero e di Nepero, 62, 101

## O

o piccolo, 209, 212

operatore lineare, 399

orbita, 334

– instabile, 341

– periodica, 340

– stabile, 341

– – asintoticamente, 341

ordinamento di Sharkowsky, 342

ortonormalizzazione, 395

oscillazione

– amplificata, 67

– smorzata, 67

## P

parabola, 468

– direttrice della, 470

– fuoco della, 470

paraboloida

– a sella, 499

– ellittico, 498

– iperbolico, 499

parte

– decimali, 68

– immaginaria

– – di un numero complesso, 36

– intera, 68

– reale

– – di un numero complesso, 36

Peano, resto secondo, 215, 216

pendenza, 152

pendolo, 211

periodo, 55, 340

permutazione, 16

piano

- complesso, 37
- di Gauss, 37
- polinomio**
  - caratteristico, 454
  - di MacLaurin, 214
  - di Taylor, 216
- potenza**
  - a esponente complesso, 252
  - a esponente razionale, 59
  - a esponente reale, 61
  - del continuo, 31
  - di un insieme, 29
- potenze a esponente razionale o reale**, 26
- predicato**, 9
- principio di induzione**, 32
- procedimento diagonale di Cantor**, 30
- prodotto**
  - cartesiano, 7
  - interno, 369, 392
  - misto, 373, 419
  - scalare, 391, 392
  - - di vettori, 369
  - - in  $\mathbb{R}^n$ , 391
  - - nel piano, 370
  - - nello spazio, 370
  - tra matrici, 404
  - vettoriale, 371, 419
- progressione geometrica**, 15
- proiezione ortogonale**, 397
- prolungamento per continuità di una funzione**, 129
- proposizione**, 9
- proprietà**, 9
  - dell'estremo superiore, 22
  - geometrica di una curva, 473
- pulsazione**, 64
- punto**
  - angoloso, 157
  - di equilibrio, 335
  - di flesso, 200
  - - a tangente orizzontale, 174
  - di massimo, 138, 172
  - locale, 172
  - di minimo, 138, 172
  - fisso, 335
  - instabile, 336
  - iperbolico, 339
  - periodico, 340
  - semi-stabile, 339

- stabile, 336
- - asintoticamente, 336
- stazionario, 174
- Q**
  - quadrica, 488
  - a centro, 492
  - degenere, 491
  - invarianti di una, 491
  - quadriche non a centro, 497
  - quantificatore
    - universale, 3
    - esistenziale, 4
- R**
  - radice  $n$ -esima
    - aritmetica, 25
    - complessa, 43
  - raggio
    - di curvatura, 196
    - polare, 40
  - rango di una matrice, 421
  - rapporto incrementale, 151
  - regola di Sarrus, 418
  - relazione
    - d'ordine, 19
    - di equivalenza, 105
  - repulsore, 338
  - resto
    - secondo Lagrange, 219
    - secondo Peano, 215, 216
  - retta tangente, 152
  - rette
    - ortogonali, 376
    - parallele, 376
  - riflessione di indici, 14
  - riflessiva, proprietà, 19, 105
  - rotazioni
    - dello spazio, 464
    - nel piano, 407
- S**
  - scala semilogaritmica, 166
  - scalari, 387
  - Schwarz, derivata di, 342
  - sella, 499
  - seno, 63
    - iperbolico, 69
  - serie, 230
    - a segni alterni, 240
    - a termini di segno variabile, 239
    - a termini non negativi, 233

- armonica, 232, 301
- - generalizzata, 234, 302
- carattere della, 231
- convergente, 231
- di funzioni, 245
- di Mengoli, 232
- di potenze, 248
- di Taylor, 245
- divergente, 231
- esponenziale, 246
- geometrica, 231
  - - nel campo complesso, 249
- irregolare, 231
- maggiorante, 234
- minorante, 234
- somma della, 231
- telescopica, 232
- termine generale della, 231
- settore**
  - coseno iperbolico, 82
  - seno iperbolico, 82
- sezione aurea**, 356
- conica, 468
- Sharkowsky, ordinamento di, 342
- simmetrica, proprietà**, 105
- sistema**
  - fisico lineare, 310
  - lineare, 426
  - omogeneo, 427
- sistemi**
  - indeterminati, 428
  - impossibili, 428
- soluzione di equilibrio**, 330
- somma**
  - della serie, 231
  - di Cauchy-Riemann, 259
  - di vettori, 362
  - parziale, 230
- somma diretta di sottospazi ortogonali**, 398
- sommatoria**, 13
- sottoinsieme**, 3
- sottospazi ortogonali**, 393
- sottospazio**
  - generato da una famiglia di vettori, 390
  - vettoriale, 387
- sottosuccessione**, 314
- spazi di funzioni**, 390
- spazio vettoriale**, 386, 392
- stabilità**
  - asintotica
    - - di un punto d'equilibrio, 336

- - di un'orbita periodica, 341
- di un punto d'equilibrio, 336
- di un'orbita periodica, 341
- successione**, 87
  - convergente, 89
  - definita per ricorrenza, 222
  - delle somme parziali, 230
  - di Cauchy, 315
  - divergente, 91
  - geometrica, 96
  - indeterminata, 92
  - infinita, 92
  - infinitesima, 92
  - irregolare, 92
  - limitata, 88, 89
    - - inferiormente, 88
    - - superiormente, 88
    - - inferiormente, 89
    - - superiormente, 89
  - monotona, 93, 181
- superficie**
  - quadrica, 488
  - rigata, 495
- T**
  - tangente**, 63
    - iperbolica, 69
  - Tartaglia, triangolo di, 17
  - tasso**
    - di incremento relativo, 166
    - di variazione, 151
  - Taylor**
    - formula di, 216
    - polinomio di, 216
    - serie di, 245
  - tensore**, 412
  - termine**
    - generale
      - - della serie, 231
    - traccia della matrice, 454
    - transitiva, proprietà, 19, 105
  - trasformazione**
    - affine, 401
    - biiettiva, 435
    - iniettiva, 435
    - lineare, 399
      - - affine, 401
      - - ortogonale, 472
      - suriettiva, 435
    - traslazione, 473
      - di indici, 14

triangolo di Tartaglia, 17

**U**

uguaglianza tra insiemi, 3

unione di due insiemi, 6

unità immaginaria, 36

**V**

valore

– assoluto, 23

– di un numero complesso, 38

– efficace, 266

– medio, 266

– principale

– del logaritmo, 251

– di una potenza, 252

variabile

– d'integrazione, 260

– di stato, 328

– dipendente, 50

– indipendente, 50

– logica, 9

– muta, 260

velocità, 153

verso di un vettore, 361

versore, 364, 391

vettore

– posizione, 367

– nel piano o nello spazio, 361

vettori, 387

–  $n$ -dimensionali: lo spazio  $\mathbb{R}^n$ . Spazi vettoriali astratti, 385

– linearmente indipendenti, 367

– colonna, 404

– di uno spazio vettoriale astratto, 387

– linearmente indipendenti, 367

– nello spazio a  $n$  dimensioni, 385

– riga, 404

vibrazione elementare, 64

volume di un solido di rotazione, 292