Valter Moretti

Dipartimento di Matematica, Università di Trento

https://sites.google.com/unitn.it/valter-moretti/home
https://webapps.unitn.it/du/en/Persona/PER0004166/Curriculum

Matematica per Filosofia

Dispense del corso – II anno

Corso tenuto presso il Dipartimento di Lettere e Filosofia, Trento
Anno accademico 2026–2027

Indice

Introduzione, nozioni e notazioni

Questo testo si riferisce a un corso di *matematica per studenti e studentesse di filosofia*, non a un corso di *filosofia della matematica*. Il suo scopo principale è introdurre al linguaggio formale della matematica, ad alcuni suoi contenuti fondamentali e alle procedure con cui si definiscono oggetti, si formulano enunciati, si costruiscono argomentazioni e si dimostrano risultati. Un obiettivo è che, al termine del percorso, chi è interessato in particolare alla filosofia della scienza sia in grado di comprendere almeno nelle linee generali il linguaggio in cui sono scritti un libro o un articolo delle cosiddette scienze dure, come la fisica o la matematica stessa. In questi ambiti la matematica costituisce oggi uno dei linguaggi fondamentali per descrivere, definire e argomentare con precisione. Un altro obiettivo forse più abbordabile è quello di introdurre le studentesse e gli studenti ad alcuni aspetti fondazionali della matematica *facendo matematica*, mostrando dove i problemi nascono a causa delle pratiche matematiche.

Poiché il corso è rivolto a studentesse e studenti di filosofia, le questioni filosofiche e storiche, così come i collegamenti con le altre scienze, che emergono nello sviluppo e nella fondazione della matematica saranno brevemente considerate quando risultino utili a comprendere la materia. Esse non costituiscono tuttavia l'oggetto principale del corso: uno studio sistematico della filosofia e della storia della matematica e dei suoi problemi filosofici fondazionali, così come la relazione tra matematica e logica, appartengono a corsi successivi e più specialistici.

Si assume che la studentessa o lo studente sappia usare argomentazioni di logica elementare, in particolare quelle che coinvolgono connettivi, quantificatori, implicazioni ed equivalenze. Tali procedure sono quelle che comunemente usiamo in modo informale nel ragionamento pratico del matematico di professione. Per una loro presentazione a livello formale si rimanda a successivi corsi di Logica. Le notazioni logiche utilizzate più frequentemente sono comunque richiamate qui sotto, così da fissare una convenzione comune per tutto il testo. La nozione di assiomatizzazione, specialmente nel caso concreto della teoria degli insiemi, verrà brevemente introdotta nel primo capitolo. Per approfondimenti si rimanda al corso di Logica e di Filosofia della Matematica.

Queste dispense sono strutturate in modo simile a molti testi di matematica: le nozioni vengono introdotte mediante *definizioni* precise, gli enunciati matematici sono formulati come *proposizioni* o *teoremi* e, quando opportuno, sono accompagnati da *dimostrazioni* formali. Questa scelta non ha soltanto una funzione espositiva, ma costituisce parte dell'intento didattico del corso: acquisire progressivamente familiarità con il linguaggio, la struttura e le forme di argomentazione proprie della matematica. *Esempi* ed *esercizi* sono parti integranti del corso e non possono essere omessi: servono sia a rendere operative le definizioni sia a verificare la comprensione effettiva dei procedimenti introdotti. Definizioni, osservazioni, esempi, proposizioni, teoremi ed esercizi condividono inoltre un unico contatore progressivo all'interno del capitolo: questa scelta rende più semplice individuare e richiamare rapidamente ciascun elemento numerato del testo. Le formule particolarmente importanti sono numerate separatamente all'interno di ciascun capitolo e vengono richiamate nel testo mediante il loro numero; le altre formule restano non numerate.

Le dispense contengono deliberatamente molto più materiale di quanto possa essere svolto integralmente durante il corso. Questa scelta permette di costruire di anno in anno percorsi parzialmente diversi, privilegiando alcuni argomenti rispetto ad altri e senza necessariamente dimostrare tutti i risultati presentati. Il testo vuole comunque offrire un riferimento generale e relativamente organico, al quale sia possibile tornare anche per le parti non affrontate direttamente a lezione.

L'ideazione e la scrittura di queste dispense sono state sviluppate anche con l'aiuto e attraverso una continua interazione con un sistema di intelligenza artificiale, impiegato nella discussione della struttura, nella formulazione e revisione del testo e nella preparazione di esempi ed esercizi. Le scelte matematiche, didattiche ed editoriali finali restano responsabilità dell'autore.

Le dispense sono articolate in tre capitoli. Il Capitolo 1 introduce la teoria degli insiemi, le relazioni e le funzioni, fino alla formulazione assiomatica ZF e alla costruzione dei numeri naturali. Il Capitolo 2 costruisce i numeri reali a partire dai naturali e introduce le prime nozioni dell'analisi matematica: limiti, continuità, derivabilità, equazioni differenziali, integrazione e misura. Il Capitolo 3 introduce alcune strutture algebriche fondamentali: monoidi, gruppi, anelli, campi, spazi vettoriali reali e complessi, spazi affini e algebre associative.

Notazioni e logica elementare

Nel seguito, quando un nuovo concetto viene introdotto e definito, il termine corrispondente è evidenziato in **neretto**. I nomi di risultati, principi, paradossi e teorie citati come tali sono invece normalmente scritti in *corsivo*. Il corsivo è anche usato per enfatizzare affermazioni o concetti.

Il simbolo := si legge "è definito come" e indica una definizione: scrivere $A := \{1, 2, 3\}$ significa che A viene definito come l'insieme $\{1, 2, 3\}$. Useremo la freccia sottile $\to$ per *funzioni* e *applicazioni*: $f : A \to B$ si legge "f da A in B" come vedremo più avanti. Lo stesso simbolo $\to$, quando compare fra formule, sarà usato anche per il connettivo di *implicazione materiale*; il contesto rende distinti i due usi.

In tutto il testo adotteremo inoltre le seguenti convenzioni. L'insieme dei *numeri naturali* è

$$\mathbb{N} := \{0, 1, 2, 3, \ldots\};$$

l'insieme dei *numeri interi* è

$$\mathbb{Z} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\};$$

l'insieme dei *numeri razionali* è

$$\mathbb{Q} := \left\{ \frac{p}{q} \,\middle|\, p \in \mathbb{Z},\ q \in \mathbb{Z},\ q \neq 0 \right\}.$$

Dunque $\mathbb{Q}$ comprende i razionali negativi, lo zero e i razionali positivi. Quando vorremo indicare soltanto i *numeri razionali strettamente positivi*, scriveremo

$$\mathbb{Q}_{>0} := \{r \in \mathbb{Q} \mid r > 0\}.$$

Useremo infine $\mathbb{R}$ per l'insieme dei *numeri reali* che definiremo con precisione nel secondo capitolo.

Useremo liberamente alcuni simboli della logica elementare. Le lettere P, Q indicano *proposizioni* del linguaggio, cioè espressioni alle quali è attribuito un valore di verità: *vero* oppure *falso*. Usando *connettivi* $\vee, \wedge, \neg$ come indicato sotto, possiamo costruire altre proposizioni complesse del linguaggio, tecnicamente si parla di *formule*. Queste formule assumono valori di verità che sono funzioni dei valori di verità delle proposizioni che le compongono.

La scrittura $F(x)$ indica invece una *formula* che dipende dalla *variabile* x; il dominio nel quale varia x sarà sempre indicato o chiaro dal contesto.

Nella tabella che segue per ciascuna notazione indichiamo sia come si legge normalmente l'enunciato, sia il suo significato logico.

Scrittura	Si legge	Significato
$P \vee Q$	"P o Q"	connettivo **disgiunzione** di P e Q, è la formula che ha valore *vero* se almeno una fra P e Q ha valore vero e *falso* negli altri casi.
$P \wedge Q$	"P e Q"	connettivo **congiunzione** di P e Q, è la formula che ha valore *vero* se sia P sia Q hanno valore *vero* e *falso* negli altri casi.
$\neg P$	"non P"	connettivo **negazione** di P, è la formula che ha valore *vero* esattamente quando P ha valore *falso* e viceversa.
$P \Rightarrow Q$	"P implica logicamente Q"; equivalentemente, "P è condizione *sufficiente* per Q"	non è qui un connettivo del linguaggio, ma una scrittura metalinguistica: significa che ogni assegnazione o interpretazione che rende vera P rende vera anche Q. Nel caso proposizionale equivale a dire che $P \to Q$ è una tautologia.
$P \iff Q$	"P e Q sono logicamente equivalenti"	è una scrittura metalinguistica: significa che valgono sia $P \Rightarrow Q$ sia $Q \Rightarrow P$; nel caso proposizionale P e Q hanno lo stesso valore di verità in ogni assegnazione.
$\forall x\, F(x)$	"per ogni x, F di x è soddisfatta"	**quantificatore universale** applicato a F. Per ogni valore ammesso di x, la formula $F(x)$ ha valore vero.
$\exists x\, F(x)$	"esiste un x tale che F di x è soddisfatta"	**quantificatore esistenziale** applicato a F. Per almeno un valore ammesso di x, la formula $F(x)$ ha valore vero.
$\exists! x\, F(x)$	"esiste uno e un solo x tale che F di x è soddisfatta"	Esiste esattamente un valore ammesso di x per il quale $F(x)$ ha valore vero.

Il connettivo di implicazione materiale

È essenziale distinguere due notazioni. La scrittura $P \Rightarrow Q$ viene usata qui nel *metalinguaggio* per affermare che P implica logicamente Q e non è, in questa convenzione, una formula ottenuta applicando un connettivo a P e Q. La scrittura

$$P \to Q,$$

invece, è una formula del *linguaggio*, analogamente a $P \vee Q$, $P \wedge Q$ e $\neg P$. Il simbolo $\to$ denota il connettivo chiamato **implicazione materiale**, definito mediante i connettivi già introdotti da

$$P \to Q := \neg P \vee Q.$$

Le **tavole di verità** dei connettivi usati più frequentemente, i quattro menzionati sopra, sono le seguenti; V indica il valore *vero* e F il valore *falso*.

Negazione

P	$\neg P$
V	F
F	V

Connettivi binari

P	Q	$P \wedge Q$	$P \vee Q$	$P \to Q$
V	V	V	V	V
V	F	F	V	F
F	V	F	V	V
F	F	F	F	V

Dalla corrispondente tavola di verità si vede che $P \to Q$ è falsa soltanto quando P è vera e Q è falsa. Nel caso proposizionale l'affermazione metalinguistica $P \Rightarrow Q$ equivale quindi a dire che la formula $P \to Q$, dove P e Q sono formule costruite da formule più elementari, è vera per ogni assegnazione dei valori di verità delle formule elementari, cioè è una *tautologia*. Per esempio, se $P = R$ e $Q = R \vee S$, allora l'implicazione materiale $R \to R \vee S$ è una tautologia. Metalinguisticamente la verità di R implica la verità di $R \vee S$.

Se $P(x)$ e $Q(x)$ sono formule dipendenti da una variabile x, per affermare nel linguaggio che ogni x che soddisfa $P(x)$ soddisfa anche $Q(x)$ scriviamo

$$\forall x \big(P(x) \to Q(x)\big).$$

Per esempio, se x varia in $\mathbb{R}$,

$$\forall x\,(x > 1 \to x > 0).$$

Se $P(x, y)$ è una formula dipendente da due variabili x, y, per affermare nel linguaggio che per ogni x esiste un unico y tale che $P(x, y)$ è soddisfatta scriviamo

$$\forall x\, \exists! y P(x, y).$$

Per esempio, se x, y variano in $\{r \in \mathbb{R} \mid r > 0\}$,

$$\forall x\, \exists! y (y^2 = x).$$

CAPITOLO 1

Teoria degli insiemi, relazioni e funzioni

1.1 Introduzione

La teoria degli insiemi prende avvio da un gesto che sembra del tutto naturale: raccogliere più oggetti e considerarli come un'unica totalità. Dietro questa operazione elementare si nascondono però domande non banali. Quando possiamo dire che una collezione costituisce davvero un oggetto? È sufficiente descrivere una proprietà perché esista l'insieme di tutte le cose che la possiedono? E che cosa accade quando la collezione che vogliamo considerare è infinita?

Queste domande spiegano perché la teoria degli insiemi occupi una posizione particolare. Da un lato fornisce il linguaggio fondazionale con cui viene formulata gran parte della matematica contemporanea, forse addirittura tutta, malgrado esistano fondazioni alternative come la *teoria delle categorie*. Dall'altro è un luogo privilegiato per osservare l'incontro fra matematica e filosofia: vi emergono continuamente le distinzioni fra oggetto e proprietà, finito e infinito, esistenza e costruzione, verità e dimostrabilità.

In questo capitolo partiremo dalle nozioni più semplici, come *appartenenza*, *inclusione*, *unione*, *intersezione* e *complemento*. L'attenzione alla scrittura formale servirà soprattutto a chiarire il ragionamento. Le identità fra insiemi verranno quindi dimostrate traducendo l'appartenenza nel linguaggio della logica elementare (che non formalizzeremo); *relazioni* e *funzioni* saranno presentate come particolari insiemi di *coppie ordinate*. Tutte le nozioni astratte saranno accompagnate da esempi concreti.

Una volta acquisito questo linguaggio, nel corso del capitolo passeremo al problema dell'*infinito* in matematica. L'idea cantoriana di confrontare due insiemi mediante corrispondenze biunivoche conduce a risultati lontani dall'intuizione finita e mostra che non tutti gli infiniti hanno la stessa grandezza. Lo stesso metodo diagonale che permette di distinguere diverse cardinalità prepara anche il terreno per comprendere i paradossi della teoria ingenua.

Il *paradosso di Russell* segna il punto di crisi: non ogni condizione linguisticamente formulabile può determinare senza restrizioni un insieme. Da qui nasce l'esigenza di un'impostazione *assiomatica*, nella quale l'esistenza degli insiemi sia regolata da principi espliciti. Studieremo l'assiomatizzazione ZF per chiudere il capitolo.

Non sono richieste conoscenze matematiche specialistiche. Gli esempi, le dimostrazioni guidate e gli esercizi servono ad acquisire familiarità con il linguaggio e con il modo di argomentare. Le soluzioni, collocate subito dopo ciascun esercizio, permettono di controllare non soltanto il risultato, ma anche la struttura dei passaggi. I riferimenti storici sono essenziali e mirano a mostrare da quali problemi siano nate le principali idee; la bibliografia finale offre alcune indicazioni per approfondirle.

1.2 La cosiddetta teoria ingenua degli insiemi

La teoria che introdurremo nelle prossime sezioni è comunemente chiamata *teoria ingenua degli insiemi*. L'aggettivo "ingenua" non significa che si tratti di una matematica approssimativa o priva di struttura. Indica piuttosto che insiemi, appartenenza e operazioni fra insiemi vengono trattati secondo il loro significato intuitivo, senza specificare fin dall'inizio un sistema formale di assiomi. Questa impostazione è quella normalmente usata nella pratica matematica ordinaria, anche da matematici, fisici, informatici, economisti, studiosi di statistica professionisti, quando non sono in gioco questioni specificamente fondazionali [5].

Storicamente, tuttavia, l'uso senza restrizioni di alcuni principi intuitivi — in particolare dell'idea che ogni proprietà determini l'insieme di tutti gli oggetti che la soddisfano — condusse alla scoperta di paradossi e alla cosiddetta *crisi dei fondamenti*. Non si tratta quindi di difficoltà che compaiono normalmente nei calcoli elementari con gli insiemi, ma di problemi che emergono quando si pretende di applicare quei principi senza limitazioni e si interroga la teoria sulle proprie basi. Più avanti analizzeremo questa crisi e la successiva rifondazione assiomatica della teoria degli insiemi, indicata con *ZF* e, quando vi si aggiunge l'*assioma della scelta*, con *ZFC*; queste sigle saranno introdotte e spiegate nel dettaglio nelle sezioni dedicate [9, 6].

È importante però anticipare che quasi tutto ciò che useremo nella trattazione ingenua — appartenenza, inclusione, unione, intersezione, complemento, prodotto cartesiano, relazioni, funzioni e le relative identità

elementari — resta perfettamente valido anche in ZF e ZFC. La rifondazione assiomatica non elimina questa matematica: precisa piuttosto quali insiemi possiamo affermare che esistano e in base a quali principi.

1.2.1 Insiemi, vuoto, singoletti e distinzione dei livelli

Nella teoria ingenua degli insiemi un **insieme** è una collezione considerata come un unico oggetto. Gli elementi possono essere numeri, persone, proposizioni o anche *altri insiemi*. Scriviamo

$$x \in A \quad \text{se } x \text{ appartiene ad } A, \qquad x \notin A \quad \text{altrimenti.}$$

Un insieme può essere descritto **per estensione**, elencandone gli elementi, oppure **per comprensione**, indicando una proprietà:

$$E := \{2, 4, 6, 8\} = \{n \in \mathbb{N} \mid n \text{ è pari} \wedge 0 < n < 10\}.$$

La barra verticale si legge "tale che". La seconda definizione di sopra mette in luce il rapporto tra linguaggio e oggetto: una formula seleziona gli elementi da un dominio già indicato. L'insieme delle vocali italiane si può definire con $V := \{a, e, i, o, u\}$. L'ordine e le ripetizioni non contano: $\{a, e\} = \{e, a\}$ e $\{a, a, e\} = \{a, e\}$.

Il principio di **estensionalità** definisce quando due insiemi sono uguali: due insiemi sono uguali quando hanno esattamente gli stessi elementi:

$$A = B \iff \forall x\,((x \in A \to x \in B) \wedge (x \in B \to x \in A)).$$

L'identità dell'insieme dipende dunque dal suo contenuto, non dal modo in cui lo descriviamo. "Numeri primi pari" e "soluzioni naturali di $x + 1 = 3$" individuano entrambi $\{2\}$.

L'**insieme vuoto** $\varnothing$ non possiede elementi *per definizione*. Il **singoletto** $\{x\}$, invece, possiede un solo elemento: x.

1.1 Definizione [inclusione]. Dati due insiemi A e B, scriviamo $A \subseteq B$ se ogni elemento di A è anche elemento di B:

$$A \subseteq B \iff \forall x\,(x \in A \to x \in B).$$

e diciamo che A è **incluso** in B o, equivalentemente, che A è **sottoinsieme** di B. ■

L'inclusione non esclude l'uguaglianza. Se $A \subseteq B$ e $A \neq B$, si può scrivere $A \subsetneq B$. Nel secondo caso si dice che A è un **sottoinsieme proprio** di B.

1.2 Osservazione. È essenziale non confondere $\in$, che è una relazione *fra elemento e insieme*, con $\subseteq$, che è relazione *fra insiemi*. ■

1.3 Esempio. Sia $A := \{\varnothing, \{\varnothing\}\}$. Per distinguere con attenzione appartenenza e inclusione, conviene verificare separatamente i seguenti tre punti:

(1) $\varnothing \in A$ è vero, perché $\varnothing$ è uno degli elementi elencati nella definizione di A;
(2) $\varnothing \subseteq A$ è vero, perché l'insieme vuoto è sottoinsieme di ogni insieme;
(3) $\{\varnothing\} \in A$ e $\{\varnothing\} \subseteq A$ sono entrambe vere, ma per ragioni diverse: la prima perché $\{\varnothing\}$ è un elemento di A, la seconda perché il suo unico elemento, $\varnothing$, appartiene ad A.

■

1.2.2 Una questione filosofica

Un insieme è un'entità autonoma, *esistente indipendentemente dal nostro linguaggio*, oppure è un mero dispositivo per parlare collettivamente di oggetti? La stessa pratica formale può essere interpretata in modi filosoficamente differenti. Senza entrare in una trattazione sistematica di filosofia della matematica, è utile distinguere almeno tre famiglie di letture.

Lettura platonista. In una lettura platonista, gli oggetti matematici — e dunque, in particolare, gli insiemi — sono oggetti astratti esistenti che non dipendono per la loro esistenza dalle nostre pratiche linguistiche, dalle convenzioni o dalle attività mentali. La verità matematica è, in questo senso, *scoperta* più che creata. Una difficoltà classica riguarda l'epistemologia: se tali oggetti sono astratti e non entrano in relazioni causali con noi (come gli oggetti fisici), occorre spiegare in che modo possiamo conoscerli [18, 19].

Lettura strutturalista. Per lo strutturalismo, il punto essenziale non è anzitutto che cosa sia un particolare oggetto matematico preso isolatamente, ma quale *posizione* esso occupi entro una struttura e quali relazioni intrattenga con le altre posizioni. L'aritmetica, per esempio, studia la struttura dei numeri naturali: un elemento iniziale, un'operazione di successore e le relazioni che ne derivano. Diverse realizzazioni insiemistiche possono rappresentare la stessa struttura [15, 16].

Un esempio semplice chiarisce il punto. Nella costruzione di von Neumann si può rappresentare il numero 2 mediante

$$2_{\mathrm{vN}} = \{\varnothing, \{\varnothing\}\},$$

mentre nella costruzione di Zermelo si può rappresentarlo mediante

$$2_{\mathrm{Z}} = \{\{\varnothing\}\}.$$

Si tratta di due insiemi diversi. Tuttavia, inseriti nelle rispettive costruzioni dei naturali, occupano lo stesso ruolo rispetto allo zero e all'operazione di successore. Una motivazione tipica dello strutturalismo è allora che l'aritmetica non dovrebbe dipendere dalla scelta di uno di questi insiemi come *il vero* numero 2: ciò che conta è il posto occupato nella struttura.

Lettura nominalista. Il nominalismo rifiuta, o cerca di evitare, l'impegno ontologico verso oggetti matematici astratti. Il problema è stabilire se l'uso indispensabile del linguaggio matematico ci obblighi anche ad ammettere, nell'ontologia della teoria, un dominio di enti matematici che esistono in senso forte come nell'interpretazione platonista. Il programma di Hartry Field costituisce un tentativo radicale di eliminare tali enti dall'ontologia, preservando invece un'ontologia di enti fisici [17]. Questa separazione è però problematica nella fisica moderna: gli stessi enti fisici sono spesso individuati, caratterizzati e definiti proprio attraverso la struttura matematica della teoria. Oggetti teorici come una Lagrangiana, un campo, uno spazio-tempo o uno stato quantistico non possiedono in generale una semplice definizione operativa indipendente dal formalismo matematico. Il tentativo di conservare un'ontologia puramente fisica eliminando quella matematica presuppone dunque una distinzione fra contenuto fisico e struttura matematica che non è affatto ovvia. (Si veda anche l'Osservazione 1.58 sotto.)

1.4 Osservazione. Queste tre etichette non individuano compartimenti perfettamente disgiunti. In particolare, lo strutturalismo può assumere una forma realista, nella quale le strutture stesse sono considerate oggetti astratti, oppure forme più vicine al nominalismo, nelle quali parlare di strutture non richiede di postularle come entità autonome [16]. Per questo è meglio considerare platonismo, strutturalismo e nominalismo come famiglie di risposte a domande ontologiche ed epistemologiche, non come tre sistemi rigidamente incompatibili. ■

Il corso userà la teoria degli insiemi come linguaggio matematico senza presupporre una risposta provvisoria o definitiva a queste questioni. La loro presenza serve a ricordare che una formalizzazione matematica e la sua interpretazione filosofica sono due livelli distinti: gli stessi simboli e le stesse dimostrazioni possono essere accettati all'interno di concezioni differenti della natura degli oggetti matematici.

1.2.3 Un'alternativa: la mereologia

La teoria degli insiemi non è l'unico linguaggio formale disponibile per parlare di una pluralità come di un'unità. La **mereologia** assume come nozione fondamentale la relazione fra parte e tutto [1, 2]. Essere elemento di ed essere parte di non vanno confusi: Socrate appartiene all'insieme dei filosofi, ma non è una sua parte; una mano è parte di un corpo, ma non gli appartiene come un elemento appartiene a un insieme. Inoltre un insieme e i suoi elementi possono essere di natura completamente diversa, mentre un intero mereologico e le sue parti sono normalmente collocati sullo stesso piano ontologico.

Da Leśniewski e dal *calcolo degli individui* di Leonard e Goodman, la mereologia è stata spesso proposta come alternativa ontologicamente più parsimoniosa alla teoria degli insiemi. Per comprendere questa idea occorre introdurre la nozione di **fusione mereologica**. Dati alcuni individui, la loro fusione è, informalmente, un intero che li ha come parti: non un nuovo oggetto astratto che li contiene come elementi, ma un oggetto composto da essi. Per esempio, l'insieme $\{a, b\}$ ha a e b come elementi, mentre la fusione mereologica di a e b, nei sistemi mereologici in cui una tale fusione è ammessa, è un intero di cui a e b sono parti. La distinzione è essenziale: appartenenza e relazione parte–tutto sono relazioni di natura diversa [1, 2].

Questa impostazione ha avuto particolare importanza in alcuni programmi nominalisti. L'idea è che, invece di introdurre insiemi o classi come nuove entità astratte, si possa parlare degli individui già ammessi e dei loro rapporti di parte e tutto. Se gli individui di partenza sono oggetti concreti, anche le loro fusioni

possono essere concepite come oggetti concreti, eventualmente composti o dispersi, senza postulare un ulteriore livello di entità astratte come gli insiemi. È in questo senso che la mereologia può apparire ontologicamente più parsimoniosa e che è stata storicamente associata al nominalismo.

L'associazione, tuttavia, non è necessaria. La mereologia, presa in sé, stabilisce principi relativi alla relazione parte–tutto, ma non determina la natura ontologica degli enti che possono essere parti o interi. Nulla vieta, per esempio, di applicare una teoria mereologica anche a oggetti astratti. È in questo senso che Achille C. Varzi sottolinea che la mereologia non implica, da sola, una posizione nominalista [1]. La sostituzione della teoria degli insiemi con la mereologia non è comunque immediata. In teoria degli insiemi il singoletto $\{a\}$ è distinto da a; in una mereologia estensionale, invece, la fusione del solo a coincide normalmente con a. Proprio il trattamento dei singoletti e dell'oggetto nullo segna una differenza profonda fra i due quadri. La nozione di funzione e il suo uso sistematico diventano più complicati nella formulazione mereologica rispetto alla teoria degli insiemi.

David Lewis, in *Parts of Classes* [3], ha mostrato quanto lontano possa spingersi una ricostruzione mereologica dell'universo delle classi. Il confronto contemporaneo è sviluppato in modo sistematico da Cotnoir e Varzi [2] e, con particolare attenzione alla mereologia classica, da Lando [4]. Per il nostro percorso la mereologia va quindi considerata come una vera alternativa concettuale e ontologica, non come una semplice traduzione automatica della formulazione rigorosa della teoria degli insiemi che vedremo più avanti e che indicheremo con ZF.

1.3 L'algebra degli insiemi

In questa sezione introdurremo le principali operazioni e strutture che emergono naturalmente dalle definizioni date.

1.3.1 Unione, intersezione, differenza

1.5 Definizione [operazioni elementari]. Per due insiemi A, B poniamo

$$\begin{aligned} A \cup B &:= \{x \mid x \in A \vee x \in B\} && \text{detto } \textbf{unione} \text{ di } A \text{ e } B, \\ A \cap B &:= \{x \mid x \in A \wedge x \in B\} && \text{detto } \textbf{intersezione} \text{ di } A \text{ e } B, \\ A \setminus B &:= \{x \in A \mid x \notin B\} && \text{detto } \textbf{differenza} \text{ di } A \text{ e } B. \end{aligned}$$

■

1.6 Osservazione. Nella Definizione 1.5 abbiamo scritto, per $A \cup B$, $x \in A \vee x \in B$ mentre avremmo potuto anche scrivere $(x \in A) \vee (x \in B)$. Tuttavia qui l'uso delle parentesi sarebbe stato pleonastico dato che l'unico modo di intendere $x \in A \vee x \in B$ è $(x \in A) \vee (x \in B)$. Per esempio $x \in (A \vee x \in B)$ non è un costrutto formale ammesso: $A \vee x \in B$ non è una formula ben formata, perché A è un termine che denota un insieme, mentre il connettivo $\vee$ può essere applicato soltanto a formule. L'uso delle parentesi ha utilità quando serve a distinguere tra due o più costrutti ben formati come $x \in A \wedge (x \in B \vee x \in C)$ e $(x \in A \wedge x \in B) \vee x \in C$. Per il momento useremo dunque in modo intuitivo l'espressione "costrutto ben formato": qualche precisazione ulteriore su che cosa significhi, in un linguaggio formale, essere un enunciato o una formula ben formata verrà data nella Sezione 1.6.1, quando introdurremo l'idea di assiomatizzazione. ■

1.7 Definizione [insieme delle parti]. Dato un insieme A, l'**insieme delle parti** di A è l'insieme di tutti i suoi sottoinsiemi:

$$\mathcal{P}(A) := \{X \mid X \subseteq A\}.$$

■

1.8 Esempio. Per definizione $\mathcal{P}(\varnothing) = \{\varnothing\}$, perché l'insieme vuoto ha come unico sottoinsieme se stesso. Se $A := \{a, b\}$, allora $\mathcal{P}(A) = \{\varnothing, \{a\}, \{b\}, \{a, b\}\}$. ■

1.9 Definizione [coppia ordinata e prodotto cartesiano]. Dati due oggetti a e b della teoria degli insiemi, la **coppia ordinata** (a, b) è definita mediante la **definizione di Kuratowski**:

$$(a, b) := \{\{a\}, \{a, b\}\}.$$

Per ogni scelta di oggetti a, b, c, d, questa codifica soddisfa la proprietà caratteristica

$$(a, b) = (c, d) \quad \Longleftrightarrow \quad (a = c \wedge b = d).$$

Dati inoltre due insiemi A e B, il **prodotto cartesiano** di A e B è

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

Quando i due fattori coincidono, si usa spesso la notazione abbreviata

$$A^2 := A \times A.$$

■

1.10 Osservazione. La differenza con gli insiemi ordinari è essenziale. Per estensionalità,

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

perché i due insiemi hanno gli stessi elementi. Invece, per le coppie ordinate,

$$(a, b) \neq (b, a)$$

in generale: per la proprietà caratteristica della codifica di Kuratowski, l'uguaglianza $(a, b) = (b, a)$ vale soltanto nel caso particolare $a = b$. ■

Possiamo generalizzare la definizione di coppia ordinata e prodotto cartesiano a N fattori.

1.11 Definizione. Sia $N \geq 1$ un intero e siano $A_1, \ldots, A_N$ insiemi.

$$A_1 \times A_2 \times \cdots \times A_N := \{(a_1, a_2, \ldots, a_N) \mid a_k \in A_k,\ k = 1, \ldots, N\}.$$

indica l'insieme delle **N-ple ordinate** di elementi $(a_1, a_2, \ldots, a_N)$ dove gli elementi $a_k \in A_k$ per $k = 1, \ldots, N$ sono scelti in tutti i modi possibili. Analogamente al caso di coppie ordinate

$$(a_1, a_2, \ldots, a_N) = (a'_1, a'_2, \ldots, a'_N) \quad \Longleftrightarrow \quad a_k = a'_k \text{ per ogni } k = 1, \ldots, N.$$

Nuovamente, se $A_1 = A_2 = \cdots = A_N$,

$$A^N := A \times \cdots \times A \quad N \text{ volte.}$$

■

Per visualizzare le relazioni fra alcuni insiemi useremo talvolta i **diagrammi di Venn**. Si disegna una regione contornata da una curva chiusa (spesso un rettangolo ma non è necessario) per il cosiddetto (insieme) **universo** U del quale tutti gli insiemi considerati saranno sottoinsiemi e, al suo interno, una regione chiusa per ciascun insieme considerato, di solito un cerchio. La zona comune ad A e B rappresenta $A \cap B$; la zona occupata da almeno uno dei due rappresenta $A \cup B$; la parte di U esterna ad A rappresenta A^c. Se gli insiemi sono finiti, si possono anche collocare nel diagramma i singoli elementi come punti. Si tratta di uno strumento intuitivo utile per vedere le relazioni fra gli insiemi, non di una definizione matematica né di un sostituto della dimostrazione formale.

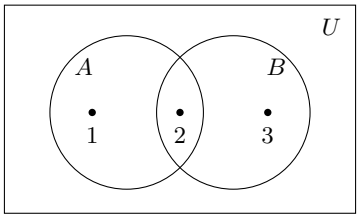
1.12 Esempio [operazioni su insiemi finiti]. Per $A := \{1, 2\}$ e $B := \{2, 3\}$ si ha

$$A \cup B = \{1, 2, 3\}, \qquad A \cap B = \{2\}, \qquad A \setminus B = \{1\},$$

e

$$A \times B = \{(1, 2), (1, 3), (2, 2), (2, 3)\}.$$

Se prendiamo come universo $U := A \cup B = \{1, 2, 3\}$, il diagramma di Venn rende immediatamente visibile perché 2 appartiene all'intersezione, mentre 1 appartiene soltanto ad A e 3 soltanto a B:



La sovrapposizione dei due cerchi contiene quindi precisamente l'elemento di $A \cap B$. Il diagramma rappresenta bene unione, intersezione e differenza; il prodotto cartesiano $A \times B$, invece, richiede di considerare coppie ordinate e non è rappresentato da queste regioni. ■

1.3.2 L'algebra di unione e intersezione

Le operazioni insiemistiche soddisfano alcune identità che, in corsi successivi, si vedranno essere strettamente analoghe alle leggi della *logica proposizionale*. Le raccogliamo in un unico enunciato, così da distinguere nettamente le identità da dimostrare dagli esempi e dagli esercizi che seguiranno.

1.13 Proposizione [identità fondamentali dell'algebra degli insiemi]. *Sia U un insieme. Per ogni $A, B, C \subseteq U$ valgono le seguenti identità:*

$$\begin{aligned}
A \cup B &= B \cup A, & A \cap B &= B \cap A && \textbf{(commutatività)},\\
(A \cup B) \cup C &= A \cup (B \cup C), & (A \cap B) \cap C &= A \cap (B \cap C) && \textbf{(associatività)},\\
A \cup A &= A, & A \cap A &= A && \textbf{(idempotenza)},\\
A \cup \varnothing &= A, & A \cap U &= A && \textbf{(elementi neutri)},\\
A \cup U &= U, & A \cap \varnothing &= \varnothing && \textbf{(elementi assorbenti)},\\
A \cup (A \cap B) &= A, & A \cap (A \cup B) &= A && \textbf{(assorbimento)},\\
A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) & & && \textbf{(distributività)},\\
A \cup (B \cap C) &= (A \cup B) \cap (A \cup C). & & &&
\end{aligned}$$

Dimostrazione. Per estensionalità, per dimostrare l'uguaglianza di due insiemi basta verificare che un elemento arbitrario appartenga al primo se e solo se appartiene al secondo. Procediamo per gruppi di identità.

Per la commutatività,

$$\begin{aligned}
x \in A \cup B &\iff (x \in A \vee x \in B) \iff (x \in B \vee x \in A) \iff x \in B \cup A,\\
x \in A \cap B &\iff (x \in A \wedge x \in B) \iff (x \in B \wedge x \in A) \iff x \in B \cap A.
\end{aligned}$$

Infatti, se x appartiene all'unione di A e B, allora deve appartenere ad A oppure a B (oppure ad entrambi). In formule,

$$x \in A \vee x \in B.$$

Ma dire che x appartiene ad A oppure a B è la stessa cosa che dire che appartiene a B oppure ad A. In formule,

$$x \in B \vee x \in A.$$

Questo significa precisamente che x appartiene all'unione di B e A. La seconda riga si legge nello stesso modo, sostituendo l'unione con l'intersezione e "oppure" con "e".

Da questo punto in poi, in dimostrazioni di questo tipo, non commenteremo più a parole ogni singola equivalenza: scriveremo direttamente le catene di formule che esprimono i passaggi logici.

Per l'associatività,

$$\begin{aligned}
x \in (A \cup B) \cup C &\iff (x \in A \vee x \in B) \vee x \in C\\
&\iff x \in A \vee (x \in B \vee x \in C) \iff x \in A \cup (B \cup C),\\
x \in (A \cap B) \cap C &\iff (x \in A \wedge x \in B) \wedge x \in C\\
&\iff x \in A \wedge (x \in B \wedge x \in C) \iff x \in A \cap (B \cap C).
\end{aligned}$$

Per l'idempotenza,

$$\begin{aligned}
x \in A \cup A &\iff (x \in A \vee x \in A) \iff x \in A,\\
x \in A \cap A &\iff (x \in A \wedge x \in A) \iff x \in A.
\end{aligned}$$

Per gli elementi neutri e assorbenti, ricordando che $A \subseteq U$ e che nessun elemento appartiene a $\varnothing$,

$$\begin{aligned}
x \in A \cup \varnothing &\iff (x \in A \vee x \in \varnothing) \iff x \in A,\\
x \in A \cap U &\iff (x \in A \wedge x \in U) \iff x \in A,\\
x \in A \cup U &\iff (x \in A \vee x \in U) \iff x \in U,\\
x \in A \cap \varnothing &\iff (x \in A \wedge x \in \varnothing) \iff x \in \varnothing.
\end{aligned}$$

Per l'assorbimento,

$$\begin{aligned}
x \in A \cup (A \cap B) &\iff x \in A \vee (x \in A \wedge x \in B) \iff x \in A,\\
x \in A \cap (A \cup B) &\iff x \in A \wedge (x \in A \vee x \in B) \iff x \in A.
\end{aligned}$$

Infine, per le due **leggi distributive**,

$$\begin{aligned} x \in A \cap (B \cup C) &\iff x \in A \wedge (x \in B \vee x \in C) \\ &\iff (x \in A \wedge x \in B) \vee (x \in A \wedge x \in C) \\ &\iff x \in (A \cap B) \cup (A \cap C), \end{aligned}$$

$$\begin{aligned} x \in A \cup (B \cap C) &\iff x \in A \vee (x \in B \wedge x \in C) \\ &\iff (x \in A \vee x \in B) \wedge (x \in A \vee x \in C) \\ &\iff x \in (A \cup B) \cap (A \cup C). \end{aligned}$$

In ogni caso i due insiemi hanno dunque gli stessi elementi. □

1.14 Osservazione [principio di doppia inclusione]. In alternativa, per dimostrare $X = Y$ si possono provare separatamente $X \subseteq Y$ e $Y \subseteq X$. Il metodo è equivalente all'argomento elemento per elemento e rende esplicite le due direzioni della dimostrazione.

Per esempio, dimostriamo con questo metodo che

$$A = A \cup A.$$

Anzitutto $A \subseteq A \cup A$: infatti, se $x \in A$, allora certamente $x \in A$ oppure $x \in A$, e quindi $x \in A \cup A$. Viceversa, $A \cup A \subseteq A$: se $x \in A \cup A$, allora $x \in A$ oppure $x \in A$, e in entrambi i casi $x \in A$. Per il principio di doppia inclusione segue dunque $A = A \cup A$.

Abbiamo così dimostrato, mediante il principio di doppia inclusione, l'idempotenza dell'unione. La stessa identità si può però ricavare anche dalle leggi già dimostrate: ponendo $B = A$ nella legge di assorbimento si ottiene

$$A \cup (A \cap A) = A,$$

e usando l'idempotenza dell'intersezione, $A \cap A = A$, segue

$$A \cup A = A.$$

In questo modo l'idempotenza dell'unione viene ricavata dalla legge di assorbimento insieme con l'idempotenza dell'intersezione. ■

1.3.3 Complemento e leggi di De Morgan

1.15 Definizione [complemento]. Fissato un insieme che diremo universo U e dato $A \subseteq U$, il **complemento** di A rispetto a U è l'insieme

$$A^c := U \setminus A = \{x \in U \mid x \notin A\}.$$

La notazione A^c è dunque sempre relativa all'universo U che si sta considerando. ■

Dalla Definizione 1.15 seguono anzitutto alcune identità elementari.

1.16 Proposizione [identità del complemento e della differenza]. *Sia U un insieme e siano i complementi intesi rispetto a U. Per ogni $A, B \subseteq U$ valgono*

$$(A^c)^c = A, \qquad A \cup A^c = U, \qquad A \cap A^c = \varnothing,$$

$$U^c = \varnothing, \qquad \varnothing^c = U, \qquad A \setminus B = A \cap B^c.$$

Dimostrazione. Poiché tutti gli insiemi coinvolti sono sottoinsiemi di U, fissiamo $x \in U$. Per il doppio complemento,

$$x \in (A^c)^c \iff x \notin A^c \iff \neg(x \notin A) \iff x \in A.$$

Per l'unione con il complemento, il principio del terzo escluso dà

$$x \in A \cup A^c \iff (x \in A \vee x \notin A),$$

che è vero per ogni $x \in U$; dunque $A \cup A^c = U$. Analogamente,

$$x \in A \cap A^c \iff (x \in A \wedge x \notin A),$$

che è falso per ogni x; quindi $A \cap A^c = \varnothing$.

Inoltre, per ogni $x \in U$, è falso che $x \notin U$, mentre è vero che $x \notin \varnothing$. Pertanto

$$U^c = \varnothing, \qquad \varnothing^c = U.$$

Infine,

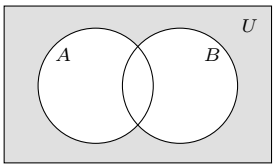
$$x \in A \setminus B \iff (x \in A \wedge x \notin B) \iff (x \in A \wedge x \in B^c) \iff x \in A \cap B^c.$$

Questo dimostra tutte le identità enunciate. □

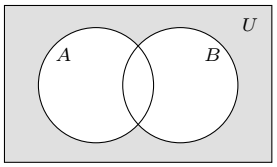
1.17 Proposizione [leggi di De Morgan]. *Sia U un insieme e siano i complementi intesi rispetto a U. Per ogni $A, B \subseteq U$ valgono*

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c.$$

I diagrammi di Venn seguenti suggeriscono la prima identità: nei due casi viene evidenziata la stessa parte di U, quella che sta fuori sia da A sia da B.



$(A \cup B)^c$



$A^c \cap B^c$

Il diagramma suggerisce l'identità, ma la dimostrazione è indipendente dalla figura.

Dimostrazione. Per la prima legge, per ogni $x \in U$,

$$\begin{aligned} x \in (A \cup B)^c &\iff x \notin A \cup B \\ &\iff \neg(x \in A \vee x \in B) \\ &\iff (x \notin A \wedge x \notin B) \\ &\iff x \in A^c \cap B^c. \end{aligned}$$

Per la seconda,

$$\begin{aligned} x \in (A \cap B)^c &\iff x \notin A \cap B \\ &\iff \neg(x \in A \wedge x \in B) \\ &\iff (x \notin A \vee x \notin B) \\ &\iff x \in A^c \cup B^c. \end{aligned}$$

I passaggi centrali sono precisamente quelli che, nel corso di Logica, si vedranno essere le due leggi di De Morgan della logica proposizionale. □

Prima di affrontare gli esercizi, conviene fissare il metodo: si traduce l'appartenenza in connettivi logici, si usa un'equivalenza proposizionale e infine si ritorna al linguaggio degli insiemi. Le Proposizioni 1.13, 1.16 e 1.17 possono inoltre essere usate direttamente per semplificare espressioni insiemistiche.

1.3.4 Esercizi sull'algebra degli insiemi

1.18 Esercizio.

1. Semplificare $A \cap (B \cup A^c)$ usando le identità appena dimostrate.
   Soluzione. Per distributività e per le identità del complemento,
   $$A \cap (B \cup A^c) = (A \cap B) \cup (A \cap A^c) = (A \cap B) \cup \varnothing = A \cap B.$$
2. Dimostrare che
   $$(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A).$$
   Soluzione. Usando la differenza come intersezione con un complemento, De Morgan e la distributività,
   $$\begin{aligned} (A \cup B) \setminus (A \cap B) &= (A \cup B) \cap (A \cap B)^c \\ &= (A \cup B) \cap (A^c \cup B^c) \\ &= (A \cap A^c) \cup (A \cap B^c) \cup (B \cap A^c) \cup (B \cap B^c) \\ &= (A \cap B^c) \cup (B \cap A^c) \\ &= (A \setminus B) \cup (B \setminus A). \end{aligned}$$

3. Dimostrare che
$$A \setminus (A \setminus B) = A \cap B.$$
Soluzione.
$$\begin{aligned} A \setminus (A \setminus B) &= A \cap (A \setminus B)^c \\ &= A \cap (A \cap B^c)^c \\ &= A \cap (A^c \cup B) \\ &= (A \cap A^c) \cup (A \cap B) \\ &= A \cap B. \end{aligned}$$

4. Semplificare $(A \cap B) \cup (A \cap B^c)$ e $(A \cup B) \cap (A \cup B^c)$.
Soluzione. Usando distributività, complemento ed elemento neutro:
$$\begin{aligned} (A \cap B) \cup (A \cap B^c) &= A \cap (B \cup B^c) = A \cap U = A, \\ (A \cup B) \cap (A \cup B^c) &= A \cup (B \cap B^c) = A \cup \varnothing = A. \end{aligned}$$

5. Stabilire se
$$A \setminus (B \cup C) = (A \setminus B) \cup (A \setminus C).$$
Se è falsa, fornire un controesempio e scrivere l'identità corretta.
Soluzione. L'identità proposta è falsa. Poniamo
$$A := \{1\}, \qquad B := \{1\}, \qquad C := \varnothing.$$
Allora
$$A \setminus (B \cup C) = \varnothing,$$
mentre
$$(A \setminus B) \cup (A \setminus C) = \varnothing \cup \{1\} = \{1\}.$$
L'identità corretta è
$$A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C).$$
Infatti $A \setminus X = A \cap X^c$ e $(B \cup C)^c = B^c \cap C^c$.

6. Siano $A := \{1, 2, 3, 4\}$, $B := \{2, 4, 6\}$ e $C := \{1, 4, 6\}$. Verificare direttamente entrambe le leggi distributive.
Soluzione. Si ha
$$B \cup C = \{1, 2, 4, 6\}, \qquad A \cap (B \cup C) = \{1, 2, 4\}.$$
D'altra parte
$$(A \cap B) \cup (A \cap C) = \{2, 4\} \cup \{1, 4\} = \{1, 2, 4\}.$$
Inoltre
$$B \cap C = \{4, 6\}, \qquad A \cup (B \cap C) = \{1, 2, 3, 4, 6\},$$
e
$$(A \cup B) \cap (A \cup C) = \{1, 2, 3, 4, 6\} \cap \{1, 2, 3, 4, 6\} = \{1, 2, 3, 4, 6\}.$$
Entrambe le leggi distributive sono quindi verificate.

1.3.5 Relazioni e proprietà delle relazioni

1.19 Definizione [relazione]. Siano A e B insiemi. Una **relazione** R da A a B è un sottoinsieme del prodotto cartesiano:
$$R \subseteq A \times B.$$
Scriviamo aRb come abbreviazione di $(a, b) \in R$. Il **dominio di una relazione** e l'**immagine** della relazione sono
$$\mathrm{dom}(R) := \{a \in A \mid \exists b \in B : \ aRb\},$$
$$\mathrm{im}(R) := \{b \in B \mid \exists a \in A : \ aRb\}.$$

■

1.20 Esempio. Siano $A := \{1, 2, 3\}$ e $B := \{a, b\}$. Consideriamo la relazione

$$R := \{(1, a), (2, a), (2, b)\}.$$

Essa ha dominio $\{1, 2\}$ e immagine $\{a, b\}$. Non mette in relazione 3 con alcun elemento; inoltre 2 è in relazione con due elementi. Perciò R non è una funzione da A a B (la nozione di funzione sarà definita formalmente nella Sezione 1.3.7). ■

1.21 Definizione [proprietà di una relazione]. Sia A un insieme. Una **relazione su** A è una relazione da A ad A, cioè un sottoinsieme di $A \times A$. Essa si dice:

- **riflessiva** se $\forall a \in A,\ aRa$;
- **simmetrica** se $\forall a, b \in A,\ aRb \to bRa$;
- **antisimmetrica** se $\forall a, b \in A,\ (aRb \wedge bRa) \to a = b$;
- **transitiva** se $\forall a, b, c \in A,\ (aRb \wedge bRc) \to aRc$.

Una relazione riflessiva, simmetrica e transitiva si chiama **relazione di equivalenza**. Se R è una relazione di equivalenza su A e $a \in A$, la **classe di equivalenza** di a è

$$[a]_R := \{b \in A \mid aRb\}.$$

Quando non vi è rischio di ambiguità scriveremo semplicemente $[a]$. Una relazione riflessiva, antisimmetrica e transitiva si chiama **ordine parziale**. ■

1.22 Esempio [classi di equivalenza e partizione]. Su un insieme di persone, la relazione "avere lo stesso colore di capelli" è riflessiva, simmetrica e transitiva: è dunque una relazione di equivalenza. Le sue classi di equivalenza sono precisamente i gruppi di persone con lo stesso colore di capelli.

Una **partizione** di un insieme A è una famiglia $\mathcal{C}$ di sottoinsiemi non vuoti di A che sono a due a due disgiunti e la cui unione è tutto A; in simboli,

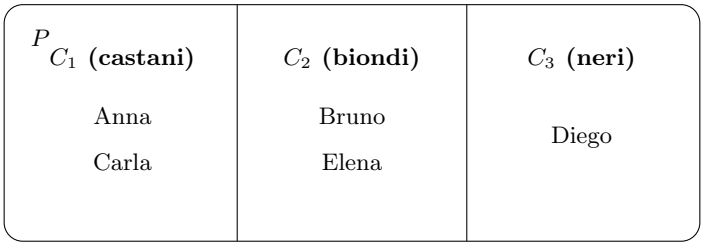
$$\forall C, D \in \mathcal{C}, \quad C \neq D \to C \cap D = \varnothing, \qquad \bigcup_{C \in \mathcal{C}} C = A.$$

Le classi di una relazione di equivalenza formano precisamente una partizione dell'insieme di partenza: ogni elemento appartiene a una e una sola classe.

Per esempio, se

$$P := \{\text{Anna}, \text{Bruno}, \text{Carla}, \text{Diego}, \text{Elena}\}$$

e Anna e Carla hanno i capelli castani, Bruno ed Elena i capelli biondi e Diego i capelli neri, le tre classi di equivalenza sono rappresentate dal seguente diagramma di Venn.



$$C_1 = [\text{Anna}]_R = [\text{Carla}]_R, \qquad C_2 = [\text{Bruno}]_R = [\text{Elena}]_R, \qquad C_3 = [\text{Diego}]_R.$$

Il bordo esterno rappresenta tutto P e i due segmenti interni lo dividono completamente in tre regioni. Non rimane alcuna parte di P fuori dalle classi: ciascuna regione è non vuota, due regioni distinte sono disgiunte e la loro unione è tutto P. Il diagramma rende quindi visibile la proprietà caratteristica di una partizione. ■

1.3.6 Esercizi sulle relazioni

1.23 Esercizio.

1. *Stesso colore di capelli.* Consideriamo l'insieme di persone

$$P := \{\text{Anna}, \text{Bruno}, \text{Carla}, \text{Diego}, \text{Elena}\}.$$

Anna e Carla hanno i capelli castani, Bruno ed Elena hanno i capelli biondi, Diego ha i capelli neri. Definiamo xRy se x e y hanno lo stesso colore di capelli.

(a) Scrivere esplicitamente R come sottoinsieme di $P \times P$.

(b) Stabilire se R è riflessiva, simmetrica, antisimmetrica e transitiva.

(c) Stabilire se R è una relazione di equivalenza e, in caso affermativo, determinarne le classi di equivalenza.

Soluzione. Le coppie della relazione sono

$$\begin{aligned} R = \{&(\text{Anna}, \text{Anna}), (\text{Anna}, \text{Carla}), (\text{Carla}, \text{Anna}), (\text{Carla}, \text{Carla}), \\ &(\text{Bruno}, \text{Bruno}), (\text{Bruno}, \text{Elena}), (\text{Elena}, \text{Bruno}), (\text{Elena}, \text{Elena}), (\text{Diego}, \text{Diego})\}. \end{aligned}$$

La relazione è riflessiva, perché ogni persona ha lo stesso colore di capelli di se stessa; è simmetrica, perché se x ha lo stesso colore di capelli di y, allora y ha lo stesso colore di capelli di x; è transitiva, perché se x e y hanno lo stesso colore e anche y e z hanno lo stesso colore, allora x e z hanno lo stesso colore. Non è antisimmetrica: per esempio Anna è in relazione con Carla e Carla con Anna, ma Anna e Carla sono persone distinte. Dunque R è una relazione di equivalenza. Le classi distinte sono

$$[\text{Anna}] = [\text{Carla}] = \{\text{Anna}, \text{Carla}\},$$

$$[\text{Bruno}] = [\text{Elena}] = \{\text{Bruno}, \text{Elena}\}, \qquad [\text{Diego}] = \{\text{Diego}\}.$$

Queste classi formano una suddivisione di P: ogni persona appartiene a una e una sola classe, determinata dal colore dei capelli.

2. *Altezza maggiore o uguale.* Sia

$$P := \{\text{Anna}, \text{Bruno}, \text{Carla}, \text{Diego}\},$$

con le seguenti altezze: Anna 165 cm, Bruno 180 cm, Carla 165 cm, Diego 172 cm. Definiamo xRy se l'altezza di x è maggiore o uguale all'altezza di y.

(a) Stabilire se R è riflessiva, simmetrica, antisimmetrica e transitiva.

(b) Spiegare perché il fatto che Anna e Carla abbiano la stessa altezza è importante per l'antisimmetria.

Soluzione. La relazione è riflessiva, perché ogni persona ha altezza maggiore o uguale alla propria. Non è simmetrica: Bruno è alto almeno quanto Anna, ma Anna non è alta almeno quanto Bruno. È transitiva, perché se l'altezza di x è almeno quella di y e quella di y è almeno quella di z, allora l'altezza di x è almeno quella di z.
Non è invece antisimmetrica. Infatti Anna e Carla hanno la stessa altezza, quindi Anna R Carla e Carla R Anna, pur essendo Anna e Carla persone distinte. La relazione "essere alto almeno quanto" non è dunque un ordine parziale sull'insieme delle persone. Se invece gli oggetti confrontati fossero direttamente le loro altezze numeriche, la relazione $\geq$ sarebbe antisimmetrica.

3. *Avere almeno un genitore in comune.* Consideriamo tre persone, Anna, Bruno e Carla. Anna e Bruno hanno la stessa madre; Bruno e Carla hanno lo stesso padre; Anna e Carla non hanno alcun genitore in comune. Definiamo xRy se x e y sono persone distinte e hanno almeno un genitore in comune.

(a) Scrivere le coppie che appartengono a R.

(b) Stabilire se R è riflessiva, simmetrica e transitiva.

Soluzione. Si ha

$$R = \{(\text{Anna}, \text{Bruno}), (\text{Bruno}, \text{Anna}), (\text{Bruno}, \text{Carla}), (\text{Carla}, \text{Bruno})\}.$$

La relazione non è riflessiva, perché abbiamo richiesto che le due persone siano distinte. È simmetrica: condividere almeno un genitore è una proprietà reciproca. Non è transitiva: Anna R Bruno e Bruno R Carla, ma Anna non è in relazione con Carla. Questo esempio mostra in modo concreto che simmetria e transitività sono proprietà indipendenti.

4. *Essere figlio di.* Siano Anna, Bruno e Carla persone tali che Bruno è figlio di Anna e Carla è figlia di Bruno. Definiamo xRy se x è figlio o figlia di y.

(a) Quali delle coppie (Bruno, Anna), (Anna, Bruno), (Carla, Bruno), (Carla, Anna) appartengono a R?

(b) La relazione è riflessiva? simmetrica? transitiva?

Soluzione. Appartengono a R le coppie (Bruno, Anna) e (Carla, Bruno). Non appartengono a R le altre due. La relazione non è riflessiva, perché nessuno è figlio di se stesso; non è simmetrica, perché se Bruno è figlio di Anna, Anna non è figlia di Bruno; non è transitiva, perché Carla è figlia di Bruno e Bruno è figlio di Anna, ma Carla non è figlia di Anna: è sua nipote.

5. *Stessa altezza.* Riprendiamo le quattro persone dell'Esercizio 1.23.2 e definiamo ora xEy se x e y hanno esattamente la stessa altezza.

(a) Mostrare che E è una relazione di equivalenza.

(b) Determinare le classi di equivalenza.

(c) Confrontare queste classi con la relazione dell'Esercizio 1.23.2.

Soluzione. La relazione è riflessiva, simmetrica e transitiva, perché l'uguaglianza delle altezze possiede tutte e tre queste proprietà. Le classi sono

$$[\text{Anna}] = [\text{Carla}] = \{\text{Anna}, \text{Carla}\}, \qquad [\text{Bruno}] = \{\text{Bruno}\}, \qquad [\text{Diego}] = \{\text{Diego}\}.$$

La relazione E raggruppa quindi le persone in base all'altezza, mentre la relazione R dell'Esercizio 1.23.2 confronta le loro altezze mediante $\geq$. Le due relazioni rispondono a domande diverse: una esprime equivalenza rispetto a una caratteristica, l'altra esprime un confronto.

1.3.7 Funzioni e proprietà delle funzioni

1.24 Definizione [funzione]. Siano A e B insiemi. Una **funzione** $f : A \to B$ è data da un sottoinsieme

$$G_f \subseteq A \times B,$$

detto **grafico** di f, tale che

$$\forall a \in A \ \exists! b \in B : \ (a, b) \in G_f.$$

Più precisamente:

- l'insieme A si chiama **dominio della funzione**;
- l'insieme B si chiama **codominio** della funzione;
- per ogni $a \in A$, l'unico elemento $b \in B$ tale che $(a, b) \in G_f$ si denota con $f(a)$ e si chiama **immagine di** a;
- l'**immagine della funzione** è il sottoinsieme del codominio definito da

$$f(A) := \{f(a) \mid a \in A\} \subseteq B.$$

Quando dominio e codominio sono fissati, si identifica spesso la funzione con il suo grafico e si scrive semplicemente $f \subseteq A \times B$. ■

1.25 Definizione [preimmagine di un insieme]. Siano A e B insiemi, sia $f : A \to B$ una funzione e sia $C \subseteq B$. La **preimmagine** di C mediante f è il sottoinsieme di A

$$f^{-1}(C) := \{a \in A \mid f(a) \in C\}.$$

Se $C = \{b\}$ è un singoletto, la preimmagine $f^{-1}(\{b\})$ si chiama anche **fibra** di f sopra b. Questa notazione non richiede che f sia invertibile e non va confusa con la funzione inversa, che sarà definita nella Definizione 1.32. ■

1.26 Osservazione. Come già detto all'inizio delle dispense, il simbolo $\exists!$ significa "esiste ed è unico". La condizione precedente dice dunque *due* cose simultaneamente:

- per ogni $a \in A$ esiste almeno un valore $b \in B$ associato ad a,
- tale valore è unico.

Per questo motivo ogni elemento del dominio ha esattamente un'immagine nel codominio. Il codominio B fa parte dei dati della funzione e non va confuso con l'immagine $f(A)$, che può essere un suo sottoinsieme proprio. La notazione

$$A \ni a \mapsto f(a) \in B$$

si legge "$a \in A$ viene mandato in $f(a) \in B$" e descrive la regola di assegnazione della funzione. ■

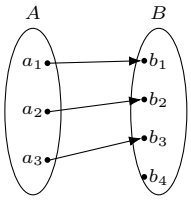
1.27 Definizione [iniettività, suriettività, biiettività]. Siano A e B insiemi e sia $f : A \to B$ una funzione. Essa si dice:

$$\begin{aligned} \textbf{iniettiva} &\iff \forall a_1, a_2 \in A, \quad f(a_1) = f(a_2) \to a_1 = a_2, \\ \textbf{suriettiva} &\iff \forall b \in B \ \exists a \in A : \ f(a) = b, \\ \textbf{biiettiva} &\iff (\text{iniettiva} \wedge \text{suriettiva}). \end{aligned}$$

■

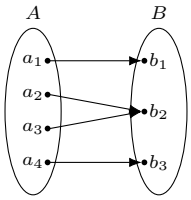
Una funzione iniettiva si chiama anche **iniezione**; una funzione suriettiva si chiama anche **suriezione**; una funzione biiettiva si chiama anche **biiezione**.

Iniettiva, non suriettiva



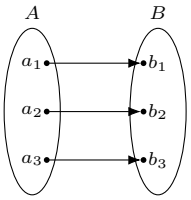
b_4 non è raggiunto.

Suriettiva, non iniettiva



$f(a_2) = f(a_3) = b_2$.

Biiettiva



ogni $b \in B$ è raggiunto una sola volta.

Poniamo

$$[0, +\infty) := \{x \in \mathbb{R} \mid x \geq 0\}.$$

1.28 Esempio [stessa formula, funzioni diverse]. Consideriamo tre funzioni definite dalla stessa formula, ma con dominio o codominio diversi:

- $f : \mathbb{R} \to \mathbb{R}$, definita da $f(x) := x^2$ per ogni $x \in \mathbb{R}$, non è iniettiva, perché $f(1) = f(-1)$, e non è suriettiva, perché nessun $x \in \mathbb{R}$ soddisfa $x^2 = -1$;
- $g : \mathbb{R} \to [0, +\infty)$, definita da $g(x) := x^2$ per ogni $x \in \mathbb{R}$, è suriettiva ma non iniettiva;
- $h : [0, +\infty) \to [0, +\infty)$, definita da $h(x) := x^2$ per ogni $x \in [0, +\infty)$, è biiettiva.

Questo esempio mostra che dominio e codominio sono parte essenziale dei dati di una funzione e determinano le sue proprietà di iniettività e suriettività. ■

1.29 Definizione [composizione di funzioni]. Siano A, B, C insiemi e siano $f : A \to B$ e $g : B \to C$ due funzioni. La **composizione** $g \circ f$, cioè la composizione ottenuta applicando prima f e poi g, è la funzione

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

definita da

$$(g \circ f)(a) := g(f(a)) \qquad (a \in A).$$

■

1.30 Proposizione [associatività della composizione]. *Siano A, B, C, D insiemi e siano $f : A \to B$, $g : B \to C$ e $h : C \to D$. Allora*

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

Dimostrazione. Per ogni $a \in A$,

$$\big(h \circ (g \circ f)\big)(a) = h(g(f(a))) = \big((h \circ g) \circ f\big)(a).$$

Le due funzioni hanno quindi lo stesso valore su ogni elemento di A, e sono uguali. Ne segue che, quando si compongono più funzioni compatibili, le parentesi possono essere omesse senza ambiguità. □

1.31 Definizione [funzione identità]. Per ogni insieme A, la **funzione identità** su A è la funzione

$$\mathrm{id}_A : A \to A, \qquad \mathrm{id}_A(a) := a \quad (a \in A).$$

■

1.32 Definizione [funzione inversa]. Siano A e B insiemi e sia $f : A \to B$ una funzione. La funzione f è **invertibile** se esiste una funzione $g : B \to A$ tale che

$$g \circ f = \mathrm{id}_A, \qquad f \circ g = \mathrm{id}_B .$$

Una funzione g con queste proprietà si chiama **funzione inversa** di f e si denota con f^{-1}. Le due identità si scrivono quindi

$$f^{-1} \circ f = \mathrm{id}_A, \qquad f \circ f^{-1} = \mathrm{id}_B,$$

ossia, elemento per elemento,

$$f^{-1}(f(a)) = a \quad \text{per ogni } a \in A, \qquad f(f^{-1}(b)) = b \quad \text{per ogni } b \in B.$$

■

1.33 Proposizione [unicità dell'inversa]. *Siano A e B insiemi e sia $f : A \to B$ una funzione invertibile. Allora la funzione inversa di f è unica.*

Dimostrazione. Siano $g, h : B \to A$ due inverse di f. Usando l'associatività della composizione,

$$g = g \circ \mathrm{id}_B = g \circ (f \circ h) = (g \circ f) \circ h = \mathrm{id}_A \circ h = h.$$

Quindi $g = h$. □

1.34 Proposizione [invertibilità e biiettività]. *Siano A e B insiemi e sia $f : A \to B$ una funzione. Allora f è invertibile se e solo se è biiettiva.*

Dimostrazione. Supponiamo dapprima che f sia invertibile. Se $f(a_1) = f(a_2)$, applicando f^{-1} ai due membri otteniamo $a_1 = a_2$, dunque f è iniettiva. Inoltre, per ogni $b \in B$, ponendo $a := f^{-1}(b)$ si ha $f(a) = b$, dunque f è suriettiva.

Viceversa, supponiamo che f sia biiettiva. Per ogni $b \in B$ esiste uno e un solo $a \in A$ tale che $f(a) = b$. Definiamo $f^{-1} : B \to A$ ponendo $f^{-1}(b) := a$. Per costruzione,

$$f^{-1}(f(a)) = a \quad (a \in A), \qquad f(f^{-1}(b)) = b \quad (b \in B),$$

ossia $f^{-1} \circ f = \mathrm{id}_A$ e $f \circ f^{-1} = \mathrm{id}_B$. □

1.35 Esempio [una funzione inversa]. Per la funzione biiettiva $h : [0, +\infty) \to [0, +\infty)$, definita da $h(x) = x^2$ per $x \in [0, +\infty)$, l'inversa è

$$h^{-1}(y) = \sqrt{y} \qquad (y \in [0, +\infty)).$$

■

1.3.8 Esercizi sulle funzioni

1.36 Esercizio.

1. Stabilire se $f : \mathbb{Z} \to \mathbb{Z}$, $f(n) := 2n + 1$, è iniettiva o suriettiva. Determinarne l'immagine.
   Soluzione. Se $f(n) = f(m)$, allora $2n + 1 = 2m + 1$, dunque $n = m$: la funzione è iniettiva. La sua immagine è
   $$f(\mathbb{Z}) = \{2n + 1 \mid n \in \mathbb{Z}\},$$
   l'insieme degli interi dispari. Non è suriettiva su $\mathbb{Z}$, perché nessun intero pari appartiene all'immagine.
2. Studiare $g : \mathbb{R} \to \mathbb{R}$, $g(x) := 2x - 3$, e, se possibile, trovare g^{-1}.
   Soluzione. La funzione $g : \mathbb{R} \to \mathbb{R}$, $g(x) := 2x - 3$, è iniettiva: $2x_1 - 3 = 2x_2 - 3$ implica $x_1 = x_2$. È suriettiva perché, dato $y \in \mathbb{R}$, scegliendo $x := (y + 3)/2$ si ha $g(x) = y$. È quindi biiettiva e
   $$g^{-1}(y) = \frac{y + 3}{2}.$$

3. Trovare dominio e immagine di
$$R := \{(x, y) \in \mathbb{R}^2 \mid y^2 = x\}.$$
È una funzione $\mathbb{R} \to \mathbb{R}$?
Soluzione. Poiché un quadrato è non negativo,
$$\mathrm{dom}(R) = [0, +\infty).$$
Ogni $y \in \mathbb{R}$ compare come seconda componente della coppia (y^2, y), quindi $\mathrm{im}(R) = \mathbb{R}$. La relazione non è una funzione $\mathbb{R} \to \mathbb{R}$: per $x < 0$ non esiste alcun y, mentre per $x > 0$ esistono due valori, $y = \sqrt{x}$ e $y = -\sqrt{x}$.
4. Provare che la composizione di funzioni iniettive è iniettiva e che la composizione di funzioni suriettive è suriettiva.
Soluzione. Siano $f : A \to B$ e $g : B \to C$. Se entrambe sono iniettive e $(g \circ f)(a_1) = (g \circ f)(a_2)$, l'iniettività di g dà $f(a_1) = f(a_2)$ e quella di f dà $a_1 = a_2$.
Se entrambe sono suriettive, per ogni $c \in C$ esiste $b \in B$ con $g(b) = c$; inoltre esiste $a \in A$ con $f(a) = b$. Quindi $(g \circ f)(a) = c$, e la composizione è suriettiva.

1.4 Nozione di cardinalità e insiemi infiniti

In questa sezione incontreremo la nozione di infinito attuale in matematica e la teoria degli infiniti sviluppata da Cantor.

1.4.1 Confronto tra cardinalità

1.37 Definizione [cardinalità]. Due insiemi A e B hanno la stessa **cardinalità** (equivalentemente, sono **equipotenti**), e si scrive $|A| = |B|$, quando esiste una funzione biiettiva $f : A \to B$ [5, 6]. ■

1.38 Definizione [confronto tra cardinalità]. Dati due insiemi A e B, scriviamo

$$|A| \leq |B| \quad \text{e diciamo che la cardinalità di } A \text{ è } \textbf{minore o uguale} \text{ a quella di } B$$

se esiste una funzione iniettiva $A \to B$. Scriviamo invece

$$|A| < |B| \quad \text{e diciamo che la cardinalità di } A \text{ è } \textbf{minore} \text{ di quella di } B$$

se $|A| \leq |B|$ ma $|A| \neq |B|$. ■

Abbiamo definito che cosa significa che *due insiemi hanno la stessa cardinalità*, ma non abbiamo definito in modo formale la nozione stessa di cardinalità e di *numero cardinale*. Dalle proprietà delle biiezioni segue che *avere la stessa cardinalità* è una relazione di equivalenza: l'identità è una biiezione, l'inversa di una biiezione è una biiezione e la composizione di biiezioni è ancora una biiezione.

Informalmente si potrebbe essere tentati di identificare la cardinalità di un insieme A con la collezione di tutti gli insiemi equipotenti ad A. Nella versione rigorosa della teoria degli insiemi ZF che vedremo più avanti, però, questa collezione è in generale troppo grande per essere un insieme: la definizione rigorosa dei numeri cardinali richiede quindi una costruzione più accurata [6]. Non ne avremo bisogno qui; useremo la notazione $|A|$ attraverso i confronti definiti sopra e daremo una definizione esplicita soltanto nel caso degli insiemi finiti.

La relazione $|A| \leq |B|$, definita mediante l'esistenza di un'iniezione $A \to B$, è riflessiva e transitiva. Considerata direttamente sugli insiemi non è però antisimmetrica: due insiemi distinti possono essere equipotenti. *Essa induce invece una relazione d'ordine parziale sulle cardinalità.* La proprietà di antisimmetria a questo livello è precisamente garantita dal seguente importante teorema, che non dimostreremo [5, 6].

1.39 Teorema [Cantor–Bernstein]. *Siano A e B insiemi. Se $|A| \leq |B|$ e $|B| \leq |A|$ allora $|A| = |B|$.*

Come preannunciato sopra, definiamo ora in modo esplicito la cardinalità degli insiemi *finiti*.

1.40 Definizione [insieme finito e insieme infinito]. Un insieme A si dice **finito** se $A = \varnothing$, oppure se esiste un intero $N \geq 1$ tale che A possa essere messo in corrispondenza biunivoca con
$$\{1, 2, \ldots, N\}.$$

Un insieme che non è finito si dice **infinito** [5, 6]. ■

1.41 Proposizione [unicità del numero degli elementi]. *Siano $N, M \geq 1$. Se esiste una biiezione*

$$\{1, \ldots, N\} \longrightarrow \{1, \ldots, M\},$$

allora $N = M$.

Dimostrazione. La dimostrazione procede per induzione su N. Se $N = 1$, una biiezione da un insieme con un solo elemento può essere suriettiva soltanto se anche il codominio ha un solo elemento, dunque $M = 1$.

Supponiamo ora vero il risultato per N e consideriamo una biiezione

$$f : \{1, \ldots, N+1\} \longrightarrow \{1, \ldots, M\}.$$

Poiché $N + 1 \geq 2$ e f è iniettiva, il codominio deve contenere almeno due elementi, dunque $M \geq 2$. Scambiando, se necessario, il valore $f(N+1)$ con M nel codominio, possiamo supporre $f(N+1) = M$. La restrizione di f a $\{1, \ldots, N\}$ è allora una biiezione su $\{1, \ldots, M-1\}$. Per l'ipotesi induttiva, $N = M - 1$, e quindi $N + 1 = M$. □

1.42 Definizione [cardinalità di un insieme finito]. Se $A = \varnothing$, poniamo $|A| := 0$. Se $A \neq \varnothing$ è finito, la Proposizione 1.41 garantisce che esiste un unico intero $N \geq 1$ per il quale A è in biiezione con $\{1, \ldots, N\}$; poniamo allora

$$|A| := N.$$

Il numero $|A|$ si chiama **cardinalità** di A. ■

La definizione precedente rende precisa, nel caso finito, l'idea intuitiva del "numero degli elementi". Due insiemi finiti hanno la stessa cardinalità, secondo la Definizione 1.37, se e solo se possiedono lo stesso numero di elementi.

1.43 Esercizio. Sia $A \neq \varnothing$ un insieme finito. Mostrare che la cardinalità di $\mathcal{P}(A)$ è

$$|\mathcal{P}(A)| = 2^{|A|}.$$

In particolare quindi vale sempre $|\mathcal{P}(A)| > |A|$ se $A \neq \varnothing$.

Soluzione. Supponiamo che $A = \{a_1, a_2, \ldots, a_N\}$, con $a_k \neq a_j$ se $j \neq k$. Allora $N = |A|$. Ogni sottoinsieme $B \subseteq A$ è caratterizzato in modo biunivoco da una N-pla $(b_1, b_2, \ldots, b_N)$ con $b_k \in \{0, 1\}$, dove $b_k = 0$ se e solo se $a_k \notin B$ e $b_k = 1$ se e solo se $a_k \in B$. In particolare, $(0, 0, \ldots, 0)$ corrisponde a $\varnothing$ e $(1, 1, \ldots, 1)$ corrisponde ad A. La cardinalità di $\mathcal{P}(A)$ è quindi la stessa dell'insieme delle N-ple $(b_1, b_2, \ldots, b_N)$ con $b_k \in \{0, 1\}$.

Per ciascuna delle N posizioni abbiamo 2 scelte indipendenti, 0 oppure 1; il numero complessivo di N-ple è dunque

$$\underbrace{2 \cdot 2 \cdots 2}_{N \text{ fattori}} = 2^N.$$

Pertanto $|\mathcal{P}(A)| = 2^N = 2^{|A|}$. Poiché $N \geq 1$, vale inoltre $2^N > N$, e quindi $|\mathcal{P}(A)| > |A|$ [5, 6].

Per insiemi *infiniti* le definizioni generali in termini di funzioni biiettive e iniettive producono risultati sorprendenti, che hanno rivoluzionato e rifondato la matematica moderna.

1.44 Esempio [un insieme con la stessa cardinalità di un suo sottoinsieme proprio]. Consideriamo i naturali e i naturali pari:

$$\mathbb{N} := \{0, 1, 2, 3, \ldots\}, \qquad P := \{0, 2, 4, 6, \ldots\}.$$

La funzione $f : \mathbb{N} \to P$ definita da $f(n) := 2n$ per $n \in \mathbb{N}$ è una biiezione. Benché $P \subsetneq \mathbb{N}$, i due insiemi hanno quindi la stessa cardinalità. ■

Questa proprietà, impossibile nel finito, motivò la nozione di insieme infinito dovuta a Dedekind.

1.45 Definizione. Un insieme è **infinito nel senso di Dedekind** se può essere messo in biiezione con un proprio sottoinsieme [5, 6]. ■

Ogni insieme infinito nel senso di Dedekind è infinito nel senso della Definizione 1.40, cioè non è finito. Il viceversa richiede però attenzione: nelle teorie assiomatiche degli insiemi che introdurremo più avanti, la

risposta dipende dagli assiomi adottati. In particolare, in ZF non si può in generale dimostrare che ogni insieme infinito sia infinito nel senso di Dedekind, mentre aggiungendo l'assioma della scelta le due nozioni coincidono. Le sigle ZF e ZFC e l'assioma della scelta saranno introdotti nelle sezioni successive [6].

1.46 Definizione [insiemi numerabili]. Un insieme è **numerabile** se è finito oppure è infinito e può essere messo in biiezione con $\mathbb{N}$ [5, 6]. ■

Su alcuni testi un insieme è detto numerabile solo nel secondo dei due casi della Definizione 1.46. Noi ci atterremo alla definizione scritta. Un insieme infinito numerabile è anche detto di **cardinalità** $\aleph_0$ (letto "aleph zero"), simbolo introdotto da Cantor. La nozione di numero cardinale può essere definita rigorosamente nella teoria assiomatica degli insiemi, ma non ne avremo bisogno qui [6].

1.4.2 Cantor e l'infinito attuale

Nota storica. Cantor *e l'infinito attuale.* Georg Cantor (1845–1918) giunse alla teoria degli insiemi infiniti studiando l'unicità delle rappresentazioni mediante serie trigonometriche. La sua innovazione fu trattare gli infiniti come totalità matematiche compiute e confrontarne le grandezze. Il contrasto con Aristotele può essere formulato in modo preciso. Nella *Fisica*, libro III, capitolo 6 (206a14–24), Aristotele distingue ciò che è in potenza da ciò che è in atto e sostiene che l'infinito non è dato come una totalità attualmente compiuta: esso esiste piuttosto in quanto un processo, per esempio la divisione di una grandezza, può sempre essere ulteriormente proseguito. Cantor, invece, tratta insiemi infiniti come totalità matematiche date e ne confronta le cardinalità [7]. ■

La definizione di avere la stessa cardinalità data per mezzo di funzioni biiettive separa la grandezza dal modo in cui gli elementi sono presentati. Non chiede "quanti elementi contengono?", ma se esista una corrispondenza perfetta fra i loro elementi. Ecco alcuni primi risultati elementari interessanti di questo approccio.

1.47 Proposizione [numerabilità degli interi]. *L'insieme $\mathbb{Z}$ è numerabile.*

Dimostrazione. Definiamo la funzione $\varphi : \mathbb{N} \to \mathbb{Z}$ ponendo

$$\varphi(0) := 0, \qquad \varphi(2k-1) := k, \qquad \varphi(2k) := -k \qquad (k \geq 1).$$

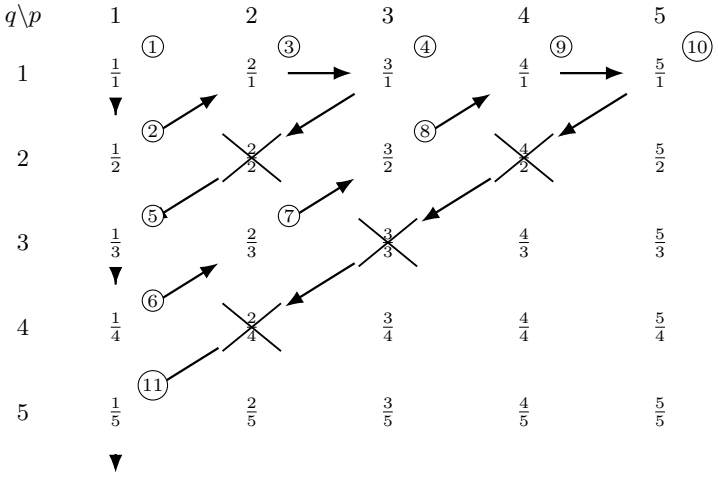
La funzione produce l'elenco

$$0, 1, -1, 2, -2, 3, -3, \ldots$$

e ogni intero compare una e una sola volta. Pertanto φ è biiettiva e $\mathbb{Z}$ è numerabile. □

1.48 Proposizione [numerabilità dei razionali]. *L'insieme $\mathbb{Q}$ è numerabile.*

Dimostrazione. L'idea si vede anzitutto per i razionali strettamente positivi, cioè per $\mathbb{Q}_{>0}$. Osserviamo esplicitamente che 0 è un numero razionale, per esempio $0 = 0/1$, ma non appartiene a $\mathbb{Q}_{>0}$ e quindi non compare nella griglia seguente: verrà aggiunto in un secondo momento. Ogni numero razionale positivo può essere scritto nella forma p/q, con $p, q \in \mathbb{N} \setminus \{0\}$, e può quindi essere collocato nella griglia. La griglia viene percorsa a zig-zag lungo le diagonali $p + q =$ costante, nel verso indicato dalle frecce della figura. In questo modo ogni coppia (p, q) viene visitata dopo un numero finito di passi. Quando una frazione non è ridotta ai minimi termini, la si visita ma non la si trascrive nell'elenco, perché rappresenta un razionale già incontrato.



Si segue la linea spezzata nel verso delle frecce. I numeri cerchiati $1, 2, 3, \ldots$ danno l'ordine dei razionali effettivamente inseriti nell'elenco. Le frazioni barrate sono comunque visitate, ma vengono saltate perché non sono ridotte ai minimi termini: per esempio $2/2 = 1/1$. Il percorso continua allo stesso modo sulle diagonali successive.

In questo modo si ottiene un elenco $r_1, r_2, r_3, \ldots$ di tutti i razionali positivi, senza ripetizioni. Inserendo poi lo zero e alternando ogni razionale positivo con il suo opposto,

$$0, r_1, -r_1, r_2, -r_2, r_3, -r_3, \ldots,$$

si ottiene un elenco di tutti gli elementi di $\mathbb{Q}$. Ciò mostra che $\mathbb{Q}$ è numerabile. □

1.4.3 Cantor e le classi di infiniti

L'approccio di Cantor dimostrò l'esistenza di una gerarchia di infiniti (attuali) basandosi sul seguente notevole risultato che generalizza al caso infinito parte dell'Esercizio 1.43.

1.49 Teorema [Cantor]. *Per ogni insieme A,*

$$|A| < |\mathcal{P}(A)|.$$

Dimostrazione. Esiste sempre un'iniezione $A \to \mathcal{P}(A)$, per esempio $a \mapsto \{a\}$. Resta da mostrare che A e $\mathcal{P}(A)$ non possono avere la stessa cardinalità; proveremo il fatto più forte che non esiste alcuna suriezione $A \to \mathcal{P}(A)$.

Supponiamo per assurdo che $f : A \to \mathcal{P}(A)$ sia suriettiva. Costruiamo

$$D := \{x \in A \mid x \notin f(x)\}.$$

Poiché $D \subseteq A$, si ha $D \in \mathcal{P}(A)$; per la suriettività di f esiste quindi $d \in A$ tale che $f(d) = D$. Ma allora

$$d \in D \quad \Longleftrightarrow \quad d \notin f(d) \quad \Longleftrightarrow \quad d \notin D,$$

che è impossibile. Dunque nessuna funzione $A \to \mathcal{P}(A)$ è suriettiva, e in particolare non esiste alcuna biiezione tra A e $\mathcal{P}(A)$. Pertanto $|A| < |\mathcal{P}(A)|$ [5, 6]. □

1.50 Osservazione [la struttura della diagonale]. Il nuovo oggetto D viene costruito differendo da $f(x)$ proprio nel punto x: se $x \in f(x)$, lo escludiamo; se $x \notin f(x)$, lo includiamo. È una tecnica di auto-riferimento controllato che ricomparirà nella logica, nei *teoremi di Gödel* e nella teoria della computabilità. ■

Applicando ripetutamente il teorema otteniamo una gerarchia senza massimo:

$$|A| < |\mathcal{P}(A)| < |\mathcal{P}(\mathcal{P}(A))| < \cdots$$

In particolare, se applichiamo il risultato all'insieme infinito numerabile $A = \mathbb{N}$ abbiamo subito che $\mathcal{P}(\mathbb{N})$ non può essere numerabile e, come diremo, si riesce a provare che $\mathcal{P}(\mathbb{N})$ ha la stessa cardinalità dell'insieme dei numeri reali $\mathbb{R}$. Il risultato generale basato sulla successione di insiemi scritta sopra è che esiste una gerarchia crescente di cardinalità infinite differenti.

1.4.4 La cardinalità di $\mathcal{P}(\mathbb{N})$ e quella di $\mathbb{R}$

Prima di confrontare $\mathcal{P}(\mathbb{N})$ con la cardinalità di $\mathbb{R}$, introduciamo una nozione e una notazione.

1.51 Definizione [successione]. Sia C un insieme. Una **successione a valori in** C è una funzione

$$s : \mathbb{N} \to C.$$

Se $s(n) = c_n$, la stessa successione si denota spesso con $(c_n)_{n\in\mathbb{N}}$. Con

$$C^{\mathbb{N}}$$

indichiamo l'insieme di tutte le successioni a valori in C, cioè l'insieme di tutte le funzioni $\mathbb{N} \to C$. ■

La definizione data è in realtà una naturale estensione della Definizione 1.11: le successioni di $C^{\mathbb{N}}$ possono vedersi come $\mathbb{N}$-ple ordinate di elementi presi nell'insieme C.

1.52 Osservazione [successione e insieme dei valori]. Non bisogna confondere la successione $(c_n)_{n\in\mathbb{N}}$ con l'insieme dei valori che essa assume,

$$\{c_n \mid n \in \mathbb{N}\}.$$

Nella successione contano l'indice, l'ordine e le eventuali ripetizioni; nell'insieme dei valori, invece, ordine e ripetizioni non contano. Per esempio, se $c_n = (-1)^n$, la successione alterna 1 e -1, mentre l'insieme dei suoi valori è soltanto $\{-1, 1\}$. ■

In particolare

$$\{0,1\}^{\mathbb{N}}$$

è l'insieme di tutte le successioni infinite di zeri e uni, ossia di tutte le funzioni $\mathbb{N} \to \{0, 1\}$.

1.53 Definizione [funzione caratteristica]. Se $S \subseteq \mathbb{N}$, la **funzione caratteristica** di S è la funzione

$$\chi_S : \mathbb{N} \to \{0,1\}, \qquad \chi_S(n) = \begin{cases} 1 & \text{se } n \in S, \\ 0 & \text{se } n \notin S. \end{cases}$$

■

1.54 Proposizione [sottoinsiemi dei naturali e successioni binarie]. *Gli insiemi $\mathcal{P}(\mathbb{N})$ e $\{0,1\}^{\mathbb{N}}$ hanno la stessa cardinalità.*

Dimostrazione. A ogni $S \subseteq \mathbb{N}$ associamo la sua funzione caratteristica χ_S. Viceversa, a ogni successione binaria $(\varepsilon_n)_{n\in\mathbb{N}}$ associamo il sottoinsieme

$$S_\varepsilon := \{n \in \mathbb{N} \mid \varepsilon_n = 1\}.$$

Le due costruzioni sono inverse l'una dell'altra: $S_{\chi_S} = S$ e $\chi_{S_\varepsilon}(n) = \varepsilon_n$ per ogni $n \in \mathbb{N}$. Esiste dunque una biiezione tra $\mathcal{P}(\mathbb{N})$ e $\{0,1\}^{\mathbb{N}}$. □

A questo punto vale il risultato conclusivo dovuto a Cantor: la cardinalità di $\{0,1\}^{\mathbb{N}}$ coincide con quella di $\mathbb{R}$ e quindi $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$. Questo risultato tecnico è complesso e non sarà oggetto d'esame.

1.55 Teorema [Cantor]. *Vale: $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$.*

1.56 Osservazione [una dimostrazione che anticipa proprietà dei reali]. La dimostrazione seguente usa provvisoriamente alcune proprietà standard dei numeri reali che verranno motivate più avanti: la convergenza di semplici serie geometriche, gli sviluppi binari e ternari e una funzione che mette in corrispondenza $\mathbb{R}$ con un intervallo. Va quindi letta come una dimostrazione svolta nel sistema usuale dei numeri reali, non come parte della loro successiva costruzione mediante sezioni di Dedekind. ■

Dimostrazione. Procediamo al confronto tra la cardinalità di $\{0,1\}^{\mathbb{N}}$ e quella di $\mathbb{R}$. A ogni successione $(\varepsilon_n)_{n\in\mathbb{N}}$, con $\varepsilon_n \in \{0, 1\}$, possiamo associare il numero reale

$$x = \sum_{n=0}^{\infty} \frac{2\varepsilon_n}{3^{n+1}} \in [0, 1].$$

Si tratta di uno sviluppo ternario che usa soltanto le cifre 0 e 2. L'applicazione è iniettiva. Infatti, supponiamo che due successioni differiscano per la prima volta all'indice k e, scambiandole se necessario,

che in quel punto la prima abbia cifra 0 e la seconda cifra 1. Il contributo della cifra in posizione k alla differenza dei due numeri è $2/3^{k+1}$, mentre tutti i termini successivi possono compensarlo per una quantità non superiore a

$$\sum_{n=k+1}^{\infty} \frac{2}{3^{n+1}} = \frac{1}{3^{k+1}}.$$

La differenza resta quindi almeno $1/3^{k+1} > 0$: due successioni diverse producono numeri reali diversi. Quindi

$$|\mathcal{P}(\mathbb{N})| = |\{0,1\}^{\mathbb{N}}| \leq |\mathbb{R}|.$$

Per l'altra direzione, a ogni $x \in [0,1]$ associamo prima il numero $x/2 \in [0,1/2]$. Ogni numero di $[0,1/2]$ possiede uno sviluppo binario e, nei casi in cui ne possiede due, scegliamo quello che non termina con una coda infinita di 1:

$$\frac{x}{2} = 0,\varepsilon_0\varepsilon_1\varepsilon_2\cdots_2, \qquad \varepsilon_n \in \{0,1\}.$$

Questa convenzione assegna a ogni x una successione binaria in modo univoco; inoltre due valori distinti di x producono successioni distinte. Otteniamo quindi un'iniezione $[0,1] \to \{0,1\}^{\mathbb{N}}$ e dunque

$$|[0,1]| \leq |\{0,1\}^{\mathbb{N}}| = |\mathcal{P}(\mathbb{N})|.$$

Resta soltanto da confrontare $[0,1]$ con $\mathbb{R}$. La funzione di inclusione $i : [0,1] \to \mathbb{R}$, $i(x) = x$, è iniettiva. Nell'altro verso, la funzione

$$t \longmapsto \frac{1}{2} + \frac{1}{\pi}\arctan t$$

è un'iniezione $\mathbb{R} \to (0,1) \subseteq [0,1]$. Un'ulteriore applicazione del Teorema 1.39 (*Cantor–Bernstein*) fornisce quindi $|[0,1]| = |\mathbb{R}|$. Combinando i risultati ottenuti, $|\mathbb{R}| = |[0,1]| = |\{0,1\}^{\mathbb{N}}| = |\mathcal{P}(\mathbb{N})|$. □

Il risultato modifica radicalmente il concetto di infinito: non una sola indeterminatezza oltre il finito, ma una pluralità ordinata di grandezze matematiche.

1.5 La crisi dei fondamenti: dalla teoria ingenua ai paradossi

Nota storica. *Il contesto della crisi dei fondamenti.* Tra la fine dell'Ottocento e i primi decenni del Novecento la visione scientifica del mondo fu attraversata da profonde trasformazioni. Hilbert presentò nel 1900 i suoi celebri *problemi* e, soprattutto negli anni Venti, sviluppò il programma fondazionale che porta il suo nome, una delle espressioni principali del *formalismo* in filosofia della matematica. Nel 1905 Einstein formulò la *teoria della relatività speciale* e diede, con lo studio quantitativo del *moto browniano*, un contributo decisivo all'affermazione della teoria molecolare della materia e alla meccanica statistica, cioè alla branca della fisica che collega il comportamento macroscopico dei sistemi alle proprietà statistiche dei loro costituenti microscopici. Nel 1915 formulò la *teoria della relatività generale* che, tra le altre cose, come cruciale conseguenza, diede luogo alla *cosmologia* come scienza moderna. Planck aveva introdotto nel 1900 il quanto d'azione, aprendo la strada alla teoria quantistica che si sviluppò in modo esplosivo nei primi trent'anni del Novecento. Nello stesso clima culturale Freud sviluppava la psicoanalisi. ■

1.5.1 Frege, Russell e la crisi dei fondamenti

In questo clima innovativo e fondazionale Gottlob Frege, considerato oggi uno dei principali padri fondatori della logica matematica moderna, della filosofia del linguaggio e della corrente analitica, pubblicò i suoi testi riguardanti il suo grande progetto, il *Logicismo*. La pretesa era dimostrare che tutta la matematica (a partire dall'aritmetica) è riducibile alla logica pura, senza bisogno di intuizioni geometriche o empiriche. Uno degli episodi decisivi della crisi dei fondamenti fu la scoperta, da parte di Bertrand Russell, dell'antinomia che porta il suo nome.

Partiamo da una considerazione tecnica. La notazione $\{x \mid P(x)\}$ suggerisce il seguente *principio ingenuo di comprensione* alla base della teoria, che oggi diciamo ingenua, degli insiemi: per ogni proprietà P sintatticamente corretta esiste l'insieme di tutti e soli gli oggetti che la soddisfano. Come ora vedremo, tale principio è troppo forte: applicato senza restrizioni conduce a una contraddizione.

Nel 1901 Bertrand Russell osservò che, adottando una formulazione della teoria degli insiemi che in qualche modo include tale principio, possiamo fare riferimento alla proprietà sintatticamente ben scritta

"non appartenere a se stesso", espressa dalla formula $P(x) := (x \notin x)$ [9]. Se P determinasse un insieme, avremmo

$$R := \{x \mid x \notin x\}.$$

Chiediamo se $R \in R$. Per definizione:

$$R \in R \quad \Longleftrightarrow \quad R \notin R. \tag{1.1}$$

La relazione (1.1) è impossibile e la teoria produce una contraddizione poi nota come *paradosso di Russell*. Non si tratta di scegliere la risposta corretta: entrambe le alternative implicano la propria negazione: il principio ingenuo di comprensione è insostenibile.

NOTA STORICA. RUSSELL *e* FREGE. RUSSELL comunicò il problema a FREGE mentre era in stampa il secondo volume dell'opera rigorosa e fondazionale della matematica basato sulla logica *Grundgesetze der Arithmetik* (1903) [9]. Il sistema logicista di FREGE ammetteva un cosiddetto *principio di estensione* sufficientemente forte da generare la contraddizione di RUSSELL. La scoperta mostrò che un programma formale rigoroso può essere compromesso dai propri principi di formazione degli oggetti. ■

1.57 Osservazione. Esiste una versione informale del paradosso di Russell detta del *paradosso del barbiere di Russell*. Essa dice: in un villaggio il barbiere rade tutti e soltanto gli uomini che non si radono da soli. Chi rade il barbiere? Se si rade, non dovrebbe radersi; se non si rade, dovrebbe radersi. Questa formulazione, in qualche modo simile al *paradosso di Epimenide*, aiuta l'intuizione, ma il suo significato è più vago: mostra al più che un tale barbiere non può esistere. Il *paradosso di Russell formulato nella costruzione di Frege* colpiva invece i criteri che si ritenevano capaci di fondare l'esistenza stessa degli oggetti matematici fondamentali. ■

Diagnosi concettuale

Il passaggio illegittimo, tacitamente ammesso in quella che oggi si chiama teoria ingenua degli insiemi, ed ammesso da Frege nella sua costruzione, è il seguente:

$$\text{condizione linguisticamente corretta} \quad \Rightarrow \quad \text{insieme esistente.}$$

La soluzione dovrà controllare tale passaggio. Una proprietà può essere descritta rispettando tutte le regole sintattiche della grammatica senza corrispondere a un insieme.

Una risposta storicamente importante all'antinomia fu sviluppata da Russell e Whitehead mediante la *teoria dei tipi*: gli oggetti e le espressioni vengono organizzati in livelli differenti, in modo da impedire forme viziose di auto-applicazione come quella coinvolta nel paradosso di Russell [8]. La distinzione, usata in altre formalizzazioni assiomatiche, fra *insiemi* e *classi proprie* è un'altra strategia concettuale e non va identificata con la soluzione russelliana.

La soluzione oggi più comunemente adottata nella pratica matematica (ma ci sono vari punti di vista) si basa invece su un differente approccio che, per essere introdotto, necessita qualche chiarimento sulla *struttura assiomatica delle teorie*.

1.5.2 I tre programmi sui fondamenti filosofici della teoria degli insiemi e della Matematica

La crisi dei fondamenti della matematica, dovuta in particolare alla scoperta delle antinomie della teoria ingenua degli insiemi e alle difficoltà emerse nei programmi di fondazione, non ricevette una risposta unica. Nel primo Novecento emersero programmi filosofici differenti, ciascuno fondato su una diversa concezione di oggetto, prova e verità matematica [19, 20]. Si vedranno questi argomenti in corsi più avanzati, ci limitiamo qui a una descrizione sommaria e necessariamente approssimativa.

Logicismo

Frege, e poi Russell e Whitehead, tentarono di ricondurre la matematica alla logica. Nella formulazione russelliana la teoria dei tipi impedisce le forme viziose di auto-applicazione [8]: individui, classi di individui e classi di classi appartengono a livelli distinti. $x \in x$ non è una proposizione lecita quando i due x occupano lo stesso livello.

Formalismo

Il programma di David Hilbert tratta le teorie come sistemi assiomatici formali e cerca prove finitarie della loro coerenza [20, 19]. La domanda non è anzitutto che cosa "siano" gli insiemi, ma se le regole conducano mai a una contraddizione. I *teoremi di incompletezza di Gödel* (1931) limitarono il programma: ogni teoria coerente, effettivamente assiomatizzata e abbastanza forte da includere l'aritmetica, contiene enunciati che non può né dimostrare né confutare; inoltre non può, sotto condizioni standard, dimostrare internamente la propria coerenza.

Intuizionismo

Per L. E. J. Brouwer la matematica nasce da *costruzioni mentali* [20, 19]. Questo punto di vista diede luogo a una rifondazione in cui le costruzioni esplicite sono d'obbligo. Dimostrare $\exists x\, P(x)$ richiede la *costruzione esplicita* di un testimone x; dimostrare $P \vee Q$ richiede una prova di uno dei disgiunti. Il *principio del terzo escluso*, $P \vee \neg P$, è una tautologia della logica classica, ma non è accettato senza restrizioni dalla logica intuizionista, in particolare quando si ragiona su totalità infinite.

1.58 Osservazione.

(a) Un aspetto fondamentale dell'*impostazione assiomatico-formale* inaugurata da Hilbert consiste nella possibilità di separare la struttura delle relazioni matematiche dal significato particolare attribuito agli enti che entrano in tali relazioni. Una teoria matematica può così essere riconosciuta attraverso la forma dei rapporti tra i suoi enti anche quando questi ricevono interpretazioni semanticamente del tutto differenti. In questo senso, una nozione sviluppata originariamente in un certo ambito può essere trasferita in un altro quando la struttura delle relazioni resta la stessa.
Un esempio particolarmente importante è offerto dalla *teoria della relatività generale*. La geometria nasce storicamente come teoria dello spazio e delle figure; nella relatività generale, invece, strutture geometriche vengono interpretate in termini di eventi fisici, intervalli spazio-temporali, moti e campo gravitazionale. Le "rette" della geometria diventano geodetiche dello spaziotempo e descrivono il moto libero dei corpi. Più profondamente, la geometria dello spaziotempo non costituisce un semplice sfondo fisso nel quale avviene la dinamica: essa stessa entra nella dinamica ed è legata alla distribuzione della materia e dell'energia.
Ciò che dal punto di vista fisico interpretiamo come divenire ed evoluzione può dunque essere riconosciuto, dal punto di vista delle relazioni matematiche fra gli enti, come una geometria. Questo passaggio concettuale diventa possibile quando non si identifica una struttura matematica con il significato originario dei suoi termini: gli enti interpretati possono cambiare completamente, mentre ciò che viene conservato è la rete delle relazioni formali che li lega. In questo senso l'impostazione assiomatico-formale rende possibile riconoscere una stessa struttura matematica in domini ontologicamente diversi.
Non si intende con ciò sostenere che la relatività generale derivi storicamente dal successivo *Programma di Hilbert* in teoria della dimostrazione. Il punto è piuttosto concettuale: l'atteggiamento hilbertiano verso le teorie assiomatiche rende intelligibile il trasferimento di una struttura matematica da un'interpretazione semantica a un'altra.

(b) Vogliamo chiarire la proposta intuizionista con un esempio esplicito. Sia $(a_n)_{n\in\mathbb{N}}$ una successione arbitraria con valori in $\{0,1\}$ e consideriamo la proposizione

$$P: \quad \exists n \in \mathbb{N}\ (a_n = 1).$$

Per dimostrare costruttivamente P occorre esibire almeno un indice n per il quale $a_n = 1$. Per dimostrare $\neg P$ occorre invece provare che un tale indice non esiste, cioè, in questo caso, che

$$\forall n \in \mathbb{N}\ (a_n = 0).$$

La proposizione P può essere provata costruttivamente esibendo un indice n per il quale $a_n = 1$. Anche $\neg P$ può, in certi casi, essere provata costruttivamente: occorre però una dimostrazione generale che $a_n = 0$ per ogni n, per esempio ricavata dalla legge che definisce la successione; non basta avere controllato un numero finito, per quanto grande, di termini tutti uguali a 0. Per una successione arbitraria della quale non possediamo ulteriori informazioni può quindi accadere che non disponiamo né di una prova di P né di una prova di $\neg P$. La logica classica ammette comunque il principio del terzo escluso $P \vee \neg P$; l'intuizionismo non considera invece tale disgiunzione automaticamente dimostrata quando non è disponibile una costruzione che stabilisca uno dei due disgiunti.

(c) Le nozioni di assioma, sistema formale e coerenza che stiamo per introdurre sono utili per comprendere i programmi fondazionali, ma non occupano in essi lo stesso ruolo. Sono centrali nel formalismo

hilbertiano; il logicismo impiega strumenti formali per perseguire la riduzione della matematica alla logica; l'intuizionismo di Brouwer, invece, pone a fondamento della matematica la costruzione mentale e non il sistema formale.

Programma	Tesi guida	Problema centrale
Logicismo	La matematica è riducibile alla logica.	Evitare paradossi e assunzioni non logiche.
Formalismo	La matematica può essere formalizzata in sistemi assiomatici.	Garantire con metodi finitari la coerenza dei sistemi formalizzati.
Intuizionismo	La verità matematica dipende dalla costruibilità.	Riformare logica e uso dell'infinito.

■

Le tre posizioni illustrate non sono semplici episodi storici: sopravvivono in discussioni contemporanee su realismo, prova assistita dal calcolatore, matematica costruttiva e pluralismo fondazionale.

1.6 La rifondazione assiomatica ZF

Prima di proseguire è utile chiarire il significato di alcune parole che ricorreranno spesso. In matematica non si cerca di dimostrare ogni affermazione a partire dal nulla: si fissano alcuni punti di partenza e si stabiliscono i modi leciti di passare da certe affermazioni ad altre.

1.6.1 Che cosa significa assiomatizzare

Useremo qui in modo ancora informale alcuni termini fondamentali. Un **assioma** è un enunciato, ben formato secondo le regole sintattiche del linguaggio in uso, che, all'interno di una teoria, viene assunto come punto di partenza e non viene dimostrato a partire da enunciati precedenti. Questo non significa necessariamente che l'assioma sia "evidente" o "indubitabile": significa soltanto che, nella teoria considerata, esso fa parte delle premesse di base.

Una **regola di inferenza** stabilisce quali passaggi da alcune formule ben formate ad altre formule ben formate sono ammessi. Quando gli assiomi di una teoria sono enunciati, un **teorema** della teoria è un enunciato derivabile dagli assiomi mediante una catena di applicazioni delle regole di inferenza. D'ora in poi, quando non vi sarà rischio di ambiguità, sottintenderemo la qualifica *ben formato*. Un **sistema assiomatico** specifica dunque, almeno in linea generale, il linguaggio nel quale si parla e quindi quali formule siano ben formate, gli assiomi da cui si parte e le regole con cui si possono produrre nuove formule e, in particolare, nuovi teoremi. Queste indicazioni servono qui a orientare il discorso; non intendono ancora costituire una definizione formale dei corrispondenti concetti logici.

L'idea generale è già presente nel ragionamento sillogistico. Dalle premesse "tutti gli A sono B" e "tutti i B sono C" si conclude "tutti gli A sono C" perché si accetta una certa forma di inferenza. In una teoria matematica gli assiomi svolgono il ruolo di premesse fissate una volta per tutte per quella teoria, mentre le dimostrazioni sono catene di inferenze che conducono dagli assiomi, eventualmente insieme a risultati già dimostrati, a nuovi risultati. In queste dispense non formalizzeremo in dettaglio tutte le regole della logica: useremo le forme elementari di ragionamento richiamate all'inizio del corso.

È importante osservare che una regola di inferenza, considerata come parte di un sistema formale, è formulata in modo puramente *sintattico*: riguarda la forma delle formule e stabilisce quali passaggi fra formule sono ammessi, indipendentemente dal significato attribuito ai simboli non logici che vi compaiono. Per esempio, la regola di inferenza nota come **modus ponens** si scrive

$$\frac{P \qquad P \to Q}{Q}.$$

Qui $\to$ è il connettivo di *implicazione materiale* introdotto nell'introduzione. La frazione non è una formula del linguaggio: rappresenta una regola che autorizza a passare sintatticamente dalle due premesse P e $P \to Q$ alla conclusione Q.

La sola ammissibilità sintattica, tuttavia, non basta: vogliamo anche che le regole siano compatibili con la semantica del linguaggio. Più sotto definiremo precisamente questa richiesta mediante la relazione di soddisfacimento. La formulazione generale varrà anche per formule che contengono occorrenze libere di variabili; quando invece premesse e conclusione sono enunciati, essa si ridurrà alla richiesta che una regola non conduca da enunciati veri a un enunciato falso.

Le regole di inferenza sono generali e, nel quadro qui adottato, vengono fissate indipendentemente dalla particolare teoria matematica che si intende assiomatizzare.

Prima dell'assiomatica moderna, tuttavia, la nozione di "assioma" aveva spesso un significato diverso. Un esempio fondamentale è la geometria di Euclide. Negli *Elementi* Euclide distingue i *postulati* dalle cosiddette *nozioni comuni*. I postulati riguardano specificamente gli oggetti e le costruzioni della geometria: per esempio, la possibilità di tracciare una retta da un punto a un altro o di prolungare indefinitamente un segmento. Le nozioni comuni, invece, hanno carattere più generale e non sono proprie soltanto della geometria: esprimono principi elementari di ragionamento, come "cose uguali a una stessa cosa sono uguali tra loro" oppure "se a cose uguali si aggiungono cose uguali, i risultati sono uguali".

La differenza, quindi, non è semplicemente tra due nomi per la stessa cosa. I postulati fissano alcune proprietà e possibilità fondamentali dello spazio geometrico; le nozioni comuni esprimono principi generali utilizzati nelle dimostrazioni. Entrambi, tuttavia, svolgono il ruolo di principi iniziali e non vengono dimostrati all'interno dell'opera. Essi non sono presentati semplicemente come formule scelte arbitrariamente da cui dedurre conseguenze, ma come affermazioni elementari considerate evidenti o comunque immediatamente accettabili. Il loro valore è dunque strettamente legato al significato dei termini coinvolti.

Con l'assiomatica moderna cambia il punto di vista. Dire che un enunciato è un assioma significa anzitutto assegnargli una certa *posizione sintattica* nel sistema: esso è assunto come punto di partenza e può essere usato nelle dimostrazioni senza essere a sua volta dimostrato all'interno della teoria. Non è necessario considerarlo una verità evidente, né attribuirgli subito un unico significato intuitivo. Da questo punto di vista il ruolo di assioma dipende da come è costruito il sistema deduttivo, non dal fatto che l'enunciato appaia vero per evidenza.

Assiomatizzare una teoria significa

- inserire la teoria in un linguaggio formale preciso per stabilire quali formule siano ben formate;
- elencare esplicitamente i punti di partenza in termini di assiomi;
- adottare regole di inferenza sintatticamente determinate e semanticamente corrette, nel senso preciso che definiremo più avanti mediante il soddisfacimento delle formule.

Questo non significa però che nell'assiomatica moderna la *semantica* scompaia. Una volta fissata una teoria, possiamo chiedere come interpretarne i simboli in strutture matematiche e se, in una data struttura, gli assiomi risultino soddisfatti. Più avanti definiremo precisamente quando una struttura costituisce un modello della teoria. La distinzione fondamentale resta questa: il fatto che un enunciato sia assunto come assioma riguarda il suo ruolo *sintattico* nel sistema; il fatto che esso sia vero in una certa interpretazione riguarda invece la *semantica*. La *geometria non euclidea* mostra bene questa separazione: modificando, per esempio, il postulato delle parallele si ottengono sistemi assiomatici differenti, ciascuno dei quali può essere studiato per le proprie conseguenze e per le strutture che ne soddisfano gli assiomi.

1.6.2 Il ruolo della coerenza

Se dagli assiomi e dalle regole ammesse si potessero inferire sia un enunciato P sia la sua negazione $\neg P$, la teoria sarebbe **contraddittoria**. Una teoria nella quale ciò *non* accade si dice **coerente**; la proprietà di non condurre a contraddizioni si chiama **coerenza**. In altre parole, la coerenza richiede che non esista alcun enunciato per il quale siano dimostrabili, all'interno della stessa teoria, tanto l'enunciato quanto la sua negazione.

Perché la richiesta di coerenza è così importante? Nella logica classica, da una contraddizione si può dedurre qualunque enunciato: questo fatto è noto come *principio di esplosione* ed è tradizionalmente espresso dalle formule latine *ex contradictione quodlibet* oppure *ex falso quodlibet*, spesso abbreviata in *ex falso*. Se una teoria permettesse di dimostrare sia P sia $\neg P$, finirebbe quindi per permettere di dimostrare qualsiasi cosa; la distinzione fra ciò che segue e ciò che non segue dagli assiomi perderebbe il suo valore. Per questo la coerenza è una condizione fondamentale perché un sistema assiomatico possa svolgere il proprio compito deduttivo.

Stabilire la coerenza di una teoria può però essere molto difficile. Spesso si dimostra soltanto una **coerenza relativa**: si mostra, per esempio, che se una certa teoria T è coerente, allora è coerente anche una teoria T' ottenuta modificando o ampliando gli assiomi. Una dimostrazione di questo tipo non prova la coerenza "assoluta" della teoria di partenza, ma trasferisce il problema: mostra che l'eventuale contraddittorietà di T' implicherebbe già una contraddizione in T. Risultati di questo genere vengono spesso ottenuti costruendo un modello o un'interpretazione di T' all'interno di una teoria assunta coerente. È precisamente in questo senso che, più avanti, incontreremo risultati di coerenza relativa riguardanti *ZF*, *l'assioma della scelta* e l'*ipotesi del continuo*.

1.59 Osservazione.

(a) La coerenza è una proprietà di un sistema deduttivo una volta fissati sia gli assiomi sia le regole di inferenza che determinano che cosa sia derivabile. Nel seguito, però, terremo fissate le regole della

logica classica e confronteremo teorie che differiscono soprattutto per gli assiomi; in questo quadro è naturale dire, in forma abbreviata, che il problema della coerenza riguarda la scelta degli assiomi della teoria studiata.

(b) La storia delle geometrie non euclidee fornisce un esempio importante della nascita della nozione moderna di coerenza relativa. Nel 1868 Eugenio Beltrami costruì interpretazioni della geometria iperbolica mediante strutture della geometria euclidea: il suo risultato forniva il modello matematico decisivo, ma non era ancora formulato come un teorema metamatematico di coerenza relativa. Negli anni Ottanta dell'Ottocento Poincaré rese particolarmente esplicita la portata logica di questo modo di procedere. L'idea geometrica può essere vista in modo molto concreto distinguendo le dimensioni: si assume come teoria ambiente la geometria euclidea tridimensionale e si rappresenta una geometria non euclidea bidimensionale su opportune superfici o domini. Le "rette" della geometria bidimensionale vengono allora interpretate come *geodetiche*: dal punto di vista della geometria euclidea tridimensionale esse sono particolari curve. Nel caso ellittico occorre distinguere la geometria sferica dalla geometria ellittica propriamente detta: sulla sfera le geodetiche sono i cerchi massimi; identificando inoltre i punti antipodali della sfera si ottiene il modello proiettivo usuale del piano ellittico, nel quale le rette sono rappresentate dai cerchi massimi con punti antipodali identificati. Nel caso iperbolico esistono realizzazioni locali come superfici a curvatura negativa nello spazio euclideo tridimensionale; non esiste invece una realizzazione isometrica globale e completa dell'intero piano iperbolico come superficie regolare in $\mathbb{R}^3$. Si possono inoltre usare modelli intrinseci bidimensionali, come quelli di Poincaré, nei quali le rette iperboliche sono rappresentate da determinate curve del piano euclideo. In questo modo si ottiene una traduzione — un "dizionario" — degli enti e degli enunciati della geometria non euclidea in enti ed enunciati della geometria euclidea. Un'eventuale contraddizione nella prima si tradurrebbe allora in una contraddizione nella seconda. Nel linguaggio moderno, questo è precisamente un argomento di coerenza relativa. Considerando inoltre un modello in cui vale il postulato delle parallele e un modello iperbolico in cui esso fallisce, si comprende in senso relativo perché il postulato delle parallele non sia deducibile dagli altri assiomi della geometria neutrale.

■

1.6.3 La teoria assiomatica degli insiemi ZF

Passiamo ora a mostrare come la crisi dei fondamenti ha prodotto una versione rigorosa della teoria degli insiemi utilizzando l'approccio assiomatico. Ernst Zermelo propose nel 1908 un'assiomatizzazione della teoria degli insiemi [10, 11, 9], poi ampliata da Abraham Fraenkel e Thoralf Skolem. La teoria così ottenuta, nella sua formulazione usuale, è indicata con **ZF**, dalle iniziali di Zermelo e Fraenkel [6].

1.60 Osservazione. Il confronto descritto nella sezione 1.5.2 tra i tre approcci *filosofici* permette di evitare un equivoco importante: *ZF è una teoria assiomatica formalizzata della teoria degli insiemi, ma adottare ZF non significa adottare il formalismo come posizione filosofica.* Gli assiomi di ZF possono essere interpretati entro prospettive filosofiche differenti. ■

Prima di entrare in qualche dettaglio della formulazione di ZF ci sono alcune osservazioni che preparano il lettore a comprendere meglio il contenuto degli assiomi.

(1) Quando nella teoria ingenua scriviamo $x \in A$, diciamo che x è un *elemento* di A e che A è un *insieme*. Nell'interpretazione usuale di ZF non vi sono due tipi distinti di oggetti, "elementi" e "insiemi": tutte le variabili individuali sono dello stesso tipo e si intendono riferite a insiemi. Il termine "elemento di" indica soltanto il ruolo che un oggetto svolge nella relazione di appartenenza con un altro oggetto.
Nel linguaggio standard del primo ordine di ZF il simbolo non logico fondamentale è la relazione binaria $\in$; le espressioni più semplici che incontreremo sono, per esempio, $x \in y$ e $x = y$. Formule più complesse si ottengono a partire da queste mediante i connettivi logici e i quantificatori; la costruzione formale verrà precisata nella sottosezione dedicata al linguaggio di ZF. Simboli e notazioni familiari come $\varnothing$, $\{x, y\}$ e $x \subseteq y$ sono invece introdotti mediante definizioni o abbreviazioni. In particolare,

$$x \subseteq y \quad \text{abbrevia} \quad \forall z\,(z \in x \to z \in y).$$

In questo senso, nella sua interpretazione usuale, ZF è una teoria di *insiemi puri*: non introduce un secondo tipo primitivo di oggetti, come gli urelementi.

(2) Il fatto che, nell'interpretazione usuale di ZF, tutti gli oggetti siano insiemi non significa che la relazione fondamentale fra essi sia una relazione di parte–tutto. In ZF la relazione primitiva è

$$x \in y,$$

cioè *appartenenza*: l'insieme x è elemento dell'insieme y. In una mereologia, invece, la relazione fondamentale è qualcosa come *"x è parte di y"*. La nozione di inclusione $x \subseteq y$ è derivata dalla relazione di appartenenza e non coincide con una relazione mereologica di parte–tutto.

(3) Gli assiomi di ZF non forniscono un catalogo completo degli insiemi. Alcuni garantiscono direttamente l'esistenza di certi insiemi o permettono di ottenerne di nuovi da insiemi già disponibili; altri, come estensionalità e fondazione, impongono condizioni generali alla struttura dell'universo insiemistico.
È utile distinguere qui un'intuizione informale dalla sintassi formale di ZF. Se partiamo da $\varnothing$ e usiamo un numero finito di volte costruzioni come il singoletto o la coppia, otteniamo esempi quali

$$\varnothing, \qquad \{\varnothing\}, \qquad \{\varnothing, \{\varnothing\}\}, \dots$$

Questi sono esempi di *insiemi ereditariamente finiti*: insiemi finiti i cui elementi sono, ricorsivamente, ancora insiemi ereditariamente finiti. Non sono però tutti gli insiemi finiti: per esempio $\{\mathbb{N}\}$ ha un solo elemento ed è quindi finito, ma contiene l'insieme infinito $\mathbb{N}$. Inoltre le parentesi graffe non costituiscono, nel linguaggio formale standard di ZF, un procedimento sintattico primitivo per "generare" tutti gli insiemi.
L'immagine intuitiva più generale è quella della **gerarchia cumulativa**: procedendo per stadi si considerano insiemi i cui elementi appartengono a stadi precedenti. *Il procedimento non si esaurisce dopo un numero finito di stadi*; l'assioma dell'infinito garantisce, in particolare, l'esistenza di un insieme infinito, e la gerarchia prosegue anche attraverso stadi transfiniti. Questa immagine è utile per comprendere l'universo di ZF, purché non venga confusa con la sintassi delle formule della teoria.

(4) Gli insiemi puri di ZF vanno distinti dagli insiemi di uso informale che appaiono nella teoria ingenua. Se, per esempio, scriviamo

$$S = \{\text{sedia}_1, \text{sedia}_2, \text{sedia}_3\},$$

intendendo letteralmente tre sedie fisiche, stiamo trattando le sedie come oggetti che possono essere elementi di un insieme. In ZF pura, invece, una sedia fisica non è uno degli oggetti su cui variano le variabili della teoria e quindi quella scrittura non descrive letteralmente un insieme di ZF.
Possiamo però *rappresentare* le tre sedie scegliendo tre insiemi puri distinti a, b, c che fungano da codici e considerando l'insieme $\{a, b, c\}$. L'associazione fra ciascuna sedia e il corrispondente codice è una *convenzione esterna di codifica*, formulata nel metalinguaggio: non appartiene al linguaggio di ZF e non è contenuta, da sola, nella relazione di appartenenza fra gli insiemi. *Quando passiamo dalla teoria ingenua a ZF, un "insieme di sedie" va dunque inteso come una rappresentazione mediante insiemi puri, non come un insieme che contenga letteralmente sedie fisiche.* Esistono teorie degli insiemi con *urelementi*, cioè oggetti che possono appartenere a insiemi senza essere a loro volta insiemi, ma la ZF standard non li ammette.

(5) Il paradosso di Russell viene disinnescato in ZF non perché esista un universo di insiemi semplicemente "più piccolo" di quello della teoria ingenua, ma perché *sono limitati i principi con cui possiamo affermare l'esistenza di un insieme.* In particolare, alla comprensione illimitata si sostituisce lo schema di **separazione**. Data una proprietà P esprimibile nel linguaggio della teoria e un insieme *già esistente* A, possiamo formare il sottoinsieme

$$\{x \in A \mid P(x)\}.$$

La proprietà P non crea quindi una totalità dal nulla: ritaglia un sottoinsieme entro un insieme dato A.
Il cosiddetto "Russell relativo ad A"

$$R_A := \{x \in A \mid x \notin x\}$$

è un insieme ammesso da separazione. Esso non produce una contraddizione. Infatti, se per assurdo fosse $R_A \in A$, dalla sua definizione seguirebbe

$$R_A \in R_A \quad \iff \quad R_A \notin R_A,$$

che è impossibile. Dunque $R_A \notin A$ (come si prova anche nell'Esercizio 1.70.7).

(6) In una teoria assiomatica come ZF, data una formula $P(x)$ scritta nel linguaggio della teoria, devono essere distinte due domande:

- se $P(x)$ sia una formula ben formata;
- se gli assiomi permettano di dimostrare l'esistenza di un insieme A i cui elementi siano tutti e soli gli x per i quali vale $P(x)$, ossia l'enunciato

$$\exists A \, \forall x \, \big(x \in A \leftrightarrow P(x)\big).$$

Il fatto che una proprietà sia espressa da una formula perfettamente ben formata non implica dunque, in ZF, che esista l'insieme di tutti e soli gli oggetti che godono di quella proprietà. Questa distinzione è una delle eredità filosofiche della crisi dei fondamenti.

1.6.4 Gli assiomi ZF

La descrizione precedente ha chiarito perché sia necessario controllare il passaggio da una proprietà alla supposta esistenza dell'insieme di tutti gli oggetti che la soddisfano. Vediamo ora più concretamente come questo controllo viene realizzato in ZF. Non svilupperemo una formalizzazione completa della teoria, ma scriveremo gli assiomi in una forma sufficientemente precisa da riconoscere il ruolo della logica del primo ordine e, nello stesso tempo, da comprenderne il significato matematico [6].

Il linguaggio di ZF

ZF è formulata nel linguaggio della logica del primo ordine. Le **variabili individuali** sono simboli come

$$x_0, x_1, x_2, \ldots$$

(o, per comodità tipografica, $x, y, z, A, B, C, \ldots$). Indichiamo con Var l'insieme di tutte queste variabili. Il linguaggio di ZF ha un solo tipo di variabili individuali: nell'interpretazione usuale esse denotano insiemi. L'espressione "primo ordine" significa che i quantificatori $\forall$ ed $\exists$ quantificano sugli oggetti del dominio, non direttamente su proprietà o relazioni fra tali oggetti.

Il simbolo di uguaglianza $=$, i connettivi logici, i quantificatori e le parentesi appartengono all'apparato logico generale. Il solo **simbolo non logico** primitivo di ZF è il simbolo di relazione binaria $\in$.

Le **formule atomiche** del linguaggio di ZF sono le espressioni

$$x = y, \qquad x \in y,$$

dove $x, y \in \mathrm{Var}$. A partire dalle formule atomiche si definiscono induttivamente tutte le **formule ben formate**, che d'ora in poi chiameremo semplicemente *formule*: se P e Q sono formule, allora lo sono anche

$$\neg P, \qquad P \wedge Q, \qquad P \vee Q, \qquad P \to Q;$$

se P è una formula e x è una variabile, allora sono formule anche

$$\forall x\, P, \qquad \exists x\, P.$$

Non è necessario che x compaia effettivamente in P. Nessun'altra espressione è una formula se non è ottenuta, in un numero finito di passi, applicando queste clausole a formule atomiche.

Occorre ora distinguere le occorrenze **libere** da quelle **legate** delle variabili. La distinzione si definisce direttamente seguendo il modo in cui una formula è costruita.

Nelle formule atomiche, come

$$x \in y \qquad \text{oppure} \qquad x = y,$$

tutte le occorrenze delle variabili sono libere. Se si costruisce una nuova formula mediante una *negazione* o un *connettivo logico*, le occorrenze delle variabili conservano il carattere libero o legato che avevano nelle formule di partenza.

La situazione cambia quando si introduce un quantificatore. Se P è una formula, allora in

$$\forall x\, P \qquad \text{e} \qquad \exists x\, P$$

le occorrenze di x che erano libere in P diventano legate dal quantificatore; le occorrenze delle altre variabili conservano invece il carattere libero o legato che avevano in P. Se x non compare affatto in P, il quantificatore non lega alcuna occorrenza di x e si dice **vacuo**. Per esempio, in

$$\forall x\, (x \in y)$$

l'occorrenza di x è legata, mentre quella di y resta libera. In

$$\forall y \forall x\, (x \in y)$$

sono legate sia l'occorrenza di x sia quella di y. Invece, in

$$\forall x\, (y \in y)$$

x non compare: il quantificatore $\forall x$ è vacuo e y resta libera.

Un **enunciato** è una formula che non contiene alcuna occorrenza libera di variabili. Dunque ogni enunciato è una formula, ma non ogni formula è un enunciato. Gli **assiomi di ZF sono enunciati**: sono formule senza occorrenze libere di variabili assunte come punti di partenza della teoria.

Nei paragrafi seguenti alcuni quantificatori esterni potranno essere espressi nella frase che precede la formula anziché essere ripetuti nella formula visualizzata; essi fanno comunque parte dell'enunciato assiomatico completo. Questa precisazione è particolarmente importante per gli schemi di separazione e rimpiazzamento. La formula usata per generare un'istanza dello schema può contenere occorrenze libere di alcune variabili, che fungono da parametri, ma essa, presa da sola, non è l'assioma. L'istanza assiomatica completa si ottiene quantificando anche tali parametri. Per esempio, se $P(x, z)$ è una formula nella quale le sole variabili che possono avere occorrenze libere sono x e z, e P non contiene occorrenze libere di B, una corrispondente istanza dello schema di separazione ha la forma

$$\forall z\, \forall A\, \exists B\, \forall x\, \big[x \in B \leftrightarrow (x \in A \wedge P(x, z))\big],$$

che non contiene occorrenze libere di variabili.

Per non appesantire le formule useremo anche il bicondizionale

$$P \leftrightarrow Q$$

come abbreviazione di

$$(P \to Q) \wedge (Q \to P).$$

Questo simbolo va distinto dalla scrittura metalinguistica $P \iff Q$ usata nelle dispense per affermare che due condizioni sono logicamente equivalenti.

Simboli familiari come $\varnothing$, $\{a, b\}$, $A \subseteq B$, $\bigcup A$ e $\mathcal{P}(A)$ non sono nuovi simboli primitivi del linguaggio: vengono introdotti mediante definizioni e abbreviazioni una volta che gli assiomi ne garantiscono l'esistenza e, quando necessario, l'unicità. Per esempio,

$$A \subseteq B$$

abbrevia

$$\forall x\, (x \in A \to x \in B).$$

Per leggibilità continueremo a usare prevalentemente lettere maiuscole $A, B, C, \ldots$ per gli insiemi considerati come totalità e lettere minuscole $x, y, z, \ldots$ per gli insiemi considerati come loro elementi. *Questa è soltanto una convenzione tipografica*: formalmente sono tutte variabili dello stesso linguaggio e, in una data interpretazione, ricevono valori nello stesso dominio M.

Interpretazione, soddisfacimento, verità e modelli

Finora abbiamo descritto soprattutto il lato sintattico di ZF: abbiamo stabilito quali simboli appartengono al linguaggio e come si costruiscono le formule. Passiamo ora alla semantica, cioè al modo in cui quei simboli e quelle formule vengono interpretati matematicamente.

Una **interpretazione** del linguaggio di ZF è una coppia

$$\mathcal{M} = (M, E),$$

dove $M \neq \varnothing$ è il dominio degli oggetti sui quali variano le variabili ed $E \subseteq M \times M$ è la relazione scelta per interpretare il simbolo formale $\in$. Gli elementi di M sono dunque gli oggetti ai quali, nell'interpretazione, possono riferirsi le variabili; la relazione E stabilisce quali coppie di tali oggetti devono essere considerate in relazione di appartenenza. Non si richiede che E coincida con la relazione di appartenenza insiemistica del metalinguaggio: M ed E possono essere scelti in modi diversi.

Fissata un'interpretazione $\mathcal{M} = (M, E)$, una **assegnazione** è una funzione

$$s : \mathrm{Var} \to M,$$

che associa a ogni variabile un elemento del dominio M. Per esempio, $s(x) = a$ significa semplicemente che, rispetto all'assegnazione s, alla variabile x è associato l'oggetto $a \in M$.

Possiamo ora definire la **relazione di soddisfacimento**. Il termine "relazione" è usato qui nel consueto senso matematico: fissata un'interpretazione $\mathcal{M}$, il soddisfacimento è una relazione fra le assegnazioni $s : \mathrm{Var} \to M$ e le formule del linguaggio, cioè un sottoinsieme del prodotto cartesiano fra l'insieme delle assegnazioni e l'insieme delle formule.

Scriviamo

$$\mathcal{M}, s \models P$$

e leggiamo: "la formula P è soddisfatta nell'interpretazione $\mathcal{M}$ rispetto all'assegnazione s". Dunque, fissata $\mathcal{M}$, la scrittura afferma che la coppia (s, P) appartiene alla relazione di soddisfacimento determinata da $\mathcal{M}$. L'interpretazione compare esplicitamente nella notazione perché, cambiando $\mathcal{M}$, può cambiare quali formule sono soddisfatte.

La definizione parte dalle formule atomiche. Poniamo

$$\mathcal{M}, s \models x = y \quad \Longleftrightarrow \quad s(x) = s(y),$$

e

$$\mathcal{M}, s \models x \in y \quad \Longleftrightarrow \quad \big(s(x), s(y)\big) \in E.$$

La prima equivalenza dice che la formula $x = y$ è soddisfatta quando l'assegnazione associa a x e a y lo stesso elemento di M. La seconda dice che la formula $x \in y$ è soddisfatta quando gli elementi $s(x)$ e $s(y)$ sono nella relazione E scelta per interpretare l'appartenenza. Nell'espressione $x \in y$ il simbolo $\in$ appartiene al linguaggio di ZF; nell'espressione $\big(s(x), s(y)\big) \in E$ il simbolo $\in$ è invece quello usuale del metalinguaggio e indica che una coppia ordinata appartiene all'insieme E.

Per esempio, sia

$$M = \{0, 1\}, \qquad E = \{(0, 1)\},$$

e sia s un'assegnazione con $s(x) = 0$ e $s(y) = 1$. Allora

$$\mathcal{M}, s \models x \in y,$$

mentre

$$\mathcal{M}, s \not\models y \in x \qquad \text{e} \qquad \mathcal{M}, s \not\models x = y.$$

Dalle formule atomiche si passa alle formule composte seguendo il significato dei connettivi logici. La scrittura $\mathcal{M}, s \not\models P$ significa semplicemente che non vale $\mathcal{M}, s \models P$. Si pone quindi

$$\mathcal{M}, s \models \neg P \quad \Longleftrightarrow \quad \mathcal{M}, s \not\models P,$$

$$\mathcal{M}, s \models P \wedge Q \quad \Longleftrightarrow \quad \mathcal{M}, s \models P \ \text{ e } \ \mathcal{M}, s \models Q,$$

$$\mathcal{M}, s \models P \vee Q \quad \Longleftrightarrow \quad \mathcal{M}, s \models P \ \text{ oppure } \ \mathcal{M}, s \models Q,$$

$$\mathcal{M}, s \models P \to Q \quad \Longleftrightarrow \quad \mathcal{M}, s \not\models P \ \text{ oppure } \ \mathcal{M}, s \models Q.$$

Restano i quantificatori. Se $a \in M$, indichiamo con $s[x \mapsto a]$ l'assegnazione che coincide con s per tutte le variabili diverse da x e associa ad x l'elemento a. Definiamo

$$\mathcal{M}, s \models \forall x\, P \quad \Longleftrightarrow \quad \text{per ogni } a \in M,\ \mathcal{M}, s[x \mapsto a] \models P,$$

e

$$\mathcal{M}, s \models \exists x\, P \quad \Longleftrightarrow \quad \text{esiste } a \in M \text{ tale che } \mathcal{M}, s[x \mapsto a] \models P.$$

Queste clausole definiscono il soddisfacimento per tutte le formule, perché ogni formula è costruita in un numero finito di passi a partire dalle formule atomiche mediante connettivi e quantificatori. Nel quadro della logica classica che adottiamo, una volta fissati $\mathcal{M}$, s e una formula P, è quindi determinato se

$$\mathcal{M}, s \models P \qquad \text{oppure} \qquad \mathcal{M}, s \not\models P.$$

Questo non significa che esista sempre un algoritmo capace di stabilire effettivamente quale dei due casi valga: una nozione può essere ben definita senza essere, in generale, decidibile mediante un procedimento finito.

Il soddisfacimento di P dipende soltanto dai valori assegnati alle variabili che hanno occorrenze libere in P. Più precisamente, se due assegnazioni s e t associano lo stesso elemento di M a ogni variabile che ha almeno un'occorrenza libera in P, allora

$$\mathcal{M}, s \models P \quad \Longleftrightarrow \quad \mathcal{M}, t \models P.$$

In particolare, se P è un enunciato, non contiene occorrenze libere di variabili e il suo soddisfacimento non dipende dall'assegnazione: dipende soltanto dall'interpretazione $\mathcal{M}$.

Possiamo perciò definire il **valore di verità di un enunciato** P nell'interpretazione $\mathcal{M}$. Scelta una qualunque assegnazione $s : \mathrm{Var} \to M$, poniamo

$$v_{\mathcal{M}}(P) := \begin{cases} V, & \text{se } \mathcal{M}, s \models P, \\ F, & \text{se } \mathcal{M}, s \not\models P. \end{cases}$$

La definizione non dipende dalla particolare assegnazione scelta, perché per un enunciato tutte le assegnazioni danno lo stesso risultato. Quando $v_{\mathcal{M}}(P) = V$, diciamo che P è **vero in** $\mathcal{M}$ e scriviamo

$$\mathcal{M} \models P.$$

Quando $v_{\mathcal{M}}(P) = F$, diciamo che P è **falso in** $\mathcal{M}$ e scriviamo

$$\mathcal{M} \not\models P.$$

Un'interpretazione $\mathcal{M}$ è un **modello** di una teoria T se tutti gli assiomi di T sono veri in $\mathcal{M}$, cioè se

$$\mathcal{M} \models A \qquad \text{per ogni assioma } A \text{ di } T.$$

In particolare, $\mathcal{M}$ è un modello di ZF se tutti gli assiomi di ZF sono veri in $\mathcal{M}$.

La stessa nozione di soddisfacimento permette ora di precisare che cosa significa dire che una regola di inferenza è corretta.

1.61 Definizione [correttezza semantica]. Una regola di inferenza si dice **semanticamente corretta** se ogni sua istanza

$$\frac{P_1, \dots, P_n}{Q}$$

ha la seguente proprietà: per ogni interpretazione $\mathcal{M} = (M, E)$ e per ogni assegnazione $s : \mathrm{Var} \to M$, se tutte le premesse sono soddisfatte rispetto a quella stessa interpretazione e a quella stessa assegnazione, allora è soddisfatta anche la conclusione. In simboli,

$$\big(\mathcal{M}, s \models P_i \text{ per ogni } i = 1, \dots, n\big) \quad \Longrightarrow \quad \mathcal{M}, s \models Q.$$

■

In altre parole, una regola semanticamente corretta *preserva il soddisfacimento*. Se $P_1, \dots, P_n, Q$ sono tutti enunciati, l'assegnazione non conta e la stessa condizione diventa

$$v_{\mathcal{M}}(P_1) = \cdots = v_{\mathcal{M}}(P_n) = V \quad \Longrightarrow \quad v_{\mathcal{M}}(Q) = V.$$

In questo caso possiamo dire semplicemente che la regola *preserva la verità*.

Possiamo infine definire la nozione di *conseguenza semantica*.

1.62 Definizione. Per formule $P_1, \dots, P_n, Q$ scriviamo

$$P_1, \dots, P_n \vDash Q$$

e diciamo che Q è **conseguenza semantica** di $P_1, \dots, P_n$ se, per ogni interpretazione $\mathcal{M} = (M, E)$ e per ogni assegnazione $s : \mathrm{Var} \to M$, il soddisfacimento di tutte le premesse implica il soddisfacimento della conclusione:

$$\big(\mathcal{M}, s \models P_i \text{ per ogni } i = 1, \dots, n\big) \quad \Longrightarrow \quad \mathcal{M}, s \models Q.$$

■

1.63 Osservazione. La scrittura $P_1, \dots, P_n \vDash Q$ non descrive un passo di derivazione sintattica e non dice *come* ricavare Q dalle premesse. Afferma invece un fatto semantico: non esistono un'interpretazione e un'assegnazione che soddisfino tutte le premesse ma non la conclusione. ■

Per esempio, la correttezza semantica del modus ponens si esprime con

$$P,\ P \to Q \vDash Q.$$

1.64 Osservazione. Nel caso di una sola premessa, $P \vDash Q$ esprime precisamente la relazione che nell'introduzione abbiamo indicato, in forma più informale, con $P \Rightarrow Q$: per ogni interpretazione $\mathcal{M}$ e per ogni assegnazione s, se $\mathcal{M}, s \models P$, allora $\mathcal{M}, s \models Q$. Se P e Q sono enunciati, ciò equivale a dire che ogni interpretazione che rende vero P rende vero anche Q. ■

Gli assiomi

Le presentazioni di ZF non sono tutte identiche nella scelta di quali principi assumere come primitivi e quali ricavare dagli altri. Per esempio, l'esistenza dell'insieme vuoto può essere formulata come assioma separato oppure dedotta, in assiomatizzazioni usuali, da altri assiomi. Qui adotteremo una presentazione pedagogica nella quale il vuoto è enunciato esplicitamente. Inoltre separazione e rimpiazzamento non sono, propriamente, singoli assiomi: sono *schemi di assiomi*, perché a ogni formula ammessa del linguaggio corrisponde un'istanza dello schema. Per questa ragione ZF contiene, in questa formulazione, infinitamente molti assiomi [6].

1. Estensionalità. L'**assioma di estensionalità** dichiara quando due insiemi sono identici: se hanno gli stessi elementi, allora sono lo stesso insieme. Una formulazione è

$$\forall A\,\forall B\,\Big[\forall x\,(x \in A \leftrightarrow x \in B) \to A = B\Big].$$

Usando le proprietà logiche dell'uguaglianza possiamo quindi scrivere, come abbiamo fatto fin dall'inizio del capitolo,

$$A = B \quad\Longleftrightarrow\quad \forall x\,(x \in A \leftrightarrow x \in B).$$

L'assioma esprime così l'idea che un insieme non possiede una struttura interna ulteriore rispetto ai suoi elementi: due insiemi con esattamente gli stessi elementi non possono essere distinti.

2. Insieme vuoto. L'assioma del **vuoto** ammette l'esistenza di un insieme privo di elementi:

$$\exists A\,\forall x\,\neg(x \in A).$$

Per estensionalità un tale insieme è unico; lo indichiamo con $\varnothing$. La formula non assume quindi che $\varnothing$ sia già un oggetto nominato nel linguaggio: prima ne afferma l'esistenza, poi introduce una notazione per l'unico insieme che non possiede elementi.

3. Coppia. L'assioma della **coppia** ammette che, dati due insiemi arbitrari a e b, esista un insieme che abbia precisamente a e b come elementi:

$$\forall a\,\forall b\,\exists C\,\forall x\,\Big[x \in C \leftrightarrow (x = a \lor x = b)\Big].$$

Per estensionalità tale insieme è unico e lo indichiamo con

$$C = \{a, b\}.$$

Se poniamo $a = b$, otteniamo il singoletto $\{a\}$.

4. Unione. L'**assioma di unione** raccoglie in un unico insieme gli elementi degli elementi di un insieme dato. Formalmente:

$$\forall A\,\exists U\,\forall x\,\Big[x \in U \leftrightarrow \exists B\,(B \in A \land x \in B)\Big].$$

L'insieme U è unico per estensionalità e viene indicato con $\bigcup A$. Per esempio, se

$$A = \big\{\{1,2\},\{2,3\}\big\},$$

allora

$$\bigcup A = \{1,2,3\}.$$

In questo senso l'unione elimina un livello di raggruppamento: non raccoglie gli elementi di A, ma gli elementi degli insiemi che appartengono ad A.

5. Insieme delle parti. L'**assioma dell'insieme delle parti** produce, per ogni insieme A, un insieme i cui elementi sono esattamente tutti i sottoinsiemi di A:

$$\forall A\,\exists P\,\forall X\,\Big[X \in P \leftrightarrow \forall x\,(x \in X \to x \in A)\Big].$$

Poiché la condizione a destra abbrevia $X \subseteq A$, indichiamo questo insieme con

$$P = \mathcal{P}(A).$$

Abbiamo già incontrato una conseguenza molto importante di questo assioma nel Teorema 1.49:

$$|A| < |\mathcal{P}(A)|.$$

L'assioma garantisce l'esistenza dell'insieme delle parti; il teorema di Cantor stabilisce poi che esso ha cardinalità strettamente maggiore di quella di A.

6. Infinito. L'**assioma dell'infinito** ammette l'esistenza di almeno un insieme che contiene il vuoto ed è chiuso rispetto all'operazione che manda x in $x \cup \{x\}$:

$$\exists I \Big[\varnothing \in I \wedge \forall x \, (x \in I \to x \cup \{x\} \in I) \Big].$$

Un insieme con questa proprietà si dice *induttivo*. L'assioma dell'infinito garantisce dunque l'esistenza di almeno un insieme induttivo. La costruzione

$$0 := \varnothing, \qquad 1 := 0 \cup \{0\} = \{0\}, \qquad 2 := 1 \cup \{1\} = \{0, 1\},$$

$$3 := 2 \cup \{2\} = \{0, 1, 2\}, \qquad \ldots$$

mostra come si ottenga $\mathbb{N}$ nella cosiddetta **rappresentazione di von Neumann** come caso speciale di insieme induttivo. Si osservi che aggiungendo altri elementi all'insieme di sopra, l'insieme finale continua ad essere induttivo. Per isolare precisamente i numeri naturali, che risultaranno formare *il più piccolo insieme induttivo*, useremo più avanti lo schema di separazione, che introduciamo ora.

7. Separazione. Lo schema di separazione ammette la costruzione di sottoinsiemi definiti da proprietà, ma soltanto a partire da un insieme già disponibile. Sia $P(x, z_1, \ldots, z_k)$ una formula nella quale le sole variabili che possono avere occorrenze libere sono $x, z_1, \ldots, z_k$. La corrispondente istanza dello schema è l'enunciato

$$\forall z_1 \cdots \forall z_k \, \forall A \, \exists B \, \forall x \Big[x \in B \leftrightarrow \big(x \in A \wedge P(x, z_1, \ldots, z_k) \big) \Big]. \tag{1.2}$$

Nella formula (1.2), se non vi sono parametri, i quantificatori $\forall z_1 \cdots \forall z_k$ si omettono. Per estensionalità l'insieme B è unico e scriviamo

$$B = \{x \in A \mid P(x, z_1, \ldots, z_k)\}.$$

La differenza rispetto al principio ingenuo di comprensione è decisiva. La formula P non autorizza a formare senza restrizioni $\{x \mid P(x)\}$: può soltanto selezionare, o *separare*, gli elementi di un insieme A già esistente. Per esempio,

$$R_A := \{x \in A \mid x \notin x\}$$

è un insieme ammesso da ZF. L'argomento di Russell applicato a R_A non produce una contraddizione, ma mostra che $R_A \notin A$.

Si parla di *schema* perché ogni formula ammessa P, con gli eventuali parametri indicati, produce una distinta istanza assiomatica; ciascuna istanza completa è un enunciato.

8. Rimpiazzamento. Lo schema di **rimpiazzamento** ammette che l'immagine di un insieme mediante una regola funzionale definibile sia ancora un insieme. Sia $P(x, y, z_1, \ldots, z_k)$ una formula nella quale le sole variabili che possono avere occorrenze libere sono $x, y, z_1, \ldots, z_k$. La corrispondente istanza dello schema è l'enunciato

$$\forall z_1 \cdots \forall z_k \, \forall A \Big[\forall x \, \big(x \in A \to \exists! y \, P(x, y, z_1, \ldots, z_k) \big) \to \exists B \, \forall y \, \big(y \in B \leftrightarrow \exists x \, (x \in A \wedge P(x, y, z_1, \ldots, z_k)) \big) \Big]. \tag{1.3}$$

Nella formula (1.3), se non vi sono parametri, i quantificatori $\forall z_1 \cdots \forall z_k$ si omettono. In termini meno formali: se a ogni elemento x di un insieme A corrisponde uno e un solo y mediante una proprietà esprimibile nel linguaggio di ZF, allora l'insieme di tutti questi valori y esiste. Anche rimpiazzamento è uno schema: formule diverse P producono istanze assiomatiche diverse, e ogni istanza completa è un enunciato.

La differenza concettuale rispetto a separazione è utile. Separazione *ritaglia* un sottoinsieme da un insieme già dato; rimpiazzamento permette invece di *sostituire* gli elementi di un insieme con i valori determinati da una regola funzionale, ottenendo un nuovo insieme che non deve essere un sottoinsieme dell'insieme di partenza.

9. Fondazione. L'assioma di **fondazione** (o regolarità) ammette che ogni insieme non vuoto A possieda almeno un elemento $y \in A$ che non abbia elementi in comune con A. Una formulazione che usa soltanto l'appartenenza è

$$\forall A \Big[A \neq \varnothing \to \exists y \, \big(y \in A \wedge \neg \exists z \, (z \in y \wedge z \in A) \big) \Big].$$

Equivalentemente, usando la notazione insiemistica già introdotta,

$$A \neq \varnothing \to \exists y \in A \, (y \cap A = \varnothing).$$

L'assioma esclude in particolare $x \in x$. Infatti, se fosse $x \in x$, per $A = \{x\}$ l'unico possibile elemento y di A sarebbe x, ma x avrebbe allora in comune con A almeno l'elemento x. Analogamente fondazione esclude cicli finiti di appartenenza, come $x \in y \in x$.

Che cosa fanno complessivamente questi assiomi? Gli assiomi non costituiscono un elenco di tutti gli insiemi esistenti. Alcuni fissano condizioni generali, come estensionalità e fondazione; altri garantiscono l'esistenza di insiemi fondamentali o permettono di passare da insiemi già disponibili a nuovi insiemi: coppie, unioni, insiemi delle parti, sottoinsiemi definiti da proprietà e immagini definite da regole funzionali. L'assioma dell'infinito assicura inoltre che l'universo non si arresti agli insiemi ereditariamente finiti.

L'idea unificante è che in ZF l'esistenza di un insieme non segue dalla sola possibilità di descriverne linguisticamente gli elementi: deve essere giustificata dagli assiomi della teoria. È precisamente questa disciplina dell'esistenza insiemistica che sostituisce il principio ingenuo di comprensione responsabile del paradosso di Russell.

1.6.5 Costruzione di $\mathbb{N}$ come più piccolo insieme induttivo

Possiamo ora rendere precisa l'affermazione sulla costruzione di $\mathbb{N}$, come più piccolo insieme induttivo, lasciata in sospeso nell'assioma dell'infinito. Indichiamo con $\mathrm{Ind}(J)$ la proprietà "J è induttivo", cioè

$$\mathrm{Ind}(J) \quad := \quad \varnothing \in J \wedge \forall x \, \big(x \in J \to x \cup \{x\} \in J \big).$$

L'assioma dell'infinito garantisce l'esistenza di almeno un insieme induttivo I. All'interno di questo insieme già esistente applichiamo separazione alla proprietà

$$P(x) : \quad \forall J \, \big(\mathrm{Ind}(J) \to x \in J \big),$$

e definiamo

$$\omega := \big\{ x \in I \mid \forall J \, (\mathrm{Ind}(J) \to x \in J) \big\}. \tag{1.4}$$

La definizione (1.4) dice, in altre parole, che ω contiene precisamente quegli elementi di I che appartengono a *ogni* insieme induttivo. È importante osservare il ruolo della separazione: non stiamo formando senza restrizioni "l'insieme di tutti gli oggetti che appartengono a ogni insieme induttivo"; stiamo ritagliando, mediante una proprietà, un sottoinsieme dell'insieme I la cui esistenza è già garantita dall'assioma dell'infinito.

Verifichiamo che ω è a sua volta induttivo. Poiché $\varnothing$ appartiene a ogni insieme induttivo e in particolare a I, si ha $\varnothing \in \omega$. Se poi $x \in \omega$, allora x appartiene a ogni insieme induttivo J; poiché ogni tale J è chiuso rispetto all'operazione $x \mapsto x \cup \{x\}$, anche $x \cup \{x\}$ appartiene a ogni insieme induttivo. Inoltre I stesso è induttivo, dunque $x \cup \{x\} \in I$. Ne segue

$$x \cup \{x\} \in \omega.$$

Quindi ω è induttivo.

Infine, se J è un qualunque insieme induttivo, dalla definizione (1.4) di ω segue immediatamente che ogni $x \in \omega$ appartiene a J; perciò

$$\omega \subseteq J.$$

Dunque ω è contenuto in ogni insieme induttivo: è precisamente il *più piccolo insieme induttivo*. Nella rappresentazione di von Neumann,

$$\omega = \{0, 1, 2, 3, \ldots\},$$

e questo insieme rappresenta $\mathbb{N}$ all'interno di ZF. La costruzione mostra bene come cooperano i due principi: *l'assioma dell'infinito fornisce almeno un insieme induttivo; lo schema di separazione ne ritaglia il nucleo comune a tutti gli insiemi induttivi.*

1.7 L'assioma della scelta e ZFC

L'**assioma della scelta** (AC dall'inglese *Axiom of Choice*) riguarda una situazione molto semplice da formulare: disponiamo di molti insiemi non vuoti e vogliamo scegliere simultaneamente un elemento da ciascuno di essi. La delicatezza non sta nella singola scelta, ma nell'esistenza di un'unica funzione che raccolga tutte le scelte quando la famiglia è arbitraria e può essere infinita. Aggiungendo AC agli assiomi di ZF si ottiene la teoria **ZFC**; la lettera C sta per *Choice* [6].

Formulazione esplicita

Sia I un insieme di indici e sia $(A_i)_{i\in I}$ una famiglia di insiemi. Con questa notazione intendiamo semplicemente che a ogni $i \in I$ è associato un insieme A_i. L'assioma della scelta afferma:

$$(\forall i \in I\ A_i \neq \varnothing) \quad \Rightarrow \quad \exists c : I \to \bigcup_{i\in I} A_i\ \forall i \in I\ c(i) \in A_i. \tag{1.5}$$

Nella formula (1.5), la funzione c si chiama **funzione di scelta**. Per ogni indice i, il valore $c(i)$ è uno degli elementi dell'insieme A_i.

È utile confrontare questa formula con l'affermazione più debole

$$\forall i \in I\ \exists x\ (x \in A_i).$$

Quest'ultima dice soltanto che ogni singolo insieme A_i possiede almeno un elemento. AC consente di raccogliere simultaneamente una scelta per ogni indice in una sola funzione c.

Il punto è ora capire che cosa ZF permette di dedurre da questa affermazione senza aggiungere AC.

1.65 Esempio [perché una famiglia finita non richiede AC]. Siano A_1, A_2, A_3 tre insiemi non vuoti. Per definizione,

$$\exists a_1\,(a_1 \in A_1), \qquad \exists a_2\,(a_2 \in A_2), \qquad \exists a_3\,(a_3 \in A_3).$$

Poiché le scelte da compiere sono soltanto tre, la normale logica del primo ordine permette di concludere

$$\exists a_1\,\exists a_2\,\exists a_3\,\big(a_1 \in A_1 \wedge a_2 \in A_2 \wedge a_3 \in A_3\big).$$

Una volta ottenuti simultaneamente a_1, a_2, a_3, le usuali costruzioni insiemistiche disponibili in ZF permettono di formare la funzione

$$c : \{1,2,3\} \longrightarrow A_1 \cup A_2 \cup A_3, \qquad c(1) = a_1, \quad c(2) = a_2, \quad c(3) = a_3.$$

Dunque *per tre insiemi non vuoti l'esistenza di una funzione di scelta è già dimostrabile in ZF*: non occorre aggiungere un nuovo assioma. Lo stesso argomento si estende, per induzione, a ogni famiglia finita di insiemi non vuoti. ■

Se invece la famiglia $(A_i)_{i\in I}$ è infinita, il ragionamento precedente non può essere prolungato semplicemente "una volta per ogni indice". Da

$$\forall i \in I\ \exists x\ (x \in A_i)$$

non possiamo ottenere una scelta simultanea ripetendo un numero infinito di volte il passaggio esistenziale usato nel caso di tre insiemi: *le formule e le dimostrazioni formali di ZF sono finite*. Questo non dimostra che una funzione di scelta non esista; mostra soltanto che l'argomento finito precedente non basta. Per una famiglia infinita arbitraria, ZF non dimostra in generale il passaggio all'esistenza di un'unica funzione c che scelga un elemento da ogni A_i. *AC afferma precisamente questo passaggio con un unico enunciato finito.*

1.66 Esempio [una famiglia infinita con una regola esplicita]. Per ogni $n \in \mathbb{N}$ poniamo

$$A_n := \{n, n+1\}.$$

Anche se la famiglia è infinita, non occorre AC: la formula

$$c(n) := n$$

definisce direttamente una funzione di scelta, perché $n \in A_n$ per ogni $n \in \mathbb{N}$. Questo mostra che *non è l'infinito, da solo, a rendere necessario AC*: alcune famiglie infinite possiedono una funzione di scelta la cui esistenza è già dimostrabile in ZF. Il problema riguarda le famiglie arbitrarie, per le quali una simile funzione non è in generale dimostrabile in ZF senza un principio ulteriore di scelta. ■

1.67 Osservazione.

(a) Una prima conseguenza importante dell'assioma della scelta riguarda il confronto tra cardinalità. In ZFC, dati due insiemi arbitrari A e B, vale sempre

$$|A| \le |B| \qquad \text{oppure} \qquad |B| \le |A|.$$

L'ordine parziale tra cardinalità diventa quindi un *ordine totale*. In ZF, senza l'assioma della scelta, questa confrontabilità generale non è dimostrabile; anzi, il principio secondo cui ogni coppia di cardinalità è confrontabile è equivalente all'assioma della scelta [6].

(b) L'assioma della scelta è un principio di esistenza, non un algoritmo. Dall'affermazione che esiste una funzione di scelta non segue che disponiamo di una formula, di una procedura effettiva o di un criterio canonico per determinare i suoi valori. Proprio questa separazione fra esistenza e costruzione spiega una parte della sua importanza filosofica e delle discussioni fondazionali che lo hanno accompagnato. ■

Un modo particolarmente concreto di usare AC usa soltanto il linguaggio delle funzioni già introdotto in questo capitolo.

1.68 Proposizione [suriezioni e inverse destre]. *In ZF, l'assioma della scelta è equivalente alla seguente affermazione: per ogni coppia di insiemi X, Y e per ogni funzione suriettiva $p : X \to Y$, esiste una funzione $s : Y \to X$ tale che*

$$p \circ s = \mathrm{id}_Y .$$

*Una funzione s con questa proprietà si chiama **inversa destra** di p.*

Dimostrazione. Supponiamo AC e sia $p : X \to Y$ suriettiva. Per ogni $y \in Y$ la fibra, cioè la preimmagine definita nella Definizione 1.25,

$$p^{-1}(\{y\}) = \{x \in X \mid p(x) = y\}$$

è non vuota. Applicando AC alla famiglia delle fibre, scegliamo per ogni y un elemento $s(y) \in p^{-1}(\{y\})$. Allora $p(s(y)) = y$ per ogni $y \in Y$, cioè $p \circ s = \mathrm{id}_Y$.

Viceversa, supponiamo che ogni suriezione ammetta un'inversa destra e sia $(A_i)_{i \in I}$ una famiglia di insiemi non vuoti. Consideriamo

$$X := \{(i, a) \mid i \in I,\ a \in A_i\}$$

e la proiezione $p : X \to I$ definita da $p(i, a) := i$. Poiché ogni A_i è non vuoto, p è suriettiva. Per ipotesi esiste $s : I \to X$ con $p \circ s = \mathrm{id}_I$. Per ogni i possiamo scrivere $s(i) = (i, a_i)$ con $a_i \in A_i$; ponendo $c(i) := a_i$ otteniamo una funzione di scelta per la famiglia $(A_i)_{i \in I}$. □

1.7.1 Perché l'assioma della scelta è fondazionalmente importante?

L'assioma della scelta non è un espediente isolato. In ZF esso è equivalente a diversi principi centrali della matematica, fra cui il *teorema del buon ordinamento* e il *lemma di Zorn* [6]. Il primo afferma che ogni insieme può essere dotato di un buon ordinamento, cioè di un ordine nel quale ogni sottoinsieme non vuoto possiede un elemento minimo. Il secondo consente, sotto opportune ipotesi, di passare da informazioni su catene ordinate all'esistenza di elementi massimali.

Questi principi intervengono in risultati molto diversi. Un esempio classico è l'affermazione che ogni spazio vettoriale possiede una base: per spazi vettoriali arbitrari la dimostrazione usuale passa attraverso il *lemma di Zorn*. Anche risultati molto controintuitivi, come il *paradosso di Banach–Tarski*, sono dimostrabili in ZFC e dipendono da principi di scelta.

Il punto fondazionale è quindi duplice. Da una parte AC permette di dimostrare molti teoremi di esistenza senza fornire necessariamente gli oggetti in forma esplicita. Dall'altra, proprio perché ha conseguenze forti e non è riducibile agli altri assiomi di ZF, la domanda se assumerlo va distinta dalla successiva questione metamatematica della sua indipendenza.

1.7.2 Indipendenza dell'assioma della scelta, l'ipotesi del continuo e pluralità dei modelli

Ora possiamo formulare separatamente la questione dell'**indipendenza**. Per l'assioma della scelta, i risultati di Gödel e Cohen mostrano, in forma di coerenza relativa, che gli altri assiomi di ZF non lo decidono: assumendo coerente ZF, non si può dimostrare in ZF né AC né la sua negazione [12, 14]. Per questo la teoria ottenuta aggiungendo AC a ZF viene indicata con ZFC.

Un secondo problema celebre è l'**ipotesi del continuo** (CH, dall'inglese *Continuum Hypothesis*): non esiste un insieme con cardinalità strettamente compresa tra quella di $\mathbb{N}$ e quella di $\mathbb{R}$. Cantor non riuscì né a

dimostrarla né a confutarla. La più piccola cardinalità non numerabile si indica con $\aleph_1$. Per il Teorema 1.55, $|\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$; introduciamo quindi la notazione cardinale

$$2^{\aleph_0} := |\{0,1\}^{\mathbb{N}}| = |\mathcal{P}(\mathbb{N})|.$$

L'ipotesi del continuo si esprime allora nella forma

$$2^{\aleph_0} = \aleph_1.$$

Si tratta dunque dell'affermazione che la cardinalità del continuo sia precisamente la cardinalità infinita immediatamente successiva a $\aleph_0$ [6].

NOTA STORICA. GÖDEL, COHEN *e l'indipendenza.* Nel 1940 KURT GÖDEL mostrò, mediante l'universo costruibile, che se ZF è coerente allora lo è anche ZF con l'assioma della scelta e CH [12]. PAUL COHEN introdusse successivamente il metodo del forcing e stabilì i risultati di indipendenza complementari [13, 14]. In particolare, assumendo la coerenza delle teorie di base, CH è indipendente da ZFC e AC è indipendente da ZF. ■

1.69 Osservazione [che cosa significa indipendenza?]. Qui un modello è, informalmente, una struttura matematica nella quale tutti gli assiomi della teoria risultano soddisfatti. Dire che CH è indipendente da ZFC non significa che sia priva di significato o vagamente "vera e falsa": significa che, se ZFC è coerente, esistono modelli di ZFC in cui CH vale e modelli di ZFC in cui CH non vale. Analogamente, l'indipendenza di AC da ZF significa che, se ZF è coerente, esistono modelli di ZF in cui AC vale e modelli di ZF in cui AC non vale. Il passaggio dalla coerenza di queste teorie all'esistenza di modelli usa il *teorema di completezza di Gödel* per la logica del primo ordine, che qui non dimostreremo. Gli assiomi della teoria di partenza, da soli, non determinano una risposta unica. ■

Questi risultati aprono almeno tre letture filosofiche:

1. cercare nuovi assiomi, giustificati dalla loro evidenza o fecondità, che decidano questioni rimaste aperte;
2. accettare un pluralismo di universi insiemistici, nessuno dei quali esclusivo;
3. sostenere che esista un unico universo degli insiemi cui la teoria intende riferirsi, anche se i nostri assiomi attuali non lo caratterizzano completamente.

Il punto di vista dei modelli diventa così centrale anche nello studio della teoria degli insiemi: non ci si chiede soltanto quali teoremi seguano dagli assiomi, ma anche quali strutture li soddisfino e come vari la verità al variare della struttura.

1.8 Esercizi di ricapitolazione

1.70 Esercizio.

1. Sia $A := \{\varnothing, \{\varnothing\}\}$. Stabilire se sono vere o false le seguenti asserzioni: $\varnothing \in A$, $\varnothing \subseteq A$, $\{\varnothing\} \subseteq A$, $A \in A$, $A \cap \varnothing = \varnothing$, $\varnothing \subseteq \varnothing$, $A \cap \{\varnothing\} = \{\varnothing\}$, $A \cap \{\{\varnothing\}\} = \{\{\varnothing\}\}$, $\varnothing \in \varnothing$, $A \setminus \{\{\{\varnothing\}\}\} = A$; $A \cup \varnothing = A$, $A \cup \{\varnothing\} = A$.
   Motivare ogni risposta.
   Soluzione. Per $A := \{\varnothing, \{\varnothing\}\}$:
   - $\varnothing \in A$ è vera: $\varnothing$ è uno dei due elementi di A;
   - $\varnothing \subseteq A$ è vera: il vuoto è sottoinsieme di ogni insieme;
   - $\{\varnothing\} \subseteq A$ è vera, perché il suo unico elemento $\varnothing$ appartiene ad A;
   - $A \in A$ è falsa: gli elementi di A sono $\varnothing$ e $\{\varnothing\}$, e A non coincide con nessuno dei due;
   - $A \cap \varnothing = \varnothing$ e $\varnothing \subseteq \varnothing$ sono vere dato che $\varnothing$ è un sottoinsieme di ogni insieme incluso $\varnothing$ stesso;
   - $A \cap \{\varnothing\} = \{\varnothing\}$ è vera dato che $\varnothing \in A$;
   - $A \cap \{\{\varnothing\}\} = \{\{\varnothing\}\}$ è vera dato che $\{\varnothing\} \in A$;
   - $\varnothing \in \varnothing$ è falsa dato che $\varnothing$ non contiene elementi;
   - $A \setminus \{\{\{\varnothing\}\}\} = A$ è vera dato che $\{\{\varnothing\}\} \notin A$;
   - $A \cup \varnothing = A$ è vera dato che $\varnothing$ non ha elementi;
   - $A \cup \{\varnothing\} = A$ è vera dato che $\varnothing \in A$;

2. Per $A := \{1, 2, 3\}$ e $B := \{2, 3, 4\}$, calcolare $A \cup B$, $A \cap B$, $A \setminus B$, $\mathcal{P}(A)$ e $A \times \{0, 1\}$.
Soluzione. Si ha

$$A \cup B = \{1, 2, 3, 4\}, \qquad A \cap B = \{2, 3\}, \qquad A \setminus B = \{1\}.$$

L'insieme delle parti è

$$\mathcal{P}(A) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Infine

$$A \times \{0, 1\} = \{(1, 0), (1, 1), (2, 0), (2, 1), (3, 0), (3, 1)\}.$$

3. Mostrare con una funzione esplicita che l'insieme dei numeri naturali dispari è numerabile.
Soluzione. Sia

$$O := \{1, 3, 5, \ldots\}.$$

La funzione $f : \mathbb{N} \to O$ definita da $f(n) := 2n + 1$ è iniettiva, perché $2n + 1 = 2m + 1$ implica $n = m$, ed è suriettiva per definizione di O. È dunque una biiezione e mostra che O è numerabile.

4. Spiegare perché il *teorema di Cantor* non si limita a dire che "non abbiamo ancora trovato" un elenco di $\mathcal{P}(A)$.
Soluzione. Il teorema non esprime una difficoltà provvisoria nella ricerca di un elenco. Considera una funzione arbitraria $f : A \to \mathcal{P}(A)$ e costruisce

$$D := \{x \in A \mid x \notin f(x)\}.$$

L'insieme D differisce da ogni $f(a)$ almeno rispetto all'elemento a:

$$a \in D \iff a \notin f(a).$$

Quindi $D \neq f(a)$ per ogni $a \in A$. Si prova così che nessuna funzione $A \to \mathcal{P}(A)$ può essere suriettiva.

5. Ricostruire la dimostrazione diagonale indicando esattamente dove viene usata l'ipotesi di suriettività.
Soluzione. Poiché $D \subseteq A$, si ha $D \in \mathcal{P}(A)$. L'ipotesi che $f : A \to \mathcal{P}(A)$ sia suriettiva viene usata esattamente per dedurre l'esistenza di $d \in A$ tale che

$$f(d) = D.$$

Applicando la definizione di D a d:

$$d \in D \iff d \notin f(d) \iff d \notin D,$$

che è impossibile. Senza suriettività non avremmo il diritto di scegliere un tale d.

6. Mostrare che la relazione $\leq$ tra cardinalità, definita ponendo

$$|A| \leq |B|$$

se esiste una funzione iniettiva $A \to B$, è una relazione d'ordine parziale: verificare riflessività, transitività e antisimmetria. Per l'antisimmetria usare il *teorema di Cantor–Bernstein*. Spiegare infine perché questo non significa che la relazione così definita sia una relazione d'ordine sugli insiemi stessi.
Soluzione. La riflessività segue dal fatto che l'identità $\mathrm{id}_A : A \to A$ è iniettiva, dunque $|A| \leq |A|$. Per la transitività, se $|A| \leq |B|$ e $|B| \leq |C|$, esistono iniezioni $f : A \to B$ e $g : B \to C$; la composizione $g \circ f : A \to C$ è ancora iniettiva, e quindi $|A| \leq |C|$.
Per l'antisimmetria, se $|A| \leq |B|$ e $|B| \leq |A|$, esistono iniezioni $A \to B$ e $B \to A$. Il *teorema di Cantor–Bernstein* garantisce allora l'esistenza di una biiezione tra A e B, cioè

$$|A| = |B|.$$

La relazione $\leq$ è dunque un ordine parziale sulle cardinalità. Non è invece un ordine sugli insiemi stessi: da iniezioni in entrambe le direzioni si conclude che i due insiemi hanno la stessa cardinalità, non che siano lo stesso insieme. Per esempio, $\mathbb{N}$ e l'insieme dei numeri naturali pari sono distinti, ma hanno la stessa cardinalità.

7. Per un insieme fissato A, porre $R_A := \{x \in A \mid x \notin x\}$. Mostrare che $R_A \notin A$ e spiegare perché non nasce il *paradosso di Russell*.
Soluzione. Per definizione, per ogni oggetto x vale

$$x \in R_A \iff (x \in A \wedge x \notin x).$$

Supponiamo per assurdo che $R_A \in A$. Sostituendo $x = R_A$ nella formula precedente si ottiene allora

$$R_A \in R_A \iff (R_A \in A \wedge R_A \notin R_A).$$

Poiché stiamo assumendo $R_A \in A$, questa equivalenza diventa

$$R_A \in R_A \iff R_A \notin R_A,$$

una contraddizione. Dunque $R_A \notin A$. Non nasce un paradosso globale: la conclusione è precisamente che il sottoinsieme costruito non appartiene all'insieme di partenza A. Il punto delicato è che $R_A \in A$ non implica direttamente $R_A \in R_A$: serve solo a rendere applicabile a R_A la definizione relativa di R_A.

8. Usando l'assioma di fondazione, mostrare per assurdo che non esiste alcun insieme x tale che $x \in x$.
Soluzione. Supponiamo per assurdo che esista un insieme x tale che $x \in x$ e consideriamo il singoletto

$$A := \{x\}.$$

L'insieme A è non vuoto. Per l'assioma di fondazione deve dunque esistere un elemento $y \in A$ tale che

$$y \cap A = \varnothing.$$

Poiché l'unico elemento di A è x, necessariamente $y = x$. Dovrebbe quindi valere

$$x \cap \{x\} = \varnothing.$$

Ma dall'ipotesi $x \in x$ e dal fatto che $x \in \{x\}$ segue

$$x \in x \cap \{x\},$$

e dunque $x \cap \{x\} \neq \varnothing$, in contraddizione con quanto trovato sopra. Pertanto non esiste alcun insieme x tale che $x \in x$.

1.9 Letture consigliate e bibliografia

Per una prima lettura matematica sono particolarmente adatti Halmos [5] ed Enderton [6]. Per il versante filosofico dell'infinito si veda Moore [7]; per la mereologia, Cotnoir e Varzi [2]. Le fonti storiche sulla crisi dei fondamenti sono raccolte in van Heijenoort [9], mentre Shapiro [19] offre un'introduzione generale alla filosofia della matematica. Per approfondire le letture ontologiche appena richiamate, Benacerraf [15] è un testo classico alla base di molte motivazioni strutturaliste; Shapiro [16] sviluppa sistematicamente lo strutturalismo e Balaguer [18] discute in modo ampio il confronto tra platonismo e antiplatonismo.

Bibliografia del Capitolo 1

[1] A. C. Varzi, "Mereology", in E. N. Zalta (a cura di), *The Stanford Encyclopedia of Philosophy*, edizione Spring 2019.

[2] A. J. Cotnoir e A. C. Varzi, *Mereology*, Oxford University Press, 2021.

[3] D. K. Lewis, *Parts of Classes*, Basil Blackwell, 1991.

[4] G. Lando, *Mereology: A Philosophical Introduction*, Bloomsbury Academic, 2017.

[5] P. R. Halmos, *Naive Set Theory*, Springer-Verlag, New York, 1974.

[6] H. B. Enderton, *Elements of Set Theory*, Academic Press, 1977.

[7] A. W. Moore, *The Infinite*, 3a ed., Routledge, 2019.

[8] B. Russell, "Mathematical Logic as Based on the Theory of Types", *American Journal of Mathematics* **30**(3) (1908), 222–262.

[9] J. van Heijenoort (a cura di), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Harvard University Press, 1967.

[10] E. Zermelo, "Untersuchungen über die Grundlagen der Mengenlehre I", *Mathematische Annalen* **65** (1908), 261–281.

[11] E. Zermelo, "Investigations in the Foundations of Set Theory I", trad. S. Bauer-Mengelberg, in J. van Heijenoort (a cura di), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Harvard University Press, Cambridge (MA), 1967, pp. 199–215.

[12] K. Gödel, *The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis with the Axioms of Set Theory*, Annals of Mathematics Studies 3, Princeton University Press, 1940.

[13] P. J. Cohen, "The Independence of the Continuum Hypothesis", *Proceedings of the National Academy of Sciences of the United States of America* **50**(6) (1963), 1143–1148.

[14] P. J. Cohen, *Set Theory and the Continuum Hypothesis*, W. A. Benjamin, New York, 1966.

[15] P. Benacerraf, "What Numbers Could Not Be", *The Philosophical Review* **74**(1) (1965), 47–73.

[16] S. Shapiro, *Philosophy of Mathematics: Structure and Ontology*, Oxford University Press, 1997.

[17] H. Field, *Science Without Numbers: A Defence of Nominalism*, Princeton University Press, 1980.

[18] M. Balaguer, *Platonism and Anti-Platonism in Mathematics*, Oxford University Press, 1998.

[19] S. Shapiro, *Thinking about Mathematics: The Philosophy of Mathematics*, Oxford University Press, 2000.

[20] P. Mancosu (a cura di), *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s*, Oxford University Press, 1998.

CAPITOLO 2

I numeri reali ed elementi di analisi matematica

2.1 Introduzione: dai numeri reali all'analisi matematica

Questo capitolo è dedicato all'insieme dei *numeri reali* $\mathbb{R}$ e alle nozioni elementari di *analisi matematica* (la versione moderna del *calcolo infinitesimale* di Newton e Leibniz) che possono essere costruite a partire da esso. Il punto di vista resterà quello fondazionale del Capitolo 1: non assumeremo i reali come una nuova specie di oggetti primitivi, ma mostreremo come essi possano essere ottenuti entro ZF a partire dai numeri naturali già costruiti. Nel Capitolo 1 abbiamo infatti visto come, nella rappresentazione di von Neumann, l'insieme dei naturali possa essere identificato con *il più piccolo insieme induttivo*

$$\omega = \{0, 1, 2, 3, \ldots\},$$

e nel seguito useremo, come di consueto, la notazione

$$\mathbb{N} = \omega.$$

A partire da $\mathbb{N}$ costruiremo progressivamente gli altri insiemi numerici come segue. Nello schema il simbolo $\rightsquigarrow$ ha soltanto valore espositivo e si legge "si passa a" o "conduce a":

$$\mathbb{N} \rightsquigarrow \mathbb{Z} \rightsquigarrow \mathbb{Q} \rightsquigarrow \mathbb{R}.$$

I primi due passaggi saranno realizzati mediante *relazioni* e quindi mediante *classi di equivalenza*; il passaggio da $\mathbb{Q}$ a $\mathbb{R}$ verrà invece effettuato mediante le cosiddette *sezioni di Dedekind*. Non intendiamo ricostruire la storia effettiva con cui i diversi sistemi numerici sono comparsi nella matematica: gli accenni storici saranno limitati al minimo necessario, mentre l'attenzione sarà rivolta alla costruzione moderna degli oggetti e alle proprietà che essa mette in luce. Esiste in particolare un'altra costruzione equivalente, più assiomatica e meno esplicitamente costruttiva, che consiste nel caratterizzare $\mathbb{R}$ mediante un elenco di dieci assiomi nella presentazione adottata in [1, cap. 8]. Tale costruzione presuppone però alcune nozioni riferite a particolari strutture algebriche, in particolare la nozione di *campo*, che vedremo nel prossimo capitolo.

Il passaggio dai razionali ai reali introduce la proprietà decisiva della *completezza*. A partire da essa studieremo le successioni reali e la nozione di limite, quindi estenderemo il concetto ai limiti di funzioni mediante la definizione ε–δ. Il limite permetterà di definire la *continuità* e successivamente la *derivabilità*, interpretando la derivata come limite del rapporto incrementale. Dalla derivata passeremo poi a una prima introduzione alle *equazioni differenziali*, con particolare attenzione al loro ruolo nella fisica meccanica e alla relazione fra condizioni iniziali e determinazione dell'evoluzione.

La parte finale del capitolo sarà dedicata all'*integrazione*. Introdurremo l'*integrale di Riemann* attraverso somme associate a suddivisioni sempre più fini di un intervallo e vedremo il legame fra integrazione e derivazione espresso dal teorema fondamentale del calcolo. Concluderemo con un primo ingresso nel linguaggio moderno della misura: definiremo una σ-algebra e una *misura* σ-additiva e ne discuteremo alcuni esempi elementari, senza sviluppare la teoria dell'integrazione di Lebesgue.

Il filo concettuale del capitolo sarà quindi

$$\begin{aligned}\mathbb{N} \rightsquigarrow \mathbb{Z} \rightsquigarrow \mathbb{Q} \rightsquigarrow \mathbb{R} \rightsquigarrow \text{completezza}\\ \rightsquigarrow \text{limite} \rightsquigarrow \text{continuità} \rightsquigarrow \text{derivabilità}\\ \rightsquigarrow \text{equazioni differenziali} \rightsquigarrow \text{integrazione} \rightsquigarrow \text{misura}.\end{aligned}$$

Come nel capitolo precedente, le definizioni e i risultati principali saranno accompagnati da esercizi risolti. Alcuni di essi sono scelti anche con un secondo scopo: far emergere, senza introdurne ancora il linguaggio astratto, proprietà delle operazioni che nel Capitolo 3 condurranno alle nozioni di monoide, gruppo, anello, campo, spazio vettoriale e algebra associativa.

2.1.1 Una precisazione fondazionale sull'aritmetica di $\mathbb{N}$

Nel Capitolo 1 abbiamo costruito l'insieme $\mathbb{N} = \omega$ a partire dagli assiomi di ZF. Per procedere abbiamo ora bisogno anche delle usuali operazioni aritmetiche sui naturali. Esse non devono essere considerate nuovi oggetti primitivi: possono essere definite in ZF a partire dalla funzione successore

$$S(n) := n \cup \{n\}$$

mediante definizioni ricorsive.

Per esempio, fissato $m \in \mathbb{N}$, la somma è caratterizzata dalle condizioni

$$m + 0 = m, \qquad m + S(n) = S(m + n),$$

mentre il prodotto è caratterizzato da

$$m \cdot 0 = 0, \qquad m \cdot S(n) = m \cdot n + m.$$

Analogamente si definisce l'ordine naturale. Non svilupperemo qui la teoria generale della ricorsione su ω; useremo le usuali proprietà di somma, prodotto e ordine dei naturali, ricordando però che anche queste operazioni e proprietà possono essere ricostruite all'interno di ZF.

2.1 Osservazione [costruire una struttura numerica]. Costruire un nuovo sistema numerico non significa soltanto specificarne gli elementi. Occorre anche definire su di esso le operazioni che intendiamo usare e verificare che esse abbiano le proprietà richieste. Quando gli elementi sono classi di equivalenza, compare inoltre un problema nuovo: le operazioni devono essere indipendenti dal rappresentante scelto per ciascuna classe. ■

2.2 Dai numeri naturali ai numeri interi

Nei numeri naturali la sottrazione non è sempre possibile. Per esempio, non esiste alcun $n \in \mathbb{N}$ tale che

$$n + 3 = 2.$$

Vogliamo costruire un nuovo sistema numerico nel quale abbia significato anche una differenza come $2 - 3$.

L'idea consiste nel rappresentare formalmente una differenza $a - b$, con $a, b \in \mathbb{N}$, mediante la coppia ordinata

$$(a, b).$$

Non possiamo però identificare direttamente gli interi con le coppie di naturali, perché coppie differenti devono rappresentare la stessa differenza. Per esempio,

$$(3, 0), \qquad (4, 1), \qquad (5, 2)$$

devono rappresentare lo stesso numero.

Definiamo dunque su $\mathbb{N} \times \mathbb{N}$ la relazione

$$(a, b) \sim (c, d) \quad \Longleftrightarrow \quad a + d = b + c.$$

La condizione esprime l'uguaglianza formale

$$a - b = c - d$$

senza richiedere che la sottrazione sia già definita come operazione sempre possibile nei naturali.

2.2 Proposizione [la relazione che costruisce gli interi]. *La relazione $\sim$ definita su $\mathbb{N} \times \mathbb{N}$ da*

$$(a, b) \sim (c, d) \quad \Longleftrightarrow \quad a + d = b + c$$

è una relazione di equivalenza.

Dimostrazione. La riflessività segue dalla commutatività della somma:

$$a + b = b + a,$$

quindi $(a, b) \sim (a, b)$.

La simmetria è immediata: se $a + d = b + c$, allora $c + b = d + a$, e quindi $(c, d) \sim (a, b)$.
Per la transitività, supponiamo

$$a + d = b + c, \qquad c + f = d + e.$$

Sommando f alla prima uguaglianza e b alla seconda otteniamo

$$a + d + f = b + c + f = b + d + e.$$

Per associatività e commutatività,

$$d + (a + f) = d + (b + e).$$

Per la cancellazione della somma nei naturali segue

$$a + f = b + e,$$

cioè $(a, b) \sim (e, f)$. □

2.2.1 Classi di equivalenza e definizione di $\mathbb{Z}$

Ricordiamo che, se R è una relazione di equivalenza su un insieme A, la classe di equivalenza di $x \in A$ è

$$[x]_R := \{y \in A \mid yRx\}.$$

L'insieme di tutte le classi di equivalenza viene indicato con

$$A/R$$

e si chiama **insieme quoziente**. Nel nostro caso definiamo

$$\mathbb{Z} := (\mathbb{N} \times \mathbb{N})/\sim.$$

Un elemento di $\mathbb{Z}$ è dunque una classe di equivalenza

$$[(a, b)].$$

Intuitivamente essa rappresenta la differenza $a - b$. Per esempio,

$$[(3, 0)] = [(4, 1)] = [(5, 2)] = \cdots$$

rappresenta l'intero 3, mentre

$$[(0, 3)] = [(1, 4)] = [(2, 5)] = \cdots$$

rappresenta l'intero -3. Lo zero è rappresentato da

$$[(0, 0)] = [(1, 1)] = [(2, 2)] = \cdots .$$

I naturali vengono rappresentati all'interno di $\mathbb{Z}$ mediante l'applicazione

$$\iota : \mathbb{N} \to \mathbb{Z}, \qquad \iota(n) := [(n, 0)].$$

Questa applicazione è iniettiva. Dopo aver identificato ogni naturale n con la sua immagine $[(n, 0)]$, scriveremo semplicemente

$$\mathbb{N} \subseteq \mathbb{Z}.$$

È importante ricordare che si tratta di un'identificazione strutturale: il naturale n costruito come insieme di von Neumann e la classe $[(n, 0)]$ non sono letteralmente lo stesso insieme.

2.2.2 Somma e prodotto degli interi

Definiamo la somma ponendo

$$[(a,b)] + [(c,d)] := [(a+c, b+d)].$$

Questa formula corrisponde all'identità intuitiva

$$(a-b) + (c-d) = (a+c) - (b+d).$$

Definiamo inoltre il prodotto ponendo

$$[(a,b)] \cdot [(c,d)] := [(ac+bd, ad+bc)],$$

coerentemente con il calcolo formale

$$(a-b)(c-d) = ac + bd - (ad+bc).$$

Poiché una classe di equivalenza possiede molti rappresentanti, dobbiamo verificare che queste definizioni non dipendano dalle coppie scelte.

2.3 Definizione [operazione ben posta sul quoziente]. Un'operazione definita mediante rappresentanti di classi di equivalenza si dice **ben posta** se il risultato dipende soltanto dalle classi e non dai particolari rappresentanti scelti. ■

2.4 Proposizione [buona definizione della somma]. *La somma su $\mathbb{Z}$ definita, sui rappresentanti, da*

$$[(a,b)] + [(c,d)] := [(a+c, b+d)]$$

è ben posta nel senso della Definizione 2.3.

Dimostrazione. Supponiamo

$$(a,b) \sim (a',b'), \qquad (c,d) \sim (c',d').$$

Allora

$$a + b' = b + a', \qquad c + d' = d + c'.$$

Sommando membro a membro,

$$a + c + b' + d' = b + d + a' + c'.$$

Per associatività e commutatività della somma nei naturali,

$$(a+c) + (b'+d') = (b+d) + (a'+c').$$

Dunque

$$(a+c, b+d) \sim (a'+c', b'+d'),$$

e quindi

$$[(a+c, b+d)] = [(a'+c', b'+d')].$$

La somma non dipende dai rappresentanti. □

2.5 Osservazione. La verifica analoga per il prodotto è più laboriosa ma segue la stessa idea: si sostituiscono i rappresentanti con coppie equivalenti e si mostra che la classe ottenuta non cambia. La proponiamo come esercizio guidato, perché anticipa un procedimento che incontreremo spesso: quando un oggetto è definito come classe di equivalenza, ogni nuova costruzione effettuata sui rappresentanti deve rispettare la relazione di equivalenza. ■

2.2.3 Esercizi risolti sugli interi

2.6 Esercizio.

1. *Rappresentazioni dello stesso intero.* Stabilire se le coppie $(7,3)$ e $(5,1)$ rappresentano lo stesso intero.
   Soluzione. Per definizione,
   $$(7,3) \sim (5,1)$$
   se e solo se
   $$7+1 = 3+5.$$
   Entrambi i membri valgono 8. Dunque
   $$[(7,3)] = [(5,1)],$$
   e le due coppie rappresentano l'intero 4.
2. *L'opposto.* Dato
   $$z = [(a,b)],$$
   mostrare che
   $$-z := [(b,a)]$$
   soddisfa
   $$z + (-z) = 0.$$
   Soluzione. Per definizione della somma,
   $$[(a,b)] + [(b,a)] = [(a+b, b+a)].$$
   Poiché $a+b = b+a$, la coppia $(a+b, b+a)$ è equivalente a $(0,0)$. Pertanto
   $$z + (-z) = [(0,0)] = 0.$$
3. *Proprietà della somma.* Verificare, usando la definizione sulle classi, che per ogni $x, y, z \in \mathbb{Z}$
   $$x + y = y + x, \qquad (x+y) + z = x + (y+z).$$
   Soluzione. Scriviamo
   $$x = [(a,b)], \qquad y = [(c,d)], \qquad z = [(e,f)].$$
   Allora
   $$x + y = [(a+c, b+d)], \qquad y + x = [(c+a, d+b)],$$
   e le due classi coincidono per la commutatività della somma nei naturali.
   Analogamente,
   $$(x+y) + z = [((a+c)+e, (b+d)+f)],$$
   mentre
   $$x + (y+z) = [(a+(c+e), b+(d+f))].$$
   Le due classi coincidono per l'associatività della somma nei naturali.
4. *Buona definizione del prodotto.* Siano
   $$(a,b) \sim (a',b'), \qquad (c,d) \sim (c',d').$$
   Mostrare che
   $$(ac+bd, ad+bc) \sim (a'c' + b'd', a'd' + b'c').$$
   Concludere che il prodotto su $\mathbb{Z}$ è ben posto.
   Soluzione. Dalle ipotesi abbiamo
   $$a + b' = b + a', \qquad c + d' = d + c'.$$
   Il modo più trasparente di organizzare il calcolo consiste nell'espandere le due condizioni necessarie all'equivalenza e usare ripetutamente distributività, associatività e commutatività nei naturali. Si ottiene
   $$(ac+bd) + (a'd' + b'c') = (ad+bc) + (a'c' + b'd'),$$
   che è precisamente la condizione richiesta. Di conseguenza la classe del prodotto non dipende dai rappresentanti scelti.

5. *Proprietà da conservare.* Verificare per gli interi, a partire dalle definizioni, almeno le seguenti proprietà:
$$x+y=y+x, \qquad (x+y)+z=x+(y+z),$$
$$xy=yx, \qquad (xy)z=x(yz),$$
$$x(y+z)=xy+xz.$$
Individuare inoltre gli elementi che svolgono il ruolo di 0 e di 1.
Soluzione. Le proprietà discendono dalle corrispondenti proprietà delle operazioni sui naturali e dalla buona definizione delle operazioni sul quoziente. Gli elementi neutri sono
$$0=[(0,0)], \qquad 1=[(1,0)].$$
Ogni intero possiede inoltre un opposto additivo. Registriamo queste proprietà: nel Capitolo 3 vedremo che la loro forma, considerata indipendentemente dal significato degli elementi, definisce strutture matematiche che compaiono in contesti molto più generali dei sistemi numerici.

2.3 Dai numeri interi ai numeri razionali

Negli interi la sottrazione è sempre possibile, ma la divisione non lo è. Per esempio, non esiste alcun $x \in \mathbb{Z}$ tale che
$$2x=1.$$
Vogliamo costruire un sistema nel quale sia possibile dividere per ogni numero diverso da zero.

Consideriamo
$$\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}).$$
Una coppia
$$(a,b), \qquad b \neq 0,$$
rappresenterà intuitivamente il rapporto a/b. Ancora una volta coppie differenti possono rappresentare lo stesso numero, per esempio
$$\frac{1}{2}=\frac{2}{4}=\frac{3}{6}.$$
Definiamo dunque
$$(a,b) \sim (c,d) \quad \Longleftrightarrow \quad ad=bc.$$

2.7 Proposizione [la relazione che costruisce i razionali]. *Sull'insieme*
$$\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$$
la relazione
$$(a,b) \sim (c,d) \quad \Longleftrightarrow \quad ad=bc$$
è una relazione di equivalenza.

Dimostrazione. La riflessività e la simmetria sono immediate. Per la transitività, supponiamo
$$ad=bc, \qquad cf=de.$$
Moltiplicando la prima uguaglianza per f e la seconda per b otteniamo
$$adf=bcf=bde.$$
Poiché $d \neq 0$ e negli interi vale la cancellazione rispetto alla moltiplicazione per un fattore non nullo, segue
$$af=be.$$
Quindi $(a,b) \sim (e,f)$. □

Definiamo allora
$$\mathbb{Q} := \big(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\big)/\!\sim.$$
La classe
$$[(a,b)]$$
rappresenta il numero razionale che usualmente scriviamo a/b.

Gli interi vengono rappresentati nei razionali mediante
$$\jmath : \mathbb{Z} \to \mathbb{Q}, \qquad \jmath(a) := [(a,1)].$$
Dopo questa identificazione scriveremo
$$\mathbb{Z} \subseteq \mathbb{Q}.$$

2.3.1 Operazioni sui razionali

Definiamo

$$[(a,b)] + [(c,d)] := [(ad+bc, bd)]$$

e

$$[(a,b)] \cdot [(c,d)] := [(ac, bd)].$$

Sono precisamente le usuali regole

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}, \qquad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}.$$

2.8 Proposizione [buona definizione delle operazioni su $\mathbb{Q}$]. *Le definizioni precedenti non dipendono dai rappresentanti scelti: se*

$$(a,b) \sim (a',b'), \qquad (c,d) \sim (c',d'),$$

allora le somme e i prodotti costruiti con le due coppie di rappresentanti appartengono alle stesse classi di equivalenza.

Dimostrazione. Dalle ipotesi seguono $ab' = a'b$ e $cd' = c'd$. Per la somma,

$$\begin{aligned}(ad+bc)b'd' &= ab'dd' + bb'cd' \\ &= a'bdd' + bb'c'd \\ &= (a'd' + b'c')bd,\end{aligned}$$

quindi $(ad+bc, bd) \sim (a'd' + b'c', b'd')$. Per il prodotto,

$$(ac)b'd' = (ab')(cd') = (a'b)(c'd) = (a'c')bd,$$

quindi $(ac, bd) \sim (a'c', b'd')$. Le due operazioni sono pertanto ben definite sulle classi. □

Una nuova proprietà compare ora rispetto agli interi. Se

$$q = [(a,b)] \neq 0,$$

allora $a \neq 0$ e possiamo definire

$$q^{-1} := [(b,a)].$$

Si ha

$$qq^{-1} = 1.$$

Ogni razionale non nullo possiede dunque un inverso rispetto alla moltiplicazione.

2.3.2 Esercizi risolti sui razionali

2.9 Esercizio.

1. *Frazioni equivalenti.* Verificare che
$$[(2,3)] = [(6,9)].$$
Soluzione. Per definizione,
$$(2,3) \sim (6,9)$$
se
$$2 \cdot 9 = 3 \cdot 6.$$
Poiché entrambi i membri valgono 18, le due coppie appartengono alla stessa classe di equivalenza.
2. *Inverso moltiplicativo.* Sia
$$q = [(3,5)].$$
Determinare un razionale r tale che $qr = 1$.
Soluzione. Poniamo
$$r = [(5,3)].$$
Allora
$$qr = [(15,15)].$$
Ma $(15,15) \sim (1,1)$, perché $15 \cdot 1 = 15 \cdot 1$. Dunque
$$qr = [(1,1)] = 1.$$

3. *Una differenza tra $\mathbb{Z}$ e $\mathbb{Q}$.* Stabilire se l'equazione

$$3x = 2$$

ammette soluzione in $\mathbb{Z}$ e in $\mathbb{Q}$.
Soluzione. In $\mathbb{Z}$ non esiste alcun intero x tale che $3x = 2$. In $\mathbb{Q}$, invece,

$$x = \frac{2}{3}$$

è una soluzione. Il passaggio da $\mathbb{Z}$ a $\mathbb{Q}$ rende dunque possibile la divisione per ogni elemento diverso da zero.

4. *Proprietà delle operazioni.* Verificare sui razionali le identità

$$x + y = y + x, \qquad (x + y) + z = x + (y + z),$$
$$xy = yx, \qquad (xy)z = x(yz),$$
$$x(y + z) = xy + xz.$$

Individuare inoltre gli elementi neutri per somma e prodotto e verificare che ogni razionale possiede un opposto additivo e ogni razionale non nullo possiede un inverso moltiplicativo.
Soluzione. Gli elementi neutri sono

$$0 = [(0, 1)], \qquad 1 = [(1, 1)].$$

L'opposto di $[(a, b)]$ è $[(-a, b)]$ e, se $a \neq 0$, l'inverso moltiplicativo è $[(b, a)]$. Le altre proprietà seguono dalle corrispondenti proprietà degli interi e dalla buona definizione delle operazioni sulle classi.
Questo esercizio riassume una parte importante della struttura aritmetica di $\mathbb{Q}$. Nel Capitolo 3 confronteremo proprietà di questo tipo con quelle possedute da altri insiemi dotati di operazioni, anche quando gli elementi non sono numeri e anche quando il prodotto non è commutativo.

2.4 Perché i numeri razionali non bastano?

A questo punto disponiamo di un sistema numerico molto ricco. In $\mathbb{Q}$ possiamo eseguire

$$+, \qquad -, \qquad \cdot, \qquad /$$

con la sola esclusione della divisione per zero. Il problema che conduce ai numeri reali è quindi differente.

Consideriamo l'equazione

$$x^2 = 2.$$

Essa non possiede soluzione razionale.

2.10 Proposizione [irrazionalità di $\sqrt{2}$]. *Non esiste alcun $q \in \mathbb{Q}$ tale che $q^2 = 2$.*

Dimostrazione. Supponiamo per assurdo che esistano interi p e $r \neq 0$, scelti senza fattori comuni, tali che

$$\left(\frac{p}{r}\right)^2 = 2.$$

Allora

$$p^2 = 2r^2.$$

Quindi p^2 è pari e pertanto p è pari. Scriviamo $p = 2k$. Sostituendo,

$$4k^2 = 2r^2,$$

da cui

$$r^2 = 2k^2.$$

Anche r è quindi pari. Ma allora p e r hanno il fattore comune 2, contro l'ipotesi che la frazione fosse ridotta ai minimi termini. La supposizione iniziale è impossibile. □

Consideriamo ora l'insieme

$$A := \{q \in \mathbb{Q} \mid q \geq 0 \text{ e } q^2 < 2\}.$$

Esso è non vuoto, perché $1 \in A$, ed è limitato superiormente, perché per esempio 2 è un maggiorante. Intuitivamente ci aspettiamo che esista un numero che rappresenti il "confine" superiore di A, precisamente il numero che chiamiamo $\sqrt{2}$. Ma tale numero non appartiene a $\mathbb{Q}$.

Il problema fondamentale di $\mathbb{Q}$ non riguarda quindi più la possibilità di eseguire un'operazione aritmetica. Riguarda invece l'esistenza di certi estremi e, come vedremo in seguito, di certi limiti.

2.11 Osservazione [il nuovo problema]. I passaggi $\mathbb{N} \rightsquigarrow \mathbb{Z}$ e $\mathbb{Z} \rightsquigarrow \mathbb{Q}$ ampliano le operazioni aritmetiche disponibili. Il passaggio $\mathbb{Q} \rightsquigarrow \mathbb{R}$ risponde invece a un problema di natura differente: vogliamo ottenere un sistema numerico privo delle lacune d'ordine che compaiono nei razionali. Questa proprietà verrà formulata mediante la nozione di completezza. ■

2.4.1 Esercizi risolti verso la completezza

2.12 Esercizio.

1. *Approssimazioni razionali di* $\sqrt{2}$. Verificare che
$$1.4^2 < 2 < 1.5^2$$
e che
$$1.41^2 < 2 < 1.42^2.$$
Soluzione. Si ha
$$1.4^2 = 1.96 < 2, \qquad 1.5^2 = 2.25 > 2,$$
e inoltre
$$1.41^2 = 1.9881 < 2, \qquad 1.42^2 = 2.0164 > 2.$$
Le approssimazioni razionali possono essere rese progressivamente più precise, ma il numero che esse suggeriscono non appartiene a $\mathbb{Q}$.
2. *Un insieme limitato.* Per
$$A = \{q \in \mathbb{Q} \mid q \geq 0 \text{ e } q^2 < 2\}$$
mostrare che A è non vuoto e trovare almeno due maggioranti razionali distinti.
Soluzione. Poiché $1^2 = 1 < 2$, si ha $1 \in A$, dunque $A \neq \varnothing$. Inoltre 2 è un maggiorante, perché se $q \geq 2$ allora $q^2 \geq 4 > 2$. Anche $3/2$ è un maggiorante, perché
$$\left(\frac{3}{2}\right)^2 = \frac{9}{4} > 2.$$
Quindi ogni elemento di A è minore di $3/2$.

2.5 Verso i numeri reali: sezioni di Dedekind

La costruzione moderna dei numeri reali può essere realizzata in modi differenti ma equivalenti dal punto di vista della struttura ottenuta. In queste dispense useremo principalmente le **sezioni di Dedekind**, perché permettono di costruire i reali direttamente come particolari sottoinsiemi di $\mathbb{Q}$ e rendono particolarmente trasparente il rapporto fra ordine e completezza.

La costruzione prende il nome da Richard Dedekind, che nel XIX secolo diede una formulazione rigorosa di questa idea. Qui, tuttavia, il nostro interesse non è storico: useremo le sezioni come una costruzione insiemistica all'interno del quadro fondazionale introdotto nel Capitolo 1.

L'idea iniziale è semplice. Ogni razionale r determina il sottoinsieme
$$A_r := \{q \in \mathbb{Q} \mid q < r\}.$$
Possiamo quindi rappresentare r mediante l'insieme di tutti i razionali che lo precedono. Il passo decisivo consisterà nel considerare anche sottoinsiemi di $\mathbb{Q}$ che possiedono le stesse proprietà strutturali degli insiemi A_r ma che non corrispondono ad alcun razionale. Uno di essi permetterà di rappresentare precisamente il punto che manca in $\mathbb{Q}$ in corrispondenza di $\sqrt{2}$.

2.5.1 Definizione di sezione di Dedekind

2.13 Definizione [sezione di Dedekind]. Una **sezione di Dedekind** è un sottoinsieme $A \subseteq \mathbb{Q}$ che soddisfa le seguenti proprietà:

(i) $A \neq \varnothing$ e $A \neq \mathbb{Q}$;

(ii) se $q \in A$ e $p < q$, allora $p \in A$;

(iii) A non possiede un elemento massimo: per ogni $q \in A$ esiste $r \in A$ tale che $q < r$.

■

La seconda proprietà dice che una sezione contiene, insieme a ogni suo elemento, tutti i razionali che lo precedono. La terza impedisce che il confine superiore della sezione sia già incluso nella parte inferiore. Una sezione descrive quindi una porzione iniziale di $\mathbb{Q}$ priva del proprio eventuale punto di separazione.

2.14 Esempio [la sezione determinata da un razionale]. Per ogni $r \in \mathbb{Q}$ poniamo

$$A_r := \{q \in \mathbb{Q} \mid q < r\}.$$

Allora A_r è una sezione di Dedekind. È non vuota e diversa da $\mathbb{Q}$; se $q < r$ e $p < q$, allora $p < r$; infine, fra due razionali distinti esiste sempre un altro razionale, perciò A_r non ha elemento massimo. ■

La proprietà appena usata — fra due razionali distinti esiste sempre un razionale — viene detta **densità** di $\mathbb{Q}$. Per esempio, se $p < q$, il numero

$$\frac{p+q}{2}$$

è razionale e soddisfa

$$p < \frac{p+q}{2} < q.$$

La densità non deve però essere confusa con la completezza: anche se fra due razionali possiamo sempre inserirne altri, in $\mathbb{Q}$ rimangono le lacune illustrate dal caso di $\sqrt{2}$.

2.15 Esempio [la sezione che rappresenta $\sqrt{2}$]. Consideriamo

$$S := \{q \in \mathbb{Q} \mid q < 0 \text{ oppure } q^2 < 2\}.$$

Questo insieme è una sezione di Dedekind. Esso rappresenterà il numero reale positivo che indicheremo con $\sqrt{2}$.

Osserviamo in particolare che nessun razionale r soddisfa

$$S = A_r.$$

Se ciò accadesse, il confine della sezione sarebbe un numero razionale; ma il confine individuato dalla condizione $q^2 < 2$ è precisamente quello che manca in $\mathbb{Q}$. ■

2.16 Osservazione [un reale come oggetto insiemistico]. La costruzione non aggiunge a ZF una nuova specie di oggetti primitivi. Una sezione è un sottoinsieme di $\mathbb{Q}$ e dunque un elemento di $\mathcal{P}(\mathbb{Q})$. L'insieme dei numeri reali potrà pertanto essere ottenuto selezionando, dentro $\mathcal{P}(\mathbb{Q})$, precisamente i sottoinsiemi che soddisfano le tre condizioni della Definizione 2.13. ■

2.5.2 Definizione dell'insieme dei numeri reali

Definiamo

$$\mathbb{R} := \{A \in \mathcal{P}(\mathbb{Q}) \mid A \text{ è una sezione di Dedekind}\}. \tag{2.1}$$

La definizione (2.1), dal punto di vista di ZF, questa è un'applicazione diretta dello schema di separazione all'insieme già esistente $\mathcal{P}(\mathbb{Q})$. In particolare, $\mathbb{R}$ è ancora un insieme puro costruito a partire dagli oggetti introdotti in precedenza.

Ogni numero reale è dunque, nella costruzione scelta, una sezione di Dedekind. Come già accaduto passando da $\mathbb{N}$ a $\mathbb{Z}$ e da $\mathbb{Z}$ a $\mathbb{Q}$, non dobbiamo confondere la rappresentazione insiemistica con l'intuizione usuale del numero: ciò che conta è che sulla nuova collezione si possano definire ordine e operazioni in modo da ottenere la struttura numerica desiderata.

I razionali si immergono nei reali mediante la funzione

$$\iota_{\mathbb{Q}} : \mathbb{Q} \longrightarrow \mathbb{R}, \qquad \iota_{\mathbb{Q}}(r) := A_r.$$

Essa è iniettiva. Infatti, se $r < s$, per densità esiste $q \in \mathbb{Q}$ con

$$r < q < s.$$

Allora $q \in A_s$ ma $q \notin A_r$, e quindi $A_r \neq A_s$.

Dopo questa identificazione scriveremo, come di consueto,

$$\mathbb{Q} \subseteq \mathbb{R}.$$

Anche qui il simbolo di inclusione esprime un'identificazione mediante una funzione iniettiva: il razionale r e la sezione A_r non sono letteralmente lo stesso insieme.

2.5.3 L'ordine nei numeri reali

La rappresentazione mediante sezioni rende l'ordine particolarmente semplice.

2.17 Definizione [ordine sui reali]. Se $A, B \in \mathbb{R}$, definiamo

$$A \leq B \quad \Longleftrightarrow \quad A \subseteq B.$$

Scriviamo $A < B$ quando $A \subsetneq B$. ■

Questa definizione coincide con l'ordine usuale sui razionali identificati con le corrispondenti sezioni di Dedekind in $\mathbb{R}$. Infatti,

$$r < s \quad \Longleftrightarrow \quad A_r \subsetneq A_s.$$

Inoltre due sezioni sono sempre confrontabili per inclusione. Supponiamo infatti che $A \not\subseteq B$ e scegliamo $a \in A \setminus B$. Se $b \in B$ fosse maggiore di a, dalla proprietà di chiusura verso il basso di B seguirebbe $a \in B$, assurdo. Dunque ogni $b \in B$ soddisfa $b < a$. Poiché $a \in A$ e A è chiusa verso il basso, segue $b \in A$. Pertanto

$$B \subseteq A.$$

L'ordine così definito è quindi totale: per ogni $A, B \in \mathbb{R}$ vale

$$A \leq B \qquad \text{oppure} \qquad B \leq A.$$

2.5.4 Somma e prodotto: come si recupera l'aritmetica

La costruzione dei reali deve conservare le operazioni già presenti nei razionali. Per la somma la definizione è particolarmente trasparente. Se $A, B \in \mathbb{R}$, poniamo

$$A + B := \{a + b \mid a \in A,\ b \in B\}.$$

Si verifica che $A + B$ è ancora una sezione di Dedekind. Inoltre, per $r, s \in \mathbb{Q}$,

$$A_r + A_s = A_{r+s},$$

perciò la nuova operazione estende la somma razionale.

Lo zero reale è la sezione

$$0 = A_0 = \{q \in \mathbb{Q} \mid q < 0\}.$$

Si possono definire anche l'opposto di una sezione, il prodotto e l'inverso di ogni sezione non nulla. Le formule complete sono più laboriose di quella della somma e non sono necessarie per il nostro obiettivo principale; ne registriamo però il risultato strutturale.

2.18 Teorema [struttura dei reali costruiti mediante sezioni]. *Con le definizioni standard di somma, opposto, prodotto, inverso degli elementi non nulli e ordine, l'insieme delle sezioni di Dedekind è un campo ordinato completo. L'applicazione*

$$r \longmapsto A_r \qquad (r \in \mathbb{Q})$$

identifica i razionali con un sottocampo ordinato dei reali e conserva le operazioni e l'ordine.

In questa formulazione, "campo ordinato completo" significa che sono possibili le quattro operazioni aritmetiche usuali, che l'ordine è totale e compatibile con somma e prodotto e che ogni sottoinsieme non vuoto e limitato superiormente possiede un estremo superiore. La nozione astratta di campo sarà definita

nel Capitolo 3; la completezza verrà dimostrata direttamente nel Teorema 2.23. Omettiamo invece le verifiche algebriche più lunghe, che si possono trovare in [2, 1].

2.19 Osservazione [che cosa distingue allora $\mathbb{R}$ da $\mathbb{Q}$?]. Dal punto di vista delle operazioni aritmetiche fondamentali, $\mathbb{Q}$ e $\mathbb{R}$ si comportano in modo molto simile: in entrambi possiamo sommare, sottrarre, moltiplicare e dividere per un elemento non nullo, e in entrambi esiste un ordine compatibile con tali operazioni. La differenza decisiva per l'analisi matematica non è dunque una nuova operazione, ma una proprietà dell'ordine: la *completezza*. ■

2.6 Estremo superiore e completezza dei numeri reali

2.20 Definizione [maggioranti, minoranti, estremo superiore e inferiore]. Sia $E \subseteq \mathbb{R}$. Un numero reale M si dice **maggiorante** di E se

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

Un maggiorante s si dice **estremo superiore** o *supremo* di E se è minore o uguale a ogni altro maggiorante di E. Scriviamo

$$s = \sup E.$$

Simmetricamente, un numero reale m si dice **minorante** di E se

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

Un minorante i si dice **estremo inferiore** o *infimo* di E se è maggiore o uguale a ogni altro minorante di E. Scriviamo

$$i = \inf E.$$

Se $\sup E \in E$, allora $\sup E$ si chiama **massimo** di E e si scrive

$$\max E = \sup E.$$

Analogamente, se $\inf E \in E$, allora $\inf E$ si chiama **minimo** di E e si scrive

$$\min E = \inf E.$$

■

La nozione di estremo superiore o inferiore non richiede che l'estremo appartenga a E: un insieme può quindi possedere *supremo* o *infimo* senza possedere, rispettivamente, massimo o minimo.

2.21 Esempio. Nell'insieme

$$E = \{x \in \mathbb{R} \mid 0 < x < 1\}$$

l'estremo superiore è 1, ma $1 \notin E$. Dunque E non possiede massimo. ■

2.22 Definizione [completezza per l'estremo superiore]. Un sistema numerico ordinato si dice **completo** rispetto all'ordine se ogni suo sottoinsieme non vuoto e limitato superiormente possiede un estremo superiore appartenente al sistema stesso. ■

La costruzione di Dedekind rende questa proprietà quasi visibile nella rappresentazione dei reali data in (2.1).

2.23 Teorema [completezza di $\mathbb{R}$]. *Ogni insieme non vuoto $\mathcal{E} \subseteq \mathbb{R}$ che sia limitato superiormente possiede un estremo superiore in $\mathbb{R}$.*

Dimostrazione. Ricordiamo che gli elementi di $\mathcal{E}$ sono essi stessi sezioni di Dedekind, cioè sottoinsiemi di $\mathbb{Q}$. Definiamo

$$S := \bigcup_{A \in \mathcal{E}} A.$$

Mostriamo anzitutto che S è una sezione di Dedekind.

Poiché $\mathcal{E}$ è non vuoto, contiene almeno una sezione A, e poiché $A \neq \varnothing$, anche $S \neq \varnothing$. Inoltre $\mathcal{E}$ è limitato superiormente: esiste quindi una sezione $B \in \mathbb{R}$ tale che

$$A \subseteq B \qquad \text{per ogni } A \in \mathcal{E}.$$

Ne segue

$$S \subseteq B.$$

Poiché $B \neq \mathbb{Q}$, anche $S \neq \mathbb{Q}$.

Se $q \in S$, esiste $A \in \mathcal{E}$ tale che $q \in A$. Se $p < q$, la chiusura verso il basso di A implica $p \in A$, dunque $p \in S$.

Infine, se $q \in S$, scegliamo ancora $A \in \mathcal{E}$ con $q \in A$. Poiché A non possiede elemento massimo, esiste $r \in A$ con $q < r$. Allora $r \in S$. Quindi anche S non possiede elemento massimo.

Abbiamo provato che $S \in \mathbb{R}$. Per ogni $A \in \mathcal{E}$ vale evidentemente $A \subseteq S$, dunque S è un maggiorante di $\mathcal{E}$. Se C è un qualunque altro maggiorante, allora

$$A \subseteq C \qquad \text{per ogni } A \in \mathcal{E},$$

e quindi

$$S = \bigcup_{A \in \mathcal{E}} A \subseteq C.$$

Pertanto S è il più piccolo dei maggioranti:

$$S = \sup \mathcal{E}.$$

□

La dimostrazione mostra in modo particolarmente netto il significato della costruzione: *la completezza dei reali emerge direttamente dal modo in cui i reali sono stati definiti come sezioni di* $\mathbb{Q}$. Il *supremo* di una famiglia non vuota e superiormente limitata di reali è semplicemente l'unione delle sezioni che la compongono.

2.24 Osservazione [perché lo stesso argomento non funziona in $\mathbb{Q}$]. Se consideriamo soltanto i razionali rappresentati dalle sezioni A_r, l'unione di una famiglia di tali sezioni può essere una sezione che non è della forma A_r per alcun $r \in \mathbb{Q}$. È precisamente ciò che accade in corrispondenza di $\sqrt{2}$. Allargando il dominio da queste sole sezioni razionali a *tutte* le sezioni di Dedekind, il punto mancante viene incluso nel nuovo sistema numerico. ■

2.6.1 Il numero reale $\sqrt{2}$

Riprendiamo la sezione

$$S = \{q \in \mathbb{Q} \mid q < 0 \text{ oppure } q^2 < 2\}.$$

Nella costruzione appena data, S è un elemento di $\mathbb{R}$. Lo indichiamo con

$$\sqrt{2}.$$

Una volta definita la moltiplicazione delle sezioni, si verifica che questo numero è positivo e soddisfa

$$(\sqrt{2})^2 = 2.$$

Ciò che in $\mathbb{Q}$ appariva come una lacuna è diventato un elemento del nuovo sistema.

È importante però non interpretare la costruzione come un artificio destinato soltanto ad aggiungere $\sqrt{2}$. La proprietà di completezza riguarda *ogni* sottoinsieme non vuoto e superiormente limitato di $\mathbb{R}$ e sarà la base generale dei risultati sui limiti.

2.6.2 Esercizi risolti sulle sezioni e sulla completezza

2.25 Esercizio.

1. *Verificare una sezione razionale.* Sia $r \in \mathbb{Q}$. Verificare direttamente le tre condizioni della definizione per

$$A_r = \{q \in \mathbb{Q} \mid q < r\}.$$

Soluzione. L'insieme è non vuoto, perché per esempio $r - 1 \in A_r$, ed è diverso da $\mathbb{Q}$, perché $r \notin A_r$. Se $q \in A_r$ e $p < q$, allora

$$p < q < r,$$

dunque $p \in A_r$. Infine, se $q \in A_r$, il razionale

$$s := \frac{q+r}{2}$$

soddisfa

$$q < s < r,$$

perciò $s \in A_r$. Quindi A_r non ha elemento massimo.

2. *Ordine e inclusione.* Siano $r, s \in \mathbb{Q}$. Mostrare che

$$r < s \quad \Longleftrightarrow \quad A_r \subsetneq A_s.$$

Soluzione. Se $r < s$ e $q \in A_r$, allora $q < r < s$, quindi $q \in A_s$ e $A_r \subseteq A_s$. L'inclusione è propria: per densità esiste $t \in \mathbb{Q}$ tale che

$$r < t < s,$$

e allora $t \in A_s$ ma $t \notin A_r$.
Viceversa, se $A_r \subsetneq A_s$, non può essere $s \leq r$, perché in tal caso si avrebbe $A_s \subseteq A_r$. Pertanto $r < s$.

3. *La sezione di $\sqrt{2}$.* Mostrare direttamente che

$$S = \{q \in \mathbb{Q} \mid q < 0 \text{ oppure } q^2 < 2\}$$

è una sezione di Dedekind.
Soluzione. L'insieme è non vuoto, perché $0 \in S$, ed è diverso da $\mathbb{Q}$, perché per esempio $2 \notin S$.
Verifichiamo la chiusura verso il basso. Sia $q \in S$ e sia $p < q$. Se $q < 0$, allora anche $p < 0$, dunque $p \in S$. Se invece $q \geq 0$, allora $q^2 < 2$. Se $p < 0$ abbiamo ancora $p \in S$; se $p \geq 0$, da $0 \leq p < q$ segue

$$p^2 < q^2 < 2,$$

e quindi $p \in S$.
Resta da mostrare che S non possiede elemento massimo. Sia $q \in S$. Se $q < 0$, allora $q/2$ è razionale e

$$q < \frac{q}{2} < 0,$$

quindi $q/2 \in S$.
Supponiamo ora $q \geq 0$. Allora $q^2 < 2$. Poniamo

$$r := q + \frac{2-q^2}{q+2}.$$

Poiché $2 - q^2 > 0$ e $q + 2 > 0$, si ha $r > q$. Inoltre

$$r = \frac{2q+2}{q+2}$$

e un calcolo diretto dà

$$2 - r^2 = \frac{2(2-q^2)}{(q+2)^2} > 0.$$

Dunque $r^2 < 2$, perciò $r \in S$. In ogni caso abbiamo trovato un elemento di S strettamente maggiore di q. Tutte le condizioni della definizione sono soddisfatte.

4. *La somma conserva i razionali.* Per $r, s \in \mathbb{Q}$, mostrare che

$$A_r + A_s = A_{r+s}.$$

Soluzione. Se $x \in A_r + A_s$, esistono $a < r$ e $b < s$ tali che $x = a + b$. Quindi

$$x = a + b < r + s,$$

e dunque $x \in A_{r+s}$.
Viceversa, sia $x < r + s$. Poniamo

$$\varepsilon := \frac{r+s-x}{2} > 0, \qquad a := r - \varepsilon, \qquad b := s - \varepsilon.$$

Poiché tutti questi numeri sono razionali, $a \in A_r$ e $b \in A_s$, e

$$a + b = r + s - 2\varepsilon = x.$$

Dunque $x \in A_r + A_s$. Le due sezioni coincidono.

5. *Un supremo come unione.* Consideriamo la famiglia di reali razionali

$$\mathcal{E} := \{A_q \mid q \in \mathbb{Q},\ q < 1\}.$$

Determinare $\sup \mathcal{E}$ nella costruzione mediante sezioni.
Soluzione. Per il Teorema 2.23,

$$\sup \mathcal{E} = \bigcup_{q<1} A_q.$$

Mostriamo che questa unione è A_1. Se x appartiene all'unione, esiste $q < 1$ tale che $x < q$, dunque $x < 1$ e quindi $x \in A_1$.
Viceversa, se $x < 1$, per densità dei razionali possiamo scegliere q tale che

$$x < q < 1.$$

Allora $x \in A_q$, e quindi x appartiene all'unione. Pertanto

$$\sup \mathcal{E} = A_1,$$

cioè, dopo l'identificazione dei razionali con le corrispondenti sezioni,

$$\sup\{q \in \mathbb{Q} \mid q < 1\} = 1.$$

6. *Una proprietà che anticipa il capitolo successivo.* Siano $A, B, C \in \mathbb{R}$. Usando la definizione della somma fra sezioni, verificare che

$$A + B = B + A$$

e che

$$(A + B) + C = A + (B + C).$$

Soluzione. Per la commutatività della somma razionale,

$$A + B = \{a + b \mid a \in A,\ b \in B\} = \{b + a \mid b \in B,\ a \in A\} = B + A.$$

Analogamente, per l'associatività della somma razionale,

$$(A + B) + C = \{(a + b) + c \mid a \in A, b \in B, c \in C\}$$

coincide con

$$\{a + (b + c) \mid a \in A, b \in B, c \in C\} = A + (B + C).$$

Abbiamo nuovamente incontrato proprietà formali delle operazioni che non dipendono dalla particolare rappresentazione degli elementi; nel Capitolo 3 questo tipo di fenomeno verrà studiato in modo astratto.

2.7 Completezza, successioni, limiti

La costruzione dei numeri reali ci ha fornito un sistema numerico nel quale ogni insieme non vuoto e limitato superiormente possiede un estremo superiore. Nel seguito vedremo come questa proprietà consenta di controllare processi di approssimazione infinita.

Il primo oggetto da studiare sarà una *successione di numeri reali*. La domanda fondamentale sarà: in quali condizioni i suoi valori si avvicinano indefinitamente a un numero reale determinato? La risposta porterà alla nozione di *limite*, che costituirà il linguaggio comune delle successive definizioni di continuità, derivata e integrale.

2.7.1 Valore assoluto come distanza

Per descrivere in modo quantitativo quanto due numeri reali siano vicini useremo il *valore assoluto*.

2.26 Definizione [valore assoluto e distanza]. Per $x \in \mathbb{R}$ definiamo il **valore assoluto** di x

$$|x| := \begin{cases} x & \text{se } x \geq 0, \\ -x & \text{se } x < 0. \end{cases}$$

La quantità

$$|x - y|$$

è detta **distanza euclidea** fra i numeri reali x e y sulla retta reale. ■

Useremo in particolare le proprietà elementari generali, valide per $x, y \in \mathbb{R}$,

$$|x| \geq 0, \qquad |xy| = |x|\,|y|,$$

e la **disuguaglianza triangolare**

$$|x + y| \leq |x| + |y|.$$

Quest'ultima esprime il fatto intuitivo che la distanza percorsa passando da un punto a un altro attraverso un punto intermedio non può essere minore della distanza diretta. Per esempio,

$$|x - L| < \varepsilon$$

significa precisamente

$$L - \varepsilon < x < L + \varepsilon.$$

Il numero positivo ε rappresenta quindi una tolleranza arbitrariamente piccola attorno a L.

2.7.2 Successioni di numeri reali

Richiamando la nozione generale di successione introdotta nella Definizione 1.51, specializziamo ora il codominio ai numeri reali.

2.27 Definizione [successione reale]. Una **successione reale**, o successione di numeri reali, è una funzione

$$x : \mathbb{N} \to \mathbb{R}.$$

Scriveremo

$$x_n := x(n)$$

e indicheremo la successione con $(x_n)_{n\in\mathbb{N}}$ o, più brevemente, con (x_n). L'indice n non va confuso con il valore x_n: il primo è un numero naturale che indica una posizione nella successione, il secondo è un numero reale. ■

2.28 Definizione [serie reale e somme parziali]. Data una successione reale $(a_n)_{n\in\mathbb{N}}$, la **serie reale** associata a (a_n) è la particolare successione reale $(s_N)_{N\in\mathbb{N}}$ definita da

$$s_N := \sum_{n=0}^{N} a_n = a_0 + a_1 + \cdots + a_N.$$

I termini s_N della serie si chiamano **somme parziali**. La serie si denota convenzionalmente con

$$\sum_{n=0}^{\infty} a_n.$$

L'indice iniziale può essere diverso da 0; per esempio, se si parte da $n = 1$, le somme parziali sono

$$s_N = \sum_{n=1}^{N} a_n.$$

■

2.7.3 La definizione ε–N di convergenza

2.29 Definizione [limite di una successione]. Sia (x_n) una successione di numeri reali e sia $L \in \mathbb{R}$. Diciamo che (x_n) **converge** a L, o che L è il limite della successione, se

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n \in \mathbb{N}, \qquad n \geq N \implies |x_n - L| < \varepsilon. \tag{2.2}$$

In tal caso scriviamo

$$\lim_{n\to\infty} x_n = L \qquad \text{oppure} \qquad x_n \longrightarrow L.$$

■

La condizione (2.2) deve essere letta con attenzione nell'ordine in cui compaiono i quantificatori. Fissata una tolleranza arbitraria $\varepsilon > 0$, dobbiamo essere in grado di trovare un indice N — che può dipendere da ε — tale che *tutti* i termini successivi, cioè quelli con $n \geq N$, distino da L meno di ε.

In termini geometrici, qualunque sia l'intervallo

$$(L - \varepsilon, L + \varepsilon)$$

centrato in L, per quanto piccolo, tutti i termini della successione tranne eventualmente un numero finito devono trovarsi al suo interno.

2.30 Osservazione [l'ordine dei quantificatori è essenziale]. La definizione non dice che esiste un unico N che funzioni contemporaneamente per ogni $\varepsilon > 0$. Dice invece

$$\forall \varepsilon > 0 \; \exists N = N(\varepsilon) \; \cdots .$$

Quando chiediamo una precisione maggiore, cioè scegliamo un ε più piccolo, può essere necessario procedere più avanti nella successione. ■

Una successione che possiede un limite reale, cioè un limite finito, si dice **convergente**. Introduciamo ora anche due particolari forme di divergenza.

2.31 Definizione [limiti infiniti di una successione]. Sia (x_n) una successione di numeri reali. Diciamo che (x_n) **diverge a** $+\infty$, e scriviamo

$$\lim_{n\to\infty} x_n = +\infty,$$

se

$$\forall M \in \mathbb{R} \; \exists N \in \mathbb{N} \; \forall n \in \mathbb{N}, \qquad n \geq N \implies x_n > M.$$

Analogamente, diciamo che (x_n) **diverge a** $-\infty$, e scriviamo

$$\lim_{n\to\infty} x_n = -\infty,$$

se

$$\forall M \in \mathbb{R} \; \exists N \in \mathbb{N} \; \forall n \in \mathbb{N}, \qquad n \geq N \implies x_n < M.$$

I simboli $+\infty$ e $-\infty$ non indicano numeri reali: queste scritture descrivono il comportamento della successione rispetto a ogni soglia reale. ■

2.32 Osservazione.

(a) Da questo punto in poi, quando scriveremo

$$\lim_{n\to\infty} x_n = L \qquad \text{con } L \in \mathbb{R},$$

intenderemo che L è un *limite finito*. Una successione che diverge a $+\infty$ o a $-\infty$ non converge dunque a un numero reale.

(b) Una successione priva di limite finito non sempre diverge a uno dei due infiniti. Per esempio,

$$x_n := (-1)^n n$$

non diverge né a $+\infty$ né a $-\infty$. Infatti, prendendo la soglia $M = 0$, per ogni N esistono indici pari $n \geq N$ per i quali $x_n > 0$ e indici dispari $n \geq N$ per i quali $x_n < 0$. I termini pari impediscono quindi la divergenza a $-\infty$, mentre quelli dispari impediscono la divergenza a $+\infty$.

■

2.33 Definizione. Diremo che una proprietà dei termini di una successione (x_n) vale **definitivamente** se esiste N tale che essa vale per ogni $n \geq N$. Per esempio, la convergenza a un limite finito L afferma che, per ogni $\varepsilon > 0$, la disuguaglianza $|x_n - L| < \varepsilon$ vale definitivamente. ■

2.34 Esempio [successione costante]. Se $x_n = c$ per ogni n, allora

$$\lim_{n\to\infty} x_n = c.$$

Infatti, per ogni $\varepsilon > 0$ e per ogni n,

$$|x_n - c| = 0 < \varepsilon.$$

Possiamo quindi scegliere, per esempio, $N = 0$. ■

2.7.4 La proprietà archimedea e il limite di $1/n$

Per verificare direttamente dalla definizione alcuni limiti abbiamo bisogno di una proprietà fondamentale dei numeri reali.

2.35 Teorema [proprietà archimedea]. $\mathbb{R}$ *soddisfa la* ***proprietà archimedea****: per ogni* $x \in \mathbb{R}$ *esiste* $n \in \mathbb{N}$ *tale che*

$$n > x.$$

Equivalentemente l'insieme $\mathbb{N}$*, considerato come sottoinsieme di* $\mathbb{R}$*, non è limitato superiormente.*

Dimostrazione. Supponiamo per assurdo che $\mathbb{N}$ sia limitato superiormente. Per la completezza di $\mathbb{R}$ esisterebbe allora

$$L := \sup \mathbb{N}.$$

Il numero $L - 1$ non può essere un maggiorante di $\mathbb{N}$, perché L è il più piccolo dei maggioranti. Esiste quindi $n \in \mathbb{N}$ tale che

$$n > L - 1.$$

Ne segue

$$n + 1 > L.$$

Ma $n + 1 \in \mathbb{N}$, in contraddizione con il fatto che L sia un maggiorante di $\mathbb{N}$. Dunque $\mathbb{N}$ non è limitato superiormente. □

2.36 Osservazione. La proprietà archimedea mostra già un primo uso concreto della completezza: da una proprietà globale dell'ordine di $\mathbb{R}$ ricaviamo il fatto elementare che i naturali superano qualunque numero reale fissato. ■

2.37 Proposizione [un limite fondamentale]. *Vale*

$$\lim_{n\to\infty} \frac{1}{n+1} = 0.$$

Dimostrazione. Sia $\varepsilon > 0$. Per la proprietà archimedea esiste $N \in \mathbb{N}$ tale che

$$N + 1 > \frac{1}{\varepsilon}.$$

Se $n \geq N$, allora

$$n + 1 \geq N + 1 > \frac{1}{\varepsilon},$$

e quindi

$$\left| \frac{1}{n+1} - 0 \right| = \frac{1}{n+1} < \varepsilon.$$

La definizione di convergenza è soddisfatta. □

2.38 Osservazione [perché usiamo $1/(n+1)$]. Poiché nelle dispense $\mathbb{N}$ contiene anche 0, la successione $1/n$ non è definita all'indice 0. Scrivere $1/(n+1)$ evita questa irrilevante eccezione e conserva lo stesso comportamento per $n \to \infty$. ■

2.39 Definizione [serie convergente]. Una serie $\sum_{n=0}^{\infty} a_n$, cioè la successione delle sue somme parziali (s_N), si dice **convergente** se (s_N) converge a un limite finito S. In tal caso si pone

$$\sum_{n=0}^{\infty} a_n := S.$$

La stessa definizione vale, con l'ovvio cambiamento degli indici, quando la serie parte da un indice diverso da 0. ■

2.40 Esempio [una serie geometrica convergente]. Consideriamo la serie geometrica con termini 2^{-n}, a partire da $n = 1$. Le sue somme parziali sono

$$s_N := \sum_{n=1}^{N} \frac{1}{2^n} = \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^N}.$$

La formula della somma geometrica finita dà

$$s_N = 1 - \frac{1}{2^N}.$$

Per induzione si verifica che $2^N \geq N + 1$ per ogni $N \in \mathbb{N}$. Dunque

$$0 < \frac{1}{2^N} \leq \frac{1}{N+1}.$$

Il limite appena dimostrato per $1/(N+1)$ implica direttamente che $2^{-N} \to 0$: dato $\varepsilon > 0$, la disuguaglianza $1/(N+1) < \varepsilon$ vale definitivamente, nel senso della Definizione 2.33, e quindi anche $2^{-N} < \varepsilon$. Ne segue

$$\lim_{N\to\infty} s_N = 1.$$

Pertanto, per la Definizione 2.39,

$$\boxed{\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1.}$$

■

2.41 Osservazione [serie convergenti e paradossi di Zenone]. La serie geometrica precedente richiama immediatamente alcuni *paradossi di Zenone*, in particolare la Dicotomia e Achille e la tartaruga. Se un percorso viene suddiviso in tratti di lunghezza

$$\frac{1}{2},\ \frac{1}{4},\ \frac{1}{8}, \ldots,$$

la matematica moderna mostra che la somma di un numero infinito di tali tratti può essere finita. In questo senso la teoria delle serie convergenti chiarisce una difficoltà matematica centrale presente nei paradossi.

Non è però opportuno identificare questo risultato con una soluzione filosofica definitiva dei paradossi. Come sottolinea Fano [11], si possono distinguere almeno il problema matematico delle somme infinite, il problema ontologico dei *supertasks*, cioè del completamento di un'infinità di sotto-processi, e questioni più generali sulla struttura dello spazio, del tempo e del moto. La letteratura contemporanea continua a discutere questi ultimi aspetti [12, 13]. La convergenza della serie risolve dunque il problema matematico della somma, ma non rende automaticamente superflue le questioni filosofiche sollevate da Zenone. ■

2.7.5 Unicità del limite

Veniamo all'importante risultato riguardante l'unicità del limite. Questo risultato ha in realtà una natura *topologica*: più avanti introdurremo la proprietà di Hausdorff, dalla quale l'unicità del limite segue in un contesto più generale.

2.42 Teorema [unicità del limite]. *Una successione reale convergente possiede un solo limite.*

Dimostrazione. Supponiamo che

$$x_n \to L \qquad \text{e} \qquad x_n \to M.$$

Vogliamo mostrare che $L = M$. Supponiamo per assurdo $L \neq M$ e poniamo

$$\varepsilon := \frac{|L - M|}{3} > 0.$$

Dalla convergenza a L esiste N_1 tale che, per $n \geq N_1$,

$$|x_n - L| < \varepsilon.$$

Dalla convergenza a M esiste N_2 tale che, per $n \geq N_2$,

$$|x_n - M| < \varepsilon.$$

Scegliamo $n \geq \max\{N_1, N_2\}$. Per la disuguaglianza triangolare,

$$|L - M| \leq |L - x_n| + |x_n - M| < 2\varepsilon = \frac{2}{3}|L - M|,$$

che è impossibile. Pertanto $L = M$. □

Il teorema giustifica l'espressione *il* limite di una successione.

2.7.6 Ogni successione convergente è limitata

2.43 Definizione [successione limitata]. Una successione reale (x_n) si dice **limitata** se esiste $M > 0$ tale che

$$|x_n| \leq M \qquad \text{per ogni } n \in \mathbb{N}.$$

■

2.44 Teorema. *Ogni successione convergente di numeri reali è limitata.*

Dimostrazione. Supponiamo $x_n \to L$. Applichiamo la definizione con $\varepsilon = 1$. Esiste N tale che, per ogni $n \geq N$,

$$|x_n - L| < 1.$$

Dalla disuguaglianza triangolare segue

$$|x_n| \leq |x_n - L| + |L| < 1 + |L| \qquad (n \geq N).$$

Resta da considerare soltanto un numero finito di termini iniziali, quelli con $n < N$. Possiamo quindi scegliere M maggiore o uguale a $1 + |L|$ e ai valori assoluti di questi termini. Allora $|x_n| \leq M$ per ogni n.□

Il viceversa è falso: una successione può essere limitata senza essere convergente. Un esempio fondamentale è $x_n = (-1)^n$.

2.7.7 Algebra dei limiti

Le operazioni aritmetiche sui reali sono compatibili con il passaggio al limite.

2.45 Teorema [operazioni con i limiti]. *Siano (x_n) e (y_n) successioni reali e siano $L, M \in \mathbb{R}$ tali che*

$$\lim_{n\to\infty} x_n = L, \qquad \lim_{n\to\infty} y_n = M.$$

Allora:

(i)
$$\lim_{n\to\infty} (x_n + y_n) = L + M = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n;$$

(ii) per ogni $\lambda \in \mathbb{R}$,
$$\lim_{n\to\infty} (\lambda x_n) = \lambda L = \lambda \lim_{n\to\infty} x_n;$$

(iii)
$$\lim_{n\to\infty} (x_n y_n) = LM = \left(\lim_{n\to\infty} x_n\right)\left(\lim_{n\to\infty} y_n\right);$$

(iv) se $M \neq 0$ ed esiste $N_0 \in \mathbb{N}$ tale che

$$y_n \neq 0 \qquad \textit{per ogni } n \geq N_0,$$

sia (q_n) una successione reale tale che

$$q_n := \frac{x_n}{y_n} \qquad (n \geq N_0);$$

Il numero finito di termini iniziali q_n con $n < N_0$ può essere scelto arbitrariamente. Allora

$$\lim_{n\to\infty} q_n = \frac{L}{M} = \frac{\lim\limits_{n\to\infty} x_n}{\lim\limits_{n\to\infty} y_n}.$$

Con questa convenzione si scrive abitualmente anche

$$\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{L}{M}.$$

Dimostriamo i primi due punti, che mostrano già il meccanismo fondamentale.

Dimostrazione. Per la somma, sia $\varepsilon > 0$. Dalla convergenza di x_n a L esiste N_1 tale che

$$|x_n - L| < \frac{\varepsilon}{2} \qquad (n \geq N_1),$$

e dalla convergenza di y_n a M esiste N_2 tale che

$$|y_n - M| < \frac{\varepsilon}{2} \qquad (n \geq N_2).$$

Per $n \geq \max\{N_1, N_2\}$,

$$|(x_n + y_n) - (L + M)| \leq |x_n - L| + |y_n - M| < \varepsilon.$$

Dunque

$$\lim_{n\to\infty} (x_n + y_n) = L + M.$$

Per il prodotto per uno scalare, se $\lambda = 0$ il risultato è immediato. Se $\lambda \neq 0$, dato $\varepsilon > 0$ scegliamo N tale che

$$|x_n - L| < \frac{\varepsilon}{|\lambda|} \qquad (n \geq N).$$

Allora

$$|\lambda x_n - \lambda L| = |\lambda|\,|x_n - L| < \varepsilon.$$

Gli altri punti seguono con stime analoghe, usando anche il fatto che una successione convergente è limitata. □

2.46 Osservazione. La seconda proprietà e la prima, combinate, mostrano che per $\alpha, \beta, L, M \in \mathbb{R}$, se

$$\lim_{n\to\infty} x_n = L, \qquad \lim_{n\to\infty} y_n = M,$$

allora

$$\lim_{n\to\infty} (\alpha x_n + \beta y_n) = \alpha L + \beta M.$$

Questa compatibilità fra somme, moltiplicazioni per numeri e passaggio al limite ricomparirà in forma più astratta nel Capitolo 3. ■

2.7.8 Successioni monotone e completezza

La relazione fra completezza e convergenza emerge in modo particolarmente chiaro per le successioni monotone.

2.47 Definizione [successione monotona]. Una successione reale (x_n) si dice **crescente** se

$$x_n \leq x_{n+1} \qquad \text{per ogni } n,$$

e **decrescente** se

$$x_{n+1} \leq x_n \qquad \text{per ogni } n.$$

In entrambi i casi si dice **monotona**. ■

2.48 Teorema [convergenza di successioni monotone limitate]. *Ogni successione reale crescente e limitata superiormente converge. Più precisamente, se (x_n) è crescente e*

$$E := \{x_n \mid n \in \mathbb{N}\},$$

allora

$$\lim_{n\to\infty} x_n = \sup E.$$

Analogamente, ogni successione reale decrescente e limitata inferiormente converge al proprio estremo inferiore, cioè al maggiore dei suoi minoranti.

Dimostrazione. Consideriamo il caso crescente. Poiché E è non vuoto e limitato superiormente, per la completezza di $\mathbb{R}$ esiste

$$L := \sup E.$$

Sia $\varepsilon > 0$. Il numero $L - \varepsilon$ non è un maggiorante di E, altrimenti sarebbe un maggiorante strettamente minore di L. Esiste quindi N tale che

$$x_N > L - \varepsilon.$$

Poiché la successione è crescente, per ogni $n \geq N$ vale

$$x_n \geq x_N > L - \varepsilon.$$

D'altra parte L è un maggiorante di E, dunque $x_n \leq L$. Pertanto, per $n \geq N$,

$$L - \varepsilon < x_n \leq L,$$

cioè

$$|x_n - L| < \varepsilon.$$

Quindi $x_n \to L$. □

2.49 Osservazione [dove entra la completezza]. Questo teorema permette di vedere esattamente dove la completezza di $\mathbb{R}$ è necessaria: la monotonia e la limitatezza producono un insieme di valori con un estremo superiore; la completezza garantisce che tale estremo appartenga a $\mathbb{R}$, e la proprietà di essere il più piccolo maggiorante permette poi di dimostrare la convergenza. ■

2.7.9 Esercizi risolti sulle successioni

2.50 Esercizio.

1. *Verifica diretta di un limite.* Mostrare mediante la definizione ε–N che

$$x_n := \frac{3}{n+1}$$

converge a 0.
Soluzione. Sia $\varepsilon > 0$. Per la proprietà archimedea possiamo scegliere N tale che

$$N + 1 > \frac{3}{\varepsilon}.$$

Se $n \geq N$, allora

$$|x_n - 0| = \frac{3}{n+1} \leq \frac{3}{N+1} < \varepsilon.$$

Quindi

$$\lim_{n\to\infty} \frac{3}{n+1} = 0.$$

2. *Il limite dipende dalla coda.* Siano (x_n) e (y_n) due successioni tali che esista N_0 con

$$x_n = y_n \qquad \text{per ogni } n \geq N_0.$$

Mostrare che (x_n) converge a L se e solo se (y_n) converge a L.
Soluzione. Supponiamo $x_n \to L$ e sia $\varepsilon > 0$. Esiste N_1 tale che

$$|x_n - L| < \varepsilon \qquad (n \geq N_1).$$

Per $n \geq \max\{N_0, N_1\}$ vale $y_n = x_n$, quindi

$$|y_n - L| < \varepsilon.$$

Dunque $y_n \to L$. L'implicazione inversa è identica scambiando le due successioni. Il limite non dipende quindi da un numero finito di termini iniziali, ma soltanto dal comportamento della successione per indici sufficientemente grandi.

3. *Una successione limitata ma divergente.* Mostrare che

$$x_n := (-1)^n$$

è limitata ma non converge.
Soluzione. La successione è limitata perché

$$|x_n| = 1 \qquad \text{per ogni } n.$$

Supponiamo per assurdo che $x_n \to L$. Scegliamo $\varepsilon = 1/2$. Per la convergenza esisterebbe N tale che, per ogni $n \geq N$,

$$|x_n - L| < \frac{1}{2}.$$

Possiamo scegliere un indice pari $2k \geq N$ e un indice dispari $2k+1 \geq N$. Avremmo allora

$$|1 - L| < \frac{1}{2}, \qquad |-1 - L| < \frac{1}{2}.$$

Per la disuguaglianza triangolare,

$$2 = |1 - (-1)| \leq |1 - L| + |L + 1| < 1,$$

assurdo. La successione non converge.

4. *Combinazioni lineari di successioni.* Supponiamo

$$x_n \to 2, \qquad y_n \to -1.$$

Calcolare

$$\lim_{n\to\infty} (3x_n - 2y_n).$$

Soluzione. Per il Teorema 2.45,

$$3x_n - 2y_n \longrightarrow 3 \cdot 2 - 2(-1) = 8.$$

Quindi

$$\lim_{n\to\infty} (3x_n - 2y_n) = 8.$$

L'esercizio mostra nuovamente che alcune operazioni formali sono compatibili con il passaggio al limite; questa proprietà acquisterà una formulazione più generale nel Capitolo 3.

5. *Una successione monotona.* Consideriamo

$$x_n := 1 - \frac{1}{n+1}.$$

Mostrare che (x_n) è crescente, limitata superiormente e determinare il suo limite.
Soluzione. Poiché

$$\frac{1}{n+2} < \frac{1}{n+1},$$

si ha

$$x_{n+1} = 1 - \frac{1}{n+2} > 1 - \frac{1}{n+1} = x_n,$$

quindi la successione è crescente. Inoltre $x_n < 1$ per ogni n, dunque è limitata superiormente da 1. Per il teorema della convergenza di successioni monotone limitate essa converge. D'altra parte

$$1 - x_n = \frac{1}{n+1} \longrightarrow 0,$$

quindi

$$\lim_{n\to\infty} x_n = 1.$$

6. *Negare correttamente la definizione di limite.* Scrivere la negazione della proposizione

$$\lim_{n\to\infty} x_n = L.$$

Soluzione. La convergenza significa

$$\forall \varepsilon > 0 \; \exists N \; \forall n \geq N, \quad |x_n - L| < \varepsilon.$$

Negando successivamente i quantificatori otteniamo:

$$\exists \varepsilon_0 > 0 \; \forall N \; \exists n \geq N \quad \text{tale che} \quad |x_n - L| \geq \varepsilon_0.$$

Dunque x_n non converge a L se esiste una distanza positiva ε_0 dalla quale la successione continua ad allontanarsi almeno una volta, per quanto avanti si scelga di iniziare a guardarla. Questo esercizio mostra quanto sia importante l'ordine dei quantificatori nella Definizione 2.29.

2.8 Limiti di funzioni reali

Passiamo ora ad estendere la nozione di limite al caso di funzioni. Abbiamo bisogno di qualche preliminare nozione topologica del tutto elementare che useremo anche in seguito.

Da questo punto in avanti, salvo diversa indicazione, le funzioni di una variabile reale che considereremo avranno la forma $f : D \to \mathbb{R}$, con $D \subseteq \mathbb{R}$. Il codominio si intenderà dunque sempre uguale a $\mathbb{R}$; il dominio D, invece, verrà specificato di volta in volta, poiché è parte essenziale della definizione di una funzione e può influire sulle proprietà che studieremo.

2.8.1 Topologia, spazi topologici e topologia euclidea in $\mathbb{R}^n$

Prima di definire il limite di una funzione conviene introdurre, in forma generale, il linguaggio topologico che useremo anche più avanti. L'idea fondamentale è specificare quali sottoinsiemi di un insieme debbano essere considerati *aperti*.

2.51 Definizione [topologia e spazio topologico]. Sia X un insieme. Una **topologia** su X è una famiglia $\mathcal{T} \subseteq \mathcal{P}(X)$ tale che:

(i) $\varnothing, X \in \mathcal{T}$;

(ii) l'unione di una famiglia arbitraria di elementi di $\mathcal{T}$ appartiene ancora a $\mathcal{T}$;

(iii) l'intersezione di un numero finito di elementi di $\mathcal{T}$ appartiene ancora a $\mathcal{T}$.

La coppia $(X, \mathcal{T})$ si chiama **spazio topologico**. Gli elementi di $\mathcal{T}$ si chiamano **insiemi aperti**; i loro complementi in X si chiamano **insiemi chiusi**.

Un **intorno** di un punto $x \in X$ è, in queste dispense, un insieme aperto $U \in \mathcal{T}$ tale che $x \in U$. Se U è un intorno di x, l'insieme $U \setminus \{x\}$ si chiama **intorno bucato** di x. Se $D \subseteq X$, un punto $x \in D$ si dice **punto interno** di D se esiste un intorno U di x tale che $U \subseteq D$. In particolare, un sottoinsieme D è aperto se e solo se tutti i suoi punti sono interni. ■

Passiamo ora agli esempi che useremo nell'analisi. Per descrivere anzitutto il caso della retta reale, fissiamo la notazione per gli intervalli. Siano $a, b \in \mathbb{R}$ con $a < b$. L'**intervallo aperto** di estremi a e b è

$$(a, b) := \{x \in \mathbb{R} \mid a < x < b\};$$

l'**intervallo chiuso**, definito anche per $a = b$, è

$$[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\};$$

(notare che $[a, a] = \{a\}$) mentre gli **intervalli semiaperti** sono

$$[a, b) := \{x \in \mathbb{R} \mid a \leq x < b\}, \qquad (a, b] := \{x \in \mathbb{R} \mid a < x \leq b\}.$$

Si usano inoltre intervalli illimitati, per esempio

$$(-\infty, b), \qquad (-\infty, b], \qquad (a, +\infty), \qquad [a, +\infty),$$

e $\mathbb{R} = (-\infty, +\infty)$. I simboli $-\infty$ e $+\infty$ non rappresentano numeri reali e quindi non sono mai elementi dell'intervallo: presso un estremo infinito si usa sempre una parentesi tonda.

2.52 Definizione [topologia euclidea di $\mathbb{R}^n$]. Per $n \geq 1$, la **topologia euclidea** di $\mathbb{R}^n$ è la topologia i cui aperti non vuoti sono le unioni arbitrarie di prodotti cartesiani di intervalli aperti

$$(a_1, b_1) \times \cdots \times (a_n, b_n), \qquad a_j < b_j.$$

Nel caso $n = 1$ si ottiene la topologia euclidea di $\mathbb{R}$: i suoi aperti non vuoti sono precisamente le unioni arbitrarie di intervalli aperti (a, b). ■

In particolare, per $D \subseteq \mathbb{R}$ e $a \in D$, il punto a è interno a D se e solo se esiste $r > 0$ tale che

$$(a - r, a + r) \subseteq D.$$

Come sottoinsiemi di $\mathbb{R}$ con la topologia euclidea, gli intervalli $(-\infty, b)$ e $(a, +\infty)$ sono aperti, mentre $(-\infty, b]$ e $[a, +\infty)$ sono chiusi; $\mathbb{R}$ è sia aperto sia chiuso.

Più avanti vedremo che gli aperti della topologia euclidea di $\mathbb{R}^n$ possono essere descritti equivalentemente come unioni arbitrarie di palle aperte. Questa topologia soddisfa inoltre la **proprietà di Hausdorff**: se $p, q \in \mathbb{R}^n$ e $p \neq q$, esistono due intorni $U \ni p$ e $V \ni q$ tali che $U \cap V = \varnothing$.

2.8.2 Dalle successioni ai limiti di funzione

La nozione di limite è stata introdotta per successioni, cioè per funzioni definite sui numeri naturali. Il passo successivo consiste nel considerare funzioni reali

$$f : D \to \mathbb{R}, \qquad D \subseteq \mathbb{R},$$

e chiedere che cosa significhi che i valori $f(x)$ si avvicinino a un numero L quando la variabile x si avvicina a un punto a.

La struttura logica della definizione sarà molto simile a quella appena studiata. Per le successioni controlliamo la vicinanza di x_n a L facendo crescere l'indice n oltre una soglia N; per le funzioni controlleremo la vicinanza di $f(x)$ a L imponendo che x sia sufficientemente vicino ad a. La coppia di parametri ε–N verrà così sostituita dalla coppia ε–δ.

Consideriamo una funzione

$$f : D \to \mathbb{R}, \qquad D \subseteq \mathbb{R}.$$

Vogliamo formalizzare l'affermazione secondo cui i valori $f(x)$ si avvicinano a un numero L quando x si avvicina a un punto a. Come per le successioni, l'idea intuitiva deve essere tradotta in una condizione quantitativa.

2.53 Definizione [punto di accumulazione e punto isolato]. Sia $D \subseteq \mathbb{R}$. Un punto $a \in \mathbb{R}$ si dice **punto di accumulazione** di D se per ogni $\delta > 0$ esiste $x \in D$, con $x \neq a$, tale che

$$|x - a| < \delta.$$

Se invece $a \in D$ e a non è un punto di accumulazione di D, diciamo che a è un **punto isolato** di D. Equivalentemente, $a \in D$ è un punto isolato di D se esiste $\delta > 0$ tale che

$$D \cap (a - \delta, a + \delta) = \{a\}.$$

■

La condizione di punto di accumulazione garantisce che possiamo avvicinarci ad a con punti del dominio *distinti* da a. Non è necessario che a appartenga a D per parlare del limite di $f(x)$ per $x \to a$. Se invece $a \in D$, i due casi sono mutuamente esclusivi: a è un punto di accumulazione di D oppure è un punto isolato di D.

2.54 Osservazione [punti interni e punti di accumulazione]. Ogni punto interno di un insieme $D \subseteq \mathbb{R}$ è un punto di accumulazione di D. Il viceversa non vale in generale, come mostra l'Esempio 2.55.■

2.55 Esempio. Consideriamo la topologia euclidea di $\mathbb{R}$ e $D = \mathbb{Q}$. Nessun punto di D è *interno* a D, tuttavia ogni punto di D è *punto di accumulazione* di D. Se invece $D := \mathbb{Q} \cup [0, 1]$, allora ogni punto di $(0, 1)$ è interno a D. ■

2.8.3 La definizione ε–δ

2.56 Definizione [limite di una funzione]. Sia $f : D \to \mathbb{R}$, con $D \subseteq \mathbb{R}$, sia a un punto di accumulazione di D e sia $L \in \mathbb{R}$. Diciamo che

$$\lim_{x \to a} f(x) = L$$

se

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in D, \qquad 0 < |x - a| < \delta \implies |f(x) - L| < \varepsilon. \tag{2.3}$$

■

La struttura logica della condizione (2.3) è parallela alla definizione di limite di una successione:

$$\forall \varepsilon > 0\ \exists \delta = \delta(\varepsilon) > 0\ \cdots .$$

La tolleranza ε riguarda i valori della funzione; il numero δ controlla quanto dobbiamo avvicinare x ad a per ottenere quella precisione.

La condizione

$$0 < |x - a|$$

è essenziale: nel calcolo del limite non imponiamo alcuna condizione sul valore $f(a)$. Il limite descrive il comportamento della funzione *vicino* ad a, non necessariamente nel punto a.

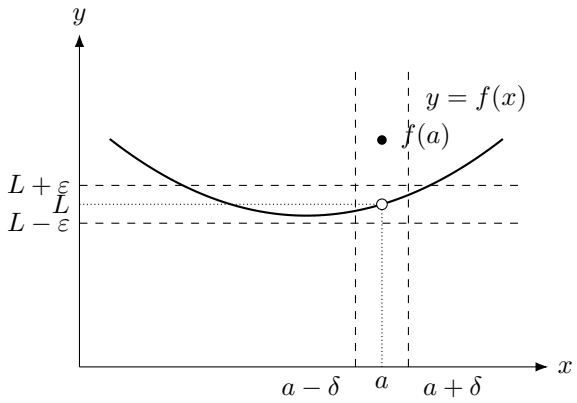


Figure 1: Interpretazione geometrica della condizione ε–δ: scegliendo x sufficientemente vicino ad a, i valori $f(x)$ vengono confinati in una fascia arbitrariamente stretta attorno a L. Il valore $f(a)$ può essere diverso da L e non interviene nella Definizione 2.56 del limite.

2.57 Osservazione [il valore nel punto non determina il limite]. Possiamo modificare $f(a)$ senza cambiare $\lim_{x\to a} f(x)$, quando questo limite esiste. La Definizione 2.56 esclude infatti il caso $x = a$. Questa distinzione diventerà decisiva nella definizione di continuità. ■

2.58 Esempio [una funzione affine]. Consideriamo su $\mathbb{R}$ la **funzione affine**

$$f(x) = 2x + 1,$$

cioè una funzione della forma $mx + q$, qui con $m = 2$ e $q = 1$. Nella terminologia scolastica si incontra talvolta anche l'espressione *funzione lineare non omogenea*. Nel seguito riserveremo invece il termine *lineare*, nel senso dell'algebra lineare, alle applicazioni che rispettano le combinazioni lineari; per una funzione affine $x \mapsto mx + q$ ciò accade precisamente quando $q = 0$.

Mostriamo che

$$\lim_{x\to a} f(x) = 2a + 1.$$

Infatti

$$|f(x) - (2a + 1)| = |2x + 1 - 2a - 1| = 2|x - a|.$$

Dato $\varepsilon > 0$, basta scegliere

$$\delta := \frac{\varepsilon}{2}.$$

Se $0 < |x - a| < \delta$, allora

$$|f(x) - (2a + 1)| < 2\delta = \varepsilon.$$

■

2.8.4 Limiti laterali

Quando il dominio di una funzione contiene punti soltanto da un lato di un numero a, o quando il comportamento della funzione cambia attraversando a, sono utili i limiti laterali. Per evitare condizioni prive di contenuto, occorre prima precisare che il dominio possiede effettivamente punti arbitrariamente vicini ad a dal lato considerato.

2.59 Definizione [punti di accumulazione e limiti laterali]. Siano $D \subseteq \mathbb{R}$ e $a \in \mathbb{R}$. Diciamo che a è un **punto di accumulazione sinistro** di D se per ogni $\delta > 0$ esiste $x \in D$ tale che

$$a - \delta < x < a.$$

Diciamo analogamente che a è un **punto di accumulazione destro** di D se per ogni $\delta > 0$ esiste $x \in D$ tale che

$$a < x < a + \delta.$$

Sia ora $f : D \to \mathbb{R}$ e sia $L \in \mathbb{R}$. Se a è un punto di accumulazione sinistro di D, scriviamo

$$\lim_{x \to a^-} f(x) = L$$

quando, nella Definizione 2.56, imponiamo anche $x < a$. Se a è un punto di accumulazione destro di D, scriviamo

$$\lim_{x \to a^+} f(x) = L$$

quando imponiamo anche $x > a$. ■

Se a è punto di accumulazione di D sia da sinistra sia da destra, allora

$$\lim_{x \to a} f(x) = L \quad \Longleftrightarrow \quad \lim_{x \to a^-} f(x) = L \text{ e } \lim_{x \to a^+} f(x) = L.$$

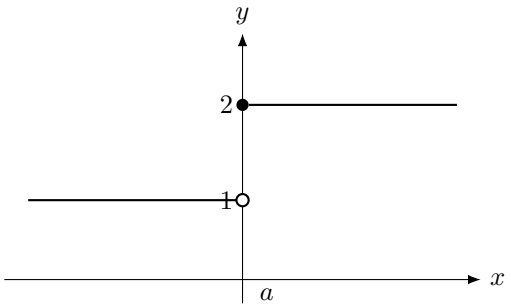


Figure 2: Una discontinuità a salto: il limite sinistro vale 1, il limite destro vale 2. Poiché i due valori sono diversi, il limite per $x \to a$ non esiste.

Osservazione. *Limiti all'infinito e limiti infiniti di funzione.* Nella Definizione 2.56 sia il punto verso cui tende la variabile sia il limite L sono reali. Si usano però comunemente anche simboli $\pm\infty$.

Sia $f : D \to \mathbb{R}$, con $D \subseteq \mathbb{R}$. Se D non è limitato superiormente e $L \in \mathbb{R}$, la scrittura

$$\lim_{x \to +\infty} f(x) = L$$

significa che per ogni $\varepsilon > 0$ esiste $A \in \mathbb{R}$ tale che, per ogni $x \in D$ con $x > A$, vale $|f(x) - L| < \varepsilon$. Analogamente si definisce $\lim_{x \to -\infty} f(x) = L$.

Se a è un punto di accumulazione di D, la scrittura

$$\lim_{x \to a} f(x) = +\infty$$

significa che per ogni $M \in \mathbb{R}$ esiste $\delta > 0$ tale che $x \in D$ e $0 < |x - a| < \delta$ implicano $f(x) > M$; per il limite $-\infty$ si inverte la disuguaglianza. Le due estensioni si possono combinare: per esempio, $\lim_{x \to +\infty} f(x) = +\infty$ significa che $f(x)$ supera ogni soglia reale quando x è sufficientemente grande. Le altre combinazioni si definiscono analogamente. I simboli $+\infty$ e $-\infty$ non sono numeri reali e queste nozioni non verranno usate sistematicamente nel seguito.

2.8.5 Operazioni con i limiti di funzione

Le regole già incontrate per le successioni si ritrovano per i limiti di funzione.

2.60 Teorema [algebra dei limiti di funzione]. *Siano $f, g : D \to \mathbb{R}$, con $D \subseteq \mathbb{R}$, sia a un punto di accumulazione di D e siano $L, M \in \mathbb{R}$. Supponiamo*

$$\lim_{x\to a} f(x) = L, \qquad \lim_{x\to a} g(x) = M.$$

Allora:

(i)
$$\lim_{x\to a}(f(x)+g(x)) = L + M = \lim_{x\to a} f(x) + \lim_{x\to a} g(x);$$

(ii) per ogni $\lambda \in \mathbb{R}$,
$$\lim_{x\to a}(\lambda f(x)) = \lambda L = \lambda \lim_{x\to a} f(x);$$

(iii)
$$\lim_{x\to a}(f(x)g(x)) = LM = \left(\lim_{x\to a} f(x)\right)\left(\lim_{x\to a} g(x)\right);$$

(iv) se $M \neq 0$ ed esiste $\delta_0 > 0$ tale che
$$g(x) \neq 0 \qquad \text{per ogni } x \in D \text{ con } 0 < |x-a| < \delta_0,$$

poniamo
$$D_g := \{x \in D \mid g(x) \neq 0\}.$$

Allora a è un punto di accumulazione di D_g e la funzione
$$\frac{f}{g} : D_g \to \mathbb{R}$$

soddisfa
$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{L}{M} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)},$$

dove il limite a sinistra è calcolato sul dominio D_g.

Le dimostrazioni seguono le stesse stime ε–δ già viste per le successioni. Per la somma, per esempio, si usa
$$|(f(x)+g(x)) - (L+M)| \leq |f(x)-L| + |g(x)-M|.$$

2.8.6 Esercizi risolti su punti e limiti di funzione

2.61 Esercizio.

1. *Punti interni, di accumulazione e isolati in un intervallo con un punto aggiunto.* Sia
$$D = [0,1] \cup \{2\} \subseteq \mathbb{R}.$$
Determinare i punti interni di D, i punti di accumulazione di D e i punti isolati di D.
Soluzione. I punti interni sono precisamente quelli di $(0,1)$, perché soltanto per essi esiste un intervallo aperto centrato nel punto e interamente contenuto in D. I punti di accumulazione sono tutti e soli i punti di $[0,1]$: anche 0 e 1 sono punti di accumulazione, benché non siano interni. Il punto 2 è invece isolato, perché, per esempio,
$$D \cap \left(\frac{3}{2}, \frac{5}{2}\right) = \{2\}.$$
Dunque i punti interni di D formano l'intervallo $(0,1)$, i punti di accumulazione formano l'intervallo $[0,1]$ e l'unico punto isolato è 2.
2. *Un punto di accumulazione che non appartiene all'insieme.* Sia
$$D = (0,1).$$
Determinare i punti interni, i punti di accumulazione e i punti isolati di D.
Soluzione. Ogni punto di $(0,1)$ è interno. I punti di accumulazione sono tutti i punti di $[0,1]$: in particolare 0 e 1 sono punti di accumulazione pur non appartenendo a D. Non vi sono punti isolati, perché ogni punto di D è circondato da altri punti di D arbitrariamente vicini.

3. *Una successione di punti con un unico punto di accumulazione.* Sia

$$D := \{0\} \cup \left\{ \frac{1}{n+1} \mid n \in \mathbb{N} \right\}.$$

Determinare i punti interni, i punti di accumulazione e i punti isolati di D.
Soluzione. L'insieme non contiene alcun intervallo aperto, quindi non ha punti interni. Il punto 0 è un punto di accumulazione, poiché

$$\frac{1}{n+1} \longrightarrow 0.$$

Ogni punto $1/(n+1)$ è invece isolato: essendo distinto dai punti adiacenti della successione, si può scegliere un intervallo sufficientemente piccolo che non contenga altri elementi di D. Non vi sono altri punti di accumulazione. Dunque 0 è l'unico punto di accumulazione, mentre tutti i punti $1/(n+1)$ sono isolati.

4. *Un limite dalla definizione.* Sia f definita su $\mathbb{R}$ da $f(x) = 3x - 1$. Dimostrare direttamente che

$$\lim_{x \to 2} f(x) = 5.$$

Soluzione. Sia $\varepsilon > 0$. Abbiamo

$$|f(x) - 5| = |(3x - 1) - 5| = 3|x - 2|.$$

Scegliamo

$$\delta := \frac{\varepsilon}{3}.$$

Allora $0 < |x - 2| < \delta$ implica

$$|f(x) - 5| < 3\delta = \varepsilon.$$

5. *Un limite quadratico.* Sia f definita su $\mathbb{R}$ da $f(x) = x^2$. Mostrare che, per ogni $a \in \mathbb{R}$,

$$\lim_{x \to a} f(x) = a^2.$$

Soluzione. Scriviamo

$$|x^2 - a^2| = |x - a|\,|x + a|.$$

Dobbiamo controllare anche il secondo fattore. Imponiamo anzitutto $|x - a| < 1$. Allora

$$|x| \leq |a| + 1$$

e quindi

$$|x + a| \leq |x| + |a| \leq 2|a| + 1.$$

Dato $\varepsilon > 0$, scegliamo

$$\delta := \min\left\{1, \frac{\varepsilon}{2|a| + 1}\right\}.$$

Se $0 < |x - a| < \delta$, allora

$$|x^2 - a^2| \leq |x - a|(2|a| + 1) < \varepsilon.$$

6. *Il valore nel punto può essere diverso dal limite.* Definiamo su $\mathbb{R}$ la funzione f ponendo

$$f(x) := \begin{cases} \dfrac{x^2 - 1}{x - 1}, & x \neq 1, \\ 7, & x = 1. \end{cases}$$

Determinare $\lim_{x \to 1} f(x)$.
Soluzione. Per $x \neq 1$,

$$\frac{x^2 - 1}{x - 1} = x + 1.$$

Pertanto

$$\lim_{x \to 1} f(x) = \lim_{x \to 1} (x + 1) = 2.$$

Il valore $f(1) = 7$ non interviene nel calcolo del limite.

7. *Un salto.* Sia f definita su $\mathbb{R}$ da

$$f(x) := \begin{cases} 1, & x < 0, \\ 2, & x \geq 0, \end{cases}$$

Determinare i limiti sinistro e destro in 0 e stabilire se esiste $\lim_{x\to 0} f(x)$.
Soluzione. Si ha

$$\lim_{x\to 0^-} f(x) = 1, \qquad \lim_{x\to 0^+} f(x) = 2.$$

Poiché i due limiti laterali sono differenti, il limite per $x \to 0$ non esiste.

2.9 Continuità

Il limite descrive il comportamento di una funzione nei pressi di un punto senza imporre alcun legame con il valore assunto nel punto stesso. La **continuità** nasce quando chiediamo che queste due informazioni coincidano.

NOTA STORICA. *La precisazione ottocentesca della continuità.* La nozione moderna di continuità si precisa nel corso dell'Ottocento. BOLZANO ne dà già nel 1817 una formulazione molto vicina a quella moderna; CAUCHY, nel *Cours d'analyse* del 1821, la inserisce sistematicamente nella teoria dei limiti. La formulazione aritmetica nel linguaggio ε–δ adottata in queste dispense si afferma pienamente con WEIERSTRASS, nelle cui lezioni del 1861 compare essenzialmente nella forma attuale [6]. La progressiva separazione fra la nozione generale di funzione e la proprietà di continuità, a cui contribuì anche DIRICHLET, è parte del processo di aritmetizzazione dell'analisi matematica ottocentesca. ■

2.62 Definizione [continuità in un punto]. Sia $D \subseteq \mathbb{R}$, sia $f : D \to \mathbb{R}$ e sia $a \in D$. Diciamo che f è **continua in** a se per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che, per ogni $x \in D$,

$$|x - a| < \delta \quad \Longrightarrow \quad |f(x) - f(a)| < \varepsilon. \tag{2.4}$$

Se f è continua in ogni punto di D, diciamo che f è continua su D. ■

2.63 Osservazione [continuità nei punti isolati e legame con il limite]. Siano $D \subseteq \mathbb{R}$, $f : D \to \mathbb{R}$ e $a \in D$. La Definizione 2.62 di continuità non richiede che a sia punto di accumulazione di D. Se a è un punto isolato di D, ogni funzione $f : D \to \mathbb{R}$ è automaticamente continua in a: scelta $\delta > 0$ in modo che

$$D \cap (a - \delta, a + \delta) = \{a\},$$

l'unico $x \in D$ con $|x - a| < \delta$ è $x = a$, e quindi

$$|f(x) - f(a)| = 0.$$

Se invece a è un punto di accumulazione di D, la condizione (2.4) equivale a

$$\boxed{f \text{ è continua in } a \quad \Longleftrightarrow \quad \lim_{x\to a} f(x) = f(a).}$$

Il legame fra continuità e limite ha dunque questo significato soltanto quando a è un punto di accumulazione del dominio. In tale caso la formula riunisce tre condizioni che, concettualmente, è utile distinguere:

(i) il valore $f(a)$ deve essere definito;
(ii) il limite $\lim_{x\to a} f(x)$ deve esistere;
(iii) il limite deve coincidere con il valore della funzione.

■

2.64 Esempio [eliminare una discontinuità rimovibile]. Consideriamo anzitutto la funzione

$$f : (-\infty, 1) \cup (1, +\infty) \to \mathbb{R}, \qquad f(x) = \frac{x^2 - 1}{x - 1}.$$

Per ogni x del suo dominio vale $f(x) = x + 1$, e quindi f è continua in ogni punto del proprio dominio. Il punto 1 non appartiene al dominio: non ha dunque senso chiedere se f sia continua in 1. Questo esempio mostra che la continuità è sempre una proprietà da riferire al dominio effettivo della funzione.

Riprendiamo ora la funzione dell'Esercizio 2.61.6, ottenuta estendendo il dominio a tutto $\mathbb{R}$ e ponendo

$$f(x) = \frac{x^2 - 1}{x - 1} \quad (x \neq 1), \qquad f(1) = 7.$$

Essa non è continua in 1, perché

$$\lim_{x \to 1} f(x) = 2 \neq 7 = f(1).$$

Se però ridefiniamo soltanto il valore nel punto ponendo $f(1) := 2$, otteniamo una funzione continua. Una discontinuità di questo tipo viene detta **rimovibile**. ■

2.65 Esempio [razionali, irrazionali e continuità]. La densità sia di $\mathbb{Q}$ sia di $\mathbb{R} \setminus \mathbb{Q}$ in $\mathbb{R}$ permette di costruire esempi molto diversi dalle funzioni elementari considerate finora.

(a) La **funzione di Dirichlet** f, definita su $\mathbb{R}$, è data da

$$f(x) := \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & x \notin \mathbb{Q} \end{cases}$$

non è continua in alcun punto. Infatti, in ogni intervallo attorno a qualunque $a \in \mathbb{R}$ esistono sia razionali sia irrazionali: possiamo quindi avvicinarci ad a attraverso punti nei quali f vale sempre 1 e attraverso punti nei quali vale sempre 0. I valori della funzione non possono dunque avvicinarsi a un unico limite.

(b) Consideriamo invece la funzione g definita su $\mathbb{R}$ da

$$g(x) := \begin{cases} x, & x \in \mathbb{Q}, \\ -x, & x \notin \mathbb{Q}. \end{cases}$$

Questa funzione è continua soltanto in 0. In 0 infatti

$$|g(x) - g(0)| = |g(x)| = |x|,$$

quindi $g(x) \to 0 = g(0)$ quando $x \to 0$.
Sia ora $a \neq 0$. Avvicinandoci ad a attraverso numeri razionali otteniamo valori di $g(x)$ che tendono ad a, mentre avvicinandoci ad a attraverso numeri irrazionali otteniamo valori che tendono a $-a$. Poiché $a \neq -a$, il limite di $g(x)$ per $x \to a$ non esiste. Dunque g non è continua in alcun punto diverso da 0. ■

2.66 Osservazione.

(a) Consideriamo prima una funzione

$$f : \mathbb{R}^2 \to \mathbb{R}.$$

Siano inoltre $(a, b) \in \mathbb{R}^2$ e $L \in \mathbb{R}$. Dire che

$$\lim_{(x,y)\to(a,b)} f(x, y) = L$$

significa che i valori di $f(x, y)$ possono essere resi arbitrariamente vicini a L prendendo (x, y) sufficientemente vicino ad (a, b), indipendentemente dalla direzione lungo la quale ci si avvicina. In termini ε–δ, per ogni $\varepsilon > 0$ deve esistere $\delta > 0$ tale che

$$0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta \quad \Longrightarrow \quad |f(x, y) - L| < \varepsilon.$$

La funzione f è continua nel punto (a, b) quando il limite esiste e coincide con il valore della funzione:

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

Equivalentemente, per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che

$$\sqrt{(x-a)^2 + (y-b)^2} < \delta \quad \Longrightarrow \quad |f(x, y) - f(a, b)| < \varepsilon.$$

Il passaggio a $\mathbb{R}^n$ è immediato: per $x = (x_1, \ldots, x_n)$ e $x_0 = (x_{0,1}, \ldots, x_{0,n})$ si usa la distanza euclidea

$$\|x - x_0\| := \sqrt{\sum_{i=1}^{n} (x_i - x_{0,i})^2},$$

e si mantengono le stesse definizioni di limite e continuità. Torneremo su questa costruzione nell'Osservazione 3.54, dove introdurremo in forma astratta la nozione di spazio metrico e chiariremo il ruolo delle norme nella definizione di una distanza.

(b) Questa distanza permette anche di estendere a $\mathbb{R}^n$ la nozione di punto interno già introdotta in $\mathbb{R}$. Data $\delta > 0$, la **palla aperta** di raggio δ e centro x_0 è l'insieme

$$B_\delta(x_0) := \{x \in \mathbb{R}^n \mid \|x - x_0\| < \delta\}.$$

Se $D \subseteq \mathbb{R}^n$ e $x_0 \in D$, diciamo che x_0 è interno a D se esiste $\delta > 0$ tale che

$$B_\delta(x_0) \subseteq D.$$

In altre parole, attorno a x_0 deve esistere un'intera palla aperta contenuta in D. Come nel caso di sottoinsiemi di $\mathbb{R}$, un insieme $D \subseteq \mathbb{R}^n$ è aperto se tutti i suoi punti sono interni. Questa formulazione sarà utile quando introdurremo le derivate parziali. Gli insiemi aperti non vuoti della topologia euclidea di $\mathbb{R}^n$ introdotta nella Definizione 2.52 possono equivalentemente essere descritti come unioni arbitrarie di palle aperte, con centri arbitrari e raggi positivi arbitrari. Anche su queste nozioni torneremo nell'Osservazione 3.54, dove saranno formulate nel contesto generale degli spazi metrici e degli spazi vettoriali normati.

■

2.9.1 Operazioni che conservano la continuità

2.67 Teorema [algebra delle funzioni continue]. *Siano $f, g : D \to \mathbb{R}$, con $D \subseteq \mathbb{R}$, continue in $a \in D$. Allora $f + g$, fg e λf, con $\lambda \in \mathbb{R}$, sono continue in a. Se $g(a) \neq 0$, anche f/g, considerata sul suo dominio naturale*

$$D_g := \{x \in D \mid g(x) \neq 0\},$$

è continua in a.

Dimostrazione. Se a è un punto isolato di D, allora è un punto isolato anche del dominio delle funzioni $f + g$, fg e λf. Se $g(a) \neq 0$, si ha inoltre $a \in D_g$ e a è isolato anche in D_g. Per l'Osservazione 2.63, tutte queste funzioni sono quindi automaticamente continue in a.

Supponiamo ora che a sia un punto di accumulazione di D. Poiché f e g sono continue in a, vale

$$\lim_{x\to a} f(x) = f(a), \qquad \lim_{x\to a} g(x) = g(a).$$

Il Teorema 2.60 dà allora

$$\lim_{x\to a}(f+g)(x) = f(a)+g(a) = (f+g)(a),$$

$$\lim_{x\to a}(fg)(x) = f(a)g(a) = (fg)(a),$$

e, per ogni $\lambda \in \mathbb{R}$,

$$\lim_{x\to a}(\lambda f)(x) = \lambda f(a) = (\lambda f)(a).$$

Quindi $f+g$, fg e λf sono continue in a.

Se infine $g(a) \neq 0$, dalla continuità di g in a, applicata con $\varepsilon = |g(a)|/2$, esiste $\delta > 0$ tale che, per $x \in D$ e $|x-a| < \delta$,

$$|g(x) - g(a)| < \frac{|g(a)|}{2}.$$

Ne segue

$$|g(x)| \geq |g(a)| - |g(x)-g(a)| > \frac{|g(a)|}{2} > 0.$$

Pertanto g non si annulla nei punti di D sufficientemente vicini ad a; in particolare a è un punto di accumulazione di D_g. Per il punto (iv) del Teorema 2.60,

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{f(a)}{g(a)},$$

e dunque f/g è continua in a sul suo dominio naturale. □

2.68 Teorema [continuità della composizione]. *Siano $D, E \subseteq \mathbb{R}$, sia $a \in D$, sia $f : D \to \mathbb{R}$ continua in a e sia $g : E \to \mathbb{R}$ continua in $f(a)$, con $f(D) \subseteq E$. Allora la funzione composta*

$$g \circ f$$

è continua in a.

Questo risultato è importante perché permette di costruire funzioni continue a partire da funzioni più semplici. In particolare, le funzioni polinomiali sono continue su tutto $\mathbb{R}$, e le funzioni razionali sono continue nei punti in cui il denominatore non si annulla.

2.69 Osservazione [una proprietà strutturale delle funzioni continue]. Se f e g sono continue su uno stesso dominio, allora lo sono anche $f+g$ e λf. La stabilità rispetto a somma e moltiplicazione per numeri reali verrà reinterpretata nel Capitolo 3 mediante una struttura algebrica astratta. ■

2.9.2 Completezza, teorema dei valori intermedi e teorema di Weierstrass

La continuità, combinata con la completezza di $\mathbb{R}$, produce risultati che formalizzano l'idea intuitiva di una curva "senza salti".

2.70 Teorema [teorema degli zeri]. *Siano $a, b \in \mathbb{R}$ con $a < b$ e sia $f : [a,b] \to \mathbb{R}$ continua. Supponiamo*

$$f(a) < 0 < f(b).$$

Allora esiste almeno un $c \in (a,b)$ tale che

$$f(c) = 0.$$

Dimostrazione. Consideriamo

$$E := \{x \in [a,b] \mid f(x) \leq 0\}.$$

L'insieme è non vuoto, perché $a \in E$, ed è limitato superiormente da b. Per la completezza di $\mathbb{R}$ esiste

$$c := \sup E.$$

Prima di tutto mostriamo che $a < c < b$. Poiché $f(a) < 0$ e f è continua in a, esiste un punto $x \in (a,b)$ sufficientemente vicino ad a per il quale $f(x) < 0$. Tale punto appartiene a E, e quindi $c \geq x > a$.

Analogamente, poiché $f(b) > 0$ e f è continua in b, esiste η con $0 < \eta < b - a$ tale che

$$f(x) > 0 \qquad \text{per ogni } x \in [a, b] \text{ con } b - \eta < x \leq b.$$

Nessun punto di questo intervallo appartiene a E; pertanto $c \leq b - \eta < b$.

Mostriamo ora che $f(c) = 0$.

Se fosse $f(c) > 0$, per continuità esisterebbe un intervallo sufficientemente piccolo attorno a c nel quale f rimane positiva. Ma, poiché c è l'estremo superiore di E, esistono punti di E arbitrariamente vicini a c da sinistra; in tali punti $f \leq 0$, contraddizione.

Se fosse $f(c) < 0$, la continuità garantirebbe invece che f rimane negativa in un piccolo intervallo attorno a c. Potremmo allora trovare un punto $x > c$ ancora appartenente a E, contro il fatto che c sia un maggiorante di E. Pertanto $f(c) = 0$. □

La conseguenza più importante è il *teorema dei valori intermedi*:

2.71 Teorema [teorema dei valori intermedi]. *Siano $a, b \in \mathbb{R}$ con $a < b$, sia $f : [a, b] \to \mathbb{R}$ continua e sia $y \in \mathbb{R}$ compreso fra $f(a)$ e $f(b)$. Allora esiste $c \in [a, b]$ tale che*

$$f(c) = y.$$

Dimostrazione. Se $y = f(a)$ oppure $y = f(b)$, basta scegliere rispettivamente $c = a$ oppure $c = b$. Supponiamo dunque che y sia strettamente compreso fra $f(a)$ e $f(b)$. Se

$$f(a) < y < f(b),$$

allora per $g(x) := f(x) - y$ vale $g(a) < 0 < g(b)$ e possiamo applicare il teorema degli zeri. Se invece

$$f(b) < y < f(a),$$

applichiamo lo stesso teorema alla funzione $-g$. In entrambi i casi esiste $c \in (a, b)$ tale che $f(c) = y$. □

Un altro importante teorema è il *teorema di Weierstrass*.

2.72 Teorema [teorema di Weierstrass]. *Siano $a, b \in \mathbb{R}$ con $a \leq b$ e sia $f : [a, b] \to \mathbb{R}$ continua. Allora f è limitata in $[a, b]$ ed esistono $x_m, x_M \in [a, b]$ per cui $f(x_m) = \min\{f(x) \mid x \in [a, b]\}$, $f(x_M) = \max\{f(x) \mid x \in [a, b]\}$.*

Dimostrazione. Se $a = b$, l'intervallo $[a, b] = \{a\}$ contiene un solo punto e la tesi è immediata. Supponiamo quindi $a < b$.

Mostriamo anzitutto che f è limitata. Poniamo

$$S := \{c \in [a, b] \mid f \text{ è limitata su } [a, c]\}.$$

L'insieme S è non vuoto: per la continuità di f in a, la funzione è limitata in un intervallo sufficientemente piccolo a destra di a. Inoltre S è limitato superiormente da $b < +\infty$. Per la completezza di $\mathbb{R}$ esiste dunque, ed è finito,

$$s := \sup S.$$

Se fosse $s < b$, la continuità in s renderebbe f limitata in un intervallo $(s - \delta, s + \delta)$, con $\delta > 0$ abbastanza piccolo. Riducendo δ, se necessario, possiamo inoltre supporre $s + \delta/2 \leq b$. Per la proprietà dell'estremo superiore possiamo scegliere $c \in S$ con $s - \delta/2 < c \leq s$. Poiché f è limitata su $[a, c]$ e anche su $(s - \delta, s + \delta)$, sarebbe limitata su $[a, s + \delta/2]$, contro la definizione di s. Dunque $s = b$. Infine, usando ancora la continuità in b e scegliendo $c \in S$ sufficientemente vicino a b, si conclude che f è limitata su tutto $[a, b]$.

Poniamo ora

$$M := \sup f([a, b]).$$

Se M non fosse assunto, avremmo $M - f(x) > 0$ per ogni $x \in [a, b]$ e la funzione

$$g(x) := \frac{1}{M - f(x)}$$

sarebbe continua su $[a, b]$, dunque limitata per quanto appena provato. Ma, per la Definizione 2.20 di estremo superiore, $f(x)$ può avvicinarsi a M quanto si vuole: per ogni $n \geq 1$ esiste $x_n \in [a, b]$ tale che

$$M - \frac{1}{n} < f(x_n) \leq M,$$

e quindi $g(x_n) > n$, assurdo. Pertanto M è assunto. Applicando lo stesso argomento a $-f$ si ottiene anche l'esistenza del minimo assoluto. □

2.73 Osservazione [una precisazione intuizionistica]. La dimostrazione del Teorema 2.72 è una dimostrazione classica. Essa usa la completezza di $\mathbb{R}$ nella forma del principio dell'estremo superiore applicato agli insiemi che compaiono nella prova. In matematica intuizionistica una tale forma di completezza non è accettata senza ulteriori ipotesi costruttive sull'insieme considerato. Anche l'ultimo passaggio è propriamente classico: dal fatto che $M - f(x)$ non sia nullo per ogni x non segue, in generale, intuizionisticamente, che il suo reciproco possa essere costruito come nella dimostrazione del Teorema 2.72. Il teorema di Weierstrass, nella formulazione qui adottata, va quindi inteso come un risultato della matematica classica. ■

2.74 Osservazione [dove riappare la completezza]. Il teorema degli zeri e il Teorema 2.72 di Weierstrass sono esempi, dopo il Teorema 2.48 di convergenza delle successioni monotone limitate, del ruolo della completezza. La continuità da sola controlla il comportamento locale della funzione; la completezza permette di trasformare questo controllo locale in risultati globali di esistenza, attraverso l'estremo superiore di opportuni insiemi. ■

2.9.3 Esercizi risolti sulla continuità

Prima di procedere, fissiamo una terminologia che useremo anche nel seguito.

2.75 Definizione [polinomio reale e grado]. Un **polinomio reale** nella variabile x è un'espressione della forma

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n, \qquad a_0, \ldots, a_n \in \mathbb{R},$$

per un certo intero $n \geq 0$. Se p non è il polinomio nullo e $a_n \neq 0$, il più grande esponente che compare con coefficiente non nullo, cioè n, si chiama **grado del polinomio**; il coefficiente a_n si chiama **coefficiente principale**. In queste dispense non assegneremo un grado al polinomio nullo. ■

2.76 Esercizio.

1. *Continuità di una funzione polinomiale.* Sia f definita su $\mathbb{R}$ da

   $$f(x) = x^2 + 3x - 1.$$

   Mostrare che f è continua in ogni $a \in \mathbb{R}$.
   Soluzione. Nell'Esercizio 2.61.5 abbiamo mostrato che $x^2 \to a^2$ quando $x \to a$, e chiaramente $x \to a$. Per il Teorema 2.60,

   $$\lim_{x \to a} (x^2 + 3x - 1) = a^2 + 3a - 1 = f(a).$$

   Quindi f è continua in a. Poiché a è arbitrario, è continua su $\mathbb{R}$.
2. *Una funzione definita a tratti.* Sia f definita su $\mathbb{R}$ da

   $$f(x) := \begin{cases} x + 1, & x < 1, \\ 3 - x, & x \geq 1. \end{cases}$$

   Stabilire se f è continua in 1.
   Soluzione. Il limite sinistro è

   $$\lim_{x \to 1^-} (x + 1) = 2,$$

   mentre il limite destro è

   $$\lim_{x \to 1^+} (3 - x) = 2.$$

   Inoltre

   $$f(1) = 3 - 1 = 2.$$

   Quindi

   $$\lim_{x \to 1} f(x) = f(1) = 2,$$

   e la funzione è continua in 1.

3. *Esistenza di una radice.* Mostrare che l'equazione

$$x^3 + x - 1 = 0$$

possiede almeno una soluzione nell'intervallo $(0, 1)$.
Soluzione. La funzione f definita su $\mathbb{R}$ da

$$f(x) = x^3 + x - 1$$

è un polinomio, dunque è continua. Inoltre

$$f(0) = -1 < 0, \qquad f(1) = 1 > 0.$$

Per il Teorema 2.70 degli zeri esiste $c \in (0, 1)$ tale che $f(c) = 0$.

2.10 Derivabilità

La continuità descrive la stabilità dei valori di una funzione rispetto a piccole variazioni dell'argomento. La **derivata** introduce una domanda più precisa: *con quale velocità varia la funzione vicino a un punto?*

NOTA STORICA. *Un breve riferimento storico.* Nel XVII secolo ISAAC NEWTON e GOTTFRIED WILHELM LEIBNIZ elaborarono, con linguaggi differenti, due formulazioni generali del calcolo differenziale e integrale [6]. NEWTON parlava di *fluenti* e delle loro *flussioni*; LEIBNIZ sviluppò un calcolo infinitesimale e introdusse notazioni come dx, dy e il segno $\int$, ancora oggi in uso.

La formulazione di queste dispense è però moderna: la derivata viene definita mediante un limite e dx, dy non sono qui numeri reali infinitesimi. Il calcolo originario di LEIBNIZ non coincide quindi con la formulazione attuale, benché molte notazioni ne conservino la traccia. Se si vogliono trattare infinitesimi non nulli come numeri, occorre estendere $\mathbb{R}$: l'**analisi non standard** di ABRAHAM ROBINSON [10] realizza questa possibilità introducendo, accanto ai reali, numeri infinitesimi e infinitamente grandi.

Un passaggio particolarmente significativo verso la fisica matematica moderna avviene con la progressiva analitizzazione della *fisica meccanica*. Con questa espressione intendiamo qui la parte della fisica che studia il moto dei corpi e le leggi che ne descrivono l'evoluzione. LAGRANGE non parte da zero: già EULER, D'ALEMBERT, i BERNOULLI, CLAIRAUT e altri matematici del Settecento avevano riformulato ampie parti della fisica meccanica mediante il calcolo differenziale e mediante equazioni che coinvolgono derivate. È particolarmente significativo il titolo stesso dell'opera di EULER del 1736, *Mechanica sive motus scientia analytice exposita* [7], che presenta programmaticamente la scienza del moto in forma analitica.

Nella *Méchanique analitique* del 1788 LAGRANGE porta questo processo a una sistemazione più radicale. Nell'*Avertissement* della prima edizione, a p. VI, dichiara programmaticamente che nell'opera non si troveranno figure e che i metodi esposti non richiedono costruzioni o ragionamenti geometrici o meccanici, ma operazioni algebriche condotte secondo una procedura regolare e uniforme [8]. Il trattato non rappresenta dunque l'origine della cosiddetta *meccanica analitica*, cioè della formulazione matematica della fisica meccanica mediante metodi analitici, ma una sua culminazione e un punto di svolta: il formalismo dell'analisi matematica viene assunto come linguaggio autosufficiente per formulare le leggi del moto.

Più in generale, l'euristica del calcolo differenziale e integrale fu una delle più potenti forze propulsive della fisica e della fisica matematica, soprattutto tra il Settecento e l'Ottocento. La fisica meccanica, la fisica dei fluidi e dei continui, la teoria matematica del calore e della termodinamica, l'elettromagnetismo, così come le formulazioni lagrangiana e hamiltoniana della fisica meccanica, furono sviluppate facendo uso essenziale di derivate, integrali e di equazioni che coinvolgono derivate. *Il successo euristico e fisico del calcolo precedette in larga misura la piena chiarificazione ottocentesca dei suoi fondamenti*: è un esempio particolarmente significativo di una pratica matematica capace di produrre nuove teorie prima di ricevere una sistemazione rigorosa completa. ■

Per due punti distinti a e $a + h$ il rapporto

$$\frac{f(a + h) - f(a)}{h}$$

misura la variazione media di f per unità di variazione della variabile. Geometricamente è il coefficiente angolare della retta secante che congiunge i punti

$$(a, f(a)) \qquad \text{e} \qquad (a + h, f(a + h)).$$

Facendo tendere h a zero, il secondo punto si avvicina al primo e, quando il limite esiste, le secanti individuano la retta tangente.

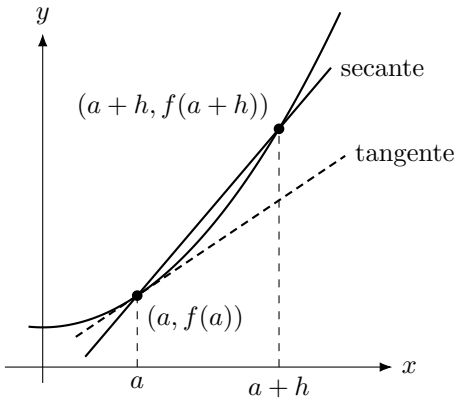


Figure 3: Il rapporto incrementale è la pendenza della secante. Quando $h \to 0$, la derivata, se esiste, è la pendenza limite della tangente.

2.10.1 Definizione di derivata

2.77 Definizione [derivata in un punto]. Sia $D \subseteq \mathbb{R}$, sia $f : D \to \mathbb{R}$ e sia x_0 un punto interno di D. Diciamo che f è **derivabile in** x_0 se esiste finito il limite

$$f'(x_0) := \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \tag{2.5}$$

Il numero $f'(x_0)$ definito dalla formula (2.5) si chiama derivata di f nel punto x_0. ■

2.78 Osservazione [derivate parziali]. Sia $D \subseteq \mathbb{R}^2$, sia $f : D \to \mathbb{R}$ e sia (a, b) un punto interno di D. La **derivata parziale** di f rispetto alla prima variabile nel punto (a, b) misura la variazione di f quando cambia la prima coordinata e la seconda resta fissata. Se il limite esiste, si pone

$$\frac{\partial f}{\partial x}(a, b) := \lim_{h \to 0} \frac{f(a + h, b) - f(a, b)}{h}.$$

Analogamente, la derivata parziale rispetto alla seconda variabile è

$$\frac{\partial f}{\partial y}(a, b) := \lim_{h \to 0} \frac{f(a, b + h) - f(a, b)}{h}.$$

Il fatto che (a, b) sia interno a D garantisce che, per h sufficientemente piccolo, i punti $(a + h, b)$ e $(a, b + h)$ appartengano ancora a D.

Per indicare in modo abbreviato le due direzioni coordinate poniamo esplicitamente

$$e_1 := (1, 0), \qquad e_2 := (0, 1).$$

Le coppie e_1 ed e_2 servono qui soltanto a indicare, rispettivamente, che si fa variare la prima oppure la seconda coordinata. Sono comuni anche le notazioni $f_x(a, b)$ e $f_y(a, b)$ per le due derivate parziali. La definizione si generalizza immediatamente a funzioni definite su sottoinsiemi di $\mathbb{R}^n$, facendo variare una coordinata alla volta e mantenendo fisse tutte le altre. Per una trattazione più ampia delle funzioni di più variabili si veda [4, cap. 9]. ■

Ponendo $x = a + h$, la stessa definizione può essere scritta nella forma

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

Quando f è derivabile in ogni punto interno di un intervallo I, possiamo considerare la **funzione derivata**

$$f' : \operatorname{int}(I) \to \mathbb{R}, \qquad x \longmapsto f'(x),$$

dove $\operatorname{int}(I)$ indica l'insieme dei punti interni di I. Se I è aperto, allora $\operatorname{int}(I) = I$.

Useremo anche la notazione di Leibniz

$$\frac{df}{dx}(x) = f'(x)$$

per la derivata di una funzione f della variabile x. In queste dispense df/dx è una notazione per la derivata e non viene interpretato come un quoziente fra due numeri df e dx.

Se la funzione derivata f' è a sua volta derivabile, definiamo la **derivata seconda** di f ponendo

$$f''(x) := (f')'(x).$$

Nella notazione di Leibniz la stessa derivata si scrive

$$\frac{d^2 f}{dx^2}(x) = f''(x).$$

Più in generale, se esistono le derivate successive, per $n \geq 1$ indicheremo con $f^{(n)}$ la **derivata di ordine** n, definita ricorsivamente da

$$f^{(1)} := f', \qquad f^{(n+1)} := (f^{(n)})'.$$

Nella notazione di Leibniz scriveremo equivalentemente

$$f^{(n)}(x) = \frac{d^n f}{dx^n}(x).$$

In particolare, $f^{(2)} = f''$.

Quando la variabile indipendente è il tempo t e la funzione è indicata con $x = x(t)$, scriveremo quindi

$$\frac{dx}{dt} = x'(t), \qquad \frac{d^2 x}{dt^2} = x''(t).$$

2.79 Esempio [la derivata di x^2]. Sia f definita su $\mathbb{R}$ da $f(x) = x^2$. Allora

$$\frac{f(a+h) - f(a)}{h} = \frac{(a+h)^2 - a^2}{h} = 2a + h.$$

Passando al limite per $h \to 0$ otteniamo

$$f'(a) = 2a.$$

Poiché a è arbitrario,

$$\boxed{(x^2)' = 2x.}$$

■

2.80 Osservazione [interpretazione fisica]. Siano $I \subseteq \mathbb{R}$ un intervallo, $s : I \to \mathbb{R}$ e t un punto interno di I. Se s descrive la posizione di un corpo al tempo t, il rapporto

$$\frac{s(t+h) - s(t)}{h}$$

è la velocità media nell'intervallo di tempo di ampiezza h. Il limite per $h \to 0$, quando esiste, è la velocità istantanea

$$v(t) = s'(t).$$

La derivata traduce quindi in termini matematici l'idea di variazione locale istantanea. ■

2.10.2 Derivabilità e continuità

La derivabilità è una condizione più forte della continuità.

2.81 Teorema [derivabilità implica continuità]. *Sia $f : D \to \mathbb{R}$, con $D \subseteq \mathbb{R}$, e sia a un punto interno di D. Se f è derivabile in a, allora f è continua in a.*

Dimostrazione. Per $h \neq 0$ scriviamo

$$f(a+h) - f(a) = \frac{f(a+h) - f(a)}{h}\, h.$$

Poiché f è derivabile in a,

$$\frac{f(a+h) - f(a)}{h} \longrightarrow f'(a) \qquad (h \to 0),$$

mentre $h \to 0$. Per il Teorema 2.60,

$$f(a+h) - f(a) \longrightarrow f'(a) \cdot 0 = 0.$$

Quindi

$$f(a+h) \longrightarrow f(a),$$

che è precisamente la continuità in a. □

Il viceversa è falso. La continuità non garantisce l'esistenza di una tangente con pendenza finita.

2.82 Esempio [il valore assoluto]. La funzione f definita su $\mathbb{R}$ da

$$f(x) = |x|$$

è continua in 0, ma non è derivabile in 0. Infatti, per $h > 0$,

$$\frac{|h| - |0|}{h} = 1,$$

mentre per $h < 0$,

$$\frac{|h| - |0|}{h} = -1.$$

I due limiti laterali del rapporto incrementale sono diversi. Dunque $f'(0)$ non esiste. ■

2.83 Osservazione [funzioni continue non derivabili in alcun punto]. Il fenomeno può essere molto più estremo di quello dell'Esempio 2.82: esistono funzioni continue su tutto $\mathbb{R}$ che non sono derivabili in alcun punto. Una costruzione classica di tali funzioni utilizza serie trigonometriche del tipo delle serie di Fourier. Non la svilupperemo in queste dispense, perché la teoria delle serie di Fourier non fa parte del programma del corso. ■

2.10.3 Regole di derivazione

Le operazioni algebriche sono compatibili anche con la derivazione. Non dimostreremo in queste dispense il teorema seguente nella sua interezza.

2.84 Teorema [regole elementari di derivazione]. *Siano $D \subseteq \mathbb{R}$, sia a un punto interno di D, siano $f, g : D \to \mathbb{R}$ derivabili in a e siano $\alpha, \beta \in \mathbb{R}$. Allora:*

(i)
$$(\alpha f + \beta g)'(a) = \alpha f'(a) + \beta g'(a);$$

(ii)
$$(fg)'(a) = f'(a)g(a) + f(a)g'(a);$$

(iii) se $g(a) \neq 0$, poniamo
$$D_g := \{x \in D \mid g(x) \neq 0\}.$$
Poiché la derivabilità implica la continuità, a è un punto interno di D_g. La funzione $f/g : D_g \to \mathbb{R}$ è derivabile in a e
$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.$$

(iv) Inoltre, siano $E \subseteq \mathbb{R}$ un insieme tale che $f(D) \subseteq E$ e $h : E \to \mathbb{R}$ una funzione derivabile in $f(a)$. Allora $h \circ f$ è derivabile in a e
$$(h \circ f)'(a) = h'(f(a))f'(a).$$

La prima regola segue direttamente dalla linearità del limite. Per il prodotto si può scrivere

$$\frac{f(a+h)g(a+h) - f(a)g(a)}{h}$$

e aggiungere e sottrarre $f(a+h)g(a)$ al numeratore. Si ottiene

$$\frac{f(a+h) - f(a)}{h}\, g(a) + f(a+h)\, \frac{g(a+h) - g(a)}{h}.$$

Poiché la derivabilità implica la continuità, $f(a+h) \to f(a)$; passando al limite segue la formula del prodotto.

La regola per la composizione viene chiamata **regola della catena**. Essa esprime un principio generale: quando una variazione agisce attraverso più passaggi successivi, le corrispondenti variazioni locali si moltiplicano.

2.85 Corollario [derivata delle potenze]. *Per ogni intero $n \geq 1$, la funzione $f(x) = x^n$ è derivabile su $\mathbb{R}$ e*

$$\frac{d}{dx} x^n = n x^{n-1}.$$

Dimostrazione. Procediamo per induzione su n. Per $n = 1$ la formula dà $(x)' = 1$. Supponiamo ora che

$$(x^n)' = n x^{n-1}.$$

Poiché $x^{n+1} = x^n \, x$, dalla regola del prodotto del Teorema 2.84 segue

$$(x^{n+1})' = (x^n)' x + x^n (x)' = n x^{n-1} x + x^n = (n+1) x^n.$$

La formula vale quindi per ogni $n \geq 1$. □

2.86 Corollario [derivata di un polinomio]. *Sia*

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n = \sum_{k=0}^{n} a_k x^k, \qquad a_0, \ldots, a_n \in \mathbb{R}.$$

Allora p è derivabile su $\mathbb{R}$ e

$$p'(x) = a_1 + 2a_2 x + \cdots + n a_n x^{n-1} = \sum_{k=1}^{n} k a_k x^{k-1}.$$

Dimostrazione. La funzione costante a_0 ha derivata nulla. Per il Corollario 2.85,

$$(a_k x^k)' = k a_k x^{k-1} \qquad (k \geq 1).$$

Applicando ripetutamente la linearità della derivazione, cioè il punto (i) del Teorema 2.84, si ottiene la formula. □

2.87 Osservazione [un'anticipazione algebrica]. La formula

$$(\alpha f + \beta g)' = \alpha f' + \beta g'$$

mostra che l'operazione di derivazione rispetta le combinazioni lineari. Nel Capitolo 3 introdurremo un linguaggio astratto nel quale questa proprietà verrà espressa dicendo che la derivazione è un'applicazione lineare su opportuni insiemi di funzioni. ■

2.10.4 Estremi locali e teoremi del valor medio

2.88 Definizione [massimo e minimo locale]. Sia $D \subseteq \mathbb{R}$, sia $f : D \to \mathbb{R}$ e sia $a \in D$. Diciamo che f possiede un **massimo locale** in a se esiste $\delta > 0$ tale che

$$f(x) \leq f(a)$$

per ogni $x \in D$ con $|x - a| < \delta$. La definizione di **minimo locale** si ottiene invertendo la disuguaglianza.

Se invece

$$f(x) \leq f(a) \qquad \text{per ogni } x \in D,$$

diciamo che f possiede in a un **massimo assoluto**, o **globale**. Analogamente si definiscono il **minimo assoluto** e il **minimo globale**. Un massimo o minimo assoluto è, in particolare, anche locale. ■

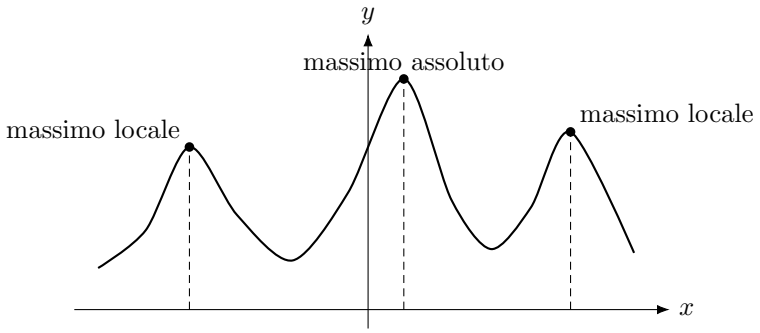


Figure 4: Una funzione può possedere più massimi locali. Nell'esempio il massimo centrale è anche il massimo assoluto, mentre gli altri due sono massimi soltanto locali, nel senso della Definizione 2.88.

2.89 Teorema [condizione necessaria di Fermat]. *Sia $f : D \to \mathbb{R}$, con $D \subseteq \mathbb{R}$, e sia a un punto interno di D. Se f è derivabile in a e possiede in a un massimo o un minimo locale, allora*

$$f'(a) = 0.$$

Dimostrazione. Supponiamo, per esempio, che a sia un massimo locale. Per $h > 0$ sufficientemente piccolo,

$$f(a+h) - f(a) \leq 0,$$

quindi

$$\frac{f(a+h) - f(a)}{h} \leq 0.$$

Per $h < 0$ sufficientemente piccolo il numeratore è ancora non positivo, ma il denominatore è negativo, dunque

$$\frac{f(a+h) - f(a)}{h} \geq 0.$$

Se il limite bilatero esiste, il limite destro è non positivo e quello sinistro è non negativo; dovendo coincidere, il loro valore comune è 0. □

Il Teorema 2.89 di Fermat, insieme al Teorema 2.72 di Weierstrass, permette di ottenere due risultati fondamentali che useremo più avanti.

2.90 Teorema [teorema di Rolle]. *Siano $a, b \in \mathbb{R}$ con $a < b$, e sia $f : [a, b] \to \mathbb{R}$ continua su $[a, b]$ e derivabile in ogni punto di (a, b). Se*

$$f(a) = f(b),$$

allora esiste almeno un punto $c \in (a, b)$ tale che

$$f'(c) = 0.$$

Dimostrazione. Per il Teorema 2.72 di Weierstrass, f assume in $[a, b]$ un massimo assoluto e un minimo assoluto. Se massimo e minimo coincidono, f è costante e quindi $f'(c) = 0$ per ogni $c \in (a, b)$. Supponiamo allora che f non sia costante. Poiché $f(a) = f(b)$, almeno uno fra il massimo assoluto e il minimo assoluto ha valore diverso da $f(a) = f(b)$; esso deve quindi essere assunto in un punto interno $c \in (a, b)$. In tale punto f possiede un estremo locale e, per il Teorema 2.89 di Fermat,

$$f'(c) = 0.$$

□

2.91 Teorema [teorema di Lagrange, o del valor medio]. *Siano $a, b \in \mathbb{R}$ con $a < b$, e sia $f : [a, b] \to \mathbb{R}$ continua su $[a, b]$ e derivabile in ogni punto di (a, b). Allora esiste almeno un punto $c \in (a, b)$ tale che*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

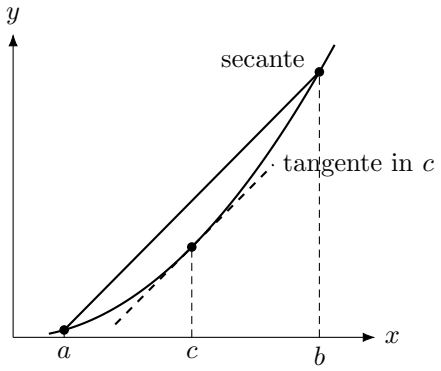


Figure 5: Interpretazione geometrica del Teorema 2.91: esiste almeno un punto $c \in (a,b)$ nel quale la tangente al grafico è parallela alla secante che congiunge $(a, f(a))$ e $(b, f(b))$.

Dimostrazione. Consideriamo la funzione

$$g(x) := f(x) - \frac{f(b)-f(a)}{b-a}(x-a).$$

Essa è continua su $[a,b]$ e derivabile su (a,b); inoltre

$$g(a) = f(a), \qquad g(b) = f(b) - \big(f(b)-f(a)\big) = f(a).$$

Per il Teorema 2.90 esiste quindi $c \in (a,b)$ tale che $g'(c) = 0$. Poiché

$$g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a},$$

la relazione $g'(c) = 0$ equivale a

$$f'(c) = \frac{f(b)-f(a)}{b-a}.$$

□

2.92 Corollario [derivata nulla e funzioni costanti]. *Siano $a, b \in \mathbb{R}$ con $a < b$, e sia $f : [a,b] \to \mathbb{R}$ continua su $[a,b]$ e derivabile in ogni punto di (a,b). Se*

$$f'(x) = 0 \qquad \text{per ogni } x \in (a,b),$$

allora f è costante su $[a,b]$.

Dimostrazione. Siano $x < y$ due punti di (a,b). Applicando il Teorema 2.91 alla restrizione di f a $[x,y]$, esiste $c \in (x,y)$ tale che

$$\frac{f(y)-f(x)}{y-x} = f'(c) = 0.$$

Poiché $y - x \neq 0$, segue $f(y) = f(x)$. Dunque f assume un unico valore costante, diciamo K, in tutto l'intervallo aperto (a,b).

Non occorre richiedere la derivabilità negli estremi, che qui non è stata definita. La continuità di f su $[a,b]$ permette infatti di estendere la costanza anche ad a e a b:

$$f(a) = \lim_{x \to a^+} f(x) = K, \qquad f(b) = \lim_{x \to b^-} f(x) = K.$$

Pertanto f è costante su tutto $[a,b]$. □

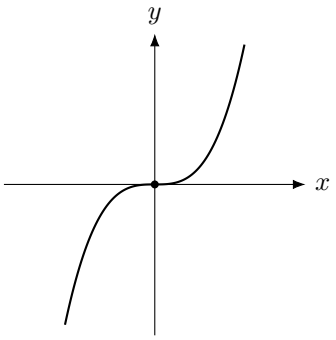
2.93 Osservazione [la condizione $f'(a) = 0$ non è sufficiente]. La condizione $f'(a) = 0$ è necessaria, ma non sufficiente per avere un estremo locale. Per esempio, per la funzione f definita su $\mathbb{R}$ da

$$f(x) = x^3$$

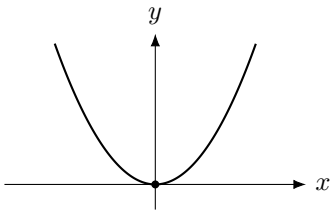
si ha $f'(0) = 0$, ma 0 non è né un massimo né un minimo locale. Il confronto con la funzione g definita su $\mathbb{R}$ da

$$g(x) = x^2$$

è istruttivo: anche $g'(0) = 0$, ma in questo caso 0 è un minimo locale (e assoluto).



$y = x^3$



$y = x^2$

Nel primo caso il grafico attraversa l'asse orizzontale continuando a crescere attraverso l'origine; nel secondo, invece, l'origine è il punto più basso del grafico. La sola condizione di annullamento della derivata non permette dunque di distinguere i due casi. ■

2.10.5 Esercizi risolti sulla derivabilità

2.94 Esercizio.

1. *Derivata dalla definizione.* Sia f definita su $\mathbb{R}$ da $f(x) = 3x + 2$. Calcolarne la derivata dalla definizione.
   Soluzione. Per ogni a,
   $$\frac{f(a+h) - f(a)}{h} = \frac{3(a+h) + 2 - (3a+2)}{h} = 3.$$
   Pertanto
   $$f'(a) = 3$$
   per ogni a, cioè $f'(x) = 3$.
2. *Una funzione continua ma non derivabile.* Sia f definita su $\mathbb{R}$ da $f(x) = |x|$. Mostrare che f è continua in 0 ma non derivabile in 0.
   Soluzione. La continuità segue da
   $$||x| - |0|| = |x| \longrightarrow 0 \qquad (x \to 0).$$
   Per la derivata, invece,
   $$\frac{|h|}{h} = \begin{cases} 1, & h > 0, \\ -1, & h < 0. \end{cases}$$
   I due limiti laterali sono differenti, quindi la derivata in 0 non esiste.
3. *Combinazioni lineari.* Siano f e g definite sullo stesso intervallo $I \subseteq \mathbb{R}$ e derivabili su I; poniamo
   $$h := 4f - 3g.$$
   Esprimere h' in termini di f' e g'.
   Soluzione. Per la prima regola del Teorema 2.84,
   $$h' = 4f' - 3g'.$$
   La stessa forma della combinazione viene conservata dalla derivazione.
4. *Derivata di un polinomio.* Sia p il polinomio su $\mathbb{R}$ definito da
   $$p(x) = x^3 - 2x^2 + 5x - 1.$$
   Calcolarne la derivata.
   Soluzione. Per il Corollario 2.86,
   $$p'(x) = 3x^2 - 4x + 5.$$

5. *Tangente a una parabola.* Sia f definita su $\mathbb{R}$ da $f(x) = x^2$. Determinare l'equazione della retta tangente al suo grafico nel punto di ascissa $a = 1$.
Soluzione. Poiché
$$f(1) = 1, \qquad f'(1) = 2,$$
la retta tangente passa per $(1, 1)$ e ha coefficiente angolare 2. Dunque
$$y - 1 = 2(x - 1),$$
ossia
$$y = 2x - 1.$$
6. *Un punto stazionario che non è un estremo.* Per la funzione f definita su $\mathbb{R}$ da $f(x) = x^3$, verificare che $f'(0) = 0$ ma che 0 non è un massimo né un minimo locale.
Soluzione. Si ha
$$f'(x) = 3x^2,$$
quindi $f'(0) = 0$. Tuttavia, per ogni $\delta > 0$, nell'intervallo $(-\delta, \delta)$ esistono punti $x < 0$ con $f(x) < 0 = f(0)$ e punti $x > 0$ con $f(x) > 0 = f(0)$. Pertanto 0 non è un estremo locale.
7. *Derivate parziali.* Sia $f : \mathbb{R}^2 \to \mathbb{R}$ definita da
$$f(x, y) = x^2 y + 3y^2.$$
Calcolare le derivate parziali rispetto a x e a y.
Soluzione. Nella derivazione rispetto a x si considera y costante, mentre nella derivazione rispetto a y si considera x costante. Pertanto
$$\frac{\partial f}{\partial x}(x, y) = 2xy, \qquad \frac{\partial f}{\partial y}(x, y) = x^2 + 6y.$$

2.11 Equazioni differenziali: l'analiticizzazione della fisica e il determinismo della fisica classica

La nozione di derivata permette di esprimere molte leggi della fisica mediante equazioni nelle quali l'incognita non è un numero, ma una funzione. Una **equazione differenziale** è precisamente un'equazione in cui compaiono una funzione incognita e una o più delle sue derivate. Quando la funzione incognita dipende da una sola variabile reale si parla di **equazione differenziale ordinaria**; quando dipende da più variabili e compaiono derivate parziali si parla invece di **equazione alle derivate parziali**. In questa sezione considereremo soltanto esempi di equazioni differenziali ordinarie. Per **soluzione di un'equazione differenziale** intendiamo una funzione, definita su un intervallo e derivabile il numero di volte necessario, che rende vera l'equazione in ogni punto dell'intervallo.

L'**ordine di un'equazione differenziale** è l'ordine più alto delle derivate che vi compaiono. Un primo esempio molto semplice è
$$f'(x) = 2x,$$
che è un'equazione differenziale del primo ordine. Le funzioni f_c, definite su $\mathbb{R}$ da
$$f_c(x) := x^2 + c, \qquad c \in \mathbb{R},$$
sono soluzioni, perché $f_c'(x) = 2x$. L'equazione da sola non sceglie dunque una singola funzione. Se imponiamo anche, per esempio,
$$f(0) = 3,$$
allora fra queste soluzioni viene selezionata $f(x) = x^2 + 3$.

Un esempio più significativo, perché mette direttamente in relazione la funzione incognita con la sua derivata, è
$$\frac{df}{dx} + kf(x) = 0, \qquad k > 0,$$
dove k è una costante assegnata. Anche questa è un'equazione differenziale del primo ordine. Assumendo note le proprietà della funzione esponenziale, per ogni $c \in \mathbb{R}$ la funzione $f_c : \mathbb{R} \to \mathbb{R}$ definita da
$$f_c(x) := ce^{-kx}$$

è una soluzione. Infatti

$$\frac{df_c}{dx} = -kce^{-kx} = -kf_c(x),$$

e quindi $\frac{df_c}{dx} + kf_c(x) = 0$. Non dimostreremo qui che queste siano tutte le soluzioni. La condizione $f(0) = c_0$ seleziona, all'interno della famiglia appena descritta, la soluzione $f(x) = c_0e^{-kx}$.

Un'equazione differenziale accompagnata da condizioni che assegnano i valori della funzione e, quando necessario, di alcune sue derivate in un istante o in un punto fissato costituisce un **problema di Cauchy**. Le condizioni assegnate si chiamano **condizioni iniziali**. Nel primo esempio il problema di Cauchy è

$$\begin{cases} f'(x) = 2x, \\ f(0) = 3. \end{cases}$$

L'idea è importante: l'equazione esprime una legge generale, mentre le condizioni iniziali individuano una particolare evoluzione fra quelle compatibili con la legge.

Il legame con la fisica meccanica è particolarmente diretto. Consideriamo un *punto materiale*, cioè un corpo idealizzato la cui posizione può essere rappresentata da un solo punto, e supponiamo per semplicità che esso si muova lungo una retta identificata con $\mathbb{R}$. La sua **legge oraria** è una funzione

$$x : I \to \mathbb{R}, \qquad t \mapsto x(t),$$

dove $I \subseteq \mathbb{R}$ è l'intervallo di tempo considerato. Essa assegna a ogni istante $t \in I$ la posizione $x(t)$. Come nell'Osservazione 2.80,

$$v(t) := \frac{dx}{dt} = x'(t)$$

è la velocità, mentre

$$a(t) := \frac{d^2x}{dt^2} = x''(t)$$

è l'accelerazione.

La seconda legge della dinamica di Newton assume allora la forma

$$m\frac{d^2x}{dt^2} = F\left(t, x(t), \frac{dx}{dt}\right). \tag{2.6}$$

Nell'equazione (2.6), $m > 0$ è la massa, nota, del punto materiale e

$$F : \mathbb{R}^3 \to \mathbb{R}$$

è una *funzione nota*, assegnata dalla legge fisica considerata, che nel caso unidimensionale descrive la *forza risultante* agente sul punto; il suo valore può dipendere dall'istante, dalla posizione e dalla velocità. L'incognita dell'equazione è quindi la legge oraria $x = x(t)$, non la funzione F.

Per determinare un moto particolare occorre specificare lo **stato iniziale**. Fissato un istante t_0, assegniamo una posizione x_0 e una velocità v_0; nel formalismo considerato, la coppia (x_0, v_0) è lo stato del punto materiale all'istante t_0. Consideriamo allora il problema di Cauchy

$$\begin{cases} mx''(t) = F\big(t, x(t), x'(t)\big), \\ x(t_0) = x_0, \\ x'(t_0) = v_0. \end{cases} \tag{2.7}$$

Sotto opportune ipotesi di regolarità su F, che qui non specifichiamo, un teorema fondamentale della teoria delle equazioni differenziali garantisce che il problema (2.7) possiede una e una sola soluzione sul suo **intervallo massimale di esistenza**, cioè sul più grande intervallo contenente t_0 al quale la soluzione può essere prolungata come soluzione dello stesso problema di Cauchy [9].

Questo risultato fornisce il nucleo matematico di una forma di *determinismo* della fisica classica: fissata la legge delle forze e fissato lo stato iniziale, l'evoluzione compatibile con quei dati è unica in ogni istante del suo intervallo massimale di esistenza. L'unicità matematica non significa però che la soluzione sia sempre esprimibile mediante una formula semplice, né che sia sempre facile calcolarla con precisione.

Le equazioni differenziali, sia ordinarie sia alle derivate parziali, svolgono un ruolo pervasivo nella fisica classica e in gran parte della fisica moderna. Il loro ruolo non consiste soltanto nel fornire un linguaggio conveniente per descrivere fenomeni già concettualmente definiti: la struttura matematica entra spesso nella formulazione stessa degli enti e delle leggi della teoria fisica.

Questo aspetto è particolarmente evidente nelle teorie fisiche moderne. Nella relatività generale, per esempio, lo spaziotempo è formulato come una varietà differenziabile dotata di una struttura metrica lorentziana, che richiameremo alla fine della sezione 3.6; grandezze come la quadrivelocità di una particella sono definite geometricamente all'interno di tale struttura. In casi di questo genere la matematica non interviene soltanto come descrizione quantitativa esterna di enti già definiti indipendentemente, ma contribuisce alla loro stessa caratterizzazione teorica.

2.12 Integrazione: dalle somme al limite

Abbiamo seguito fin qui una catena concettuale precisa:

$$\text{completezza} \rightsquigarrow \text{limite} \rightsquigarrow \text{continuità} \rightsquigarrow \text{derivata}.$$

La derivata descrive un comportamento *locale*: misura la variazione di una funzione in un punto attraverso il limite di rapporti incrementali. L'integrale affronta un problema complementare: descrivere una *accumulazione globale* ottenuta sommando molti contributi piccoli.

NOTA STORICA. *Antecedenti del calcolo integrale.* L'idea di determinare aree e volumi mediante procedimenti di approssimazione è molto più antica del calcolo del XVII secolo. Il *metodo di esaustione*, sviluppato nella matematica greca e portato a risultati particolarmente profondi da ARCHIMEDE, permetteva di determinare aree e volumi confrontando una figura curvilinea con successioni di figure più semplici. Un esempio celebre è la *Quadratura della parabola*, nella quale ARCHIMEDE ottiene l'area di un segmento parabolico attraverso una costruzione che, nel linguaggio moderno, richiama la somma di una serie geometrica [6].

Nel XVII secolo BONAVENTURA CAVALIERI sviluppò il *metodo degli indivisibili*: una figura piana veniva pensata attraverso una famiglia di segmenti paralleli e un solido attraverso una famiglia di sezioni piane. Pur essendo concettualmente diverso dalla definizione moderna di integrale, questo metodo costituì una tappa importante verso il calcolo integrale. Con NEWTON e LEIBNIZ i problemi di quadratura e di accumulazione vennero infine collegati sistematicamente ai problemi di variazione e di tangente; il segno $\int$ introdotto da LEIBNIZ conserva ancora oggi, nella sua forma, il richiamo a una somma. ■

L'esempio geometrico fondamentale è l'area sotto il grafico di una funzione non negativa. Se dividiamo un intervallo in parti piccole, su ciascuna parte possiamo approssimare la regione con un rettangolo. Sommando le aree dei rettangoli otteniamo un'approssimazione dell'area totale; rendendo la suddivisione sempre più fine, queste somme possono convergere a un numero ben determinato.

2.12.1 Suddivisioni e somme di Riemann

2.95 Definizione [suddivisione di un intervallo]. Siano $a, b \in \mathbb{R}$ con $a < b$. Una **suddivisione** dell'intervallo $[a, b]$ è una famiglia finita di punti

$$P: \quad a = x_0 < x_1 < \cdots < x_n = b.$$

Poniamo

$$\Delta x_k := x_k - x_{k-1} \qquad (k = 1, \dots, n)$$

e chiamiamo **ampiezza** o **norma** della suddivisione il numero

$$\|P\| := \max_{1 \le k \le n} \Delta x_k.$$

■

In ogni intervallo $[x_{k-1}, x_k]$ scegliamo un punto

$$\xi_k \in [x_{k-1}, x_k].$$

Se $f : [a, b] \to \mathbb{R}$ è una funzione, il numero

$$S(f; P, \xi) := \sum_{k=1}^{n} f(\xi_k)\, \Delta x_k \tag{2.8}$$

definito in (2.8) si chiama **somma di Riemann**. Quando $f \ge 0$, ogni termine è l'area di un rettangolo di base Δx_k e altezza $f(\xi_k)$.

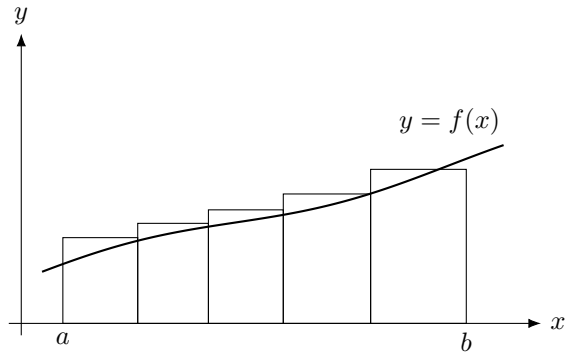


Figure 6: Una somma di Riemann approssima l'accumulazione globale mediante rettangoli. Rendendo la suddivisione più fine, si cerca un valore limite indipendente dalle scelte dei punti ξ_k.

2.96 Definizione [integrale di Riemann]. Siano $a, b \in \mathbb{R}$ con $a < b$. Una funzione limitata $f : [a, b] \to \mathbb{R}$ si dice **Riemann-integrabile** se esiste un numero $I \in \mathbb{R}$ con la seguente proprietà: per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che, per ogni suddivisione P con

$$\|P\| < \delta$$

e per ogni scelta dei punti $\xi_k \in [x_{k-1}, x_k]$, vale

$$|S(f; P, \xi) - I| < \varepsilon.$$

In tal caso scriviamo

$$I = \int_a^b f(x)\, dx.$$

■

La Definizione 2.96 ripropone la stessa idea incontrata più volte nel capitolo: un oggetto viene individuato come limite quando possiamo renderne arbitrariamente piccola la distanza da tutte le approssimazioni sufficientemente fini. In questo caso il parametro che controlla l'approssimazione è la norma $\|P\|$ della suddivisione.

2.97 Osservazione [caratterizzazione mediante somme inferiori e superiori]. Siano $a, b \in \mathbb{R}$ con $a < b$ e sia $f : [a, b] \to \mathbb{R}$ limitata. Per la Riemann-integrabilità di f esiste una formulazione equivalente, spesso chiamata formulazione di Darboux. Fissata una suddivisione

$$P : \quad a = x_0 < x_1 < \cdots < x_n = b,$$

poniamo, per ogni sottointervallo,

$$m_k := \inf_{x \in [x_{k-1}, x_k]} f(x), \qquad M_k := \sup_{x \in [x_{k-1}, x_k]} f(x),$$

e definiamo la **somma inferiore** e la **somma superiore**

$$L(f, P) := \sum_{k=1}^{n} m_k\, \Delta x_k, \qquad U(f, P) := \sum_{k=1}^{n} M_k\, \Delta x_k.$$

Chiamiamo **funzione a gradini** una funzione che, rispetto a un'opportuna suddivisione finita, è costante nell'interno di ciascun sottointervallo; i valori nei punti della suddivisione possono essere assegnati separatamente. Geometricamente, quando $f \geq 0$, le quantità $L(f, P)$ e $U(f, P)$ sono le aree delle funzioni a gradini che assumono rispettivamente i valori m_k e M_k, e stanno quindi sotto e sopra il grafico di f; nel caso generale si tratta di aree con segno.

La funzione limitata $f : [a, b] \to \mathbb{R}$ è Riemann-integrabile se e solo se

$$\boxed{\sup_P L(f, P) = \inf_P U(f, P)}$$

dove il sup e l'inf sono presi su tutte le suddivisioni finite di $[a, b]$. Il valore comune è precisamente

$$\int_a^b f(x)\, dx.$$

Equivalentemente, il *supremo* degli integrali delle funzioni a gradini che stanno sotto f coincide con l'*infimo* degli integrali delle funzioni a gradini che stanno sopra f. Questa formulazione rende particolarmente trasparente l'idea di racchiudere il valore dell'integrale fra approssimazioni inferiori e superiori. ■

2.98 Teorema [Riemann-integrabilità delle funzioni continue]. *Siano $a, b \in \mathbb{R}$ con $a < b$. Ogni funzione continua*

$$f : [a, b] \to \mathbb{R}$$

è Riemann-integrabile.

Non dimostreremo qui questo risultato. La continuità su un intervallo chiuso e limitato garantisce un controllo uniforme delle oscillazioni della funzione; questo permette di mostrare che tutte le somme di Riemann associate a suddivisioni sufficientemente fini sono arbitrariamente vicine fra loro e quindi individuano un unico valore.

È importante però osservare che la continuità è una condizione *sufficiente*, non necessaria: le funzioni continue $f : [a, b] \to \mathbb{R}$ non esauriscono affatto la famiglia delle funzioni Riemann-integrabili definite su $[a, b]$. Per esempio, ogni funzione a gradini $f : [a, b] \to \mathbb{R}$ è Riemann-integrabile; in particolare la funzione

$$g(x) := \begin{cases} 0, & a \le x < c, \\ 1, & c \le x \le b, \end{cases} \qquad a < c < b,$$

ha una discontinuità di salto in c ma è Riemann-integrabile su $[a, b]$, con

$$\int_a^b g(x)\,dx = b - c.$$

Più in generale, anche molte funzioni discontinue $f : [a, b] \to \mathbb{R}$ sono Riemann-integrabili.

2.99 Definizione [funzione monotona]. Siano $a, b \in \mathbb{R}$ con $a \le b$ e sia $f : [a, b] \to \mathbb{R}$. La funzione f si dice **crescente** se

$$x \le y \quad \Longrightarrow \quad f(x) \le f(y) \qquad \text{per ogni } x, y \in [a, b],$$

e **decrescente** se

$$x \le y \quad \Longrightarrow \quad f(x) \ge f(y) \qquad \text{per ogni } x, y \in [a, b].$$

In entrambi i casi f si dice **monotona**. Con questa convenzione, dunque, una funzione costante è sia crescente sia decrescente. Alcuni testi usano invece i termini "crescente" e "decrescente" per indicare le corrispondenti disuguaglianze strette; quando vorremo esprimere questa proprietà useremo esplicitamente le espressioni *strettamente crescente* e *strettamente decrescente*. ■

2.100 Teorema [Riemann-integrabilità delle funzioni monotone]. *Siano $a, b \in \mathbb{R}$ con $a < b$. Ogni funzione monotona*

$$f : [a, b] \to \mathbb{R}$$

è Riemann-integrabile.

Dimostrazione. Supponiamo dapprima che f sia crescente e dividiamo $[a, b]$ in n sottointervalli di uguale ampiezza

$$\Delta x = \frac{b - a}{n}.$$

Su ciascun sottointervallo il minimo di f è assunto all'estremo sinistro e il massimo all'estremo destro. Pertanto, indicando con P_n questa suddivisione,

$$\begin{aligned} U(f, P_n) - L(f, P_n) &= \frac{b - a}{n} \sum_{k=1}^{n} \big(f(x_k) - f(x_{k-1})\big) \\ &= \frac{b - a}{n} \big(f(b) - f(a)\big). \end{aligned}$$

Questa quantità tende a 0 per $n \to \infty$. Per la caratterizzazione mediante somme inferiori e superiori, f è quindi Riemann-integrabile. Se f è decrescente, si applica lo stesso argomento a $-f$. □

2.101 Osservazione [estensione dell'integrale di Riemann a $\mathbb{R}^n$]. La teoria dell'integrale di Riemann non è limitata alle funzioni di una sola variabile. Se

$$Q = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n, \qquad a_j < b_j \quad (j = 1, \dots, n)$$

è un rettangolo n-dimensionale e $f : Q \to \mathbb{R}$ è limitata, si può suddividere Q in rettangoli più piccoli, scegliere in ciascuno di essi un punto campione e formare somme di Riemann nelle quali il valore della funzione viene moltiplicato per il volume del rettangolo corrispondente. Facendo tendere a zero la dimensione massima degli elementi della suddivisione si ottiene, quando il limite esiste e non dipende dalle scelte effettuate, l'integrale multiplo

$$\int_Q f(x)\,dx.$$

Anche la formulazione mediante somme inferiori e superiori si estende nello stesso modo. In particolare, ogni funzione continua $f : Q \to \mathbb{R}$ su un rettangolo chiuso e limitato Q è Riemann-integrabile. Non svilupperemo qui la teoria in più variabili; ci basta osservare che le idee fondamentali sono le stesse del caso di $[a, b] \subseteq \mathbb{R}$. ■

2.102 Esempio [l'integrale della funzione identità]. Poiché $f : [0, 1] \to \mathbb{R}$, $f(x) = x$, è continua, è Riemann-integrabile. Possiamo quindi calcolarne l'integrale facendo tendere a zero la norma di una particolare successione di suddivisioni. Dividiamo $[0, 1]$ in n intervalli della stessa lunghezza e scegliamo come punto campione l'estremo destro:

$$x_k = \frac{k}{n}, \qquad \Delta x_k = \frac{1}{n}.$$

Per $f(x) = x$ la somma di Riemann è

$$S_n = \sum_{k=1}^{n} \frac{k}{n}\frac{1}{n} = \frac{1}{n^2}\sum_{k=1}^{n} k = \frac{n(n+1)}{2n^2} = \frac{1}{2}\left(1 + \frac{1}{n}\right).$$

Poiché $1/n \to 0$,

$$\int_0^1 x\,dx = \frac{1}{2}.$$

■

2.12.2 Proprietà elementari dell'integrale

Per funzioni Riemann-integrabili valgono alcune proprietà fondamentali, che seguono facilmente dalla Definizione 2.96 e dalle corrispondenti proprietà delle somme di Riemann. Anzitutto, siano $a, b \in \mathbb{R}$ con $a < b$. Se $f, g : [a, b] \to \mathbb{R}$ sono Riemann-integrabili e $\alpha, \beta \in \mathbb{R}$, allora vale la **linearità dell'integrale**:

$$\int_a^b (\alpha f(x) + \beta g(x))\,dx = \alpha \int_a^b f(x)\,dx + \beta \int_a^b g(x)\,dx.$$

È utile fissare anche le convenzioni relative agli estremi di integrazione. Poniamo

$$\int_a^a f(x)\,dx := 0$$

e, se $b < a$,

$$\int_a^b f(x)\,dx := -\int_b^a f(x)\,dx.$$

Di conseguenza, ogni volta che entrambi gli integrali sono definiti, **scambiare gli estremi di integrazione cambia il segno**:

$$\boxed{\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx.}$$

Con questa convenzione l'**additività rispetto al dominio di integrazione** assume una forma indipendente dall'ordine dei punti. Se $a, b, c \in \mathbb{R}$ e f è Riemann-integrabile su un intervallo chiuso che li contiene, allora

$$\boxed{\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.}$$

In particolare, quando $a < c < b$, la formula esprime la scomposizione dell'integrale sull'intervallo $[a, b]$ nei due intervalli adiacenti $[a, c]$ e $[c, b]$.

2.103 Proposizione [integrabilità del valore assoluto]. *Siano $a, b \in \mathbb{R}$ con $a < b$ e sia $f : [a, b] \to \mathbb{R}$ Riemann-integrabile. Allora anche $|f| : [a, b] \to \mathbb{R}$ è Riemann-integrabile.*

Non dimostreremo questo fatto: non è difficile ricavarlo confrontando le oscillazioni di f e di $|f|$ nelle somme di Darboux.

Infine, se $a < b$, $f, g : [a, b] \to \mathbb{R}$ sono Riemann-integrabili e

$$f(x) \leq g(x) \qquad \text{per ogni } x \in [a, b],$$

allora vale la **monotonia dell'integrale**:

$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$

In particolare, la Proposizione 2.103 garantisce che l'integrale di $|f|$ è definito. Poiché

$$-|f(x)| \leq f(x) \leq |f(x)| \qquad \text{per ogni } x \in [a, b],$$

dalla monotonia si ottiene la stima

$$\boxed{\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx.}$$

Questa disuguaglianza sarà usata nella dimostrazione del teorema fondamentale del calcolo.

La linearità sarà particolarmente significativa nel Capitolo 3: l'operazione che associa a una funzione il suo integrale rispetta le combinazioni lineari, proprio come accade per la derivazione.

2.12.3 Il teorema fondamentale del calcolo

Derivazione e integrazione sono state introdotte a partire da problemi differenti: la derivata misura una variazione locale, mentre l'integrale descrive un'accumulazione globale. Uno dei risultati centrali dell'analisi matematica mostra che le due operazioni sono strettamente collegate.

NOTA STORICA. *Alle origini del teorema fondamentale.* Il riconoscimento che i problemi di tangente e di quadratura sono inversi l'uno dell'altro precede NEWTON e LEIBNIZ. Risultati molto vicini al teorema fondamentale compaiono nel Seicento, in particolare in TORRICELLI, GREGORY e BARROW; GREGORY ne pubblicò una prima dimostrazione e BARROW ne diede una formulazione geometrica particolarmente importante. Con NEWTON e LEIBNIZ il principio diventa però uno dei cardini del nuovo calcolo differenziale e integrale: derivazione e integrazione vengono riconosciute sistematicamente come operazioni inverse [6]. La formula oggi detta *formula di* NEWTON–LEIBNIZ verrà dedotta qui sotto come conseguenza del teorema fondamentale del calcolo. ■

2.104 Teorema [teorema fondamentale del calcolo]. *Siano $a, b \in \mathbb{R}$ con $a < b$ e sia $f : [a, b] \to \mathbb{R}$ continua. Definiamo*

$$F(x) := \int_a^x f(t)\,dt.$$

Allora F è derivabile in ogni punto interno $x \in (a, b)$ e

$$\boxed{F'(x) = f(x).}$$

Dimostrazione. Fissiamo $x \in (a, b)$ e poniamo

$$\rho := \min\{x - a, b - x\} > 0.$$

Per $0 < h < \rho$ si ha $x + h \in [a, b]$ e

$$\frac{F(x + h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f(t)\,dt.$$

Sottraendo $f(x)$ otteniamo

$$\frac{F(x + h) - F(x)}{h} - f(x) = \frac{1}{h} \int_x^{x+h} \big(f(t) - f(x)\big)\,dt.$$

Per la continuità di f in x, dato $\varepsilon > 0$ esiste $\delta > 0$, che possiamo scegliere con $\delta < \rho$, tale che

$$|f(t) - f(x)| < \varepsilon$$

per ogni $t \in [a, b]$ con $|t - x| < \delta$. In questo modo, se $|h| < \delta$, anche $x + h \in [a, b]$. Se $0 < h < \delta$, usando la stima precedente otteniamo

$$\begin{aligned}\left|\frac{F(x+h) - F(x)}{h} - f(x)\right| &= \frac{1}{h}\left|\int_x^{x+h} \big(f(t) - f(x)\big)\,dt\right| \\ &\le \frac{1}{h}\int_x^{x+h} |f(t) - f(x)|\,dt \\ &< \frac{1}{h}\int_x^{x+h} \varepsilon\,dt = \varepsilon.\end{aligned}$$

Per $h < 0$ poniamo $k := -h > 0$. Allora

$$\frac{F(x-k) - F(x)}{-k} = \frac{1}{k}\int_{x-k}^{x} f(t)\,dt.$$

Sottraendo $f(x)$ e usando ancora la stima precedente otteniamo

$$\begin{aligned}\left|\frac{F(x-k) - F(x)}{-k} - f(x)\right| &= \frac{1}{k}\left|\int_{x-k}^{x} \big(f(t) - f(x)\big)\,dt\right| \\ &\le \frac{1}{k}\int_{x-k}^{x} |f(t) - f(x)|\,dt.\end{aligned}$$

Se $0 < k < \delta$, per ogni $t \in [x - k, x]$ vale $|t - x| < \delta$ e dunque $|f(t) - f(x)| < \varepsilon$. Pertanto

$$\left|\frac{F(x-k) - F(x)}{-k} - f(x)\right| < \frac{1}{k}\int_{x-k}^{x} \varepsilon\,dt = \varepsilon.$$

Abbiamo quindi ottenuto la stessa stima per incrementi positivi e negativi. Ne segue

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = f(x),$$

cioè $F'(x) = f(x)$. □

La formula esprime una relazione profonda: accumulare i valori di una funzione e poi misurare la variazione istantanea dell'accumulazione restituisce la funzione di partenza.

2.105 Proposizione [formula di Newton–Leibniz]. *Siano $u, v \in \mathbb{R}$ con $u \le v$, sia $I = [u, v]$, siano $a, b \in I$ e sia $f : I \to \mathbb{R}$ continua. Sia inoltre $G : I \to \mathbb{R}$ una **primitiva** di f, cioè una funzione continua su I, derivabile in* $\mathrm{int}(I)$ *e tale che*

$$G'(x) = f(x) \qquad \text{per ogni } x \in \mathrm{int}(I).$$

Allora

$$\boxed{\int_a^b f(x)\,dx = G(b) - G(a).} \tag{2.9}$$

La formula vale qualunque sia l'ordine di a e b. Inoltre, se $G_1, G_2 : I \to \mathbb{R}$ sono due primitive della stessa funzione f, allora esiste una costante $C \in \mathbb{R}$ tale che

$$G_1(x) - G_2(x) = C \qquad \text{per ogni } x \in I.$$

In altre parole, due primitive della stessa funzione differiscono per una costante.

Dimostrazione. Se $a = b$, la formula segue immediatamente da

$$\int_a^a f(x)\,dx = 0 = G(a) - G(a).$$

Supponiamo quindi $a < b$ e definiamo

$$F(x) := \int_a^x f(t)\,dt.$$

La funzione F è continua su $[a, b]$. Infatti, poiché f è continua, per il Teorema 2.72 di Weierstrass esiste $M \geq 0$ tale che $|f(t)| \leq M$ per ogni $t \in [a, b]$. Se $x \leq y$, usando l'additività rispetto al dominio e la stima del modulo dell'integrale otteniamo

$$|F(y) - F(x)| = \left| \int_x^y f(t)\,dt \right| \leq M(y - x).$$

Scambiando x e y si ottiene in generale

$$|F(y) - F(x)| \leq M|y - x|,$$

che prova la continuità di F. Per il Teorema 2.104,

$$F'(x) = f(x) = G'(x).$$

Quindi

$$(G - F)'(x) = 0 \qquad \text{per ogni } x \in (a, b).$$

Poiché $G - F$ è continua su $[a, b]$, è derivabile su (a, b) e ha derivata nulla in ogni punto di (a, b), il Corollario 2.92 implica che $G - F$ è costante su $[a, b]$. Poiché

$$F(a) = \int_a^a f(t)\,dt = 0,$$

si ottiene

$$G(x) - F(x) = G(a),$$

cioè

$$F(x) = G(x) - G(a).$$

Ponendo $x = b$ segue

$$\int_a^b f(x)\,dx = G(b) - G(a).$$

Se invece $b < a$, usando la proprietà di inversione degli estremi della Sottosezione 2.12.2,

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx = -\big(G(a) - G(b)\big) = G(b) - G(a).$$

La formula vale quindi in entrambi i casi.

Resta da verificare l'ultima affermazione. Siano G_1 e G_2 due primitive di f su I e poniamo

$$H := G_1 - G_2.$$

La funzione H è continua su I, derivabile nei punti interni di I e, in ogni punto interno,

$$H' = G_1' - G_2' = f - f = 0.$$

Se $u = v$, l'intervallo I contiene un solo punto e H è evidentemente costante. Se $u < v$, il Corollario 2.92, applicato a H su $[u, v]$, implica ancora che H è costante su I. Esiste quindi $C \in \mathbb{R}$ tale che

$$G_1(x) - G_2(x) = C \qquad \text{per ogni } x \in I,$$

come volevamo dimostrare. □

2.106 Osservazione [primitive e costanti]. Siano $u, v \in \mathbb{R}$ con $u \leq v$, sia $I = [u, v]$, sia $f : I \to \mathbb{R}$ e sia $G : I \to \mathbb{R}$ una primitiva di f, nel senso della Proposizione 2.105. Allora, per ogni $c \in \mathbb{R}$, anche la funzione $G + c$ è una primitiva di f. Infatti $G + c$ è continua su I, è derivabile in $\mathrm{int}(I)$ e

$$(G + c)'(x) = G'(x) = f(x) \qquad \text{per ogni } x \in \mathrm{int}(I).$$

Insieme all'ultima affermazione della Proposizione 2.105, questo mostra che, fissata una primitiva G, tutte e sole le primitive di f su I sono le funzioni $G + c$, con $c \in \mathbb{R}$. ■

La formula (2.9) trasforma il problema di calcolare un integrale nel problema di trovare una primitiva della funzione integranda. Il valore $G(b) - G(a)$ non dipende dalla primitiva scelta, perché l'aggiunta di una costante si cancella nella differenza.

2.12.4 Esercizi risolti sull'integrazione

2.107 Esercizio.

1. *Una funzione costante.* Siano $a, b \in \mathbb{R}$ con $a < b$ e sia $c \in \mathbb{R}$ costante. Calcolare mediante somme di Riemann
$$\int_a^b c\,dx.$$
Soluzione. Per ogni suddivisione e ogni scelta dei punti ξ_k,
$$S(f; P, \xi) = \sum_{k=1}^{n} c\,\Delta x_k = c\sum_{k=1}^{n}(x_k - x_{k-1}) = c(b-a).$$
La somma non dipende nemmeno dalla suddivisione. Pertanto
$$\int_a^b c\,dx = c(b-a).$$

2. *Usare il teorema fondamentale.* Calcolare
$$\int_0^1 2x\,dx.$$
Soluzione. Poiché $G(x) = x^2$ soddisfa $G'(x) = 2x$, la Proposizione 2.105 (formula di Newton–Leibniz) dà
$$\int_0^1 2x\,dx = G(1) - G(0) = 1.$$

3. *Un polinomio di grado superiore.* Calcolare, con la stessa procedura dell'Esercizio 2.107.2,
$$\int_0^1 \left(4x^3 - 3x^2 + 2x + 1\right) dx.$$
Soluzione. Cerchiamo una primitiva del polinomio. Integrando termine per termine otteniamo
$$G(x) = x^4 - x^3 + x^2 + x,$$
poiché
$$G'(x) = 4x^3 - 3x^2 + 2x + 1.$$
La Proposizione 2.105 dà quindi
$$\begin{aligned}\int_0^1 \left(4x^3 - 3x^2 + 2x + 1\right) dx &= G(1) - G(0)\\ &= (1 - 1 + 1 + 1) - 0\\ &= 2.\end{aligned}$$

4. *Linearità dell'integrale.* Siano $a, b \in \mathbb{R}$ con $a < b$ e siano $f, g : [a, b] \to \mathbb{R}$ Riemann-integrabili. Supponiamo
$$\int_a^b f(x)\,dx = 2, \qquad \int_a^b g(x)\,dx = -1.$$
Calcolare
$$\int_a^b \big(3f(x) - 2g(x)\big)\,dx.$$
Soluzione. Per la linearità,
$$\int_a^b (3f - 2g) = 3\int_a^b f - 2\int_a^b g = 3 \cdot 2 - 2(-1) = 8.$$
Questo esercizio mostra la stessa forma algebrica già incontrata per il limite e per la derivazione.

2.13 Un primo sguardo alla teoria della misura

L'integrale di Riemann è costruito dividendo un intervallo in sottointervalli e usando la loro lunghezza. Questo suggerisce una domanda più generale: *che cosa significa assegnare in modo coerente una grandezza a un insieme?*

Per $a, b \in \mathbb{R}$ con $a \leq b$, per un intervallo della retta ci aspettiamo naturalmente

$$\text{lunghezza}([a, b]) = b - a.$$

Se due insiemi disgiunti vengono uniti, ci aspettiamo che le loro grandezze si sommino. Nell'analisi matematica moderna è però essenziale poter considerare non soltanto un numero finito di insiemi, ma anche successioni numerabili di insiemi. Questo conduce alla nozione di σ-additività.

2.13.1 Perché non considerare semplicemente tutti i sottoinsiemi?

Prima di definire una misura occorre specificare su quali sottoinsiemi essa sia definita. La collezione degli insiemi ammessi deve essere stabile rispetto alle operazioni insiemistiche fondamentali e anche rispetto a unioni numerabili.

2.108 Definizione [σ-algebra]. Sia X un insieme. Una **σ-algebra** $\mathcal{A}$ su X è una famiglia di sottoinsiemi di X tale che:

(i) $X \in \mathcal{A}$;

(ii) se $A \in \mathcal{A}$, allora
$$X \setminus A \in \mathcal{A};$$

(iii) se
$$A_1, A_2, A_3, \ldots \in \mathcal{A},$$
allora
$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.$$

La coppia $(X, \mathcal{A})$ si chiama **spazio misurabile**, e gli elementi di $\mathcal{A}$ si chiamano **insiemi misurabili**. ■

Dalla Definizione 2.108 segue che $\varnothing \in \mathcal{A}$ e che $\mathcal{A}$ è chiusa anche rispetto alle intersezioni numerabili, per le leggi di De Morgan:

$$\bigcap_{n=1}^{\infty} A_n = X \setminus \bigcup_{n=1}^{\infty} (X \setminus A_n).$$

La lettera greca σ richiama precisamente la presenza di operazioni *numerabili*, cioè indicizzate dai numeri naturali.

2.109 Esempio. Per ogni insieme X, la famiglia

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

di tutti i sottoinsiemi di X è una σ-algebra. Anche

$$\{\varnothing, X\}$$

è una σ-algebra, la più piccola possibile. ■

2.110 Osservazione [intersezioni e σ-algebre generate]. Sia X un insieme. Se $\{\mathcal{A}_i\}_{i \in I}$ è una famiglia di σ-algebre su X, allora

$$\bigcap_{i \in I} \mathcal{A}_i$$

è ancora una σ-algebra su X: infatti X appartiene a ciascuna $\mathcal{A}_i$, e le proprietà di chiusura rispetto al complemento e alle unioni numerabili vengono conservate passando all'intersezione.

Questa osservazione permette di associare una σ-algebra a una famiglia arbitraria $\mathcal{E} \subseteq \mathcal{P}(X)$. Poiché $\mathcal{P}(X)$ è una σ-algebra che contiene $\mathcal{E}$, esistono σ-algebre su X contenenti $\mathcal{E}$; la loro intersezione è la più piccola fra esse. Essa si chiama **σ-algebra generata da $\mathcal{E}$** e si indica con

$$\sigma(\mathcal{E}) := \bigcap \{\mathcal{A} \subseteq \mathcal{P}(X) \mid \mathcal{A} \text{ è una } \sigma\text{-algebra su } X \text{ e } \mathcal{E} \subseteq \mathcal{A}\}.$$

■

2.111 Definizione [σ-algebra di Borel]. Sia $(X, \mathcal{T})$ uno spazio topologico. La **σ-algebra di Borel** di X è la σ-algebra generata dagli aperti:

$$\mathcal{B}(X) := \sigma(\mathcal{T}).$$

I suoi elementi si chiamano **insiemi boreliani**.

Nel caso di $\mathbb{R}^n$ useremo la topologia euclidea già introdotta nella Definizione 2.52. Ricordiamo che i suoi aperti possono essere descritti come unioni arbitrarie di prodotti di intervalli aperti, oppure, equivalentemente, come unioni arbitrarie di palle aperte. La corrispondente σ-algebra di Borel sarà indicata con

$$\mathcal{B}(\mathbb{R}^n).$$

In particolare, $\mathcal{B}(\mathbb{R}^n)$ è la più piccola σ-algebra che contiene tutti gli aperti della topologia euclidea di $\mathbb{R}^n$.
■

2.13.2 Misure e σ-additività

2.112 Definizione [serie a termini non negativi nei reali estesi]. Indichiamo con

$$[0, +\infty] := [0, +\infty) \cup \{+\infty\}$$

l'insieme dei numeri reali non negativi al quale è stato aggiunto il simbolo $+\infty$. Sia $(a_n)_{n\geq 1}$ una successione con valori in $[0, +\infty]$.

Se almeno un termine vale $+\infty$, poniamo

$$\sum_{n=1}^{\infty} a_n := +\infty.$$

Supponiamo invece che tutti gli a_n siano reali e poniamo

$$s_N := \sum_{n=1}^{N} a_n.$$

La successione delle somme parziali è crescente. Se è limitata superiormente, definiamo

$$\sum_{n=1}^{\infty} a_n := \sup_{N\geq 1} s_N \in [0, +\infty).$$

Se non è limitata superiormente, diciamo che la serie **diverge a** $+\infty$ e poniamo

$$\sum_{n=1}^{\infty} a_n := +\infty.$$

Equivalentemente, in quest'ultimo caso, per ogni $M \in \mathbb{R}$ esiste N tale che $s_N > M$. ■

2.113 Definizione [misura σ-additiva]. Sia $(X, \mathcal{A})$ uno spazio misurabile. Con la notazione introdotta nella Definizione 2.112, una **misura** è una funzione

$$\mu : \mathcal{A} \longrightarrow [0, +\infty]$$

tale che

$$\mu(\varnothing) = 0$$

e tale che, per ogni successione $A_1, A_2, \ldots$ di insiemi misurabili a due a due disgiunti,

$$A_i \cap A_j = \varnothing \qquad (i \neq j),$$

valga

$$\boxed{\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).} \tag{2.10}$$

L'identità (2.10) esprime la proprietà che si chiama **σ-additività** o additività numerabile. ■

La σ-additività implica immediatamente l'additività finita. Se A e B sono disgiunti, basta porre

$$A_1 = A, \qquad A_2 = B, \qquad A_n = \varnothing \quad (n \geq 3)$$

per ottenere

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

2.114 Esempio [misura di conteggio]. Sia X un insieme e prendiamo $\mathcal{A} = \mathcal{P}(X)$. Definiamo

$$\mu(A) := \begin{cases} |A|, & A \text{ finito}, \\ +\infty, & A \text{ infinito}. \end{cases}$$

Questa è la **misura di conteggio**. Se gli insiemi A_n sono a due a due disgiunti, il numero degli elementi della loro unione è precisamente la somma dei numeri degli elementi dei singoli insiemi, intendendo il valore $+\infty$ quando l'unione è infinita. Pertanto μ è σ-additiva. ■

2.115 Esempio [misura di Lebesgue in $\mathbb{R}^n$]. Per ogni $n \geq 1$ esiste una σ-algebra $\mathcal{L}^n$ di sottoinsiemi di $\mathbb{R}^n$, detta **σ-algebra di Lebesgue**, e una misura

$$\lambda_n : \mathcal{L}^n \longrightarrow [0, +\infty]$$

detta **misura di Lebesgue**. Essa estende le usuali nozioni geometriche di lunghezza, area e volume. Per esempio, per un rettangolo n-dimensionale

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n], \qquad a_j \leq b_j \quad (j = 1, \ldots, n)$$

si ha

$$\lambda_n(R) = \prod_{j=1}^{n} (b_j - a_j).$$

La σ-algebra $\mathcal{L}^n$ contiene la σ-algebra boreliana della topologia euclidea,

$$\mathcal{B}(\mathbb{R}^n) \subseteq \mathcal{L}^n,$$

e contiene inoltre ogni sottoinsieme di un insieme boreliano di misura di Lebesgue nulla. Più precisamente, $\mathcal{L}^n$ è il *completamento* di $\mathcal{B}(\mathbb{R}^n)$ rispetto alla misura di Lebesgue: in questo modo si aggiungono ai boreliani tutti i sottoinsiemi degli insiemi nulli e gli insiemi ottenuti da essi mediante le operazioni di σ-algebra. In generale esistono quindi insiemi misurabili secondo Lebesgue che non sono boreliani.

La costruzione rigorosa di $\mathcal{L}^n$ e di λ_n va oltre gli obiettivi di queste dispense. Nel caso $n = 1$ ritroviamo la lunghezza sulla retta e, in particolare, per $a \leq b$,

$$\lambda_1([a, b]) = b - a.$$

■

2.116 Osservazione.

(a) Una misura μ tale che

$$\mu(X) = 1$$

si chiama **misura di probabilità**. La teoria moderna della probabilità può quindi essere formulata nel linguaggio della teoria della misura: gli eventi sono insiemi misurabili e la loro probabilità è una misura σ-additiva.

(b) Più in generale, in $\mathbb{R}^n$ non si richiede che la misura di Lebesgue sia definita su ogni elemento di $\mathcal{P}(\mathbb{R}^n)$. In presenza di principi di scelta sufficientemente forti esistono sottoinsiemi di $\mathbb{R}^n$ che non sono misurabili secondo Lebesgue. Per questo λ_n è definita sulla particolare σ-algebra $\mathcal{L}^n$, che contiene $\mathcal{B}(\mathbb{R}^n)$ ma non coincide con l'intero insieme delle parti $\mathcal{P}(\mathbb{R}^n)$. Non svilupperemo qui questo tema, che si collega alle questioni fondazionali incontrate nel Capitolo 1.

■

2.13.3 Dalla misura all'integrale: soltanto l'idea

Prima di procedere introduciamo una notazione molto utile.

2.117 Definizione [funzione indicatrice]. Siano X un insieme e $A \subseteq X$. La **funzione indicatrice** di A è la funzione

$$\mathbf{1}_A : X \longrightarrow \mathbb{R}, \qquad \mathbf{1}_A(x) := \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Essa registra quindi l'appartenenza di un punto all'insieme A mediante i valori 1 e 0. ■

La teoria della misura conduce a una nozione di integrale più generale di quella di Riemann. L'idea di base può essere già vista per funzioni molto semplici. Usando la funzione indicatrice della Definizione 2.117, consideriamo

$$s : X \to \mathbb{R}, \qquad s(x) = \sum_{k=1}^{n} a_k \, \mathbf{1}_{A_k}(x),$$

dove $a_k \geq 0$ e gli insiemi A_k sono misurabili e a due a due disgiunti. Si pone

$$\int_X s \, d\mu := \sum_{k=1}^{n} a_k \, \mu(A_k).$$

In altre parole, il valore della funzione su ciascuna regione viene moltiplicato per la misura della regione e i contributi vengono sommati. A partire da questa idea, mediante opportuni passaggi al limite, si costruisce l'**integrale di Lebesgue**.

Non svilupperemo ulteriormente questa teoria. Per i nostri scopi è sufficiente comprendere il cambiamento di prospettiva:

Riemann: suddividere il dominio in intervalli
↓
Lebesgue: misurare insiemi e sommare per livelli.

2.13.4 Esercizi risolti su σ-algebre e misure

2.118 Esercizio.

1. *Una piccola σ-algebra.* Sia
$$X = \{1, 2, 3\}$$
e consideriamo
$$\mathcal{A} = \{\varnothing, \{1\}, \{2, 3\}, X\}.$$
Verificare che $\mathcal{A}$ è una σ-algebra su X.
Soluzione. L'insieme X appartiene a $\mathcal{A}$. I complementi rispetto a X sono
$$X \setminus \varnothing = X, \qquad X \setminus \{1\} = \{2, 3\},$$
$$X \setminus \{2, 3\} = \{1\}, \qquad X \setminus X = \varnothing,$$
e appartengono tutti a $\mathcal{A}$. Poiché $\mathcal{A}$ contiene soltanto quattro insiemi, ogni unione numerabile dei suoi elementi coincide con uno di essi. Dunque $\mathcal{A}$ è una σ-algebra.
2. *σ-additività della misura di conteggio.* In $X = \mathbb{N}$ consideriamo gli insiemi
$$A_n = \{n\}.$$
Verificare la σ-additività della misura di conteggio per questa successione.
Soluzione. Gli insiemi A_n sono a due a due disgiunti e
$$\bigcup_{n \in \mathbb{N}} A_n = \mathbb{N}.$$
Pertanto
$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n \right) = \mu(\mathbb{N}) = +\infty.$$

D'altra parte

$$\sum_{n\in\mathbb{N}} \mu(A_n) = \sum_{n\in\mathbb{N}} 1 = +\infty.$$

I due membri coincidono.

3. *Dalla σ-additività all'additività finita.* Siano $A, B \in \mathcal{A}$ disgiunti e sia μ una misura. Dedurre dalla Definizione 2.113 che

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

Soluzione. Applichiamo la σ-additività alla successione

$$A_1 = A, \qquad A_2 = B, \qquad A_n = \varnothing \quad (n \geq 3).$$

Poiché $\mu(\varnothing) = 0$,

$$\mu(A \cup B) = \mu(A) + \mu(B) + 0 + 0 + \cdots = \mu(A) + \mu(B).$$

2.13.5 Il paradosso di Banach–Tarski

Il *paradosso di Banach–Tarski*, già menzionato nel Capitolo 1 a proposito dell'assioma della scelta, è un teorema che mostra in modo particolarmente netto perché non si possa pretendere di assegnare un volume a *ogni* sottoinsieme dello spazio conservando tutte le proprietà naturali della misura [14].

Per rendere preciso il punto relativo al volume, usiamo la distanza euclidea su $\mathbb{R}^n$ introdotta nell'Osservazione 2.66 del Capitolo 2. Per $n \geq 1$, una **isometria euclidea** di $\mathbb{R}^n$ è una funzione biiettiva

$$g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

tale che

$$\|g(x) - g(y)\| = \|x - y\| \qquad \text{per ogni } x, y \in \mathbb{R}^n. \tag{2.11}$$

La condizione (2.11) esprime precisamente il fatto che un'isometria conserva tutte le distanze. Traslazioni, rotazioni e riflessioni sono esempi di isometrie. Nel seguito di questa sottosezione consideriamo il caso $n = 3$. La misura di Lebesgue tridimensionale, che indichiamo con λ_3, è *invariante sotto le isometrie*: se $A \subseteq \mathbb{R}^3$ è misurabile secondo Lebesgue e g è un'isometria di $\mathbb{R}^3$, allora anche $g(A)$ è misurabile e

$$\lambda_3(g(A)) = \lambda_3(A). \tag{2.12}$$

Nel Capitolo 3 vedremo che le isometrie di $\mathbb{R}^3$, con l'operazione di composizione, formano un gruppo. La formula (2.12) dice dunque che la misura di Lebesgue è invariante rispetto a tutte le trasformazioni appartenenti a questo gruppo.

Nella formulazione classica del *paradosso di Banach–Tarski* si considera una *palla solida* $B \subseteq \mathbb{R}^3$ (non soltanto la sua superficie). Il teorema afferma che B può essere decomposta in un numero finito di sottoinsiemi a due a due disgiunti che, mediante opportune rotazioni e traslazioni, possono essere ricomposti in due palle, ciascuna congruente alla palla iniziale B. Le trasformazioni usate nella ricomposizione sono quindi particolari isometrie di $\mathbb{R}^3$.

A prima vista il risultato sembra contraddire la conservazione del volume. Il punto essenziale è però che la decomposizione coinvolge necessariamente *insiemi non misurabili secondo Lebesgue*: non è possibile che tutti i pezzi della decomposizione siano misurabili. Infatti, se lo fossero, l'additività finita della misura e l'invarianza (2.12) imporrebbero, dopo la ricomposizione, la relazione

$$\lambda_3(B) = 2\,\lambda_3(B). \tag{2.13}$$

Poiché una palla ordinaria ha misura di Lebesgue finita e strettamente positiva, la relazione (2.13) è impossibile. Il teorema non produce quindi una contraddizione nella teoria della misura: mostra piuttosto che i pezzi utilizzati non appartengono al dominio della misura di Lebesgue.

Nella dimostrazione usuale interviene inoltre l'*assioma della scelta*: esso permette di effettuare scelte non costruttive necessarie alla costruzione dei pezzi. L'apparente paradosso nasce dunque dall'incontro fra due fatti già discussi in queste dispense: da un lato l'assioma della scelta consente l'esistenza di insiemi molto lontani dall'intuizione geometrica ordinaria; dall'altro la misura di Lebesgue non può essere estesa a tutti i sottoinsiemi di $\mathbb{R}^3$ mantenendo contemporaneamente le proprietà naturali di additività e di invarianza rispetto a tutte le trasformazioni del gruppo delle isometrie euclidee. Non si tratta perciò di una procedura fisica mediante la quale una sfera materiale possa essere duplicata, ma di un risultato matematico sulla struttura degli insiemi e sulle possibilità di definirne coerentemente il volume.

2.14 Conclusione del capitolo

Il percorso seguito in questo capitolo è partito dalla costruzione fondazionale dei sistemi numerici e ha condotto progressivamente alle prime nozioni dell'analisi matematica:

$$\begin{aligned}\mathbb{N} \rightsquigarrow \mathbb{Z} \rightsquigarrow \mathbb{Q} \rightsquigarrow \mathbb{R} \rightsquigarrow{}& \text{ completezza}\\ \rightsquigarrow{}& \text{ limiti} \rightsquigarrow \text{ continuità} \rightsquigarrow \text{ derivate}\\ \rightsquigarrow{}& \text{ equazioni differenziali} \rightsquigarrow \text{ integrali} \rightsquigarrow \text{ misure.}\end{aligned}$$

Un tratto comune è apparso più volte: non abbiamo studiato soltanto oggetti, ma anche *operazioni* e proprietà delle operazioni. Somma, prodotto, combinazioni lineari, passaggio al limite, derivazione e integrazione mostrano regolarità formali che possono essere separate dalla particolare natura dei numeri o delle funzioni coinvolte.

Nel Capitolo 3 cambieremo quindi prospettiva. Partiremo proprio da proprietà già incontrate concretamente in queste costruzioni e le isoleremo in forma astratta, introducendo strutture come monoidi, gruppi, anelli e campi, anche nel caso non commutativo, per arrivare poi agli spazi vettoriali e alle algebre associative.

2.15 Letture consigliate e bibliografia

Per una lettura moderna e accessibile dell'analisi matematica su $\mathbb{R}$ si veda Abbott [3]; Rudin [4] offre una trattazione classica e più avanzata. Per la costruzione dei numeri reali mediante sezioni si può consultare anche il testo di Dedekind [2]. Per un'introduzione moderna alla misura e all'integrazione di Lebesgue, oltre il livello richiesto in queste dispense, si veda Axler [5]. Per gli accenni storici sul calcolo si veda Edwards [6]; per il passaggio alla cosiddetta meccanica analitica, cioè alla formulazione matematica della fisica meccanica mediante metodi analitici, si vedano anche Euler [7] e la prima edizione dell'opera di Lagrange [8]. Per una trattazione moderna della fisica meccanica e della meccanica analitica si veda Moretti [9]. Per l'analisi non standard, Robinson [10]. Per una discussione moderna dei paradossi di Zenone, con particolare attenzione alla distinzione fra il problema matematico delle serie convergenti e le ulteriori questioni filosofiche, si vedano Fano [11] e Huggett [12]; per un contributo recente sul problema dei supertasks nel caso di Achille si veda Garrett [13].

Bibliografia del Capitolo 2

[1] Edward R. Scheinerman, *From Counting to Continuum: What Are Real Numbers, Really?*, Cambridge University Press, 2024

[2] R. Dedekind, *Essays on the Theory of Numbers*, trad. W. W. Beman, Dover Publications, New York, 1963. L'opera comprende *Continuity and Irrational Numbers*, pubblicato originariamente nel 1872.

[3] S. Abbott, *Understanding Analysis*, 2a ed., Springer, New York, 2015.

[4] W. Rudin, *Principles of Mathematical Analysis*, 3a ed., McGraw–Hill, New York, 1976.

[5] S. Axler, *Measure, Integration & Real Analysis*, Springer, Cham, 2020.

[6] C. H. Edwards Jr., *The Historical Development of the Calculus*, Springer, New York, 1979.

[7] L. Euler, *Mechanica sive motus scientia analytice exposita*, 2 voll., Ex Typographia Academiae Scientiarum, Petropoli, 1736.

[8] J.-L. Lagrange, *Méchanique analitique*, Vve Desaint, Paris, 1788; si veda in particolare l'*Avertissement*, p. VI.

[9] V. Moretti, *Meccanica Analitica: Meccanica Classica, Meccanica Lagrangiana, Hamiltoniana, Teoria della Stabilità, Relatività Speciale*, 2a ed., Springer, Cham, 2024.

[10] A. Robinson, *Non-standard Analysis*, North-Holland, Amsterdam, 1966.

[11] V. Fano, "Il paradosso di Zenone", *APhEx*, n. 2, EUT Edizioni Università di Trieste, 2010.

[12] N. Huggett, *Everywhere and Everywhen: Adventures in Physics and Philosophy*, Oxford University Press, New York, 2010; in particolare capp. 2–3.

[13] Z. Garrett, “Achilles’ To Do List”, *Philosophies*, 9(4), 104, 2024, doi:10.3390/philosophies9040104.

[14] S. Banach e A. Tarski, “Sur la décomposition des ensembles de points en parties respectivement congruentes”, *Fundamenta Mathematicae*, 6, 244–277, 1924.

CAPITOLO 3

Strutture algebriche fondamentali

3.1 Introduzione: dalle operazioni sui numeri alle strutture algebriche

Nei capitoli precedenti abbiamo incontrato molte operazioni senza considerarle ancora come oggetti di studio autonomi. Sui numeri abbiamo usato somma e prodotto; sulle funzioni abbiamo sommato e moltiplicato valori; in $\mathbb{R}^n$ abbiamo usato n-ple di numeri reali e la distanza euclidea. Più avanti in questo capitolo introdurremo esplicitamente la somma di n-ple e la moltiplicazione per scalari. Con le matrici comparirà inoltre un prodotto che in generale non è commutativo.

Il punto di vista dell'**algebra astratta** consiste nel separare alcune proprietà delle operazioni dalla natura particolare degli oggetti sui quali esse agiscono. Per esempio, l'identità

$$(a+b)+c = a+(b+c)$$

vale in $\mathbb{Z}$, in $\mathbb{Q}$ e in $\mathbb{R}$. La proprietà importante, in questo caso, non è il fatto che a, b, c siano numeri di un tipo particolare, ma il modo in cui la somma si comporta quando si combinano tre elementi.

Una **struttura algebrica** è, in questo capitolo, un insieme dotato di una o più *operazioni* (che di fatto sono sempre funzioni) che soddisfano determinate proprietà. Procederemo sempre dal concreto all'astratto. Gli insiemi numerici già costruiti saranno i primi esempi delle strutture commutative; solo dopo introdurremo esempi nei quali la commutatività fallisce.

3.1 Definizione [morfismi e isomorfismi di strutture]. Siano date due strutture matematiche dello stesso tipo.

- Un **morfismo** dalla prima struttura alla seconda è una funzione fra i rispettivi insiemi sottostanti che conserva le operazioni e le relazioni che definiscono la struttura. Il significato concreto di "conserva" dipende quindi dal tipo di struttura considerata e verrà precisato caso per caso.
- Un **isomorfismo** è un morfismo biiettivo il cui inverso è ancora un morfismo. Due strutture fra le quali esiste un isomorfismo si dicono **isomorfe**.

■

Due strutture isomorfe possono essere costituite da elementi di natura molto diversa, ma dal punto di vista della struttura considerata hanno lo stesso comportamento: un isomorfismo permette di tradurre elementi, operazioni e relazioni dell'una negli elementi, nelle operazioni e nelle relazioni dell'altra senza perdere informazione. Nelle strutture algebriche che incontreremo, l'isomorfismo sarà dunque una particolare applicazione che preserva le operazioni e possiede un inverso che le preserva a sua volta.

La catena concettuale principale sarà

$$\textit{operazione associativa con unità} \rightsquigarrow \textit{monoide} \rightsquigarrow \textit{gruppo} \rightsquigarrow \textit{anello} \rightsquigarrow \textit{campo}.$$

A questa catena affiancheremo gli *spazi vettoriali reali e complessi* e gli *spazi affini*. Infine vedremo che un'*algebra associativa reale o complessa* combina due tipi di struttura: è contemporaneamente uno spazio vettoriale sul campo considerato e un insieme dotato di un prodotto associativo compatibile con la moltiplicazione per scalari.

3.2 Osservazione. Il termine "algebra" è usato in matematica in più sensi. Nel linguaggio scolastico indica spesso il calcolo con lettere e polinomi. In questo capitolo parleremo invece di *strutture algebriche*: insiemi dotati di operazioni che soddisfano assiomi espliciti. Alla fine del capitolo introdurremo anche una struttura specifica chiamata precisamente *algebra associativa*. ■

3.2 Operazioni associative e monoidi

3.2.1 Operazioni binarie

Partiamo da una nozione già implicitamente usata molte volte. Se A è un insieme, un'operazione che prende due elementi di A e restituisce ancora un elemento di A è una funzione

$$A \times A \to A.$$

3.3 Definizione [operazione binaria]. Una **operazione binaria** su un insieme A è una funzione

$$\star : A \times A \to A.$$

Per $a, b \in A$ indichiamo normalmente il valore $\star(a, b)$ con

$$a \star b.$$

■

Il fatto che il codominio sia ancora A significa che, applicando l'operazione a due elementi di A, non si esce dall'insieme. Per esempio,

$$+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$$

è un'operazione binaria. La sottrazione è anch'essa un'operazione binaria su $\mathbb{Z}$, ma non su $\mathbb{N}$, perché per esempio $2 - 5 \notin \mathbb{N}$.

3.4 Definizione [associatività, commutatività, elemento neutro]. Sia A un insieme e sia $\star : A \times A \to A$ un'operazione binaria.

- L'operazione è **associativa** se, per ogni $a, b, c \in A$,

$$(a \star b) \star c = a \star (b \star c).$$

- L'operazione è **commutativa** se, per ogni $a, b \in A$,

$$a \star b = b \star a.$$

- Un elemento $e \in A$ è un **elemento neutro** per $\star$ se, per ogni $a \in A$,

$$e \star a = a \star e = a.$$

■

Associatività e commutatività sono proprietà differenti. L'associatività riguarda il modo di inserire le parentesi quando compaiono tre o più elementi; la commutatività riguarda invece la possibilità di scambiare l'ordine di due elementi.

3.5 Esempio [somma e sottrazione]. La somma degli interi è associativa e commutativa e possiede elemento neutro 0. La sottrazione, invece, non è né associativa né commutativa. Per esempio,

$$(5 - 3) - 1 = 1, \qquad 5 - (3 - 1) = 3,$$

e

$$5 - 3 \neq 3 - 5.$$

■

Veniamo ora a un fatto elementare di grande importanza in tutta la teoria delle strutture algebriche.

3.6 Proposizione [unicità dell'elemento neutro]. *Siano A un insieme e $\star : A \times A \to A$ un'operazione binaria. Se $\star$ possiede un elemento neutro, tale elemento è unico.*

Dimostrazione. Supponiamo che e ed e' siano entrambi elementi neutri. Poiché e è neutro,

$$e \star e' = e',$$

mentre, poiché e' è neutro,

$$e \star e' = e.$$

Dunque $e = e'$. □

3.2.2 Monoidi

Le tre proprietà appena introdotte non vengono sempre richieste tutte insieme. La struttura più semplice che useremo richiede associatività ed elemento neutro, ma non necessariamente commutatività.

3.7 Definizione [monoide]. Un **monoide** è un insieme M dotato di un'operazione binaria associativa

$$\star : M \times M \to M$$

che possiede un elemento neutro $e \in M$. Se inoltre l'operazione è commutativa, il monoide si dice **commutativo**. ■

- **Morfismi.** Se $(M, \star, e_M)$ e $(N, \diamond, e_N)$ sono monoidi, un **morfismo di monoidi** è una funzione $f : M \to N$ tale che
$$f(a \star b) = f(a) \diamond f(b) \qquad \text{e} \qquad f(e_M) = e_N.$$
- **Isomorfismi.** Un **isomorfismo di monoidi** è un morfismo di monoidi biiettivo il cui inverso è ancora un morfismo di monoidi. In questo caso i due monoidi si dicono isomorfi.

3.8 Esempio [monoidi numerici]. L'insieme $\mathbb{N}$ con la somma è un monoide commutativo:

$$(\mathbb{N}, +, 0).$$

Anche $\mathbb{N}$ con il prodotto è un monoide commutativo:

$$(\mathbb{N}, \cdot, 1).$$

Lo stesso vale per $\mathbb{Z}$, $\mathbb{Q}$ e $\mathbb{R}$ con la somma o con il prodotto, scegliendo l'elemento neutro appropriato. ■

3.9 Esempio [un monoide non commutativo: concatenazione di parole]. Sia A un insieme di lettere, che useremo come alfabeto per costruire parole. Per esempio, sia $A := \{a, b\}$ e consideriamo l'insieme A^* di tutte le **parole**, cioè delle sequenze finite costruite con le lettere di A, inclusa la **parola vuota** ε, che è la sequenza di lunghezza zero. Definiamo come operazione binaria in A^* la *concatenazione*: si scrive una parola subito dopo l'altra. Per esempio,

$$ab \star baa = abbaa.$$

La parola vuota è l'elemento neutro, per esempio

$$ab \star \varepsilon = \varepsilon \star ab = ab.$$

La concatenazione è associativa. In generale, però,

$$ab \star ba = abba \neq baab = ba \star ab.$$

Dunque $(A^*, \star, \varepsilon)$ è un *monoide non commutativo*. ■

3.10 Esercizio [riconoscere un monoide]. Stabilire quali delle seguenti strutture sono monoidi:

(a) $(\mathbb{Z}, +)$;
(b) $(\mathbb{Z}, -)$;
(c) $(\mathbb{R}, \cdot)$;
(d) $(\mathbb{R} \setminus \{0\}, \cdot)$.

Soluzione. (a) È un monoide: la somma è associativa e 0 è neutro. (b) Non è un monoide perché la sottrazione non è associativa e non possiede un elemento neutro bilatero. (c) È un monoide: il prodotto è associativo e 1 è neutro. (d) È ancora un monoide: il prodotto di due reali non nulli è non nullo, il prodotto è associativo e 1 appartiene all'insieme.

3.3 Gruppi

Il passaggio dal monoide al gruppo consiste nel richiedere che ogni elemento possa essere "annullato" mediante un elemento inverso.

3.11 Definizione [gruppo]. Un **gruppo** è un monoide $(G, \star, e)$ tale che per ogni $a \in G$ esiste un elemento $a^{-1} \in G$ per il quale

$$a \star a^{-1} = a^{-1} \star a = e.$$

L'elemento a^{-1} si chiama **inverso** di a. Se l'operazione è commutativa, il gruppo si dice **gruppo abeliano** o gruppo commutativo. ■

- **Morfismi.** Se $(G, \star)$ e $(H, \diamond)$ sono gruppi, un **morfismo di gruppi**, o **omomorfismo di gruppi**, è una funzione $f : G \to H$ tale che

  $$f(a \star b) = f(a) \diamond f(b) \qquad \text{per ogni } a, b \in G.$$

  Da questa proprietà segue automaticamente che f manda l'elemento neutro di G in quello di H e che $f(a^{-1}) = f(a)^{-1}$.
- **Isomorfismi.** Un **isomorfismo di gruppi** è un morfismo di gruppi biiettivo; il suo inverso è automaticamente ancora un morfismo di gruppi. In questo caso i due gruppi si dicono isomorfi.

Quando l'operazione è una somma si usa normalmente la notazione additiva: l'elemento neutro si indica con 0 e l'inverso di a con $-a$. In questo caso

$$a + (-a) = 0.$$

Quando l'operazione è un prodotto, l'elemento neutro si indica normalmente con 1 e l'inverso con a^{-1}.

3.12 Esempio [gruppi numerici]. Sono gruppi abeliani

$$(\mathbb{Z}, +), \qquad (\mathbb{Q}, +), \qquad (\mathbb{R}, +).$$

L'insieme $\mathbb{N}$ con la somma non è un gruppo, perché un naturale positivo non possiede opposto in $\mathbb{N}$. Con il prodotto sono gruppi abeliani

$$(\mathbb{Q} \setminus \{0\}, \cdot), \qquad (\mathbb{R} \setminus \{0\}, \cdot),$$

perché ogni elemento non nullo possiede un reciproco. Invece $(\mathbb{Z} \setminus \{0\}, \cdot)$ non è un gruppo: per esempio 2 non possiede inverso moltiplicativo intero. ■

3.13 Proposizione [unicità dell'inverso e cancellazione]. *Sia $(G, \star, e)$ un gruppo. L'inverso di ogni elemento di G è unico. Inoltre, per ogni $a, b, c \in G$, valgono le leggi di cancellazione:*

$$a \star b = a \star c \quad \Longrightarrow \quad b = c,$$

$$b \star a = c \star a \quad \Longrightarrow \quad b = c.$$

Dimostrazione. Supponiamo che b e c siano entrambi inversi di a. Allora

$$b = b \star e = b \star (a \star c) = (b \star a) \star c = e \star c = c.$$

Per la cancellazione, da $a \star b = a \star c$ moltiplichiamo a sinistra per a^{-1} e usiamo l'associatività:

$$a^{-1} \star (a \star b) = a^{-1} \star (a \star c),$$

quindi $b = c$. L'altra legge si dimostra nello stesso modo moltiplicando a destra. □

3.3.1 Un gruppo non commutativo: le permutazioni di tre oggetti

I gruppi non commutativi compaiono naturalmente quando l'operazione consiste nell'eseguire trasformazioni una dopo l'altra. Consideriamo l'insieme

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

Una permutazione di X è una funzione biiettiva $X \to X$. Le permutazioni si possono comporre, e la composizione di funzioni è associativa. L'identità id_X non sposta alcun elemento e ogni permutazione possiede una funzione inversa. Le sei permutazioni di tre oggetti formano quindi un gruppo, indicato con

$$S_3.$$

Consideriamo le due permutazioni

$$\sigma = (12), \qquad \tau = (23),$$

dove (12) significa "scambia 1 e 2 e lascia fisso 3", mentre (23) scambia 2 e 3. Adottiamo la convenzione usuale secondo cui in $\sigma \circ \tau$ si applica prima τ e poi σ. Allora

$$(\sigma \circ \tau)(1) = 2, \qquad (\tau \circ \sigma)(1) = 3.$$

Di conseguenza

$$\sigma \circ \tau \neq \tau \circ \sigma.$$

Il gruppo S_3 non è commutativo.

3.14 Osservazione. La non commutatività non è un difetto della struttura: esprime il fatto che l'ordine con cui si eseguono certe trasformazioni può cambiare il risultato. Questa idea ricompare in molti ambiti della matematica e della fisica matematica. ■

3.15 Esercizio [composizione di permutazioni]. Sia $X := \{1, 2, 3\}$ e siano $\sigma, \tau : X \to X$ le permutazioni $\sigma = (12)$ e $\tau = (23)$. Determinare l'immagine di $1, 2, 3$ mediante $\sigma \circ \tau$ e mediante $\tau \circ \sigma$.
Soluzione. Applicando prima τ e poi σ,

$$1 \mapsto 1 \mapsto 2, \qquad 2 \mapsto 3 \mapsto 3, \qquad 3 \mapsto 2 \mapsto 1.$$

Quindi $\sigma \circ \tau$ manda $(1, 2, 3)$ in $(2, 3, 1)$. Viceversa,

$$1 \mapsto 2 \mapsto 3, \qquad 2 \mapsto 1 \mapsto 1, \qquad 3 \mapsto 3 \mapsto 2,$$

perciò $\tau \circ \sigma$ manda $(1, 2, 3)$ in $(3, 1, 2)$. Le due composizioni sono diverse.

3.16 Esempio [il gruppo delle isometrie euclidee]. Riprendiamo la nozione di isometria euclidea definita nel Capitolo 2 mediante la distanza euclidea su $\mathbb{R}^n$, introdotta nell'Osservazione 2.66. Per $n = 2$ poniamo

$$\mathrm{Isom}(\mathbb{R}^2) := \{g : \mathbb{R}^2 \to \mathbb{R}^2 \mid g \text{ è un'isometria}\}.$$

Dunque ogni $g \in \mathrm{Isom}(\mathbb{R}^2)$ è biiettiva e conserva la distanza euclidea:

$$\|g(x) - g(y)\| = \|x - y\| \qquad \text{per ogni } x, y \in \mathbb{R}^2. \tag{3.1}$$

L'insieme $\mathrm{Isom}(\mathbb{R}^2)$ è un gruppo rispetto alla composizione. Infatti l'identità soddisfa la condizione (3.1); la composizione di due isometrie la soddisfa ancora; infine, poiché un'isometria è biiettiva, la sua inversa esiste ed è ancora un'isometria. Questo gruppo contiene, fra le altre trasformazioni, tutte le traslazioni, le rotazioni e le riflessioni del piano. In generale non è commutativo. ■

3.17 Osservazione. La stessa costruzione vale in $\mathbb{R}^3$ e produce il gruppo $\mathrm{Isom}(\mathbb{R}^3)$. Senza entrare in una classificazione precisa, è utile ricordare che le isometrie euclidee dello spazio tridimensionale possono essere descritte geometricamente come composizioni di traslazioni, rotazioni e, quando necessario, riflessioni rispetto a un piano. Non svilupperemo qui questa descrizione in forma sistematica. Per una trattazione più approfondita, anche in relazione alla meccanica, si veda Moretti [10]. È precisamente il gruppo $\mathrm{Isom}(\mathbb{R}^3)$ che abbiamo incontrato nel Capitolo 2 discutendo l'invarianza della misura di Lebesgue e il paradosso di Banach–Tarski. ■

Nota storica. *Gruppi e simmetrie.* La teoria dei gruppi si sviluppò nell'Ottocento a partire soprattutto dallo studio delle permutazioni delle radici delle equazioni algebriche. Il nome di Évariste Galois è centrale in questa storia: il suo lavoro mostrò che la possibilità di risolvere certe equazioni mediante radicali è legata alla struttura di particolari gruppi di permutazioni. Successivamente il concetto di gruppo si emancipò dal problema originario e divenne uno dei linguaggi fondamentali per descrivere simmetrie e trasformazioni [4, 5].

Un passaggio fondamentale nella matematica pura fu il *Programma di Erlangen*, proposto da Felix Klein nel 1872. L'idea, espressa qui in forma moderna e molto sintetica, è che una geometria possa essere studiata attraverso un gruppo di trasformazioni e attraverso le proprietà che restano invarianti sotto tali trasformazioni. Il concetto di gruppo diventa così uno strumento per organizzare e confrontare differenti geometrie [6, 7, 5].

In fisica il rapporto fra gruppi di simmetria e dinamica assume una forma particolarmente profonda. Nella formulazione lagrangiana della fisica meccanica, il *teorema di* Noether stabilisce, sotto opportune ipotesi, che a una famiglia continua di simmetrie dell'azione corrisponde una grandezza conservata lungo le soluzioni delle equazioni del moto. Per esempio, l'invarianza rispetto alle traslazioni temporali è collegata alla conservazione dell'energia. Il principio che lega simmetrie e quantità conservate possiede formulazioni e generalizzazioni anche nelle teorie quantistiche e quantistico-relativistiche, dove i gruppi di simmetria costituiscono una parte essenziale della struttura matematica della teoria [8, 9, 10, 15]. ■

3.4 Anelli

Per descrivere contemporaneamente somma e prodotto occorre una struttura con due operazioni binarie. Gli interi costituiscono l'esempio fondamentale.

3.18 Definizione [anello]. Un **anello** è un insieme R dotato di due operazioni binarie, indicate con $+$ e $\cdot$, tali che:

(i) $(R, +)$ è un gruppo abeliano, con elemento neutro 0;

(ii) il prodotto è associativo e possiede un elemento neutro 1;

(iii) il prodotto è distributivo rispetto alla somma:

$$a(b+c) = ab + ac, \qquad (a+b)c = ac + bc$$

per ogni $a, b, c \in R$.

Se inoltre

$$ab = ba$$

per ogni $a, b \in R$, l'anello si dice **commutativo**. ■

- **Morfismi.** Se R e S sono anelli nel senso adottato in queste dispense, un **morfismo di anelli**, o **omomorfismo di anelli**, è una funzione $f : R \to S$ tale che, per ogni $a, b \in R$,

$$f(a+b) = f(a) + f(b), \qquad f(ab) = f(a)f(b), \qquad f(1_R) = 1_S.$$

  Da queste condizioni segue anche $f(0_R) = 0_S$.
- **Isomorfismi.** Un **isomorfismo di anelli** è un morfismo di anelli biiettivo; il suo inverso è ancora un morfismo di anelli. In questo caso i due anelli si dicono isomorfi.

3.19 Osservazione. La richiesta di esistenza di un elemento neutro moltiplicativo non è sempre assunta in letteratura. In queste dispense, tuttavia, la definizione di anello includerà l'esistenza di un elemento neutro moltiplicativo. Supporremo inoltre $0 \neq 1$, escludendo così l'anello banale formato da un solo elemento. ■

3.20 Esempio [gli interi]. $(\mathbb{Z}, +, \cdot)$ è un anello commutativo. La somma forma un gruppo abeliano; il prodotto è associativo, ha elemento neutro 1 ed è commutativo; valgono le proprietà distributive. Anche $\mathbb{Q}$ e $\mathbb{R}$ sono anelli commutativi.

$\mathbb{N}$, con le operazioni usuali, non è invece un anello secondo la nostra definizione, perché $(\mathbb{N}, +)$ non è un gruppo: mancano gli opposti dei numeri positivi. ■

3.21 Proposizione [prime conseguenze degli assiomi di anello]. *Sia R un anello. Per ogni $a, b \in R$,*

$$0a = a0 = 0, \qquad (-a)b = -(ab), \qquad a(-b) = -(ab),$$

e quindi

$$(-a)(-b) = ab.$$

Dimostrazione. Poiché $0 + 0 = 0$, per distributività

$$a0 = a(0+0) = a0 + a0.$$

Sommando a entrambi i membri l'opposto di $a0$ otteniamo $a0 = 0$. Analogamente $0a = 0$. Inoltre

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

perciò $(-a)b$ è l'opposto di ab, cioè $(-a)b = -(ab)$. Le altre identità seguono nello stesso modo. □

3.4.1 Matrici: un anello non commutativo

Una **matrice reale** 2×2 (la definizione si estende a $n \times n$ in modo ovvio) è una tabella quadrata di quattro numeri reali

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Indichiamo con $M_2(\mathbb{R})$ l'insieme di tutte queste matrici. La somma si esegue componente per componente. Il prodotto è definito da

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} p & q \\ r & s \end{pmatrix} = \begin{pmatrix} ap + br & aq + bs \\ cp + dr & cq + ds \end{pmatrix}.$$

Con queste operazioni $M_2(\mathbb{R})$ è un anello. Lo zero è la **matrice nulla**, i cui elementi sono tutti 0, e l'unità è la **matrice identità**

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Il prodotto *non* è commutativo. Per esempio, poniamo

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Allora

$$AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \qquad BA = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

quindi $AB \neq BA$.

3.22 Esercizio [una semplice identità di anello]. Dimostrare direttamente dagli assiomi che $(-1)a = -a$ per ogni elemento a di un anello.
Soluzione. Poiché $1 + (-1) = 0$, per distributività

$$a + (-1)a = 1a + (-1)a = (1 + (-1))a = 0a = 0.$$

Dunque $(-1)a$ è l'opposto additivo di a, e quindi $(-1)a = -a$.

Nota storica. *Dall'aritmetica alla nozione di anello.* La nozione moderna di anello si formò tra la fine dell'Ottocento e l'inizio del Novecento attraverso lo studio di numeri algebrici, polinomi e sistemi di trasformazioni. Richard Dedekind, David Hilbert, Emmy Noether e altri contribuirono alla progressiva astrazione delle proprietà comuni a questi esempi. Il termine e la definizione moderna si stabilizzarono nel contesto dell'algebra strutturale del Novecento [4]. ■

3.5 Campi e il caso speciale dei numeri complessi

Negli interi possiamo sommare, sottrarre e moltiplicare senza uscire da $\mathbb{Z}$, ma non possiamo sempre dividere: l'equazione

$$2x = 1$$

non ha soluzione intera. Nei razionali e nei reali, invece, ogni numero non nullo possiede un inverso moltiplicativo. Questa proprietà conduce alla nozione di campo.

3.5.1 La nozione di campo

3.23 Definizione [campo]. Un **campo** è un anello commutativo K con $0 \neq 1$ nel quale ogni elemento non nullo possiede un inverso moltiplicativo. In altri termini,

$$(K \setminus \{0\}, \cdot)$$

è un gruppo abeliano. ■

- **Morfismi.** Se K e L sono campi, un **morfismo di campi** è un morfismo di anelli $f : K \to L$; dunque conserva somma, prodotto e unità. Un tale morfismo è necessariamente iniettivo.
- **Isomorfismi.** Un **isomorfismo di campi** è un morfismo di campi biiettivo. In questo caso i due campi si dicono isomorfi.

3.24 Esempio [campi numerici]. $\mathbb{Q}$ e $\mathbb{R}$ sono campi. $\mathbb{Z}$ non è un campo, perché, per esempio, 2 non possiede un inverso moltiplicativo in $\mathbb{Z}$. ■

La nozione di campo isola dunque la struttura algebrica necessaria per eseguire le quattro operazioni aritmetiche usuali, con la sola eccezione inevitabile della divisione per zero.

3.25 Proposizione [assenza di divisori dello zero in un campo]. *Siano a, b elementi di un campo K. Se*

$$ab = 0,$$

allora $a = 0$ oppure $b = 0$.

Dimostrazione. Se $a = 0$ non c'è nulla da dimostrare. Se $a \neq 0$, esiste a^{-1} e, moltiplicando $ab = 0$ per a^{-1},

$$b = 1b = (a^{-1}a)b = a^{-1}(ab) = a^{-1}0 = 0.$$

□

3.26 Esercizio [equazioni lineari in un campo]. Sia K un campo e siano $a, b \in K$ con $a \neq 0$. Mostrare che l'**equazione lineare** nell'incognita $x \in K$

$$ax = b$$

ha una e una sola soluzione e determinarla.
Soluzione. Moltiplicando per a^{-1} otteniamo necessariamente

$$x = a^{-1}b.$$

Questo valore è effettivamente una soluzione perché

$$a(a^{-1}b) = (aa^{-1})b = b.$$

L'unicità segue dalla cancellazione nel gruppo moltiplicativo $K \setminus \{0\}$.

3.5.2 I numeri complessi

I numeri reali formano un campo, ma alcune equazioni polinomiali molto semplici non hanno soluzioni reali. L'esempio fondamentale è

$$x^2 + 1 = 0,$$

perché $x^2 \geq 0$ per ogni $x \in \mathbb{R}$. I numeri complessi estendono $\mathbb{R}$ in modo da introdurre una soluzione di questa equazione.

3.5.3 Costruzione dei numeri complessi

Consideriamo l'insieme $\mathbb{R}^2$ delle coppie ordinate (a, b) di numeri reali. Definiamo

$$\begin{aligned} (a, b) + (c, d) &:= (a + c, b + d), \\ (a, b)(c, d) &:= (ac - bd, ad + bc). \end{aligned} \tag{3.2}$$

Le due operazioni in (3.2) sono rispettivamente la somma e il prodotto che useremo su $\mathbb{R}^2$.

3.27 Definizione [numeri complessi]. L'insieme $\mathbb{R}^2$ dotato della somma e del prodotto definiti nella formula (3.2) si chiama **insieme dei numeri complessi** e si indica con $\mathbb{C}$. Poniamo

$$1 := (1,0), \qquad i := (0,1).$$

Ogni numero complesso (a,b) si scrive allora in modo unico nella forma

$$a + bi.$$

I numeri complessi della forma bi, con $b \in \mathbb{R} \setminus \{0\}$, si chiamano **numeri immaginari puri**. Dato il numero complesso

$$z := a + bi,$$

con $a, b \in \mathbb{R}$, a si chiama **parte reale** di z e b si chiama **parte immaginaria** di z. ■

Dalla formula (3.2) del prodotto segue

$$i^2 = (0,1)(0,1) = (-1,0) = -1.$$

Quindi i è una soluzione dell'equazione $z^2 + 1 = 0$.

Identifichiamo ogni reale a con il complesso $(a,0) = a + 0i$. In questo modo

$$\mathbb{R} \subseteq \mathbb{C}$$

e le operazioni complesse estendono quelle reali.

3.28 Definizione [coniugato e modulo]. Se $z = a + bi \in \mathbb{C}$, il suo **coniugato** è

$$\overline{z} := a - bi,$$

e il suo **modulo** è

$$|z| := \sqrt{a^2 + b^2}.$$

■

Si verifica immediatamente che

$$z\overline{z} = a^2 + b^2 = |z|^2.$$

Se $z \neq 0$, allora $a^2 + b^2 > 0$ e quindi

$$z^{-1} = \frac{\overline{z}}{|z|^2} = \frac{a - bi}{a^2 + b^2}.$$

Questa formula mostra concretamente perché ogni numero complesso non nullo possiede un inverso.

3.29 Proposizione. *Con la somma e il prodotto definiti nella formula* (3.2)*, $\mathbb{C}$ è un campo commutativo.*

Dimostrazione. Le proprietà associative, commutative e distributive si verificano mediante calcolo diretto a partire dalle corrispondenti proprietà dei numeri reali. L'elemento neutro additivo è $0 = (0,0)$, l'opposto di $a + bi$ è $-a - bi$, e l'elemento neutro moltiplicativo è $1 = (1,0)$. La formula

$$(a + bi)^{-1} = \frac{a - bi}{a^2 + b^2}$$

fornisce l'inverso di ogni elemento non nullo. Pertanto $\mathbb{C}$ è un campo; il prodotto è commutativo perché lo è il prodotto dei reali. □

3.30 Esempio. Calcoliamo l'inverso di $z = 2 + i$:

$$z^{-1} = \frac{2 - i}{2^2 + 1^2} = \frac{2 - i}{5}.$$

Infatti

$$(2 + i)\frac{2 - i}{5} = \frac{(2 + i)(2 - i)}{5} = \frac{5}{5} = 1.$$

■

3.31 Esercizio [un'equazione quadratica complessa]. Risolvere in $\mathbb{C}$

$$z^2 + 2z + 2 = 0.$$

Soluzione. Completiamo il quadrato:

$$z^2 + 2z + 2 = (z+1)^2 + 1.$$

Dunque

$$(z+1)^2 = -1 = i^2,$$

e quindi

$$z + 1 = \pm i.$$

Le due soluzioni sono

$$z = -1 + i, \qquad z = -1 - i.$$

3.5.4 Il teorema fondamentale dell'algebra

La Definizione 2.75 si estende direttamente ai coefficienti complessi: un **polinomio complesso** nella variabile z è un'espressione

$$p(z) = a_0 + a_1 z + \cdots + a_n z^n, \qquad a_0, \ldots, a_n \in \mathbb{C}.$$

Se p non è nullo e $a_n \neq 0$, il suo grado è n. Un numero $z_0 \in \mathbb{C}$ è una **radice di un polinomio** p se $p(z_0) = 0$.

Se z_0 è una radice di p, diciamo che ha **molteplicità** $m \geq 1$ se

$$p(z) = (z - z_0)^m q(z)$$

per un polinomio q tale che $q(z_0) \neq 0$. In particolare, una radice di molteplicità 1 si dice **semplice**; se $m > 1$ si dice **multipla**.

3.32 Proposizione [teorema del fattore]. *Sia*

$$p(z) = a_0 + a_1 z + \cdots + a_n z^n$$

un polinomio complesso di grado $n \geq 1$ e sia $z_0 \in \mathbb{C}$. Allora $p(z_0) = 0$ se e solo se esiste un polinomio q di grado $n-1$ tale che

$$p(z) = (z - z_0) q(z).$$

Dimostrazione. Se $p(z) = (z - z_0)q(z)$, ponendo $z = z_0$ si ottiene immediatamente $p(z_0) = 0$. Viceversa, supponiamo $p(z_0) = 0$. Per ogni $k \geq 1$ vale

$$z^k - z_0^k = (z - z_0)\big(z^{k-1} + z^{k-2} z_0 + \cdots + z_0^{k-1}\big).$$

Pertanto

$$p(z) = p(z) - p(z_0) = \sum_{k=1}^{n} a_k (z^k - z_0^k)$$

possiede il fattore $z - z_0$. Il fattore rimanente è un polinomio di grado $n - 1$. □

3.33 Teorema [teorema fondamentale dell'algebra]. *Ogni polinomio non costante a coefficienti complessi possiede almeno una radice in $\mathbb{C}$.*

La dimostrazione richiede strumenti che vanno oltre gli scopi di queste dispense. Per la Proposizione 3.32, ogni radice produce un fattore lineare e abbassa di uno il grado del polinomio. Possiamo quindi applicare ripetutamente il teorema fondamentale dell'algebra. Ne segue che ogni polinomio complesso di grado $n \geq 1$ può essere scritto nella forma

$$p(z) = a_n (z - z_1)(z - z_2) \cdots (z - z_n),$$

dove le radici sono contate con la loro molteplicità.

Si dice per questo che $\mathbb{C}$ è **algebricamente chiuso**. Il contrasto con $\mathbb{R}$ è netto: il polinomio $x^2 + 1$ non possiede radici reali, mentre in $\mathbb{C}$ si fattorizza come

$$z^2 + 1 = (z - i)(z + i).$$

3.34 Osservazione [due precisazioni sul teorema fondamentale dell'algebra].

(a) Il teorema riguarda il campo $\mathbb{C}$. Se $K \subseteq \mathbb{C}$ è un sottocampo, un polinomio a coefficienti in K non deve, in generale, possedere una radice in K. Per esempio,

$$x^2 - 2 \in \mathbb{Q}[x]$$

non possiede radici razionali, mentre $x^2 + 1 \in \mathbb{R}[x]$ non possiede radici reali. La proprietà espressa dal teorema vale precisamente per i campi algebricamente chiusi; non tutti i sottocampi di $\mathbb{C}$ lo sono.

(b) Il teorema garantisce l'esistenza delle radici e, applicato ripetutamente, mostra che un polinomio complesso di grado n possiede esattamente n radici in $\mathbb{C}$ se le si conta con la loro molteplicità. Questo non significa però che esista sempre una formula esplicita che permetta di scriverle a partire dai coefficienti. Una **formula per radicali** è un'espressione ottenuta dai coefficienti mediante un numero finito di somme, sottrazioni, prodotti, divisioni ed estrazioni di radici. Per le equazioni polinomiali di grado al più 4 esistono formule generali di questo tipo. Il *teorema di Abel–Ruffini* afferma invece che, per il grado $n \geq 5$, non esiste una formula generale per radicali valida per tutti i polinomi di quel grado. Ciò non esclude che particolari polinomi di grado 5 o superiore siano risolvibili per radicali.

Il teorema fondamentale dell'algebra e il teorema di Abel–Ruffini rispondono dunque a domande diverse: il primo riguarda l'*esistenza* delle radici in $\mathbb{C}$, il secondo la possibilità di esprimerle, in generale, mediante radicali. Storicamente, Paolo Ruffini fu il primo a sviluppare estesamente l'idea dell'impossibilità di una soluzione generale per radicali delle equazioni di grado superiore al quarto; una dimostrazione conclusiva per l'equazione generale di quinto grado fu data da Niels Henrik Abel nel 1824. Poco dopo, Évariste Galois fornì un quadro strutturale molto più generale, collegando la risolubilità per radicali alle proprietà di opportuni gruppi di permutazioni delle radici, oggi detti gruppi di Galois [5]. Questa prospettiva anticipa quindi, in modo particolarmente significativo, il ruolo dei gruppi che studieremo in questo capitolo. ■

3.35 Esercizio [uso del teorema fondamentale dell'algebra]. Sia $p(z)$ un polinomio complesso di grado 5. Che cosa garantisce il teorema fondamentale dell'algebra e che cosa non garantisce riguardo alla forma delle radici?

Soluzione. Garantisce anzitutto l'esistenza di almeno una radice complessa. Dividendo il polinomio per il corrispondente fattore lineare si ottiene un polinomio di grado 4; ripetendo il procedimento, si conclude che

$$p(z) = a_5(z - z_1)(z - z_2)(z - z_3)(z - z_4)(z - z_5),$$

con $z_k \in \mathbb{C}$, contando più volte le eventuali radici multiple. Non afferma che le cinque radici debbano essere distinte e, soprattutto, non fornisce una formula generale per radicali che permetta di esprimerle in funzione dei coefficienti. L'assenza di una tale formula generale per il grado 5 è precisamente il contenuto del teorema di Abel–Ruffini; singoli polinomi di quinto grado possono comunque essere risolvibili per radicali.

Nota storica. *Dai numeri immaginari al teorema fondamentale dell'algebra.* Numeri oggi chiamati immaginari e complessi comparvero nel XVI secolo nel lavoro sulle equazioni di terzo e quarto grado, inizialmente come quantità formali difficili da interpretare. Nel Settecento Euler contribuì decisamente a stabilizzarne il calcolo e la notazione i; nell'Ottocento la rappresentazione geometrica nel piano rese il loro significato molto più trasparente. Il teorema fondamentale dell'algebra fu oggetto di numerosi tentativi di dimostrazione nel Settecento; la dissertazione di Gauss del 1799 occupa un posto centrale nella storia delle sue dimostrazioni. Il risultato mostra che l'estensione dai reali ai complessi è, per le equazioni polinomiali in una variabile, particolarmente completa [5]. ■

3.5.5 Funzioni olomorfe e analitiche

Usando l'identificazione $\mathbb{C} \simeq \mathbb{R}^2$ della Definizione 3.27, e scrivendo $z = x + iy$ per il punto $(x, y) \in \mathbb{R}^2$, consideriamo su $\mathbb{C}$ la topologia euclidea di $\mathbb{R}^2$ introdotta nella Definizione 2.52. In particolare, per $z_0 \in \mathbb{C}$ e $r > 0$, il disco, o palla, aperto di centro z_0 e raggio r è

$$B_r(z_0) := \{z \in \mathbb{C} \mid |z - z_0| < r\}.$$

3.36 Definizione [limite di una funzione complessa]. Siano $A \subseteq \mathbb{C}$, sia $z_0 \in \mathbb{C}$ un punto di accumulazione di A, cioè un punto tale che per ogni $\delta > 0$ esista $z \in A$ con

$$0 < |z - z_0| < \delta,$$

sia $F : A \to \mathbb{C}$ e sia $L \in \mathbb{C}$. Diciamo che

$$\lim_{z \to z_0} F(z) = L$$

se per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che, per ogni $z \in A$,

$$0 < |z - z_0| < \delta \quad \Longrightarrow \quad |F(z) - L| < \varepsilon.$$

Si tratta della consueta definizione di limite in uno spazio euclideo, applicata al piano complesso $\mathbb{C} \simeq \mathbb{R}^2$.■

Sia ora $A \subseteq \mathbb{C}$ un insieme aperto e sia

$$f : A \to \mathbb{C}.$$

Poiché ogni punto di A è interno, possiamo considerare il limite del rapporto incrementale in ciascun punto del dominio.

3.37 Definizione [derivata complessa e funzione olomorfa]. Sia $A \subseteq \mathbb{C}$ aperto, sia $f : A \to \mathbb{C}$ e sia $z_0 \in A$. Diciamo che f è **derivabile in senso complesso in** z_0 se esiste il limite

$$f'(z_0) := \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0},$$

dove z tende a z_0 nella topologia euclidea di $\mathbb{C} \simeq \mathbb{R}^2$ e $z \neq z_0$. Equivalentemente,

$$f'(z_0) = \lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h},$$

con $h \in \mathbb{C}$ e $z_0 + h \in A$.

La funzione f si dice **olomorfa** in A se è derivabile in senso complesso in ogni punto di A. ■

Per le derivate complesse successive useremo la stessa convenzione introdotta nel caso reale: quando esistono,

$$f^{(n)}(z) = \frac{d^n f}{dz^n}(z), \qquad n \geq 1.$$

La formula è formalmente la stessa usata nella Definizione 2.77 per le funzioni reali di variabile reale. La differenza essenziale è che ora l'incremento h è complesso e può tendere a 0 da qualunque direzione del piano. L'esistenza del limite richiede che il rapporto incrementale tenda allo stesso valore indipendentemente dalla direzione, cioè dall'angolo con cui h si avvicina a 0: nella formulazione ε–δ, fissato $\varepsilon > 0$, deve esistere un unico $\delta > 0$ che funzioni per tutti gli incrementi complessi h con $0 < |h| < \delta$, indipendentemente dal loro argomento. L'esistenza della derivata complessa è quindi una condizione molto più rigida della derivabilità reale.

3.38 Definizione [successioni e serie complesse convergenti]. Una **successione complessa** è una successione $(a_n)_{n \in \mathbb{N}}$ a valori in $\mathbb{C}$. Ricordando il modulo definito nella Definizione 3.28, diciamo che (a_n) converge ad $A \in \mathbb{C}$ se

$$\lim_{n \to \infty} |a_n - A| = 0.$$

Questa è precisamente la convergenza rispetto alla distanza euclidea di $\mathbb{C} \simeq \mathbb{R}^2$.

In analogia con le Definizioni 2.28 e 2.39, data una successione complessa $(a_n)_{n \in \mathbb{N}}$, la **serie complessa**

$$\sum_{n=0}^{\infty} a_n$$

è la successione complessa delle somme parziali

$$s_N := \sum_{n=0}^{N} a_n.$$

La serie si dice **convergente** se la successione (s_N) converge in $\mathbb{C}$, cioè se esiste $S \in \mathbb{C}$ tale che

$$\lim_{N \to \infty} |s_N - S| = 0.$$

In tal caso si scrive

$$\sum_{n=0}^{\infty} a_n = S.$$

Come nel caso reale, l'indice iniziale può essere diverso da 0. ■

3.39 Teorema [olomorfia e analiticità]. *Sia $A \subseteq \mathbb{C}$ aperto e sia $f : A \to \mathbb{C}$ olomorfa. Allora:*

(i) f possiede derivate complesse di ogni ordine in ogni punto di A;

(ii) per ogni $z_0 \in A$ esiste $r > 0$ tale che

$$B_r(z_0) := \{z \in \mathbb{C} \mid |z - z_0| < r\} \subseteq A,$$

e, per ogni $z \in B_r(z_0)$,

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n.$$

*La serie a secondo membro si chiama **serie di Taylor** di f centrata in z_0.*

Non dimostreremo questo teorema: la prova richiede strumenti propri di un corso di analisi complessa, in particolare l'integrazione lungo curve nel piano complesso.

Una funzione $f : A \to \mathbb{C}$ si dice **analitica** se, attorno a ogni punto $z_0 \in A$, coincide in un opportuno disco aperto con una serie di potenze centrata in z_0. Il Teorema 3.39 mostra una proprietà fondamentale dell'analisi complessa: ogni funzione olomorfa è analitica. Vale anche il viceversa, perché una serie di potenze può essere derivata termine a termine all'interno del suo disco di convergenza. Per questo motivo, nel caso complesso, i termini *olomorfa* e *analitica* vengono spesso usati come sinonimi, benché le due nozioni siano concettualmente distinte nelle rispettive definizioni.

3.40 Osservazione [funzioni analitiche reali]. Esiste una nozione analoga per funzioni reali. Se $D \subseteq \mathbb{R}$ è aperto, una funzione $g : D \to \mathbb{R}$ si dice **analitica reale** se per ogni $x_0 \in D$ esiste $r > 0$ con $(x_0 - r, x_0 + r) \subseteq D$ tale che

$$g(x) = \sum_{n=0}^{\infty} \frac{g^{(n)}(x_0)}{n!}(x - x_0)^n$$

per ogni $x \in (x_0 - r, x_0 + r)$. In questo caso la funzione coincide localmente con la propria serie di Taylor.

Nel caso reale, però, possedere derivate di ogni ordine non basta per essere analitica. Per esempio, la funzione

$$g(x) := \begin{cases} e^{-1/x^2}, & x \neq 0, \\ 0, & x = 0, \end{cases}$$

assumendo note le proprietà della funzione esponenziale, si può mostrare che possiede derivate di ogni ordine su $\mathbb{R}$ e che tutte le sue derivate in 0 sono nulle; la sua serie di Taylor in 0 è quindi identicamente nulla, mentre $g(x) > 0$ per $x \neq 0$. Questo contrasto mette in evidenza la particolare rigidità della derivabilità complessa. ■

Nota storica. Gauss, Riemann e Weierstrass. L'analisi complessa si sviluppò nell'Ottocento attraverso contributi di natura diversa e complementare. Gauss utilizzò sistematicamente la rappresentazione geometrica dei numeri complessi e ottenne risultati fondamentali sulle funzioni complesse, anche se una parte delle sue idee rimase a lungo inedita. Cauchy pose al centro della teoria l'integrazione nel piano complesso e stabilì risultati che collegano in modo profondo integrali, derivate e sviluppi in serie. Riemann sviluppò un punto di vista fortemente geometrico, nel quale le funzioni complesse e le loro proprietà vengono studiate anche attraverso superfici e trasformazioni conformi. Weierstrass contribuì invece alla sistemazione rigorosa della teoria con un'impostazione analitica basata in particolare sulle serie di potenze. Le formulazioni moderne dell'analisi complessa combinano questi diversi punti di vista [5]. ■

3.6 Spazi vettoriali reali e complessi e spazi affini; cenni sugli spazi di Hilbert e sulle varietà riemanniane

Nel Capitolo 2 abbiamo già usato elementi di $\mathbb{R}^n$, cioè n-ple di numeri reali, soprattutto per definire distanze, palle aperte, limiti e continuità. Introduciamo ora su $\mathbb{R}^n$ due operazioni naturali. Se

$$x = (x_1, \dots, x_n), \qquad y = (y_1, \dots, y_n),$$

definiamo

$$x + y := (x_1 + y_1, \dots, x_n + y_n),$$

e, per ogni $\lambda \in \mathbb{R}$,

$$\lambda x := (\lambda x_1, \dots, \lambda x_n).$$

La nozione di spazio vettoriale reale astrae precisamente le proprietà essenziali di queste due operazioni.

3.41 Definizione [spazio vettoriale reale]. Uno **spazio vettoriale reale** è un insieme V dotato di

- un'**operazione di somma** $V \times V \to V$, $(u, v) \mapsto u + v$;
- un'**operazione di moltiplicazione per scalari** $\mathbb{R} \times V \to V$, $(\lambda, v) \mapsto \lambda v$,

tali che $(V, +)$ sia un gruppo abeliano e, per ogni $u, v \in V$ e $\lambda, \mu \in \mathbb{R}$,

$$\lambda(u + v) = \lambda u + \lambda v,$$

$$(\lambda + \mu)v = \lambda v + \mu v,$$

$$(\lambda\mu)v = \lambda(\mu v), \qquad 1v = v.$$

Gli elementi di V si chiamano **vettori**; gli elementi di $\mathbb{R}$ usati nella moltiplicazione si chiamano **scalari**.■

- **Morfismi.** Se V e W sono spazi vettoriali reali, un **morfismo di spazi vettoriali**, più comunemente chiamato **applicazione lineare**, è una funzione $T : V \to W$ tale che

$$T(u + v) = T(u) + T(v), \qquad T(\lambda v) = \lambda T(v)$$

  per ogni $u, v \in V$ e $\lambda \in \mathbb{R}$.
- **Isomorfismi.** Un **isomorfismo di spazi vettoriali** è un'applicazione lineare biiettiva; la sua inversa è automaticamente lineare. In questo caso V e W si dicono isomorfi come spazi vettoriali.

Il termine "vettore" non implica necessariamente una freccia nello spazio fisico. Un vettore è, in questo contesto, semplicemente un elemento di uno spazio vettoriale nel senso della Definizione 3.41 (o della sua analoga versione complessa).

3.42 Esempio [lo spazio $\mathbb{R}^n$]. Con la somma e la moltiplicazione per scalari componente per componente, $\mathbb{R}^n$ è uno spazio vettoriale reale. Il vettore nullo è

$$0 = (0, \dots, 0).$$

■

3.43 Esempio [funzioni come vettori]. Sia X un insieme non vuoto e sia

$$\mathcal{F}(X, \mathbb{R}) := \{f \mid f : X \to \mathbb{R}\}.$$

Definiamo

$$(f + g)(x) := f(x) + g(x), \qquad (\lambda f)(x) := \lambda f(x),$$

con $\lambda \in \mathbb{R}$. Allora $\mathcal{F}(X, \mathbb{R})$ è uno spazio vettoriale reale. In questo esempio i "vettori" sono funzioni. ■

3.6.1 Spazi vettoriali complessi

La stessa idea si può formulare usando numeri complessi come scalari. Uno **spazio vettoriale complesso** è definito esattamente come uno spazio vettoriale reale, sostituendo $\mathbb{R}$ con $\mathbb{C}$: la moltiplicazione per scalari è quindi un'operazione

$$\mathbb{C} \times V \to V, \qquad (\lambda, v) \mapsto \lambda v,$$

e negli assiomi $\lambda, \mu \in \mathbb{C}$. Per esempio, $\mathbb{C}^n$ con le operazioni componente per componente è uno spazio vettoriale complesso. Lo stesso insieme $\mathbb{C}$ è uno spazio vettoriale complesso di dimensione 1 e uno spazio vettoriale reale di dimensione 2.

- **Morfismi.** Fra spazi vettoriali complessi, i morfismi sono le **applicazioni lineari complesse**: funzioni $T : V \to W$ che conservano la somma e la moltiplicazione per scalari complessi.
- **Isomorfismi.** Gli isomorfismi sono le applicazioni lineari complesse biiettive; in questo caso i due spazi si dicono isomorfi come spazi vettoriali complessi.

3.44 Definizione [combinazione lineare]. Sia V uno spazio vettoriale reale oppure complesso. Dati vettori $v_1, \dots, v_k \in V$ e scalari $\lambda_1, \dots, \lambda_k$ appartenenti rispettivamente a $\mathbb{R}$ oppure a $\mathbb{C}$, un vettore della forma

$$\lambda_1 v_1 + \cdots + \lambda_k v_k$$

si chiama **combinazione lineare** di $v_1, \dots, v_k$. ■

3.45 Definizione [indipendenza lineare, base e dimensione]. Sia V uno spazio vettoriale reale oppure complesso e siano $v_1, \dots, v_k \in V$. I vettori $v_1, \dots, v_k$ sono **linearmente indipendenti** se

$$\lambda_1 v_1 + \cdots + \lambda_k v_k = 0$$

implica

$$\lambda_1 = \cdots = \lambda_k = 0.$$

Una famiglia di vettori è una **base** di V se è linearmente indipendente e ogni vettore di V può essere scritto come combinazione lineare dei vettori della famiglia. Se V possiede una base finita con n elementi, il numero n si chiama **dimensione** di V. Se invece V non possiede alcuna base finita, si dice che è **infinito-dimensionale**. ■

3.46 Esempio [la base usuale di $\mathbb{R}^2$]. I vettori

$$e_1 = (1, 0), \qquad e_2 = (0, 1)$$

formano una base di $\mathbb{R}^2$, perché ogni $(x, y) \in \mathbb{R}^2$ si scrive in modo unico come

$$(x, y) = xe_1 + ye_2.$$

Quindi $\mathbb{R}^2$ ha dimensione 2. ■

3.47 Esercizio [una base non canonica]. Mostrare che

$$v_1 = (1, 1), \qquad v_2 = (1, -1)$$

formano una base di $\mathbb{R}^2$ e scrivere $(3, 1)$ come loro combinazione lineare.
Soluzione. Cerchiamo $\lambda, \mu \in \mathbb{R}$ tali che

$$\lambda(1, 1) + \mu(1, -1) = (3, 1).$$

Otteniamo il sistema

$$\lambda + \mu = 3, \qquad \lambda - \mu = 1.$$

Sommando le due equazioni, $2\lambda = 4$, quindi $\lambda = 2$ e $\mu = 1$. Dunque

$$(3, 1) = 2v_1 + v_2.$$

Per verificare l'indipendenza, se $\lambda v_1 + \mu v_2 = 0$ otteniamo

$$\lambda + \mu = 0, \qquad \lambda - \mu = 0,$$

da cui $\lambda = \mu = 0$. I due vettori sono indipendenti e generano tutto $\mathbb{R}^2$, quindi formano una base.

3.6.2 Trasformazioni lineari e matrici

Una volta fissata la struttura di spazio vettoriale, è naturale considerare le applicazioni che rispettano le operazioni lineari.

3.48 Definizione [trasformazione lineare ed endomorfismo]. Siano V e W spazi vettoriali sullo stesso campo, reale oppure complesso. Una **trasformazione lineare**, o applicazione lineare, da V a W è precisamente un morfismo di spazi vettoriali, cioè un'applicazione

$$T : V \to W$$

tale che, per ogni $u, v \in V$ e per ogni scalare λ,

$$T(u + v) = T(u) + T(v), \qquad T(\lambda v) = \lambda T(v).$$

Equivalentemente, per ogni scalari λ, μ,

$$T(\lambda u + \mu v) = \lambda T(u) + \mu T(v).$$

Quando $V = W$, una trasformazione lineare di V in se stesso viene anche chiamata **endomorfismo**. ■

Se V ha dimensione finita e fissiamo una base ordinata

$$\mathcal{E} = (e_1, \dots, e_n),$$

ogni vettore $v \in V$ si scrive in modo unico come

$$v = x_1 e_1 + \cdots + x_n e_n.$$

Per conoscere una trasformazione lineare T basta allora conoscere le immagini dei vettori della base. Per ogni j possiamo infatti scrivere

$$T(e_j) = a_{1j} e_1 + \cdots + a_{nj} e_n.$$

I coefficienti a_{ij} formano la **matrice di T rispetto alla base $\mathcal{E}$**:

$$[T]_{\mathcal{E}} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

Le coordinate di $T(v)$ si ottengono moltiplicando questa matrice per il vettore colonna delle coordinate di v:

$$[T(v)]_{\mathcal{E}} = [T]_{\mathcal{E}} [v]_{\mathcal{E}}. \tag{3.3}$$

La matrice non è dunque la trasformazione lineare stessa: è la sua rappresentazione dopo avere scelto una base.

3.49 Esempio [una trasformazione lineare di $\mathbb{R}^2$]. Consideriamo

$$T : \mathbb{R}^2 \to \mathbb{R}^2, \qquad T(x, y) = (2x + y, x - y).$$

Rispetto alla base usuale $e_1 = (1, 0)$, $e_2 = (0, 1)$ abbiamo

$$T(e_1) = (2, 1), \qquad T(e_2) = (1, -1),$$

e quindi

$$[T] = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}.$$

Infatti

$$\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x + y \\ x - y \end{pmatrix}.$$

■

Se $S, T : V \to V$ sono trasformazioni lineari, anche la composizione $S \circ T$ è lineare. Il punto essenziale è che, fissata una stessa base finita $\mathcal{E}$, la composizione delle trasformazioni diventa il prodotto delle matrici che le rappresentano. Infatti, per ogni $v \in V$,

$$\begin{aligned} [(S \circ T)(v)]_{\mathcal{E}} &= [S(T(v))]_{\mathcal{E}} \\ &= [S]_{\mathcal{E}} [T(v)]_{\mathcal{E}} \\ &= [S]_{\mathcal{E}} [T]_{\mathcal{E}} [v]_{\mathcal{E}}. \end{aligned}$$

D'altra parte, applicando la formula (3.3) direttamente alla trasformazione composta $S \circ T$, si ha

$$[(S \circ T)(v)]_{\mathcal{E}} = [S \circ T]_{\mathcal{E}} [v]_{\mathcal{E}}.$$

Poiché questo vale per ogni v, segue

$$\boxed{[S \circ T]_{\mathcal{E}} = [S]_{\mathcal{E}} [T]_{\mathcal{E}}.} \tag{3.4}$$

La composizione di trasformazioni lineari corrisponde dunque esattamente alla moltiplicazione delle matrici che le rappresentano. In particolare, l'ordine dei fattori è lo stesso ordine della composizione: si applica prima T e poi S, e la matrice risultante è $[S]_{\mathcal{E}} [T]_{\mathcal{E}}$.

Le trasformazioni lineari $V \to V$ possono inoltre essere sommate e moltiplicate per scalari ponendo

$$(S + T)(v) := S(v) + T(v), \qquad (\lambda T)(v) := \lambda T(v).$$

Con queste operazioni e con il prodotto dato dalla composizione, l'insieme delle trasformazioni lineari di V in se stesso costituisce un esempio fondamentale di quella che più avanti chiameremo **algebra associativa**. In generale tale algebra non è commutativa.

3.50 Esempio [la composizione non è commutativa]. In $\mathbb{R}^2$ consideriamo le trasformazioni lineari

$$P(x,y) = (x,0), \qquad S(x,y) = (y,x).$$

La prima proietta sull'asse delle ascisse, mentre la seconda scambia le due coordinate. Abbiamo

$$(P \circ S)(x,y) = (y,0), \qquad (S \circ P)(x,y) = (0,x).$$

Per esempio, sul vettore $(1,0)$,

$$(P \circ S)(1,0) = (0,0), \qquad (S \circ P)(1,0) = (0,1).$$

Dunque

$$P \circ S \neq S \circ P.$$

In termini di matrici,

$$[P] = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad [S] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

e infatti

$$[P][S] = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = [S][P].$$

Questo esempio mostra concretamente perché il prodotto in un'algebra di trasformazioni lineari non debba essere commutativo. ■

3.6.3 Prodotto scalare, norma e teorema di Pitagora

Uno spazio vettoriale, da solo, non possiede ancora nozioni di lunghezza o di angolo. Nel caso reale queste nozioni possono essere introdotte mediante un prodotto scalare.

3.51 Definizione [prodotto scalare reale]. Sia V uno spazio vettoriale reale. Un **prodotto scalare** su V è un'applicazione

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

tale che, per ogni $u, v, w \in V$ e $\lambda \in \mathbb{R}$,

$$\langle u, v \rangle = \langle v, u \rangle,$$

$$\langle \lambda u + v, w \rangle = \lambda \langle u, w \rangle + \langle v, w \rangle,$$

e

$$\langle v, v \rangle \geq 0, \qquad \langle v, v \rangle = 0 \iff v = 0.$$

■

Su $\mathbb{R}^n$ il prodotto scalare usuale è

$$\langle x, y \rangle := \sum_{j=1}^{n} x_j y_j. \tag{3.5}$$

Il prodotto scalare (3.5) permette di definire la **norma**

$$\|v\| := \sqrt{\langle v, v \rangle}. \tag{3.6}$$

La formula (3.6) definisce quindi la lunghezza di un vettore. La nozione generale di norma, che non richiede necessariamente la presenza di un prodotto scalare, sarà precisata nell'Osservazione 3.54 all'inizio della sottosezione successiva. Due vettori u e v si dicono **ortogonali**, e si scrive $u \perp v$, se

$$\langle u, v \rangle = 0.$$

3.52 Teorema [teorema di Pitagora]. *Sia V uno spazio vettoriale reale dotato di prodotto scalare $\langle \cdot, \cdot \rangle$ e sia*

$$\|w\| := \sqrt{\langle w, w \rangle} \qquad (w \in V)$$

la norma associata. Se $u, v \in V$ sono ortogonali, cioè $\langle u, v\rangle = 0$, allora

$$\boxed{\|u+v\|^2 = \|u\|^2 + \|v\|^2.} \tag{3.7}$$

Dimostrazione. Poiché $\langle u, v\rangle = 0$ e il prodotto scalare è simmetrico,

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + \|v\|^2.\end{aligned}$$

□

L'identità (3.7) mostra in forma astratta che il teorema di Pitagora non dipende essenzialmente dal disegno di un triangolo nel piano: segue dalla struttura di prodotto scalare e dalla nozione di ortogonalità.

Nel caso complesso la definizione viene modificata nel modo seguente. Adotteremo la convenzione secondo cui il prodotto scalare è lineare nel primo argomento e antilineare nel secondo.

3.53 Definizione [prodotto scalare complesso]. Sia V uno spazio vettoriale complesso. Un **prodotto scalare complesso** è un'applicazione

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$$

tale che, per ogni $u, v, w \in V$ e $\lambda \in \mathbb{C}$,

$$\langle \lambda u + v, w\rangle = \lambda \langle u, w\rangle + \langle v, w\rangle,$$

$$\langle u, v\rangle = \overline{\langle v, u\rangle},$$

e

$$\langle v, v\rangle \in \mathbb{R}, \qquad \langle v, v\rangle \geq 0, \qquad \langle v, v\rangle = 0 \iff v = 0.$$

La seconda proprietà è detta **simmetria hermitiana**. Da essa e dalla linearità nel primo argomento segue l'antilinearità nel secondo:

$$\langle u, \lambda v + w\rangle = \overline{\lambda}\, \langle u, v\rangle + \langle u, w\rangle.$$

In particolare, $\langle v, v\rangle$ è sempre un numero reale non negativo. ■

Anche nel caso complesso si definiscono

$$\|v\| := \sqrt{\langle v, v\rangle}$$

e l'ortogonalità mediante $\langle u, v\rangle = 0$; il teorema di Pitagora mantiene la stessa forma.

3.6.4 Un cenno agli spazi di Hilbert e al loro uso nelle teorie quantistiche

3.54 Osservazione [spazi metrici e spazi vettoriali normati]. Le nozioni appena introdotte si inseriscono in un quadro più generale. Un **spazio metrico** è una coppia (X, d), dove X è un insieme e

$$d : X \times X \to [0, +\infty)$$

è una funzione, detta **distanza**, tale che, per ogni $x, y, z \in X$,

$$d(x, y) = 0 \iff x = y, \qquad d(x, y) = d(y, x),$$

e vale la **disuguaglianza triangolare**

$$d(x, z) \leq d(x, y) + d(y, z).$$

Se V è uno spazio vettoriale reale o complesso, una **norma** su V è una funzione

$$\|\cdot\| : V \to [0, +\infty)$$

tale che, per ogni $u, v \in V$ e ogni scalare λ,

$$\|v\| = 0 \iff v = 0, \qquad \|\lambda v\| = |\lambda|\, \|v\|, \qquad \|u+v\| \leq \|u\| + \|v\|.$$

Uno spazio vettoriale dotato di una norma si chiama **spazio vettoriale normato**. Ogni norma induce una distanza ponendo

$$d(u, v) := \|u - v\|,$$

cosicché ogni spazio vettoriale normato è, in particolare, uno spazio metrico.

La norma definita in (3.6) è indotta da un prodotto scalare mediante

$$\|v\| = \sqrt{\langle v, v\rangle},$$

ma non ogni norma nasce in questo modo. Per esempio, su $\mathbb{R}^2$ la funzione

$$\|(x, y)\|_1 := |x| + |y|$$

è una norma, ma non è indotta da alcun prodotto scalare. Le norme indotte da un prodotto scalare soddisfano infatti ulteriori proprietà specifiche.

Fra queste è fondamentale la **disuguaglianza di Cauchy–Schwarz**:

$$|\langle u, v\rangle| \le \|u\|\,\|v\|.$$

Da essa segue, per la norma indotta dal prodotto scalare, la disuguaglianza triangolare

$$\|u + v\| \le \|u\| + \|v\|.$$

In uno spazio vettoriale normato generale, invece, la disuguaglianza triangolare fa parte direttamente della definizione di norma e non presuppone l'esistenza di un prodotto scalare. ■

La distanza associata a una norma è dunque

$$d(u, v) := \|u - v\|.$$

Una successione (v_n) di vettori si chiama **successione di Cauchy** se, per ogni $\varepsilon > 0$, esiste $N \in \mathbb{N}$ tale che

$$\|v_n - v_m\| < \varepsilon \qquad \text{per ogni } n, m \ge N.$$

Uno spazio vettoriale reale o complesso dotato di norma si chiama **spazio di Banach** se è **completo** rispetto alla distanza associata alla norma, cioè se ogni successione di Cauchy converge a un elemento dello spazio. Se la norma è indotta da un prodotto scalare, uno spazio di Banach si chiama **spazio di Hilbert**. In altre parole, uno spazio di Hilbert è uno spazio vettoriale completo rispetto alla norma derivata da un prodotto scalare. Gli spazi vettoriali normati di dimensione finita sono automaticamente spazi di Banach; se la norma è indotta da un prodotto scalare, sono quindi spazi di Hilbert. L'estensione concettualmente nuova riguarda soprattutto gli spazi infinito-dimensionali. Non svilupperemo qui questa teoria, che richiede strumenti di analisi matematica più avanzati.

Gli spazi di Hilbert nacquero dallo sviluppo dell'analisi matematica fra la fine dell'Ottocento e l'inizio del Novecento. A partire da precedenti spazi di funzioni e successioni studiati, fra gli altri, da Hilbert e Riesz, John von Neumann formulò alla fine degli anni Venti la nozione astratta moderna di spazio di Hilbert e la impiegò sistematicamente nella meccanica quantistica.

Von Neumann pose questa struttura al centro della formulazione matematica della meccanica quantistica nei suoi *Mathematische Grundlagen der Quantenmechanik* del 1932 [13, 14]. Nella formulazione standard delle teorie quantistiche, gli spazi di Hilbert complessi costituiscono una struttura matematica fondamentale: stati, osservabili e simmetrie vengono formulati mediante enti definiti a partire da tali spazi e da operatori. Nel caso più semplice, un operatore lineare associa vettori a vettori rispettando le combinazioni lineari. Per una trattazione moderna e sistematica di queste strutture si veda Moretti [15].

3.6.5 Spazi affini, geometria euclidea, geometria analitica ed estensioni moderne

La struttura di spazio vettoriale permette anche di introdurre una nozione geometrica fondamentale. Sia V uno spazio vettoriale (reale o complesso). Uno **spazio affine** (rispettivamente reale o complesso) associato a V è un insieme non vuoto A di punti dotato di un'operazione

$$A \times V \to A, \qquad (P, v) \mapsto P + v,$$

che descrive l'azione dei vettori come traslazioni. Per ogni $P \in A$ e $v, w \in V$ valgono

$$P + 0 = P, \qquad (P + v) + w = P + (v + w).$$

Inoltre, per ogni coppia di punti $P, Q \in A$ esiste un unico vettore $v \in V$ tale che

$$Q = P + v. \tag{3.8}$$

- **Morfismi.** Siano A e B spazi affini associati rispettivamente agli spazi vettoriali V e W sullo stesso campo. Un **morfismo affine**, o **applicazione affine**, è una funzione $F : A \to B$ per la quale esiste un'applicazione lineare $L : V \to W$ tale che

$$F(P + v) = F(P) + L(v) \qquad \text{per ogni } P \in A,\ v \in V.$$

- **Isomorfismi.** Un **isomorfismo affine** è un morfismo affine biiettivo la cui inversa è ancora affine. In questo caso i due spazi affini si dicono isomorfi.

Le trasformazioni

$$T_v : A \to A, \qquad T_v(P) := P + v,$$

formano il **gruppo delle traslazioni**, isomorfo al gruppo additivo $(V, +)$.

Uno spazio affine non possiede un'origine privilegiata. Ha senso sottrarre due punti per ottenere un vettore: la scrittura $Q - P = v$ è, per definizione, una forma equivalente della relazione (3.8). Non ha invece senso, senza aver scelto un'origine, sommare due punti come se fossero vettori. La struttura affine conserva nozioni come rette e parallelismo, ma non contiene ancora una nozione di distanza o di angolo.

In questo senso lo spazio affine reale è la formulazione moderna della struttura affine dello *spazio di Euclide*, generalizzata a dimensione arbitraria e privata della struttura metrica. Aggiungendo al corrispondente spazio vettoriale reale un prodotto scalare si introducono lunghezze, angoli e ortogonalità, e si recupera in particolare il teorema di Pitagora.

Coordinate affini e coordinate cartesiane ortonormali. Supponiamo ora che A sia uno spazio affine reale associato a uno spazio vettoriale reale V di dimensione finita n. Fissiamo un punto $O \in A$, che scegliamo come **origine**, e una base ordinata

$$\mathcal{E} = (e_1, \dots, e_n)$$

di V, cioè dello spazio vettoriale delle traslazioni. Chiameremo la scelta

$$(O; e_1, \dots, e_n)$$

un **sistema di coordinate affini** di origine O. Nel contesto puramente affine non è richiesta alcuna ortogonalità, perché una struttura affine, da sola, non contiene ancora una nozione di angolo. Le rette

$$\ell_i := \{O + te_i \mid t \in \mathbb{R}\}, \qquad i = 1, \dots, n,$$

sono gli assi coordinati; i vettori $e_1, \dots, e_n$ ne fissano le direzioni e la scala.

Per ogni punto $P \in A$, il vettore $P - O \in V$ possiede, poiché $\mathcal{E}$ è una base, una e una sola decomposizione

$$P - O = x_1 e_1 + \cdots + x_n e_n, \qquad \text{ossia} \qquad P = O + x_1 e_1 + \cdots + x_n e_n. \tag{3.9}$$

I numeri reali $x_1, \dots, x_n$ che compaiono nella formula (3.9) sono le **coordinate affini** di P rispetto a O e alla base $\mathcal{E}$. Si ottiene così l'applicazione

$$\Phi_{O,\mathcal{E}} : A \longrightarrow \mathbb{R}^n, \qquad P \longmapsto (x_1, \dots, x_n).$$

Questa applicazione è biiettiva: l'unicità delle coordinate dà l'iniettività, mentre per ogni $(x_1, \dots, x_n) \in \mathbb{R}^n$ il punto

$$O + x_1 e_1 + \cdots + x_n e_n$$

ha precisamente quelle coordinate e dà quindi la suriettività. La scelta di un'origine e di una base identifica dunque, in modo non canonico, lo spazio affine A con $\mathbb{R}^n$. Questo passaggio costituisce il fondamento matematico della **geometria analitica**: una volta fissato un sistema di coordinate, problemi geometrici possono essere tradotti in problemi su n-ple di numeri reali e viceversa; storicamente questo punto di vista è associato in particolare all'opera di Cartesio [5].

Se sullo spazio vettoriale V è inoltre fissato un prodotto scalare e la base $(e_1, \dots, e_n)$ è ortonormale, parleremo più precisamente di **sistema di coordinate cartesiane ortonormali**. In tal caso la biiezione $\Phi_{O,\mathcal{E}}$ conserva anche la struttura metrica: la distanza fra due punti corrisponde alla consueta distanza euclidea fra le loro coordinate in $\mathbb{R}^n$. Nel solo contesto affine, invece, il sistema di coordinate fornisce la biiezione con $\mathbb{R}^n$ ma non, di per sé, una struttura metrica.

Dalle coordinate alle varietà differenziabili. La costruzione precedente identifica globalmente uno spazio affine reale di dimensione n, dopo la scelta di un'origine e di una base, con $\mathbb{R}^n$. Una generalizzazione fondamentale consiste nel richiedere soltanto che uno spazio possa essere descritto *localmente* mediante coordinate in $\mathbb{R}^n$. Senza entrare nella definizione tecnica completa, una **varietà differenziabile** di dimensione n è uno spazio nel quale, attorno a ogni punto, si possono introdurre coordinate a valori in un aperto di $\mathbb{R}^n$, in modo che i cambiamenti fra sistemi di coordinate sovrapposti siano funzioni differenziabili. In generale non esiste un unico sistema di coordinate valido su tutto lo spazio.

Questo punto di vista ha una radice storica decisiva nella memoria di Bernhard Riemann *Ueber die Hypothesen, welche der Geometrie zu Grunde liegen*, presentata come lezione di abilitazione a Göttingen nel 1854 e pubblicata postuma nel 1868 [16, 17]. Riemann introdusse l'idea di studiare spazi di dimensione arbitraria nei quali la geometria metrica può variare da punto a punto, aprendo la strada alla moderna geometria differenziale.

Su una varietà differenziabile si può assegnare, in ogni punto, un prodotto scalare $g(u, v)$ sul corrispondente spazio vettoriale tangente di elementi $u, v, \ldots$, variabile in modo differenziabile con il punto. Quando questi prodotti scalari sono definiti positivi si ottiene una **metrica riemanniana**. Nelle teorie relativistiche dello *spaziotempo*, che è pensato come varietà differenziabile a 4 dimensioni i cui punti sono gli *eventi*, la struttura metrica fondamentale è invece di tipo diverso. Si usa infatti una **metrica lorentziana**, che non è definita positiva. Senza svilupparne qui la teoria, ciò significa in particolare che, per un vettore tangente v, la quantità $g(v, v)$ può essere positiva, negativa oppure nulla; quale dei due segni corrisponda alle direzioni temporali dipende dalla convenzione di segnatura adottata. Per questo non si tratta di una distanza nel senso metrico ordinario, che è sempre non negativa: la struttura fondamentale è la metrica lorentziana.

Questa peculiarità permette di distinguere matematicamente direzioni di tipo temporale, luminoso e spaziale e di definire i *coni di luce*. La struttura lorentziana codifica quindi le possibili *relazioni causali* fra eventi e consente di tradurre geometricamente la causalità fisica. Le varietà differenziabili dotate di metriche lorentziane costituiscono quindi il quadro geometrico fondamentale delle moderne teorie relativistiche; in particolare, nella relatività generale lo spaziotempo è formulato come una varietà differenziabile lorentziana. Per una trattazione moderna e rigorosa di varietà, geometria lorentziana, struttura causale e relatività generale si veda Moretti [18].

Nota storica. *Da* Euclide *a* Hilbert. La geometria degli *Elementi* di Euclide era organizzata attraverso definizioni, postulati e nozioni comuni, ma la sua struttura assiomatica lasciava implicite diverse assunzioni. Alla fine dell'Ottocento David Hilbert ne propose una riformulazione assiomatica molto più esplicita nei *Grundlagen der Geometrie* del 1899. Nella formulazione originaria gli assiomi, circa venti, sono organizzati in cinque gruppi riguardanti incidenza, ordine, congruenza, parallelismo e continuità; il numero preciso varia nelle successive edizioni e nel modo di contare alcune formulazioni [11, 12]. Questa revisione fu uno dei passaggi decisivi verso il metodo assiomatico moderno già discusso nel Capitolo 1. ■

Nota storica. *La nozione astratta di spazio vettoriale.* Il linguaggio dei vettori si sviluppò nell'Ottocento attraverso la geometria, i sistemi di equazioni lineari, il lavoro di Grassmann e, in forme diverse, quello di Hamilton. La definizione assiomatica moderna di spazio vettoriale emerse gradualmente quando si riconobbe che gli stessi procedimenti lineari potevano essere applicati non soltanto a frecce geometriche o n-ple di numeri, ma anche a polinomi, funzioni e molti altri oggetti [4, 3]. ■

3.7 Algebre associative

3.7.1 Algebre associative reali e complesse

Possiamo ora combinare il prodotto tipico di un anello con la struttura lineare di uno spazio vettoriale.

3.55 Definizione [algebra associativa reale o complessa]. Un'**algebra associativa reale** è uno spazio vettoriale reale A dotato anche di un prodotto

$$A \times A \to A, \qquad (a, b) \mapsto ab,$$

associativo e bilineare: per ogni $a, b, c \in A$ e $\lambda \in \mathbb{R}$,

$$\begin{aligned} (a+b)c &= ac + bc, \\ a(b+c) &= ab + ac, \\ (\lambda a)b &= \lambda(ab) = a(\lambda b). \end{aligned} \tag{3.10}$$

Un'**algebra associativa complessa** si definisce nello stesso modo partendo da uno spazio vettoriale complesso e richiedendo le stesse identità per ogni $\lambda \in \mathbb{C}$.

Se esiste un elemento $1 \in A$ tale che $1a = a1 = a$ per ogni a, l'algebra si dice **unitaria**. Se inoltre $ab = ba$ per ogni a, b, si dice **commutativa**. ■

- **Morfismi.** Se A e B sono algebre sullo stesso campo K, un **morfismo di algebre** è un'applicazione $\varphi : A \to B$ che è lineare e conserva il prodotto; cioè, per ogni $a, b \in A$ e $\lambda \in K$,

$$\varphi(a+b) = \varphi(a) + \varphi(b), \qquad \varphi(\lambda a) = \lambda\varphi(a), \qquad \varphi(ab) = \varphi(a)\varphi(b).$$

  Se le algebre sono unitarie e si richiede inoltre $\varphi(1_A) = 1_B$, parleremo di **morfismo unitario di algebre**.
- **Isomorfismi.** Un **isomorfismo di algebre** è un morfismo di algebre biiettivo; il suo inverso è ancora un morfismo di algebre. In questo caso le due algebre si dicono isomorfe. Per algebre unitarie, un isomorfismo preserva necessariamente anche l'unità.

Le identità (3.10) esprimono la *bilinearità* del prodotto: fissato uno dei due fattori, il prodotto dipende linearmente dall'altro.

3.56 Esempio [$\mathbb{C}$ come algebra reale]. $\mathbb{C}$ è uno spazio vettoriale reale di dimensione 2, con base

$$1, \qquad i.$$

Il prodotto complesso è associativo, commutativo e bilineare rispetto agli scalari reali. Pertanto $\mathbb{C}$ è un'algebra associativa reale, commutativa e unitaria. È inoltre un campo. ■

3.57 Definizione [quaternioni]. L'**algebra dei quaternioni** $\mathbb{H}$ è lo spazio vettoriale reale di dimensione 4 con base

$$1, \qquad i, \qquad j, \qquad k,$$

dotato del prodotto bilineare determinato dal fatto che 1 è l'unità e dalle relazioni

$$i^2 = j^2 = k^2 = -1, \qquad ij = k = -ji, \qquad jk = i = -kj, \qquad ki = j = -ik.$$

Ogni quaternione si scrive quindi in modo unico come

$$q = a + bi + cj + dk, \qquad a, b, c, d \in \mathbb{R}.$$

Il prodotto così definito è associativo ma non commutativo. Definendo il coniugato di q mediante

$$\overline{q} := a - bi - cj - dk,$$

si ha

$$q\overline{q} = \overline{q}q = a^2 + b^2 + c^2 + d^2.$$

Di conseguenza, se $q \neq 0$, allora

$$q^{-1} = \frac{\overline{q}}{a^2 + b^2 + c^2 + d^2}.$$

Dunque ogni elemento non nullo di $\mathbb{H}$ è invertibile. ■

3.58 Esercizio [i quaternioni e le matrici di Pauli]. Le **matrici di Pauli** sono le matrici complesse

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

dove i indica l'unità immaginaria di $\mathbb{C}$. Verificare che le matrici

$$I, \qquad -\mathrm{i}\sigma_1, \qquad -\mathrm{i}\sigma_2, \qquad -\mathrm{i}\sigma_3$$

soddisfano le stesse relazioni moltiplicative di $1, i, j, k$ nell'algebra dei quaternioni. Concludere che l'applicazione reale lineare

$$\Phi : \mathbb{H} \longrightarrow M_2(\mathbb{C})$$

definita da

$$\Phi(1) = I, \qquad \Phi(i) = -\mathrm{i}\sigma_1, \qquad \Phi(j) = -\mathrm{i}\sigma_2, \qquad \Phi(k) = -\mathrm{i}\sigma_3$$

realizza $\mathbb{H}$ come una sottoalgebra reale di $M_2(\mathbb{C})$.

Qui il verbo *realizzare* ha un significato preciso: vuol dire rappresentare concretamente gli elementi di $\mathbb{H}$ mediante matrici in modo da conservare le operazioni dell'algebra. Più precisamente, Φ è un isomorfismo di algebre reali fra $\mathbb{H}$ e la sua immagine $\Phi(\mathbb{H})$; quest'ultima è una sottoalgebra reale di $M_2(\mathbb{C})$.

Soluzione. Si verifica direttamente che

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = I,$$

e che

$$\sigma_1\sigma_2 = \mathrm{i}\sigma_3, \qquad \sigma_2\sigma_3 = \mathrm{i}\sigma_1, \qquad \sigma_3\sigma_1 = \mathrm{i}\sigma_2.$$

Invertendo l'ordine dei fattori cambia il segno. Pertanto

$$(-\mathrm{i}\sigma_r)^2 = -I \qquad (r = 1, 2, 3),$$

e, per esempio,

$$(-\mathrm{i}\sigma_1)(-\mathrm{i}\sigma_2) = -\sigma_1\sigma_2 = -\mathrm{i}\sigma_3,$$

mentre

$$(-\mathrm{i}\sigma_2)(-\mathrm{i}\sigma_1) = \mathrm{i}\sigma_3.$$

Si ottengono analogamente le altre relazioni cicliche. Le immagini di $1, i, j, k$ soddisfano dunque precisamente le relazioni che definiscono il prodotto in $\mathbb{H}$. Per bilinearità, Φ conserva il prodotto. Inoltre le quattro matrici $I, -\mathrm{i}\sigma_1, -\mathrm{i}\sigma_2, -\mathrm{i}\sigma_3$ sono linearmente indipendenti su $\mathbb{R}$, quindi Φ è iniettiva. La sua immagine è pertanto una sottoalgebra reale di $M_2(\mathbb{C})$ isomorfa a $\mathbb{H}$.

3.59 Osservazione [il teorema di Frobenius]. Nel linguaggio appena introdotto, dire che due algebre reali sono isomorfe significa dunque che esiste fra esse una biiezione reale lineare che conserva il prodotto. Con questa precisazione, possiamo formulare un importante risultato di classificazione.

Esiste una notevole rigidità per le algebre associative reali di dimensione finita nelle quali ogni elemento non nullo è invertibile. Il **teorema di Frobenius** afferma che, se A è un'algebra associativa unitaria reale di dimensione finita, $0_A \neq 1_A$, e ogni elemento non nullo di A è invertibile, allora A è isomorfa, come algebra reale, a una e una sola tra

$$\mathbb{R}, \qquad \mathbb{C}, \qquad \mathbb{H}.$$

Il caso di $\mathbb{H}$ mostra in particolare che non si assume la commutatività del prodotto. ■

3.60 Esempio [matrici complesse]. $M_2(\mathbb{C})$, costruito come $M_2(\mathbb{R})$ sostituendo $\mathbb{R}$ con $\mathbb{C}$, è uno spazio vettoriale complesso con somma e moltiplicazione per scalari componente per componente. Il prodotto di matrici è associativo e bilineare. Dunque $M_2(\mathbb{C})$ è un'algebra associativa complessa unitaria. In generale

$$AB \neq BA,$$

perciò è quindi un esempio fondamentale di algebra non commutativa. ■

3.61 Esempio [polinomi]. L'insieme $\mathbb{R}[x]$ di tutti i polinomi a coefficienti reali è uno spazio vettoriale reale. Con il prodotto usuale di polinomi è un'algebra associativa reale, commutativa e unitaria. ■

3.62 Esempio [funzioni complesse]. Se X è un insieme, lo spazio $\mathcal{F}(X, \mathbb{C})$ delle funzioni $X \to \mathbb{C}$ diventa un'algebra commutativa complessa definendo il prodotto punto per punto:

$$(fg)(x) := f(x)g(x).$$

La funzione costante 1 è l'elemento unità. ■

3.63 Esercizio [bilinearità del prodotto di matrici]. Siano $A, B, C \in M_2(\mathbb{C})$ e $\lambda \in \mathbb{C}$. Verificare, usando le usuali proprietà del prodotto di matrici, che

$$(A + B)C = AC + BC, \qquad (\lambda A)B = \lambda(AB).$$

Soluzione. Scrivendo

$$A = (a_{ij}), \quad B = (b_{ij}), \quad C = (c_{ij}),$$

l'elemento di posto (i,j) di $(A+B)C$ è

$$\sum_k (a_{ik}+b_{ik})c_{kj} = \sum_k a_{ik}c_{kj} + \sum_k b_{ik}c_{kj},$$

che è precisamente l'elemento (i,j) di $AC+BC$. Analogamente, l'elemento (i,j) di $(\lambda A)B$ è

$$\sum_k (\lambda a_{ik})b_{kj} = \lambda \sum_k a_{ik}b_{kj},$$

che coincide con l'elemento (i,j) di $\lambda(AB)$.

3.7.2 Accenni ad applicazioni della nozione di algebra in fisica quantistica

La nozione di algebra associativa, soprattutto sul campo complesso, si è rivelata uno degli strumenti concettuali più importanti per sviluppare il formalismo matematico delle teorie quantistiche. Fin dalle prime formulazioni della meccanica quantistica emerse infatti il ruolo centrale di prodotti non commutativi fra operatori.

Negli anni Trenta Murray e von Neumann svilupparono una teoria di particolari algebre di operatori su spazi di Hilbert, oggi chiamate *algebre di von Neumann*. Nelle formulazioni operatoriali avanzate esse permettono di organizzare matematicamente osservabili e altre grandezze della teoria. Un ulteriore livello di astrazione conduce alle C^**-algebre*: in questa impostazione non si assume come dato di partenza un particolare spazio di Hilbert, anche se opportune rappresentazioni su spazi di Hilbert riemergono naturalmente. La formulazione algebrica è particolarmente importante nella teoria quantistica dei campi. Per una trattazione moderna e sistematica degli spazi di Hilbert, delle algebre di operatori e della formulazione algebrica delle teorie quantistiche si veda Moretti [15].

3.8 Come sono collegate le strutture introdotte?

Le strutture introdotte in questo capitolo non costituiscono un elenco di oggetti indipendenti: ciascuna seleziona alcune proprietà e ne aggiunge altre. È utile tenere presente la seguente mappa concettuale.

Un gruppo è un monoide nel quale ogni elemento è invertibile. Un anello possiede due operazioni: rispetto alla somma è un gruppo abeliano, mentre rispetto al prodotto è un monoide; le due operazioni sono collegate dalla distributività. Un campo è un anello commutativo nel quale ogni elemento non nullo è invertibile anche rispetto al prodotto.

Gli spazi vettoriali seguono una direzione parzialmente diversa: possiedono una somma interna fra vettori e una moltiplicazione esterna per scalari appartenenti al campo di base, qui $\mathbb{R}$ oppure $\mathbb{C}$. Un'algebra associativa reale o complessa aggiunge al corrispondente spazio vettoriale un prodotto interno fra vettori, associativo e bilineare.

Possiamo riassumere alcuni esempi fondamentali:

struttura	esempio
monoide commutativo	$(\mathbb{N},+)$
gruppo abeliano	$(\mathbb{Z},+)$
gruppo non commutativo	S_3
anello commutativo	$\mathbb{Z}$
anello non commutativo	$M_2(\mathbb{R})$
campo	$\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$
spazio vettoriale reale	$\mathbb{R}^n$
spazio vettoriale complesso	$\mathbb{C}^n$
algebra commutativa	$\mathbb{R}[x]$, $\mathbb{C}$
algebra non commutativa	$M_2(\mathbb{R})$

3.64 Osservazione [astrazione e perdita di informazione]. Passare da un esempio concreto a una struttura astratta significa deliberatamente dimenticare molte proprietà dell'esempio e conservarne soltanto alcune. Considerare $\mathbb{R}$ come gruppo additivo significa, per esempio, ignorare temporaneamente il prodotto, l'ordine e la completezza. La stessa totalità matematica può quindi essere studiata come portatrice di strutture differenti, a seconda delle operazioni e delle proprietà che si vogliono mettere in evidenza. ■

3.9 Conclusione del capitolo

3.9.1 Un progressivo processo di astrazione

Il percorso del capitolo può essere riassunto come un progressivo processo di astrazione. Siamo partiti da operazioni già note sui numeri e abbiamo isolato alcune proprietà: associatività, esistenza di un elemento neutro, esistenza di inversi, distributività e compatibilità con la moltiplicazione per scalari.

Le stesse proprietà possono essere realizzate da oggetti molto differenti. Gli elementi di un gruppo possono essere numeri o permutazioni; gli elementi di un anello possono essere interi o matrici; i vettori di uno spazio vettoriale possono essere n-ple, funzioni o polinomi; i punti di uno spazio affine sono organizzati dalle traslazioni associate a uno spazio vettoriale. In questo senso una struttura algebrica descrive non tanto *che cosa sono* i suoi elementi, quanto *come possono essere combinati* e quali leggi soddisfano tali combinazioni.

Questa separazione fra natura degli oggetti e struttura delle relazioni operative è una delle idee centrali della matematica moderna. Essa riprende, in un contesto differente, il punto di vista già incontrato nel Capitolo 1 parlando di assiomatizzazione: proprietà espresse formalmente possono essere realizzate da domini di oggetti molto diversi.

3.9.2 Uno sguardo oltre: teoria delle categorie e fondamenti

Un ulteriore passo verso un punto di vista strutturale è offerto dalla *teoria delle categorie*. Una categoria comprende **oggetti** e **morfismi** fra oggetti. Se $f : A \to B$ e $g : B \to C$ sono morfismi, è definito il morfismo composto $g \circ f : A \to C$; la composizione è associativa e ogni oggetto A possiede un morfismo identità $\mathrm{id}_A : A \to A$ che è neutro rispetto alla composizione. L'attenzione si sposta così dalla costituzione interna degli oggetti alle trasformazioni che li collegano e alle proprietà preservate da tali trasformazioni.

Le nozioni introdotte in questo capitolo forniscono immediatamente esempi concreti.

3.65 Esempio [alcune categorie di strutture algebriche].

- La **categoria degli insiemi**, indicata usualmente con **Set**, ha come oggetti gli insiemi e come morfismi tutte le funzioni fra insiemi.
- La **categoria dei gruppi** ha come oggetti i gruppi e come morfismi gli omomorfismi di gruppi.
- La **categoria degli anelli** ha come oggetti gli anelli, nel senso unitario adottato in queste dispense, e come morfismi i morfismi di anelli che preservano l'unità.
- Fissato un campo K, la **categoria degli spazi vettoriali su** K ha come oggetti gli spazi vettoriali su K e come morfismi le applicazioni lineari.
- Fissato un campo K, la **categoria delle algebre associative su** K ha come oggetti le algebre associative su K e come morfismi i morfismi di algebre.

In ciascuno di questi esempi la composizione usuale di funzioni è ancora un morfismo dello stesso tipo e l'applicazione identità è un morfismo. Gli *isomorfismi* della categoria sono precisamente i morfismi invertibili: ritroviamo così, in un unico linguaggio, le nozioni di isomorfismo introdotte separatamente per le diverse strutture. ■

3.66 Osservazione [una precisazione sulle categorie considerate]. In ZF la totalità di tutti gli insiemi, e quindi anche quella di tutti i gruppi o di tutti gli anelli, non è a sua volta un insieme. Per questo categorie come **Set** e **Grp** vengono dette **categorie grandi**. In queste dispense useremo tale linguaggio in modo informale e introduttivo; una trattazione più avanzata può evitare il problema fissando un opportuno universo di oggetti. ■

3.67 Definizione [funtore]. Date due categorie $\mathcal{C}$ e $\mathcal{D}$, un **funtore**

$$F : \mathcal{C} \to \mathcal{D}$$

assegna a ogni oggetto A di $\mathcal{C}$ un oggetto $F(A)$ di $\mathcal{D}$ e a ogni morfismo

$$f : A \to B$$

di $\mathcal{C}$ un morfismo

$$F(f) : F(A) \to F(B)$$

di $\mathcal{D}$, in modo da preservare identità e composizione:

$$F(\mathrm{id}_A) = \mathrm{id}_{F(A)}, \qquad F(g \circ f) = F(g) \circ F(f).$$

■

3.68 Esempio [il funtore dimenticante]. Consideriamo la categoria dei gruppi e la categoria degli insiemi. A ogni gruppo $(G, \star)$ associamo semplicemente il suo insieme sottostante G, cioè dimentichiamo l'operazione di gruppo. A ogni omomorfismo di gruppi

$$\varphi : G \to H$$

associamo la stessa funzione φ, considerata soltanto come funzione fra gli insiemi sottostanti. Si ottiene così un **funtore dimenticante**

$$U : \mathbf{Grp} \to \mathbf{Set}.$$

Il termine "dimenticante" indica precisamente che il funtore conserva gli oggetti e le applicazioni sottostanti, ma trascura parte della struttura aggiuntiva. Analogamente si possono considerare funtori dimenticanti dalla categoria degli anelli, degli spazi vettoriali o delle algebre associative alla categoria degli insiemi. ■

La teoria delle categorie non è semplicemente un'altra teoria degli insiemi e, nel lavoro matematico ordinario, viene spesso sviluppata entro una fondazione insiemistica. Esistono però fondazioni categoriali, per esempio quelle basate su opportune categorie di insiemi o sulla *teoria dei topos*, nelle quali la matematica può essere organizzata senza assumere ZF come linguaggio fondazionale di partenza. In questo senso l'approccio categoriale offre una prospettiva fondazionale alternativa, o complementare, a quella insiemistica presentata nel Capitolo 1 [19].

3.10 Esercizi finali di ricapitolazione

3.69 Esercizio.

1. *Classificare gli insiemi numerici.* Per ciascuna delle seguenti strutture stabilire se è un monoide, un gruppo, un anello o un campo, limitandosi alle nozioni pertinenti:

   $$(\mathbb{N}, +), \qquad (\mathbb{Z}, +), \qquad (\mathbb{Z}, +, \cdot), \qquad (\mathbb{Q}, +, \cdot), \qquad (\mathbb{R}, +, \cdot).$$

   Soluzione. $(\mathbb{N}, +)$ è un monoide commutativo ma non un gruppo, perché i naturali positivi non hanno opposto in $\mathbb{N}$. $(\mathbb{Z}, +)$ è un gruppo abeliano. Con somma e prodotto, $\mathbb{Z}$ è un anello commutativo ma non un campo, perché per esempio 2 non ha inverso moltiplicativo intero. $\mathbb{Q}$ e $\mathbb{R}$ sono campi, e quindi in particolare anche anelli commutativi.

2. *Un'operazione non associativa.* Mostrare con un controesempio che la media aritmetica

   $$a \star b := \frac{a + b}{2}$$

   non è un'operazione associativa su $\mathbb{R}$.
   Soluzione. Prendiamo $a = 0$, $b = 0$, $c = 2$. Allora

   $$(a \star b) \star c = 0 \star 2 = 1,$$

   mentre

   $$a \star (b \star c) = 0 \star 1 = \frac{1}{2}.$$

   Poiché i risultati sono differenti, $\star$ non è associativa.

3. *Gruppo non commutativo.* Nel gruppo S_3 delle permutazioni di $\{1, 2, 3\}$ siano $\sigma = (12)$ e $\tau = (23)$. Spiegare perché il fatto che $\sigma\tau \neq \tau\sigma$ non contraddice nessun assioma di gruppo.
   Soluzione. Gli assiomi di gruppo richiedono chiusura, associatività, elemento neutro e inversi, ma non richiedono la commutatività. Un gruppo nel quale l'operazione è commutativa è un caso particolare, chiamato gruppo abeliano. S_3 soddisfa gli assiomi di gruppo ma non quello aggiuntivo di commutatività.

4. *Matrici e non commutatività.* Con

   $$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

   calcolare AB e BA.

Soluzione. Si ottiene

$$AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \qquad BA = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Quindi $AB \neq BA$. Questo mostra concretamente che $M_2(\mathbb{R})$ è un anello e un'algebra associativa non commutativa.

5. *Inverso di un numero complesso.* Calcolare l'inverso di $z = 3 - 4i$.
Soluzione. Il coniugato è $\overline{z} = 3 + 4i$ e

$$|z|^2 = 3^2 + (-4)^2 = 25.$$

Pertanto

$$z^{-1} = \frac{3 + 4i}{25}.$$

6. *Fattorizzazione complessa.* Fattorizzare in $\mathbb{C}$ il polinomio

$$p(z) = z^2 + 4.$$

Soluzione. L'equazione $z^2 = -4$ ha soluzioni $z = 2i$ e $z = -2i$. Dunque

$$z^2 + 4 = (z - 2i)(z + 2i).$$

7. *Indipendenza lineare.* Stabilire se i vettori

$$v_1 = (1, 2), \qquad v_2 = (2, 4)$$

sono linearmente indipendenti in $\mathbb{R}^2$.
Soluzione. Poiché

$$v_2 = 2v_1,$$

si ha

$$2v_1 - v_2 = 0$$

con coefficienti non entrambi nulli. I due vettori sono quindi linearmente dipendenti.

8. *Un'algebra di funzioni.* Sia X un insieme e siano $f, g : X \to \mathbb{R}$. Con le operazioni punto per punto, verificare la distributività

$$f(g + h) = fg + fh.$$

Soluzione. Per ogni $x \in X$,

$$\big(f(g + h)\big)(x) = f(x)(g(x) + h(x)) = f(x)g(x) + f(x)h(x) = (fg + fh)(x).$$

Poiché le due funzioni assumono lo stesso valore in ogni x, sono uguali.

9. *Riconoscere la struttura necessaria.* Quale delle strutture introdotte è sufficiente per giustificare ciascuna delle seguenti operazioni?
   (a) interpretare senza ambiguità $a \star b \star c$ senza specificare le parentesi;
   (b) risolvere $a \star x = b$ mediante un inverso;
   (c) usare contemporaneamente somma e prodotto con distributività;
   (d) dividere per ogni elemento non nullo;
   (e) formare combinazioni $\lambda_1 v_1 + \cdots + \lambda_k v_k$ con coefficienti reali.

   Soluzione. (a) È sufficiente un monoide: l'associatività permette di omettere le parentesi, mentre l'elemento neutro non è necessario per questa sola operazione. (b) Serve una struttura di gruppo. (c) Serve un anello. (d) Serve un campo. (e) Serve uno spazio vettoriale reale.

10. *Confronto fra campo e algebra.* Spiegare perché $\mathbb{C}$ è contemporaneamente un campo e un'algebra associativa reale, mentre $M_2(\mathbb{R})$ è un'algebra associativa reale ma non un campo.
Soluzione. $\mathbb{C}$ è un campo perché il prodotto è commutativo e ogni elemento non nullo possiede inverso. È anche uno spazio vettoriale reale di dimensione 2, e il prodotto complesso è associativo e bilineare: quindi è un'algebra associativa reale. $M_2(\mathbb{R})$ è anch'esso uno spazio vettoriale reale con un prodotto associativo e bilineare, ma il prodotto di matrici non è commutativo; inoltre esistono matrici non nulle non invertibili. Per entrambe le ragioni non è un campo.

11. *Pitagora in* $\mathbb{R}^2$. Con il prodotto scalare usuale verificare che

$$u = (1, 1), \qquad v = (1, -1)$$

sono ortogonali e controllare direttamente il teorema di Pitagora.
Soluzione. Si ha

$$\langle u, v \rangle = 1 \cdot 1 + 1 \cdot (-1) = 0,$$

quindi $u \perp v$. Inoltre

$$\|u\|^2 = 2, \qquad \|v\|^2 = 2,$$

mentre $u + v = (2, 0)$ e dunque

$$\|u + v\|^2 = 4 = 2 + 2 = \|u\|^2 + \|v\|^2.$$

12. $\mathbb{C}$ *come spazio vettoriale.* Qual è la dimensione di $\mathbb{C}$ considerato come spazio vettoriale su $\mathbb{C}$? E come spazio vettoriale su $\mathbb{R}$?
Soluzione. Come spazio vettoriale complesso, $\mathbb{C}$ ha dimensione 1: una base è $\{1\}$. Come spazio vettoriale reale ha dimensione 2: una base è $\{1, i\}$, perché ogni $z = a + bi$ si scrive in modo unico come

$$z = a\,1 + b\,i, \qquad a, b \in \mathbb{R}.$$

13. *Il gruppo delle traslazioni.* In uno spazio affine A associato a uno spazio vettoriale V, mostrare che per

$$T_v(P) := P + v$$

vale $T_v \circ T_w = T_{v+w}$ e individuare identità e inverso.
Soluzione. Per ogni $P \in A$,

$$(T_v \circ T_w)(P) = T_v(P + w) = (P + w) + v = P + (w + v).$$

Poiché il gruppo additivo di uno spazio vettoriale è commutativo, $w + v = v + w$, e quindi

$$T_v \circ T_w = T_{v+w}.$$

L'identità è T_0 e l'inverso di T_v è T_{-v}. Le traslazioni formano dunque un gruppo isomorfo a $(V, +)$.

14. *Caratterizzare le isometrie.* Mostrare che una funzione

$$g : \mathbb{R}^n \to \mathbb{R}^n$$

è un'isometria se e solo se è suriettiva e conserva le distanze, cioè se e solo se è suriettiva e

$$\|g(x) - g(y)\| = \|x - y\| \qquad \text{per ogni } x, y \in \mathbb{R}^n.$$

Soluzione. Se g è un'isometria, per definizione è biiettiva e conserva le distanze; in particolare è suriettiva. Viceversa, supponiamo che g sia suriettiva e conservi le distanze. Se $g(x) = g(y)$, allora

$$0 = \|g(x) - g(y)\| = \|x - y\|,$$

e quindi $x = y$. La conservazione delle distanze implica dunque che g è iniettiva. Essendo anche suriettiva, g è biiettiva; insieme alla conservazione delle distanze, questo mostra che g è un'isometria.

3.11 Letture consigliate e bibliografia

Per una prima introduzione all'algebra astratta, con molti esempi ed esercizi, è particolarmente accessibile Pinter [1]. Artin [2] offre una trattazione più ampia e strutturale di gruppi, anelli, campi e algebra lineare. Per gli spazi vettoriali si può consultare Axler [3]. Per la storia dell'algebra astratta e per il passaggio dalle equazioni alle strutture si vedano Kleiner [4] e Stillwell [5]. Per il Programma di Erlangen si veda Klein [6, 7]; per il rapporto fra simmetrie continue e leggi di conservazione, Noether [8, 9] e, in una trattazione moderna della fisica meccanica, Moretti [10]. Per gli spazi di Hilbert e il loro ruolo nelle teorie quantistiche si vedano von Neumann [13, 14] e, per una trattazione contemporanea che comprende anche algebre di operatori e formulazione algebrica, Moretti [15]. Per le origini della geometria delle varietà si veda la memoria di Riemann [16, 17]; per una trattazione moderna di varietà differenziabili, geometria lorentziana e relatività generale si veda Moretti [18]. Per una prospettiva categoriale sui fondamenti si vedano Lawvere e Rosebrugh [19].

Bibliografia del Capitolo 3

[1] C. C. Pinter, *A Book of Abstract Algebra*, 2a ed., Dover Publications, Mineola, 2010.

[2] M. Artin, *Algebra*, 2a ed., Pearson, Boston, 2011.

[3] S. Axler, *Linear Algebra Done Right*, 3a ed., Springer, Cham, 2015.

[4] I. Kleiner, *A History of Abstract Algebra*, Birkhäuser, Boston, 2007.

[5] J. Stillwell, *Mathematics and Its History*, 3a ed., Springer, New York, 2010.

[6] F. Klein, *Vergleichende Betrachtungen über neuere geometrische Forschungen*, Deichert, Erlangen, 1872.

[7] F. Klein, "A Comparative Review of Recent Researches in Geometry", trad. M. W. Haskell, *Bulletin of the New York Mathematical Society* **2** (1893), 215–249.

[8] E. Noether, "Invariante Variationsprobleme", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918, pp. 235–257.

[9] E. Noether, "Invariant Variation Problems", trad. M. A. Tavel, *Transport Theory and Statistical Physics* **1**(3) (1971), 186–207, doi:10.1080/00411457108231446.

[10] V. Moretti, *Meccanica Analitica: Meccanica Classica, Meccanica Lagrangiana, Hamiltoniana, Teoria della Stabilità, Relatività Speciale*, 2a ed., Springer, Cham, 2024.

[11] D. Hilbert, *Grundlagen der Geometrie*, Teubner, Leipzig, 1899.

[12] D. Hilbert, *The Foundations of Geometry*, traduzione autorizzata di E. J. Townsend, The Open Court Publishing Company, Chicago, 1902.

[13] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin, 1932.

[14] J. von Neumann, *Mathematical Foundations of Quantum Mechanics: New Edition*, trad. R. T. Beyer, a cura di N. A. Wheeler, Princeton University Press, Princeton, 2018, ISBN 978-0-691-17856-1.

[15] V. Moretti, *Fundamental Mathematical Structures of Quantum Theory: Spectral Theory, Foundational Issues, Symmetries, Algebraic Formulation, Quantum Measurement Theory*, 2a ed. riveduta e ampliata, Springer, Cham, in corso di pubblicazione, 2026.

[16] B. Riemann, "Ueber die Hypothesen, welche der Geometrie zu Grunde liegen", *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, vol. 13, 1868; memoria presentata come lezione di abilitazione nel 1854 e pubblicata postuma da R. Dedekind.

[17] B. Riemann, "On the Hypotheses Which Lie at the Bases of Geometry", trad. W. K. Clifford, *Nature* **8** (1873), 14–17 e 36–37.

[18] V. Moretti, *Geometric Methods in Mathematical Physics II: Tensor Analysis on Manifolds and General Relativity*, Springer, Cham, in corso di pubblicazione, 2026.

[19] F. W. Lawvere e R. Rosebrugh, *Sets for Mathematics*, Cambridge University Press, Cambridge, 2003.

Indice dei simboli

La pagina indicata è quella nella quale la notazione viene introdotta o fissata esplicitamente per la prima volta nelle dispense. Sono raccolte qui le notazioni con un significato stabile nel testo; non vengono invece elencati i semplici nomi di variabili locali.

Simbolo	Significato	Pag.
$:=$	definizione	6
$\to$	applicazione; nel contesto logico anche implicazione materiale	6
$\vee,\ \wedge,\ \neg,\ \to$	connettivi logici	6
$\Rightarrow,\ \Longleftrightarrow$	implicazione ed equivalenza nel metalinguaggio	6
$\forall,\ \exists,\ \exists!$	quantificatori universale, esistenziale ed esistenza unica	6
$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$	insiemi numerici fondamentali	6
$\in,\ \notin$	appartenenza e non appartenenza	9
$\subseteq,\ \subsetneq$	inclusione e inclusione propria	9
$\varnothing$	insieme vuoto	9
$\cup,\ \cap,\ \setminus$	unione, intersezione e differenza	11
A^c	complemento di A rispetto all'universo fissato	14
$\mathcal{P}(A)$	insieme delle parti di A	11
$A \times B$	prodotto cartesiano	11
$[a]_R,\ [a]$	classe di equivalenza di a	17
$g \circ f$	composizione di funzioni	20
id_A	funzione identità su A	21
f^{-1}	funzione inversa	21
$\lvert A\rvert$	cardinalità di un insieme	22
$\aleph_0$	cardinalità degli insiemi infiniti numerabili	24
$\mathcal{M}, s \models P$	soddisfacimento di una formula	35

Simbolo	Significato	Pag.
$(c_n)_{n\in\mathbb{N}},\ C^{\mathbb{N}}$	successione a valori in C e insieme di tutte tali successioni	26
χ_S	funzione caratteristica di $S \subseteq \mathbb{N}$	26
$\sup A,\ \inf A$	estremo superiore ed estremo inferiore	58
$\lvert x\rvert$	valore assoluto; distanza sulla retta mediante $\lvert x-y\rvert$	61
$\lim$	limite di una successione o di una funzione	62
$\displaystyle\sum_{n=0}^{\infty} a_n$	serie; successione delle somme parziali	62
$\mathcal{T}$	topologia su un insieme X	70
$B_r(x_0),\ r>0$	palla aperta di centro x_0 e raggio r	78
$\lVert x-y\rVert$	distanza euclidea in $\mathbb{R}^n$	78
$f'(x),\ \dfrac{df}{dx}$	derivata	83
$f^{(n)}(x),\ \dfrac{d^n f}{dx^n}(x)$	derivata di ordine n	84
$\dfrac{\partial f}{\partial x},\ \dfrac{\partial f}{\partial y}$	derivate parziali	83
$\displaystyle\int_a^b f(x)\,dx$	integrale di Riemann	93

Indice dei simboli (continua).

Simbolo	Significato	Pag.
$\mathcal{A}$	σ-algebra di insiemi misurabili	100
$\sigma(\mathcal{E})$	σ-algebra generata dalla famiglia $\mathcal{E}$	100
$\mathcal{B}(X)$	σ-algebra di Borel di X	101
μ	misura σ-additiva	101
$\mathcal{L}^n,\ \lambda_n$	σ-algebra e misura di Lebesgue in $\mathbb{R}^n$	102
$\mathbf{1}_A$	funzione indicatrice dell'insieme A	103
$M_2(\mathbb{R})$	matrici reali 2×2	113
$\mathbb{C},\ i$	numeri complessi e unità immaginaria	115
$\overline{z},\ \lvert z\rvert$	coniugato e modulo di un numero complesso	115
$[T]_{\mathcal{E}},\ [v]_{\mathcal{E}}$	matrice di una trasformazione e coordinate rispetto alla base $\mathcal{E}$	122
$\langle u, v\rangle$	prodotto scalare	123
$\lVert v\rVert$	norma; in §3.6.3 norma indotta dal prodotto scalare	123
$d(u,v) = \lVert u - v\rVert$	distanza associata a una norma	124
$u \perp v$	ortogonalità	123
$\mathbb{H}$	algebra dei quaternioni	128

Simbolo	Significato	Pag.
$\mathrm{Isom}(\mathbb{R}^n)$	gruppo delle isometrie euclidee	104
T_v	traslazione associata al vettore v in uno spazio affine	125
$\Phi_{O,\mathcal{E}}$	applicazione delle coordinate affini $A \to \mathbb{R}^n$	126
Set, **Grp**	categorie degli insiemi e dei gruppi	131
$F : \mathcal{C} \to \mathcal{D}$	funtore fra categorie	131
$U : \mathbf{Grp} \to \mathbf{Set}$	funtore dimenticante dai gruppi agli insiemi	131

Indice dei termini