

Sapienza, Università di Roma
a.a. 2008/09



Note per il corso di

Analisi

versione 1.1 (October 3, 2009)

L.Lamberti e C.Mascia

Corsi di laurea in Fisica & Fisica ed Astrofisica



Licenza © 2008 Lamberto Lamberti & Corrado Mascia

Distribuzione Creative Commons

Tu sei libero di riprodurre, stampare, inoltrare via mail, fotocopiare, distribuire questa opera alle seguenti condizioni:

- * **Attribuzione:** devi attribuire la paternità dell'opera nei modi indicati dall'autore o da chi ti ha dato l'opera in licenza,
- * **Non commerciale:** non puoi usare quest'opera per fini commerciali,
- * **Non opere derivate:** Non puoi alterare o trasformare quest'opera, né usarla per crearne un'altra.

(Licenza Creative Commons *Attribuzione - Non commerciale - Non opere derivate 3.0*)

Testo completo: <http://creativecommons.org/licenses/by-nc-nd/3.0/>)

Indice

Capitolo 1. I numeri reali	1
1. Numeri naturali, interi, razionali e irrazionali	1
2. Ordine e disequaglianze	11
3. La struttura metrica: il modulo	14
4. La verità sui reali	17
5. Estremo superiore ed estremo inferiore	20
Capitolo 2. Funzioni: anno zero	23
1. Ingredienti di base	23
2. Operazioni elementari su grafici	28
3. Funzioni invertibili e funzioni monotone	32
4. Classi di funzioni più o meno comuni	36
5. Problemi di massimo e minimo	41
Capitolo 3. Incontri ravvicinati con i limiti: le successioni	43
1. Limite di successioni	45
2. Il limite entra in società	52
3. Calcolo di alcuni limiti	58
4. Successioni monotone	60
5. Serie numeriche	61
Capitolo 4. Le funzioni continue	71
1. Limite di funzioni	71
2. Continuità	79
3. Esempi di discontinuità	84
4. Teoremi sulle funzioni continue	85
5. Gli intervalli incapsulati: “ <i>divide et impera</i> ”	89
Capitolo 5. Derivate, derivate e derivate	93
1. Definizione di derivata	96
2. Regole fondamentali di derivazione	101
3. Derivate successive	106
4. Il Teorema di Lagrange	107
Capitolo 6. Analisi locale e analisi globale	117
1. Punti stazionari	117

2. Analisi al microscopio	120
3. Comportamento asintotico	122
4. Funzioni convesse	126
5. A caccia di massimi e minimi assoluti	129
Capitolo 7. Ordini di grandezza e la formula di Taylor	139
1. Verso lo zero e ad un passo dall'infinito	139
2. Il Teorema di de l'Hôpital	147
3. La formula di Taylor	152
4. Espressioni del resto	157
Capitolo 8. L'integrale	161
1. L'area di un sottografico e la definizione di integrale	161
2. Istruzioni per l'uso	171
3. Il Teorema della media integrale	178
4. Il Teorema fondamentale del calcolo integrale	180
Capitolo 9. Zoologia dell'integrazione	185
1. Metodo di sostituzione	186
2. Integrazione per parti	191
3. Integrazione di funzioni razionali	195
Capitolo 10. I numeri complessi	203
1. Nascita dei numeri complessi e loro infanzia	203
2. Successioni, serie e continuità nei complessi	207
3. L'esponenziale complesso	212
Capitolo 11. Equazioni differenziali	215
1. Equazioni lineari del I ordine a coefficiente costante	215
2. Equazioni lineari del II ordine a coefficienti costanti omogenee	219
3. Equazioni lineari del II ordine a coefficienti costanti non omogenee	226
Capitolo 12. Lo spazio reale multidimensionale	231
1. Serve una struttura metrica	231
2. Quattro salti in \mathbb{R}^d : successioni di punti	236
3. Informazioni di base sulla topologia di \mathbb{R}^d	241
4. <i>Natura non facit saltus</i> : le curve	245
Capitolo 13. Funzioni di più variabili	257
1. Questioni elementari	257
2. Le derivate parziali	265
3. Integrali curvilinei	270

CAPITOLO 1

I numeri reali

Il primo obiettivo che ci proponiamo in queste Note è di introdurre i numeri reali, indispensabili per tutto quello che vedremo nel seguito. Lo scopo non è di proporre una costruzione rigorosa dei reali (che va ricercata in qualche altro testo), quanto di presentare il concetto di numero reale in maniera “intuitivamente convincente”, in modo da potere dedurne le proprietà fondamentali.

1. Numeri naturali, interi, razionali e irrazionali

Per arrivare ai numeri reali, seguiremo un percorso classico: dai numeri *naturali* ai *razionali*, passando per gli *interi*, e giungendo, alla fine, ai *reali*. Più in là presenteremo un ulteriore passo in avanti che conduce ai *numeri complessi*.

Numeri naturali. I numeri naturali nascono da uno dei problemi primordiali dell'uomo (specie se insonne): il “conteggio”. Essi nascono dall'esigenza di rispondere alla domanda: “quanti?": $0, 1, 2, 3, \dots$. L'insieme di questi numeri, detti **numeri naturali**, si indica con il simbolo \mathbb{N} :

$$\text{numeri naturali: } \quad \mathbb{N} = \{0, 1, 2, 3, \dots\}.$$

Nell'insieme \mathbb{N} sono definite le operazioni di addizione e moltiplicazione, per cui valgono le ben note *leggi commutative, associative, distributiva*.

I numeri naturali hanno due difetti fondamentali, che ci costringono, in qualche modo, a guardare più in là, cercando nuovi insiemi di numeri.

1. *Difetto “algebrico”*. Le operazioni inverse, sottrazione e divisione, non sono sempre possibili nell'insieme dei numeri naturali: non è possibile sottrarre 2 da 1 o dividere 1 con 2 e ottenere un numero naturale. La soluzione “alla Salomone” (cioè dividere in due metà) nei naturali non è sempre praticabile: nel caso di un numero dispari, il concetto di “metà” non è rappresentato da nessun numero naturale.

Il desiderio di risolvere il difetto “algebrico” ci porterà, tra poche righe, ad introdurre prima i numeri interi relativi e poi quelli razionali.

2. *Difetto “metrico”*. Il secondo problema è collegato ad una questione fondamentale: la *misura*. Il concetto di misura si basa sulla procedura seguente:

- si sceglie un’unità di misura (ad esempio, il metro campione);
- si contano il numero di copie dell’oggetto unitario che ricoprono l’oggetto da misurare.

Nella maggior parte dei casi, la lunghezza del segmento da misurare non è pari ad multiplo intero del segmento unitario. Come fare? Semplice: si divide in sotto-segmenti il segmento unitario e si contano quante “parti” della sotto-unità servono per misurare il nuovo oggetto. Questo è il concetto di frazione, cioè di numero razionale. Come vedremo, però, bisognerà prendere atto della realtà delle cose: l’essere razionali non basta...

Numeri interi relativi. Per rendere possibili senza restrizioni le operazioni di sottrazione, si estende il concetto degli interi “negativi”, ottenendo l’insieme dei **numeri interi relativi** (o semplicemente **interi**):

$$\text{numeri interi: } \mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$$

L’insieme \mathbb{Z} è una estensione insiemistica di \mathbb{N} , nel senso che contiene l’insieme dei numeri naturali, ed inoltre è possibile definire l’operazione di somma in \mathbb{Z} in modo che valgano le stesse proprietà che valevano in \mathbb{N} . C’è anche qualcosa in più: grazie all’introduzione dei numeri negativi, oltre a sommare, è sempre possibile sottrarre. In maniera più precisa, possiamo affermare che ogni elemento di \mathbb{Z} ha un *elemento inverso* rispetto all’operazione di somma:

$$\forall k \in \mathbb{Z} \quad \exists h \in \mathbb{Z} \quad \text{tale che} \quad k + h = 0.$$

Il numero h è quello che si ottiene cambiando di segno k . La presenza in \mathbb{Z} di un’operazione di somma per cui vale la proprietà commutativa, l’esistenza di 0 e l’esistenza dell’elemento inverso, fanno di \mathbb{Z} un **gruppo commutativo** (o **gruppo abeliano**).

Numeri razionali. Non abbiamo ancora risolto completamente il difetto “algebrico” di \mathbb{N} , dato che anche nell’insieme \mathbb{Z} non è sempre possibile dividere. Per questo motivo, introduciamo un terzo insieme di numeri: l’insieme dei **numeri razionali**, cioè numeri che sono rapporto di numeri interi

$$\text{numeri razionali: } \mathbb{Q} = \left\{ \frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{N}, q \neq 0 \right\}.$$

Attenzione! I numeri razionali possono essere scritti in molti modi diversi: ad esempio,

$$\frac{2}{1} = \frac{4}{2} = \frac{142}{71} = \dots$$

In genere si preferisce avere un’unica rappresentazione del numero e, per questo, si richiede che q sia positivo e che p e q abbiano massimo comune divisore pari a 1. Ci

si riconduce sempre ad un'espressione di questo genere attraverso la "semplificazione" dei fattori comuni.

Per costruzione valgono le inclusioni insiemistiche

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

Inoltre, nell'insieme dei numeri razionali \mathbb{Q} sono ben definite tutte le *operazioni razionali*, cioè addizione, moltiplicazione, sottrazione e divisione, eccetto la divisione per zero. La maniera precisa di formulare il fatto che sia possibile "dividere per un qualsiasi numero razionale non nullo" consiste nell'affermare che ogni elemento di \mathbb{Q} , tranne 0, ha un *elemento inverso* rispetto all'operazione di prodotto:

$$\forall x \in \mathbb{Q}, x \neq 0 \quad \exists y \in \mathbb{Q} \quad \text{tale che} \quad xy = 1.$$

L'elemento y si indica con $\frac{1}{x}$ oppure con x^{-1} .

La presenza di somma e prodotto (e relative proprietà) fa di \mathbb{Q} un *campo*.

ESERCIZIO 1.1. *Dimostrare che in \mathbb{Q} vale la legge di annullamento del prodotto: se $ab = 0$, allora almeno uno degli elementi a, b è nullo.*

Soluzione. Infatti, se $a \neq 0$, allora moltiplicando $ab = 0$ per $\frac{1}{a}$ a destra e a sinistra dell'uguale, si ottiene $\frac{1}{a}ab = \frac{1}{a}0$, da cui segue $b = 0$.

Risolto il problema delle operazioni inverse, rimane il difetto "metrico". È ragionevole sospettare che la definizione di \mathbb{Q} possa risolvere in un colpo solo sia il problema della divisione che quello della misurazione di lunghezze... ma è così?

Rappresentazione grafica dei razionali. Disegniamo una retta R (o *asse numerico*) e procediamo secondo questa scaletta.

(i) Scegliamo un punto di R come rappresentante di 0, che chiameremo *origine*, e un altro punto arbitrario come rappresentante del numero 1. Definiamo la direzione da 0 a 1 come *direzione positiva*, in questo modo, la retta R diviene una retta *orientata* (d'ora in poi, penseremo di aver scelto 1 alla destra di 0).

(ii) Replicando copie del segmento unitario nella direzione positiva di R (cioè alla destra di 1), otteniamo una famiglia di punti che indichiamo con $2, 3, \dots$. Detto in altre parole, rappresentiamo tutti i numeri naturali come punti sulla retta R .

(iii) Ripetendo lo stesso procedimento nella direzione negativa (alla sinistra di 0), otteniamo, allo stesso modo, una rappresentazione per gli interi negativi. Con questo procedimento, insieme a quanto fatto in (ii), arriviamo ad una rappresentazione i numeri interi sulla retta R .

(iv) Rappresentiamo ora i numeri razionali $\frac{p}{q}$ su R per cui $p, q \in \mathbb{N}$ con p più piccolo di q . Dato che il numero $\frac{p}{q}$ è, moralmente, " p volte $1/q$ ", si divide l'intervallo unitario

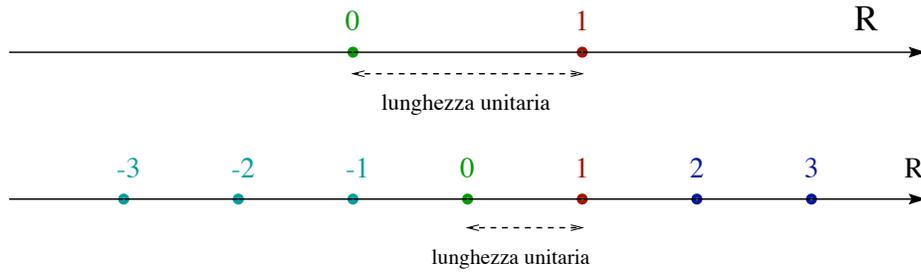


FIGURA 1. Dall'alto verso il basso: (a) i punti 0 e 1 sulla retta R determinano il segmento di lunghezza unitaria; (b) la rappresentazione dei numeri interi su R .

in q parti di uguale lunghezza e si prende come rappresentante di $\frac{p}{q}$ l'estremo destro del p -esimo intervallo.

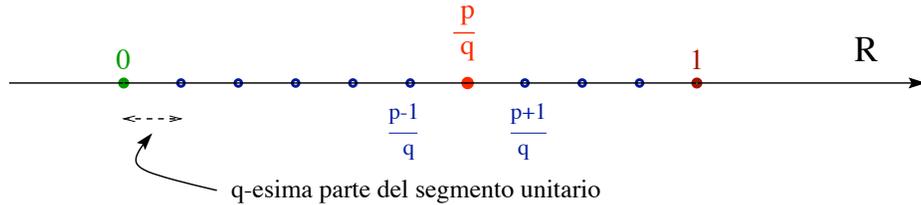


FIGURA 2. La rappresentazione su R di un numero razionale $\frac{p}{q}$ per cui $p, q \in \mathbb{N}$ e $p < q$.

(v) Nel caso di un qualsiasi numero razionale positivo p/q , si determinano p' e r positivi, con r più piccolo di q tali che

$$\frac{p}{q} = p' + \frac{r}{q},$$

e si ripete l'operazione spiegata nel passo precedente nel segmento di estremi p' e $p'+1$. Analogamente per i numeri razionali negativi.

Il significato geometrico della somma di numeri razionali è facile: se $x, y \in \mathbb{Q}$, il punto in R che corrisponde a $x + y$ corrisponde al punto che si ottiene applicando una copia del segmento di estremi 0 e y sul punto x in modo da far coincidere la copia del punto 0 con x .

Ordine, modulo e distanza nei numeri razionali. Una volta rappresentati i numeri razionali su una retta R che è orientata, è possibile mettere ordine in \mathbb{Q} .

DEFINIZIONE 1.2. Ordinamento in \mathbb{Q} . Se $x, y \in \mathbb{Q}$, allora x è minore di y (o, equivalentemente, y è maggiore di x), se, nella rappresentazione su R , x si trova alla sinistra di y . In tal caso si scrive

$$x < y.$$

Se x è minore di y o uguale ad y , si scrive $x \leq y$

$$x \leq y \iff x < y \text{ oppure } x = y.$$

Un numero $x \in \mathbb{Q}$ è **positivo**¹ se $x > 0$ ed è **negativo** se $x < 0$. Se x è positivo o è zero, si dice che è **non negativo** e si scrive $x \geq 0$ e, analogamente, se x è negativo o è zero, è **non positivo** e si scrive $x \leq 0$.

A questo punto, è possibile introdurre un oggetto di fondamentale importanza: il *modulo*.

DEFINIZIONE 1.3. Modulo e distanza in \mathbb{Q} . Dato $x \in \mathbb{Q}$, il **modulo** di x (detto anche **valore assoluto** o **norma**) si indica con $|x|$ ed è definito da

$$|x| := \begin{cases} x & x \geq 0, \\ -x & x < 0 \end{cases}$$

Dati due numeri razionali $x, y \in \mathbb{Q}$, si chiama **distanza** di x da y il numero $|x - y|$.

Dalla definizione, segue che

$$|x| \geq 0 \quad \forall x \in \mathbb{Q} \quad \text{e} \quad |x| = 0 \iff x = 0.$$

Geometricamente, il numero (non negativo) $|x|$ rappresenta la lunghezza del segmento in R che ha per estremi il punto x ed il punto 0; analogamente $|x - y|$ è la lunghezza del segmento che ha per estremi il punto x ed il punto y .

Non tutte le lunghezze sono razionali. Passiamo al collaudo di \mathbb{Q} (rappresentato sulla retta R), per la misurazione di lunghezze. Armiamoci di una fettuccia

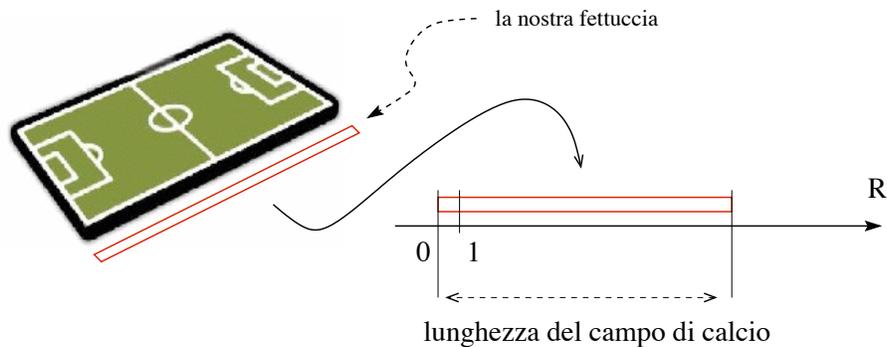


FIGURA 3. Il procedimento di misurazione: un campo di calcio.

¹Qualcuno usa una terminologia leggermente diversa: con il termine “positivo” indica un numero alla destra di 0, eventualmente anche 0 stesso, mentre un numero positivo, ma non zero, viene detto “strettamente positivo”. Analogo discorso per i numeri negativi e per la relazione d’ordine. Se y è alla destra di x ed eventualmente è x dice che “ y è maggiore di x ”; se y è alla destra di x , ma non coincide con x , dice che “ y è strettamente maggiore di x ”. Basta mettersi d’accordo.

di stoffa (o di fantasia) e procediamo nel modo più semplice possibile: se vogliamo misurare la lunghezza di un certo oggetto (un tavolo, un campo di calcio, quel che sia...), fissiamo un estremo della fettuccia ad una estremità dell'oggetto da misurare ed estendiamo fino all'altro estremo, tagliamo la fettuccia in concomitanza con il secondo estremo e riportiamo la fettuccia lungo la retta R . Collochiamo il primo estremo in corrispondenza del punto 0, stendiamo la fettuccia in tutta la sua lunghezza nella direzione positiva e vediamo il secondo estremo dove va a finire. **Se** quest'ultimo finisce in corrispondenza di un numero razionale, quel numero è la lunghezza desiderata... Non è un errore di stampa il fatto che il “Se” sia scritto in neretto: questo procedimento non sempre funziona: alcune lunghezze non corrispondono a nessun numero razionale!

Già i matematici greci scoprirono che esistono segmenti la cui lunghezza non è un numero razionale, cioè esistono punti della retta R che non corrispondono a nessun numero razionale: in simboli, $R \setminus \mathbb{Q} \neq \emptyset$. L'esempio più elementare di lunghezza non razionale è la lunghezza ℓ della diagonale di un quadrato di lato unitario. Infatti, per

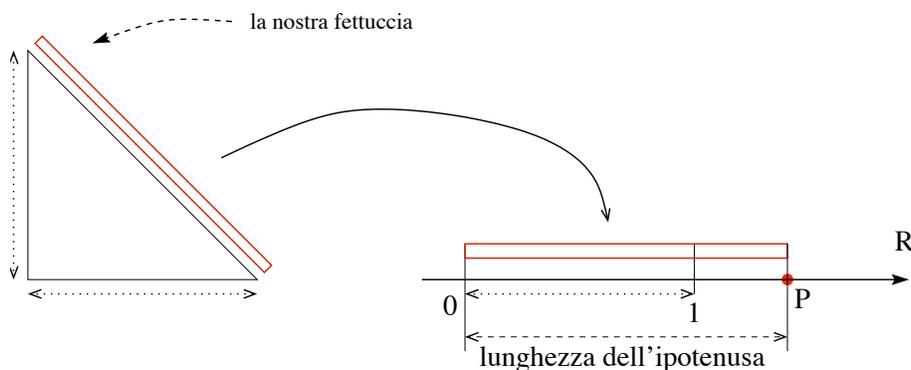


FIGURA 4. Il procedimento di misurazione dell'ipotenusa di un triangolo rettangolo di lato unitario: il punto P non corrisponde a nessun numero razionale!

il teorema di Pitagora, $\ell^2 = 2$, ma *nessun numero razionale elevato al quadrato dà per risultato il valore 2*.²

ESERCIZIO 1.4. *Dimostrare che \sqrt{p} non è razionale per ogni numero p primo. Lo stesso per $\sqrt[n]{p}$ con $n \in \{2, 3, 4, \dots\}$. Quali altre classi di numeri irrazionali sai immaginare a partire da questo esempio?*

²Per dimostrare questa affermazione, supponiamo, al contrario, che $\ell = p/q$ per opportuni p, q interi positivi. Senza restrizione, possiamo supporre che p e q non abbiano fattori comuni (altrimenti si possono semplificare). Allora si deve avere $p^2 = 2q^2$, quindi p^2 deve essere un numero pari. Dato che il quadrato di un numero dispari è dispari, anche p deve essere pari, cioè della forma $p = 2r$ con r intero positivo. Sostituendo, si ottiene $4r^2 = 2q^2$, e, semplificando il fattore 2, $2r^2 = q^2$. Ne segue che q^2 è pari e, di conseguenza, lo è anche q . Quindi p e q avrebbero un fattore comune in contraddizione con la nostra ipotesi. Pertanto ℓ non può essere razionale.

L'introduzione dei numeri razionali, che sembrava così promettente, non ha risolto il difetto “metrico” che avevamo già trovato in \mathbb{N} . Infatti è possibile costruire oggetti la cui lunghezza non è misurabile con un elemento di \mathbb{Q} . Prossima tappa: estendere \mathbb{Q} in modo da ottenere un insieme (con le stesse proprietà algebriche e metriche di \mathbb{Q}) in cui sia possibile misurare tutte le lunghezze possibili. Questa estensione è l'oggetto che chiamiamo *insieme dei numeri reali*.

Descrizione (naïf) dei numeri reali. Dato che i numeri razionali non sono sufficienti per le misurazioni, è necessario “inventare” nuovi numeri che permettano di misurare tutti i possibili segmenti. Prendiamo il toro per le corna e dichiariamo che: “ogni punto della retta R è un numero”, che chiameremo **numero reale**:

insieme dei numeri reali: $\mathbb{R} := \mathbb{Q} \cup \{\text{punti di } R \text{ che non sono in } \mathbb{Q}\}.$

Un elemento di \mathbb{R} che non sia in \mathbb{Q} si dice **numero irrazionale**. I numeri reali quindi, per definizione, coincidono con quelli della retta reale R . Esattamente come detto e fatto in precedenza, pensiamo la retta \mathbb{R} con orientamento da sinistra verso destra. La scelta del simbolo \mathbb{R} (che sostituisce R) sta a ricordare che stiamo pensando i punti della retta come oggetti per cui sono definite operazioni di somma e prodotto.

Questa definizione di numero reale grida vendetta: è intuitiva e andrebbe precisata rigorosamente. A questo livello, però, ci accontentiamo di questa versione naïf.³

Si tratta ora di capire quale rappresentazione possiamo dare ad un qualsiasi numero reale, cosa significano le operazioni di somma e prodotto in \mathbb{R} e i concetti di ordine e distanza?

Per la somma (e quindi la differenza) basta ricordare il significato della somma di razionali come punti sulla retta. Se x, y sono due numeri reali, per determinare dove si trovi sulla retta \mathbb{R} il punto $x + y$, basta procedere come segue. Rappresentiamo y come una freccia che parte da 0 e arriva nel punto corrispondente y . Per ottenere $x + y$ basta fare un “cut'n'paste” della freccia da 0 a y : se ne fa una copia e si trasla in modo da far coincidere il punto di partenza della freccia con x . Il nuovo punto di arrivo della freccia determina la posizione di $x + y$. Nel caso della differenza $x - y$, bisogna invertire la freccia che rappresenta y .

Le operazioni di prodotto e divisione in \mathbb{Q} possono essere estese, per approssimazione, ad \mathbb{R} , ma non ci soffermeremo qui sulla questione. Ci limitiamo a comunicare che valgono le stesse proprietà elencate per i numeri razionali, che, per completezza, riportiamo qui di seguito:

leggi associative: $a + (b + c) = (a + b) + c$ e $a(bc) = (ab)c$ per ogni $a, b, c \in \mathbb{R}$;

³L'idea intuitiva di numero reale come punto dell'asse numerico è stata alla base della matematica per lunghissimo tempo. Solo più tardi, nel XIX secolo, tale ipotesi è stata giustificata in modo rigoroso.

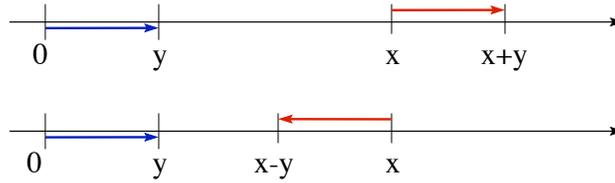


FIGURA 5. Somma e differenza di numeri reali.

leggi commutative: $a + b = b + a$ e $ab = ba$ per ogni $a, b \in \mathbb{R}$;

esistenza dell'elemento neutro per la somma: $a + 0 = a$ per ogni $a \in \mathbb{R}$;

esistenza dell'elemento neutro per il prodotto: $a \cdot 1 = a$ per ogni $a \in \mathbb{R}$;

esistenza dell'opposto per la somma: $a + (-a) = 0$ per ogni $a \in \mathbb{R}$;

esistenza dell'inverso per il prodotto: $a \cdot \frac{1}{a} = 1$ per ogni $a \in \mathbb{R}$, $a \neq 0$;

legge distributiva: $a(b + c) = ab + ac$,

annullamento del prodotto: se $ab = 0$, allora almeno uno tra a , b è 0.

Anche ordinamento e distanza si possono estendere da \mathbb{Q} ad \mathbb{R} .

DEFINIZIONE 1.5. Ordinamento in \mathbb{R} . Se $x, y \in \mathbb{R}$, allora x è minore di y (o y è maggiore di x), se x si trova alla sinistra di y . In tal caso si scrive $x < y$.

Per i simboli \leq e \geq , e per i termini positivo/negativo/non negativo/non positivo si utilizza lo stesso significato già visto per i numeri razionali.

DEFINIZIONE 1.6. Modulo e distanza in \mathbb{R} . Dato $x \in \mathbb{R}$, il modulo di x (detto anche valore assoluto o norma) si indica con $|x|$ ed è definito da

$$|x| := \begin{cases} x & x \geq 0, \\ -x & x < 0 \end{cases}$$

Dati due numeri reali $x, y \in \mathbb{R}$, si chiama distanza di x da y il numero $|x - y|$.

Tutte le proprietà che abbiamo descritto fanno di \mathbb{R} (così come lo era \mathbb{Q}) un campo totalmente ordinato dotato di metrica... più qualcosa... Mentre \mathbb{Q} è, in un certo senso, “bucato”, \mathbb{R} non lo è... Cosa vuol dire rigorosamente che “ \mathbb{R} non ha buchi”? Ci dedicheremo tra una manciata di pagine a spiegare in maniera più precisa come trasformare questa idea intuitiva in un oggetto matematicamente chiaro.

Intervalli limitati ed illimitati. Dati $a < b$, il segmento in \mathbb{R} di estremi a, b si chiama intervallo. Se gli estremi a, b sono inclusi nell'intervallo, l'intervallo si dice chiuso, se invece vengono esclusi si dice aperto:

$$\text{intervallo aperto:} \quad (a, b) := \{x \in \mathbb{R} : a < x < b\},$$

$$\text{intervallo chiuso:} \quad [a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}.$$

In entrambi i casi il valore $b - a$ è la **lunghezza** dell'intervallo (o **misura** dell'intervallo).

Si possono considerare anche intervalli semiaperti (o semichiusi) includendo uno solo dei due estremi: $(a, b]$ oppure $[a, b)$. Anche le semirette sono usualmente considerate “intervalli” e si indicano con

$$(a, +\infty) := \{x \in \mathbb{R} : x > a\} \quad [a, +\infty) := \{x \in \mathbb{R} : x \geq a\},$$

e varianti. Anche l'insieme \mathbb{R} può essere pensato come intervallo e, in tal caso, viene indicato con $(-\infty, +\infty)$.

Se necessario, per distinguere il caso degli intervalli ad estremi in \mathbb{R} da quello delle semirette, si parla, nel primo caso, di *intervalli limitati*, nel secondo di *intervalli illimitati*.

OSSERVAZIONE 1.7. Gli intervalli (limitati e illimitati) sono tutti e soli i sottoinsiemi I di \mathbb{R} che godono della proprietà seguente:

$$(x_1, x_2) \subset I \quad \forall x_1, x_2 \in I, x_1 < x_2.$$

Questa proprietà si esprime dicendo che *gli intervalli sono insiemi connessi*.

Il piano ed altre realtà. Dati due insiemi A e B si indica con $A \times B$ l'insieme

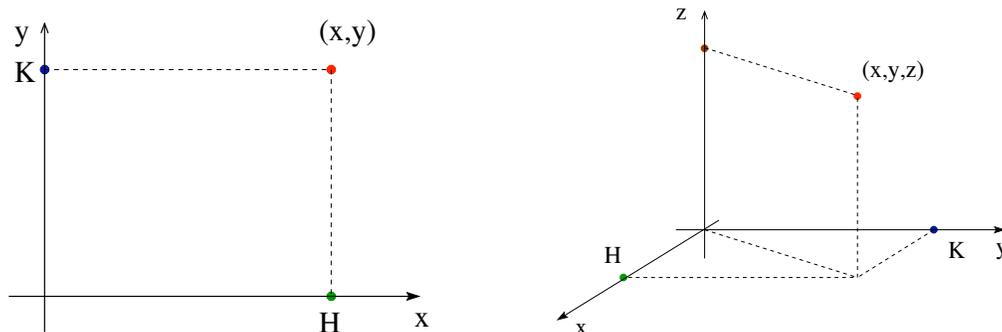
$$\text{prodotto cartesiano di } A \text{ e } B: \quad A \times B = \{(x, y) : x \in A, y \in B\},$$

Ad esempio, a quale insieme appartiene l'oggetto “giorno dell'anno”? Due numeri lo individuano: il giorno del mese (che appartiene all'insieme $I := \{1, 2, 3, \dots, 31\}$) e il numero del mese (che appartiene all'insieme $J := \{1, 2, 3, \dots, 12\}$), quindi è un elemento del prodotto cartesiano $I \times J$. Ad esempio, la data 3 aprile corrisponde all'elemento $(3, 4)$ dell'insieme $I \times J$.

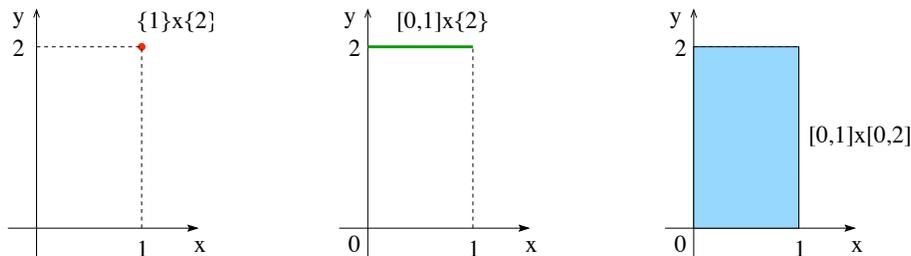
Con il simbolo $\mathbb{R} \times \mathbb{R}$ o con \mathbb{R}^2 si indica il prodotto cartesiano dell'insieme \mathbb{R} con sé stesso, ossia l'insieme costituito dalle coppie ordinate (x, y) dove $x, y \in \mathbb{R}$:

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x, y \in \mathbb{R}\}.$$

La sottolineatura della parola “ordinate” sta a segnalare che, in generale, l'elemento (x, y) è diverso da (y, x) . Ad esempio, $(1, 2) \neq (2, 1)$ (il primo febbraio è diverso dal 2 gennaio!). Per rappresentare l'insieme \mathbb{R}^2 si utilizza in genere un piano. Scegliete due rette orientate di riferimento, ortogonali fra loro e battezzatele, rispettivamente “asse x ” e “asse y ”. Per disegnare l'elemento (x_0, y_0) , si segna sull'asse x il punto H corrispondente al numero reale x_0 e sull'asse y quello K corrispondente al numero reale y_0 . Dopo di che si tracciano la retta per H e parallela all'asse y (quindi ortogonale all'asse x) e la retta per K e parallela all'asse x . Il punto intersezione rappresenta il punto (x, y) . I numeri x e y sono detti **coordinate** di (x, y) . Pensando $\mathbb{R} \times \mathbb{R}$ come

FIGURA 6. Il piano cartesiano \mathbb{R}^2 e lo spazio cartesiano \mathbb{R}^3 .

un piano, come si rappresentano gli insiemi $\{1\} \times \{2\}$, $[0, 1] \times \{2\}$ e $[0, 1] \times [0, 2]$? Rispettivamente, un punto, un segmento ed un rettangolo (Fig. 7).

FIGURA 7. Gli insiemi $\{1\} \times \{2\}$, $[0, 1] \times \{2\}$ e $[0, 1] \times [0, 2]$.

Il piano reale \mathbb{R}^2 è dotato in maniera naturale dell'operazione di somma: si tratta della somma definita componente per componente:

$$(1) \quad (x_1, y_1) + (x_2, y_2) := (x_1 + x_2, y_1 + y_2).$$

Non è evidente, al contrario, come dare una nozione di prodotto tra due punti⁴ del piano \mathbb{R}^2 . E' possibile, invece, introdurre la nozione di prodotto per uno scalare: dato $\lambda \in \mathbb{R}$, si definisce

$$(2) \quad \lambda(x, y) := (\lambda x, \lambda y).$$

Con le definizioni (1) e (2), il piano \mathbb{R}^2 viene dotato della struttura di spazio vettoriale e i suoi elementi vengono detti **vettori**.

Analogamente $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$, o \mathbb{R}^3 , indica l'insieme di terne ordinate di numeri reali (x, y, z) . L'insieme \mathbb{R}^3 si rappresenta con lo spazio, tramite una scelta di tre assi

⁴Definire il prodotto componente per componente, non è una buona idea (perché?). Una possibilità per introdurre il concetto di prodotto nel piano è alla base della definizione di numero complesso, che verrà ripresa più avanti.

coordinati ortogonali tra loro. In generale, \mathbb{R}^n (dove $n \in \mathbb{N}$) indica l'insieme delle n -uple del tipo (x_1, \dots, x_n) dove $x_1, \dots, x_n \in \mathbb{R}$. In sintesi

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

I casi $n = 2$ e $n = 3$ possono essere rappresentati geometricamente come un piano e tutto lo spazio, rispettivamente. In questa rappresentazione, i valori della coppia/terna corrispondono alle coordinate cartesiane di un punto. Nel caso di \mathbb{R}^n , i valori della n -pla possono essere interpretati come coordinate cartesiane n -dimensionali, ma poco si può visualizzare a meno di non possedere dei superpoteri.

Nell'insieme \mathbb{R}^n sono definite le operazioni di somma e di prodotto per uno scalare, dati $x = (x_1, \dots, x_n), \xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$

$$(3) \quad x + \xi = (x_1, \dots, x_n) + (\xi_1, \dots, \xi_n) := (x_1 + \xi_1, \dots, x_n + \xi_n).$$

$$(4) \quad \lambda(x_1, \dots, x_n) := (\lambda x_1, \dots, \lambda x_n).$$

In questo modo, \mathbb{R}^n acquisisce la struttura di **spazio vettoriale** e, anche in questo caso, i suoi elementi vengono detti **vettori**.

E' possibile definire anche in \mathbb{R}^n i concetti di modulo e distanza.

DEFINIZIONE 1.8. Modulo e distanza in \mathbb{R}^n . Dato $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, il modulo di x si indica con $|x|_n$ ed è definito da

$$(5) \quad |x|_n := \sqrt{x_1^2 + \dots + x_n^2} = (x_1^2 + \dots + x_n^2)^{1/2}$$

Dati due vettori $x, y \in \mathbb{R}^n$, si chiama **distanza** di x da y il numero $|x - y|_n$.

E' possibile, in realtà, introdurre anche altre definizioni di modulo, ma, per ora, ci atterremo a quella data in (5).

2. Ordine e disequaglianze

Nella pratica spesso è difficile determinare con precisione una quantità x . Ben più facile è ottenere una *stima di x* , cioè mostrare che x è compreso tra una certa quantità a e un'altra quantità b . Non solo! In molte situazioni, la stima di x è un'informazione sufficiente per la soluzione del problema. Le disequaglianze sono, perciò, un oggetto fondamentale nell'uso dei numeri reali. In questa Sezione richiamiamo alcune regole elementari a riguardo.

Regole di base per le disequazioni

(i) La somma di numeri positivi è positiva: $a > 0, b > 0 \Rightarrow a + b > 0$.

(ii) Il prodotto ab è positivo se e solo se a e b sono di segno concorde ed è negativo se e solo se sono di segno discorde

$$ab > 0 \iff a > 0, b > 0 \text{ oppure } a < 0, b < 0$$

$$ab < 0 \iff a > 0, b < 0 \text{ oppure } a < 0, b > 0.$$

(iii) Le disequazioni possono essere sommate⁵: $a > b, c > d \Rightarrow a + c > b + d$.

(iv) Se si moltiplica una disequazione per un numero positivo il segno $>$ rimane invariato:

$$a > b, c > 0 \Rightarrow ac > bc.$$

(v) Vale la *proprietà transitiva*: $a < b, b < c \Rightarrow a < c$.

Le regole precedenti valgono anche sostituendo al simbolo $<$ il simbolo \leq .

ESERCIZIO 2.1. *Dimostrare le regole di base delle disequazioni.*

Soluzione. Ecco le soluzioni di alcune delle proprietà.

(iii) $(a + c) - (b + d) = (a - b) + (c - d)$ è positivo perché somma di positivi.

(iv) Infatti $ac - bc = (a - b)c$ è positivo perché prodotto di numeri positivi. Moltiplicare per $c < 0$, a quale disequazione porta?

(v) La transitività segue subito da $a - c = (a - b) + (b - c)$ e dalla proprietà (i).

ESERCIZIO 2.2. *Dimostrare la seguente disequazione*

$$(1 + h)^n \geq 1 + nh \quad \forall h > 0, n \in \mathbb{N}.$$

Soluzione. Partiamo da un paio di situazioni semplici:

$$n = 2 : (1+h)^2 = 1+2h+h^2 \geq 1+2h, \quad n = 3 : (1+h)^3 = 1+3h+3h^2+h^3 \geq 1+3h.$$

Cosa succede in generale? Dato che $(1 + h)^n = (1 + h) \dots (1 + h)$ (il termine $(1 + h)$ compare n volte), quando si fa il prodotto di 1 per n volte si ottiene 1, quando si moltiplica il termine 1 in tutti i termini tranne uno, si ottiene h , quindi $(1 + h)^n = 1 + h + \dots + h + \dots$. In quanti modi possibili compare il termine h ? Uno per ciascuno dei termini $(1 + h)$, cioè n . In definitiva

$$(1 + h)^n = 1 + nh + \dots \text{altri termini} \dots$$

Gli "...altri termini..." sono positivi perché prodotto di positivi, quindi $(1 + h)^n \geq 1 + nh$.

Oltre a sommare/moltiplicare disequazioni, è chiaro che si può ambire ad applicare altre operazioni. Vediamo il caso dell'elevazione a potenza e delle radici n -esime.

Per ogni $a, b \in \mathbb{R}$, vale $a^2 - b^2 = (a + b)(a - b)$. Se $a, b > 0$, allora $a + b > 0$ e quindi

$$a, b > 0, \quad a > b \Rightarrow a^2 > b^2.$$

⁵Sottrarre le disuguaglianze ottenendo $a - c > b - d$, invece, non è legittimo. Ad esempio, $2 > 1$ e $3 > 1$, ma $2 - 3 = -1 < 0 = 1 - 1$.

Quindi l'operazione "elevare al quadrato" preserva l'ordine dei numeri positivi⁶: se i numeri sono ordinati, anche i loro quadrati lo sono. Dalla relazione

$$a - b = \frac{1}{a + b}(a^2 - b^2)$$

valida per ogni $a, b > 0$, si deduce che vale anche il viceversa: se $a, b > 0$ e $a^2 > b^2$, allora $a > b$. La stessa proprietà vale per qualsiasi potenza $n \in \mathbb{N}$.

PROPOSIZIONE 2.3. *Siano $a, b > 0$ e $n \in \mathbb{N}$. Allora*

$$a^n > b^n \iff a > b.$$

DIMOSTRAZIONE. Dato che, nella scomposizione

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

il secondo fattore a destra è positivo, $a^n - b^n$ ha lo stesso segno di $a - b$. \square

Dato $x \geq 0$, con il simbolo \sqrt{x} si indica la **radice quadrata** di x , cioè il numero *non negativo* il cui quadrato è x . Con questa convenzione, dato che $|c| \geq 0$ e $|c|^2 = c^2$,

$$\sqrt{c^2} = |c| \quad \forall c \in \mathbb{R},$$

(un errore comune è scrivere $\sqrt{c^2} = c$, che, invece, è vera solo per $c \geq 0$).

Applicando il risultato precedente ad $a = \sqrt{x}$ e $b = \sqrt{y}$ con $x, y > 0$, troviamo che

$$\text{se } x, y > 0, \quad \text{allora } x > y \iff \sqrt{x} > \sqrt{y}.$$

Più in generale, se $x \geq y \geq 0$, allora $\sqrt{x} \geq \sqrt{y}$. Quindi è legittimo prendere le radici quadrate di entrambi i membri di una disequazione *quando i termini sono non negativi*. Lo stesso si può dire per le radici n -esime:

$$\text{se } x, y > 0, \quad \text{allora } x > y \iff \sqrt[n]{x} > \sqrt[n]{y}.$$

Le proprietà che abbiamo descritto rientrano in una problematica più generale. Siano $x, y \in \mathbb{R}$ con $x < y$. Che succede se trasformiamo i due valori secondo una certa regola (ad esempio, l'elevazione alla potenza n o la radice n -esima)? Proprietà di questo genere sono dette *proprietà di monotonia* e verranno riprese più in là.

Se n è dispari, la conclusione della Proposizione 2.3 vale per ogni $a, b \in \mathbb{R}$. Infatti, dalle regole di base per le disequazioni, se n è dispari, allora $a^n > 0$ se e solo se $a > 0$. Quindi, nel caso $b < 0 < a$ la proprietà è evidente. Se, invece, a e b sono di segno concorde, basta notare che il termine $a^{n-1} + a^{n-2}b + \dots + ab^{n-2} + b^{n-1}$ è sempre positivo, e ragionare come nella dimostrazione della Proposizione 2.3. Riassumendo:

$$\text{se } n \text{ è dispari, allora } a^n > b^n \iff a > b.$$

⁶Attenzione, lo stesso NON è vero per numeri reali qualsiasi! Ad esempio, $-2 < -1$, ma $(-2)^2 = 4 > 1 = (-1)^2$.

3. La struttura metrica: il modulo

Ricordiamo che, dato $x \in \mathbb{R}$, il modulo di x è definito da

$$|x| := \begin{cases} x, & x \geq 0, \\ -x, & x < 0. \end{cases}$$

L'innocuo simboletto $|\cdot|$ gode di tre proprietà che gli conferiscono poteri strabilianti.

PROPOSIZIONE 3.1 (Proprietà del modulo). *Valgono le seguenti proprietà:*

- (i) $|a| \geq 0$ per ogni $a \in \mathbb{R}$ e $|a| = 0$ se e solo se $a = 0$;
- (ii) il prodotto dei moduli è il modulo del prodotto: $|ab| = |a||b|$ per ogni $a, b \in \mathbb{R}$;
- (iii) vale la disuguaglianza triangolare:

$$|a + b| \leq |a| + |b| \quad \forall a, b \in \mathbb{R}.$$

DIMOSTRAZIONE. La prima proprietà è banale. Per la seconda, basta tenere conto della regola dei segni per il prodotto di numeri reali: ad esempio, consideriamo il caso $a < 0 < b$, allora $ab < 0$ e quindi $|ab| = -ab = a(-b) = |a||b|$. Analogamente per gli altri casi. Resta da dimostrare la disuguaglianza triangolare (iii). Distinguiamo i casi $a + b \geq 0$ e $a + b < 0$. Nel primo caso, la disuguaglianza afferma $a + b \leq |a| + |b|$, che discende direttamente da $a \leq |a|$ e $b \leq |b|$ e dalla somma di queste due disequazioni. Nell'altro caso, la disuguaglianza diviene $-(a + b) \leq |a| + |b|$, che discende dalla somma delle disequazioni $-a \leq |a|$ e $-b \leq |b|$. \square

ESERCIZIO 3.2. *Verificare la validità dell'uguaglianza: $|x| = \max\{x, -x\}$.*

Soluzione. Se $x \geq 0$, allora $-x \leq 0 \leq x$, quindi $\max\{x, -x\} = x = |x|$; se $x < 0$, allora $x < 0 < -x$, quindi $\max\{x, -x\} = -x = |x|$.

Dato che il modulo è indipendente dal segno, cioè $|-x| = |x|$, la disuguaglianza triangolare vale anche con il segno $-$ al posto di $+$, cioè vale anche $|a - b| \leq |a| + |b|$. E' facile trovare esempi che mostrano che per certe scelte di a e b la disuguaglianza $|a - b| \leq |a| - |b|$ è falsa.

Applicando due volte la disuguaglianza triangolare, si ottiene

$$|a + b + c| = |(a + b) + c| \leq |a + b| + |c| \leq |a| + |b| + |c|.$$

Allo stesso modo, si ottiene la disuguaglianza più generale

$$|a_1 + a_2 + \cdots + a_n| \leq |a_1| + |a_2| + \cdots + |a_n| \quad \forall a_1, \dots, a_n \in \mathbb{R}, .$$

o, equivalentemente (tramite il simbolo di sommatoria⁷)

$$(6) \quad \left| \sum_{k=1}^n a_k \right| \leq \sum_{k=1}^n |a_k| \quad \forall a_1, \dots, a_n \in \mathbb{R},$$

Alcune conseguenze dirette della disuguaglianza triangolare sono molto utili.

COROLLARIO 3.3. *Per ogni $a, b \in \mathbb{R}$, valgono*

$$(i) \quad |a| \leq |a \pm b| + |b|, \quad (ii) \quad \left| |a| - |b| \right| \leq |a \pm b|,$$

dove il simbolo \pm indica che la disuguaglianza è vera sia con $+$ che con $-$.

DIMOSTRAZIONE. Dimostriamo le versioni di (i) e (ii) con il segno $+$. Dato che

$$|a| = |(a + b) + (-b)| \leq |a + b| + |-b| = |a + b| + |b|,$$

la (i) è vera. Per la (ii), basta riutilizzare la (i) ottenendo $|a| - |b| \leq |a + b|$. Scambiando il ruolo di a e di b , si deduce anche $|b| - |a| \leq |a + b|$. La disequazione (ii) segue da $\pm(|a| - |b|) \leq |a + b|$ e da $|x| = \max\{x, -x\}$. \square

ESERCIZIO 3.4. *Dimostrare che, per ogni $a > 0$,*

$$|x| \leq a \quad \iff \quad -a \leq x \leq a.$$

Soluzione. Dato che $|x| = \max\{-x, x\}$, la disuguaglianza $|x| \leq a$ è equivalente a dire che sia x che $-x$ sono minori od uguali di a , cioè $x \leq a$ e $-x \leq a$. Quindi $|x| \leq a$ è equivalente a $-a \leq x \leq a$. D'altronde, se si pensa ad $|x|$ come distanza di x da 0...

DEFINIZIONE 3.5. Insiemi limitati. *Un sottoinsieme $E \subset \mathbb{R}$ è limitato se esiste $R > 0$ tale che E è contenuto in un intervallo del tipo $[-R, R]$:*

$$E \subset \mathbb{R} \text{ limitato} \quad \iff \quad \exists R > 0 \text{ tale che } |x| \leq R \quad \forall x \in E.$$

Ad esempio, gli intervalli $[a, b]$, (a, b) , $(a, b]$ e $[a, b)$ sono limitati, mentre tutte le semirette sono illimitate.

ESERCIZIO 3.6. *Dire quali tra i seguenti insiemi sono limitati e quali non lo sono:*

$$\{0\} \cup \{1\}, \quad (-\infty, 0) \cup [1, 2], \quad [-1, 1] \cap [0, +\infty), \quad \mathbb{N}.$$

⁷La somma di n numeri x_1, \dots, x_n viene indicata con $\sum_{k=1}^n x_k := x_1 + x_2 + \dots + x_n$. Ad esempio $\sum_{k=1}^5 k = 1 + 2 + 3 + 4 + 5 = 15$, $\sum_{k=1}^4 k^2 = 1^2 + 2^2 + 3^2 + 4^2 = 30$.

La distanza nella retta reale. Come si è detto, il numero $|x|$ rappresenta la distanza del punto x dall'origine 0 della retta reale \mathbb{R} . Quindi, tramite il modulo, si può dare senso alla “distanza tra numeri reali”: dati $a, b \in \mathbb{R}$,

$$\text{distanza di } a \text{ da } b: \quad |a - b|.$$

Come impareremo ad apprezzare pian piano, il fatto che sia possibile introdurre in \mathbb{R} una nozione di distanza permette di sviluppare una teoria estremamente ricca ed interessante, definendo in modo rigoroso il concetto di *limite*. Ma non corriamo troppo...

Dalle proprietà del modulo discendono direttamente alcune proprietà (anche queste più che utili) per la distanza:

- (a) *positività*: la distanza di a da b è non negativa ed è nulla se e solo se $a = b$;
- (b) *simmetria*: $|a - b| = |b - a|$, cioè la distanza di a da b è uguale a quella di b da a ;
- (c) *diseguaglianza triangolare*: vale

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

ESERCIZIO 3.7. Dimostrare le proprietà (a)-(b)-(c) della distanza utilizzando le proprietà (i)-(ii)-(iii) del modulo.

DEFINIZIONE 3.8. Dato $x_0 \in \mathbb{R}$ e $r > 0$, si chiama **intorno di x_0 di raggio r** e si indica con $I(x_0; r)$ (o con $I_r(x_0)$), l'intervallo dei numeri $x \in \mathbb{R}$ tali che $x_0 - r < x < x_0 + r$:

$$I(x_0; r) = (x_0 - r, x_0 + r).$$

In maniera equivalente, si può dire che $I(x_0; r)$ è l'insieme dei numeri reali che distano da x_0 meno del raggio r . Questo insieme, sulla retta reale, si rappresenta come un segmento di lunghezza $2r$, centrato nel punto x_0 . Dalle proprietà del modulo di un numero reale discende che la condizione “ x appartiene all'intorno di x_0 di raggio r ” si può esprimere scrivendo la condizione $|x - x_0| < r$:

$$|x - x_0| < r \quad \iff \quad x \in I(x_0; r).$$

Dato che valgono (perché?)

$$|x - x_0| < r \quad \iff \quad -r < x - x_0 < r \quad \iff \quad x_0 - r < x < x_0 + r.$$

un intorno può essere scritto in maniere diverse, tutte equivalenti,

$$\begin{aligned} I(x_0; r) &= (x_0 - r, x_0 + r) = \{x : x_0 - r < x < x_0 + r\} \\ &= \{x : -r < x - x_0 < r\} = \{x : |x - x_0| < r\}. \end{aligned}$$

4. La verità sui reali

Prologo. È giunto il momento di riprendere la questione della *completezza* dei numeri reali. Partiamo da un esempio illustrativo che spiega la situazione: disegniamo due curve nel piano così come in Figura 8. La domanda è: *queste due curve si intersecano oppure no?*

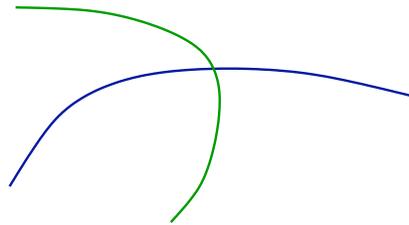


FIGURA 8. Due curve nel piano.

cano oppure no? Un sondaggio del 2003 dà queste percentuali di risposta: il 78% degli intervistati dice “SI, SEMPRE”, il 5% dice “QUALCHE VOLTA”, il 3% dice “QUASI MAI”, il 10% risponde “NON SO” e il 4% fugge scappando per timore di fare brutta figura. Evidentemente la risposta è racchiusa in quello che succede vicino al punto di incrocio delle due curve. Proponiamo tre maniere diverse di ragionare.

Versione “atomistica” (o “alla Democrito”). In questa versione, si immagina che le curve siano costituite da punti equidistanti (a distanza tanto piccola che l’occhio non è in grado di distinguerli, e vede solo un linea apparentemente continua). Con un ingrandimento di scala si vede bene che (a meno di casi particolarmente fortunati) le due curve, discretizzate in tanti atomi, non si incontrano. La risposta in questo caso è

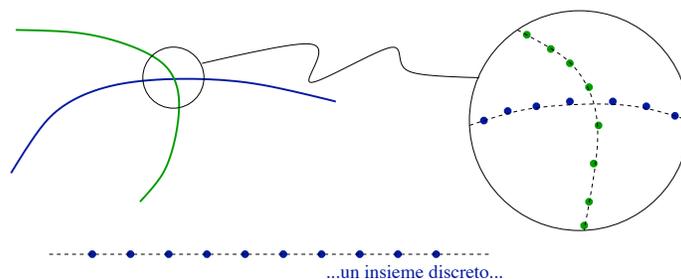


FIGURA 9. Versione “quantizzata”: una retta è un’unione di punti a distanza fissata.

“(QUASI) MAI”.

Versione “razionale”. Questa volta, immaginiamo le curve come rette, formate dall’unione di soli punti razionali, deformate. Qui il disegno non è facile: dato che in ogni intervallo cadono infiniti punti di \mathbb{Q} , nessun ingrandimento permette di riconoscere

“a occhio” se ci sia intersezione oppure no. Però sappiamo già che in alcuni casi non c’è intersezione: ad esempio, non esistono numeri razionali x tali che $x^2 = 2$, cioè le curve nel piano (x, y) (con x, y razionali) definite da $y = x^2$ e $y = 2$ non si intersecano. Il seguace di questa corrente di pensiero risponde “QUALCHE VOLTA”.

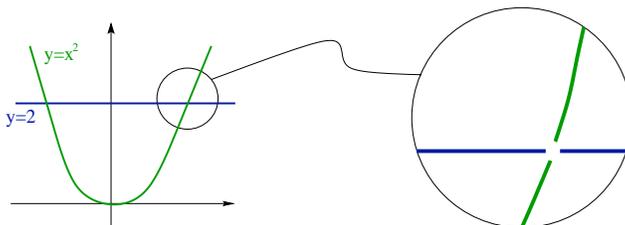


FIGURA 10. Versione “razionale”: la parabola $y = x^2$ e la retta $y = 2$ nel piano (x, y) con $x, y \in \mathbb{R}$ non si intersecano mai.

Versione “reale”. Infine c’è la versione reale: rette e curve costituiscono un “continuo” di punti senza interruzione. Dunque le due curve si intersecano sempre. Ogni ingrandimento del punto di incontro delle curve dà sempre e comunque lo stesso tipo di figura.

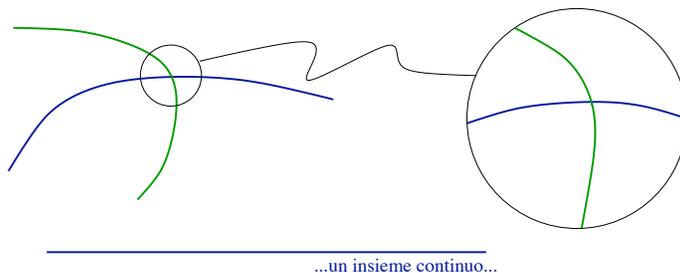


FIGURA 11. Versione “reale”: l’idea intuitiva di “continuo”.

Qual’è la risposta “esatta”? Tutte allo stesso tempo... Quello che conta, infatti, è decidere fin dall’inizio qual’è il tipo di visione che vogliamo prediligere e seguirla con coerenza e chiarezza. Scegliere una strada significa decidere qual’è l’ambiente base (discreto, razionale, continuo) con cui lavoriamo chiarendo bene quali siano le proprietà di cui gode. Tali proprietà vengono tradotte in *assiomi* (o *postulati*) che si dichiarano veri al principio. *La nostra scelta è di lavorare con l’insieme dei numeri reali.* I motivi sono tanti, primo fra tutti il fatto che la percezione del “continuo”, cioè di un universo “senza buchi” (seppure sbagliata a livello microscopico!), è estremamente naturale.

Occorre ora rendere chiaro cosa voglia dire con precisione la frase “l’insieme dei numeri reali non ha buchi”. Ci sono vari modi per esprimere attraverso un assioma la *completezza dei numeri reali*. Qui scegliamo di utilizzare quali assiomi di completezza

di \mathbb{R} la validità di due proprietà: il *postulato degli intervalli incapsulati* e il *principio di Archimede*. Per versioni equivalenti e maggiori chiarimenti sull'assioma di continuità bisogna ricorrere ad altri libri⁸.

Gli intervalli incapsulati. Un intervallo chiuso e limitato in \mathbb{R} è un insieme del tipo $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ con $a, b \in \mathbb{R}$.

Postulato degli intervalli incapsulati. *Per ogni successione di intervalli $I_0, I_1, \dots, I_n, \dots$ chiusi e limitati che siano incapsulati, cioè tali che $I_{n+1} \subset I_n$ per ogni $n \in \mathbb{N}$, esiste sempre almeno un punto $x_0 \in \mathbb{R}$ tale che $x_0 \in I_n$ per ogni n .*

Il postulato degli intervalli incapsulati (detto anche *principio di Cantor*) esprime il fatto che, se si prende una sequenza di intervalli chiusi e limitati in \mathbb{R} , ciascuno dei quali sia contenuto nel precedente (e quindi contenga il successivo), allora c'è sempre almeno un numero reale contenuto in tutti quanti gli intervalli.

L'assioma si può anche esprimere affermando che:

$$I_0, \dots, I_n, \dots \text{ intervalli chiusi e limitati, } I_{n+1} \subset I_n \quad \forall n \in \mathbb{N} \quad \Rightarrow \quad \bigcap_{n \in \mathbb{N}} I_n \neq \emptyset.$$

In \mathbb{Q} , il postulato degli intervalli incapsulati è falso! Infatti basta rappresentare \mathbb{Q} con punti di una retta orientata e “circondare” il punto $\ell = \sqrt{2}$ con una sequenza di intervalli incapsulati con lunghezza che diventa sempre più piccola in modo che l'intersezione degli intervalli sia data dal solo punto ℓ . Dato che $\ell \notin \mathbb{Q}$, non c'è nessun punto di \mathbb{Q} che sia nell'intersezione di questa sequenza di intervalli incapsulati.

L'assioma di Archimede. C'è un'altra proprietà che abbiamo usato senza grande preoccupazione perché estremamente intuitiva: se riproduciamo copie del segmento unitario della retta reale \mathbb{R} ... “arriviamo all'infinito”, cioè copriamo tutta la retta. In altre parole ogni elemento di \mathbb{R} è sempre contenuto in un intervallo che ha per estremi due numeri interi. Evidente, vero? Sì, ma tanto quanto la proprietà di “non avere buchi”, quindi, anche per questo fatto abbiamo bisogno di un assioma *ad hoc*, cioè una regola che stabiliamo valida una volta per tutte e che, quindi, potremo utilizzare tutte le volte che ci verrà utile.

Assioma di Archimede. *Per ogni numero reale a , esiste un numero naturale n più grande di a : in simboli,*

$$\forall a \in \mathbb{R}, \quad \exists n \in \mathbb{N}, \quad \text{tale che } a < n.$$

⁸Si veda, ad esempio, E.Giusti, “Esercizi e Complementi di Analisi Matematica, vol. I”, Bollati Boringhieri, cap.2.

Una conseguenza dell'assioma di Archimede verrà utilizzata migliaia di volte nel seguito. Eccola.

PROPOSIZIONE 4.1. *Vale la seguente implicazione*

$$x \leq \varepsilon \quad \forall \varepsilon > 0 \quad \Rightarrow \quad x \leq 0.$$

DIMOSTRAZIONE. Supponiamo per assurdo $x > 0$. Per la proprietà di Archimede, esiste $n \in \mathbb{N}$ tale che $\frac{1}{x} < n$. Dato che $x > 0$, anche $1/x > 0$, quindi n è diverso da zero. Moltiplicando $\frac{1}{x} < n$ per x/n (che è positivo), si deduce che

$$\exists n \in \mathbb{N} \quad \text{tale che} \quad 0 < \frac{1}{n} < x$$

che contraddice l'ipotesi. □

5. Estremo superiore ed estremo inferiore

Tra le conseguenze notevoli della completezza dei numeri reali, una delle più significative è la possibilità di definire due oggetti fondamentali dell'analisi matematica: l'*estremo superiore* e l'*estremo inferiore* di un sottoinsieme dell'asse reale.

DEFINIZIONE 5.1. *Un valore $\Lambda \in \mathbb{R}$ è un maggiorante di E se tutti gli elementi $y \in E$ verificano $y \leq \Lambda$; un valore $\lambda \in \mathbb{R}$ è un minorante di E se tutti gli elementi $y \in E$ verificano $\lambda \leq y$. Se esiste almeno un maggiorante per l'insieme $E \subset \mathbb{R}$, E si dice limitato superiormente; se esiste almeno un minorante per l'insieme $E \subset \mathbb{R}$, E si dice limitato inferiormente.*

In sostanza, un maggiorante Λ è una stima per eccesso di tutti gli elementi dell'insieme E e un minorante λ ne è una stima per difetto. Se, in qualche modo, siamo in grado di procurarci un minorante λ ed un maggiorante Λ , sappiamo già che l'insieme E è un sottoinsieme dell'intervallo chiuso $[\lambda, \Lambda]$.

Non è difficile convincersi che un insieme è limitato (secondo la Definizione 3.5) se e solo se è limitato superiormente e limitato inferiormente. Fatelo.

Il problema è che per uno stesso insieme E è possibile fornire stime diverse! Ad esempio, se

$$E := \left\{ \frac{1}{n} : n = 1, 2, 3, \dots \right\} = \left\{ 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots \right\},$$

è vero sia che $E \subset [0, 1]$ che $E \subset [-100, 100]$, ma, evidentemente la prima delle due stime è preferibile all'altra perché più "precisa". La domanda successiva è più che ovvia: è possibile determinare una stima che sia "ottimale"? Quando un maggiorante Λ (o un minorante λ) è elemento dell'insieme E è chiaro che Λ (o λ) dà la migliore stima per eccesso (per difetto). In questo caso si parla di *massimo* (o di *minimo*).

DEFINIZIONE 5.2. Massimo e minimo. Il valore $M \in \mathbb{R}$ è il massimo di E , e si scrive $M = \max E$, se

(i) M è un maggiorante di E , (ii) M è un elemento di E .

Analogamente, il valore $m \in \mathbb{R}$ è il minimo di E , e si scrive $M = \min E$, se

(i) m è un minorante di E , (ii) m è un elemento di E .

Quindi, se siamo informati (da qualche agenzia investigativa segreta) del fatto che $M = \max E$, sappiamo che tutto l'insieme E è limitato dall'alto da M e che questa è la stima migliore possibile.

ESERCIZIO 5.3. *Dimostrare che, se M è il massimo dell'insieme E , allora:*

se M' è un altro maggiorante di E , vale $M \leq M'$.

Il problema è che ci sono sottoinsiemi di \mathbb{R} che non hanno massimo o minimo, o nessuno dei due! Ad esempio, \mathbb{R} stesso, per il principio di Archimede, non ha né massimo né minimo. Se $E = \{1, 1/2, 1/3, \dots\}$, è facile convincersi che $1 = \max E$ e che E non ha elemento minimo. L'intervallo aperto $(0, 1)$ non ha invece né massimo né minimo. Quindi dato un qualsiasi sottoinsieme di \mathbb{R} non è detto che abbia senso scrivere $\max E$ e/o $\min E$. Bisogna, perciò, introdurre dei nuovi oggetti che siano ben definiti anche quando il massimo e/o il minimo non esistono. L'idea è semplice: sostituire la condizione (ii) della Definizione 5.2 (troppo restrittiva!), con la condizione dell'Esercizio 5.3.

DEFINIZIONE 5.4. Estremi superiore e inferiore. Il valore $\Lambda \in \mathbb{R}$ è l'estremo superiore di E , e si scrive $\Lambda = \sup E$, se

(i) Λ è un maggiorante di E , (ii) ogni maggiorante L di E verifica $\Lambda \leq L$.

Analogamente, il valore $\lambda \in \mathbb{R}$ è l'estremo inferiore di E , e si scrive $\Lambda = \inf E$, se

(i) λ è un minorante di E , (ii) ogni minorante ℓ di E verifica $\ell \leq \lambda$.

La proprietà (ii) dell'estremo superiore Λ garantisce che *non esiste una stima per eccesso migliore di Λ* . Ogni altro possibile maggiorante dell'insieme, necessariamente è maggiore (o uguale) a Λ . Similmente, la proprietà (ii) dell'estremo inferiore λ garantisce che *non esiste una stima per difetto migliore di λ* . In altre parole, l'estremo superiore è il più piccolo dei maggioranti e l'estremo inferiore è il più piccolo dei minoranti.

Dalle definizioni segue che, se E ammette massimo M (o minimo m), questo valore è anche l'estremo superiore (o estremo inferiore) di E . Infatti se $M = \max E$, dato che $M \in E$, si ha $M \leq L$ per ogni L maggiorante di E e quindi la condizione (ii) dell'estremo superiore è soddisfatta.

A prima vista passare dalla definizione di massimo/minimo a quella di estremo superiore/inferiore sembra una vera e propria truffa. Infatti, la definizione di estremo superiore discende da questa strategia:

- dato l'insieme E , costruirne l'insieme $F = \{\Lambda \in \mathbb{R}; x \leq \Lambda \forall x \in E\}$ dei maggioranti,
- dichiarare che l'estremo superiore di E è il minimo dell'insieme F : $\sup E = \min F$.

Ma (attenzione!) ci sono due tranelli in questa definizione: il primo è che l'insieme F potrebbe essere anche vuoto (a questo problema penseremo tra poco), il secondo, più grave, è che abbiamo appena dichiarato che non abbiamo nessuna garanzia che un sottoinsieme di \mathbb{R} abbia minimo! Dunque, chi ci assicura che $\min F$, e quindi $\sup E$, esista? Una proprietà fondamentale, che discende dall'assioma di continuità dei numeri reali, è che se esiste almeno un maggiorante (o minorante) dell'insieme E , allora *esiste sempre l'estremo superiore (o inferiore)*.

TEOREMA 5.5. Esistenza degli estremi superiore e inferiore. *Sia $E \subset \mathbb{R}$ un insieme non vuoto. Allora*

- (i) *se E è limitato superiormente, esiste $\Lambda = \sup E \in \mathbb{R}$;*
- (ii) *se E è limitato inferiormente, esiste $\lambda = \inf E \in \mathbb{R}$.*

Vedremo la dimostrazione di questo risultato più avanti.

ESEMPIO 5.6. Gli intervalli $(-1, 1)$, $[-1, 1)$, $(-1, 1]$ e $[-1, 1]$ hanno tutti estremo superiore uguale a 1 e estremo inferiore uguale a -1 , ma solo due di loro ammettono massimo e due ammettono minimo (quali?).

Rimane da decidere il da farsi nel caso in cui l'insieme dei maggioranti e/o quello dei minoranti siano vuoti:

- se non esistono maggioranti, E è **illimitato superiormente**, si scrive $\sup E = +\infty$;
- se non esistono minoranti E è **illimitato inferiormente**, si scrive $\inf E = -\infty$.

I simboli $+\infty$ e $-\infty$ non corrispondono a nessun numero reale: non si tratta di punti sulla retta reale! Nel seguito, se parlerà di “ x finito” nel caso in cui sia utile sottolineare che si tratta di un numero reale e non di uno dei simboli $\pm\infty$.

ESERCIZIO 5.7. *Dimostrare che $\inf \left\{ \frac{1}{n} : n = 1, 2, 3, \dots \right\} = 0$.*

Le definizioni di maggiorante, minorante, massimo, minimo, estremo superiore/inferiore, limitato superiormente/inferiormente, illimitato superiormente/inferiormente si basano sull'ordinamento di \mathbb{R} , cioè sul simbolo \leq (e varianti). Quindi non hanno estensioni ai sottoinsiemi del piano \mathbb{R}^2 , dello spazio \mathbb{R}^3 o di qualsivoglia altro oggetto privo di ordine!

CAPITOLO 2

Funzioni: anno zero

1. Ingredienti di base

In tutti i campi della scienza compaiono, in modo del tutto naturale, oggetti chiamati *funzioni*: la pressione di un gas ideale è funzione della densità e della temperatura, la posizione di una particella in movimento è funzione del tempo, il volume e la superficie di un cilindro sono funzioni del raggio e dell'altezza, etc. etc.. In generale, quando certe quantità a, b, c, \dots , dette *variabili dipendenti*, sono determinate da altre quantità x, y, z, \dots , dette *variabili indipendenti*, si dice che a, b, c, \dots “sono funzioni di” x, y, z, \dots o, in modo equivalente che a, b, c, \dots “dipendono da” x, y, z, \dots . L'idea è semplice: cambiando il valore delle variabili indipendenti, cambia il valore delle variabili dipendenti.

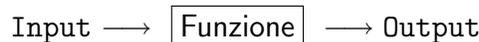
Ecco alcuni esempi tanto per cominciare.

- i. L'area A di un quadrato di lato ℓ è data da $A = \ell^2$, quindi la variabile dipendente *area* A è funzione della variabile indipendente *lato* ℓ .
- ii. Il Teorema di Pitagora afferma: *la lunghezza ℓ dell'ipotenusa è pari alla radice quadrata della somma dei quadrati delle lunghezze a e b dei cateti*, o, in formule, $\ell = \sqrt{a^2 + b^2}$. In questo caso, la lunghezza dell'ipotesa è una funzione delle lunghezze dei cateti: la variabile dipendente è ℓ , mentre le variabili indipendenti sono a e b .
- iii. Esistono anche oggetti che associano ad una sola variabile indipendente t , due variabili dipendenti x e y . Ad esempio, la funzione

$$x = t + 1, \quad y = 2 - t^3.$$

Interpretando x e y come coordinate di un punto P nel piano e t come il tempo, queste equazioni descrivono la posizione di P al tempo t , cioè il *moto* del punto P .

In generale, una **funzione** è una legge che associa ad ogni dato valore di una variabile (indipendente) un unico valore di un'altra variabile (dipendente). In termini più informatici, si può pensare alla variabile indipendente come **Input** della funzione e alla variabile dipendente come **Output**.



Nella prima parte di queste Note, approfondiremo il caso delle *funzioni che associano ad un numero reale un altro numero reale* (vedi esempio **i.**). In questa situazione, si usa la notazione

$$f : I \subset \mathbb{R} \rightarrow \mathbb{R},$$

che esprime che la funzione f è definita per valori $x \in I$ dove I è un assegnato sottoinsieme della retta reale \mathbb{R} e che la f trasforma x nel valore $y = f(x)$ di \mathbb{R} . Quindi, per definire una funzione occorre conoscere:

- i valori della variabile indipendente per cui la funzione f è considerata (l'insieme I);
- in quale insieme “vive” la variabile dipendente (qui l'insieme dei numeri reali);
- la regola definita dalla funzione f .

Volete qualche altro esempio? Eccovene un paio:

$$f(x) = x^2 \quad x \in \mathbb{R} \quad \text{oppure} \quad f(x) = \sqrt{1 - x^2} \quad -1 \leq x \leq 1$$

Frequentemente useremo i seguenti vocaboli, con cui ci si familiarizza col tempo.

Piccolo glossario per le funzioni

- x : variabile indipendente
- y : variabile dipendente
- I : dominio di definizione (o campo di esistenza)
- \mathbb{R} (di arrivo): codominio
- $f(x)$: immagine di x (o trasformato di x)
- x : (una) pre-immagine, o contro-immagine, di $f(x)$
- $f(I) = \{y \in \mathbb{R} : y = f(x) \text{ per qualche } x \in I\}$:
insieme immagine o immagine (di I tramite f)¹

OSSERVAZIONE 1.1. L'assegnazione di una funzione include anche la definizione del dominio della funzione. Funzioni con la stessa espressione analitica, ma differente dominio di definizione sono da considerarsi funzioni diverse! Ad esempio, la funzione $f(x) = x^2$ per $0 < x < 2$ non coincide con la funzione $g(x) = x^2$ per $x \in \mathbb{R}$, dato che il loro dominio di definizione è diverso.

Nell'esempio appena descritto, però, le funzioni f e g coincidono in $(0, 2)$, cioè nell'insieme in cui è definita la funzione f . In questo caso, si utilizza la definizione seguente.

DEFINIZIONE 1.2. Restrizione ed estensione. La funzione $f : I_f \subset \mathbb{R} \rightarrow \mathbb{R}$ è una restrizione della funzione $g : I_g \subset \mathbb{R} \rightarrow \mathbb{R}$ (e g è una estensione di f) se l'insieme di

¹In inglese, si parla di *range* della funzione f .

definizione di g contiene quello di f e le due funzioni coincidono dove sono definite entrambe, cioè se

$$I_f \subset I_g \quad e \quad f(x) = g(x) \quad \forall x \in I_f.$$

Usualmente, se una funzione viene assegnata dandone l'espressione analitica, ma senza specificarne l'insieme di definizione, si intende che la funzione è considerata nell'insieme più grande in cui le operazioni richieste sono lecite. Ad esempio, la funzione $f(x) = x^3 + 1$ si considera definita in $I = \mathbb{R}$, mentre la funzione $f(x) = \frac{1}{x}$ è definita in $I = \mathbb{R} \setminus \{0\}$.

Grafico di funzioni. Per individuare proprietà delle funzioni è utile realizzarne una rappresentazione grafica: il **grafico della funzione** f è il sottoinsieme del piano

$$\Gamma := \{(x, y) \in \mathbb{R}^2 : x \in I, y = f(x)\}.$$

Per iniziare, vediamo alcuni esempi.

(i) y è una “funzione affine” di x , cioè la funzione f è un polinomio di grado 1:

$$f(x) = ax + b \quad \text{per qualche } a, b \in \mathbb{R}, \quad a \neq 0.$$

Come è noto dalla geometria elementare, il grafico è una retta nel piano.

(ii) y è inversamente proporzionale a x ,

$$y = \frac{1}{x}.$$

Questa funzione è definita per $x \neq 0$ dato che la divisione per zero non ha senso. Il grafico rappresenta una *iperbole (rettangolare)*.

(iii) y è il quadrato di x ,

$$f(x) = x^2$$

come è ben noto questa funzione ha per grafico una parabola (vedi Fig.1(a)). Lo stesso vale per le funzioni del tipo $f(x) = ax^2 + bx + c$, con $a, b, c \in \mathbb{R}$, $a \neq 0$.

(iv) y è uguale a $|x|$. Dato che, per definizione

$$|x| := \begin{cases} x, & x \geq 0, \\ -x, & x < 0 \end{cases}$$

il grafico della funzione è composto da due semirette (vedi Fig.1(b)).

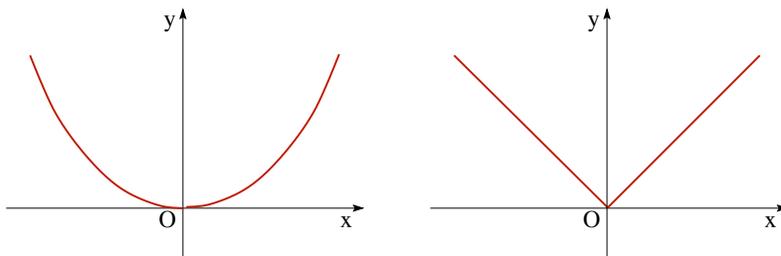


FIGURA 1. (a) La parabola $y = x^2$; (b) Il modulo: $y = |x|$.

Polinomi. Il tipo più semplice di funzione si ottiene utilizzando le sole operazioni di somma e moltiplicazione: un **polinomio** (di grado n) è una funzione della forma

$$y = a_0 + a_1x + \cdots + a_nx^n \quad a_n \neq 0.$$

dove a_0, a_1, \dots, a_n (con $a_n \neq 0$) sono $n+1$ numeri reali assegnati. Quindi $y = 3x+1$, $y = x^2 - 2x + 5$, $5x^{47} + 47x^5$ sono esempi di polinomi.

ESERCIZIO 1.3. *Disegnare i grafici delle seguenti funzioni*

$$f(x) = 1, \quad f(x) = x, \quad f(x) = 2x + 1, \quad f(x) = 2x^2 + x + 1.$$

E' una buona idea quella di sperimentare al calcolatore come siano fatti i grafici di polinomi. In particolare è interessante contare il “numero di oscillazioni” delle funzioni al variare del grado, dove per “numero di oscillazioni” si intende il numero delle zone in cui il grafico “sale” e di quelle in cui “scende”. Qual è la regola generale?

Funzioni razionali. I rapporti di polinomi sono dette **funzioni razionali**

$$y = \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m} \quad \text{con } a_i, b_j \in \mathbb{R}, \quad (b_j \text{ non tutti nulli})$$

e sono definite per tutti i valori di x per cui il denominatore è diverso da zero. Per quanto riguarda gli zeri della funzione, questi sono tutti e soli gli zeri del polinomio a numeratore (l'unico modo per ottenere zero da un rapporto è che il numeratore sia zero). Lo studio del segno si traduce invece in un sistema di disequazioni.

Una buona classe per iniziare lo studio delle funzioni razionali è

$$f(x) = \frac{ax + b}{cx + d} \quad a, b, c, d \in \mathbb{R}.$$

Ad esempio, consideriamo la funzione

$$f(x) = \frac{2x + 3}{x + 1}.$$

L'insieme di definizione è $I = \{x : x \neq -1\}$, inoltre la funzione è positiva per $x > 1$ e per $x \leq -3/2$ e negativa nel resto dell'insieme. Il grafico è in Figura 2(a).

Un altro esempio di funzione razionale facile è $f(x) = 1/x^2$ (vedi Figura 2(b)).

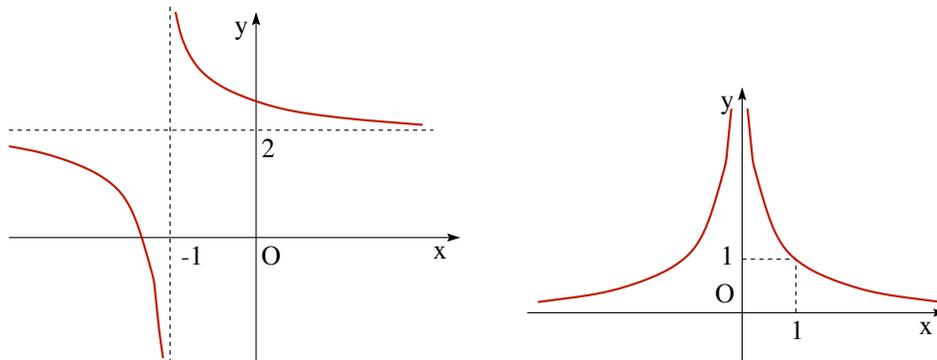


FIGURA 2. (a) La funzione $y = (2x + 3)/(x + 1)$; (b) la funzione $y = 1/x^2$.

Funzioni trigonometriche. Non è possibile in poche righe ricordare tutto il necessario sulle funzioni trigonometriche. Qui ci limitiamo alle proprietà principali. Le funzioni trigonometriche di base sono $\sin x$ e $\cos x$ le cui proprietà fondamentali sono:

- entrambe sono definite per ogni valore reale x ;
- $\cos 0 = 1$ e $\sin 0 = 0$;
- per ogni $x \in \mathbb{R}$, si hanno $\cos(x + 2\pi) = \cos x$ e $\sin(x + 2\pi) = \sin x$;
- per ogni $x \in \mathbb{R}$, vale la relazione $\cos^2 x + \sin^2 x = 1$;
- per ogni $\alpha, \beta \in \mathbb{R}$, valgono le formule di somma e sottrazione

$$\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta \quad \sin(\alpha \pm \beta) = \sin \alpha \cos \beta \pm \cos \alpha \sin \beta$$

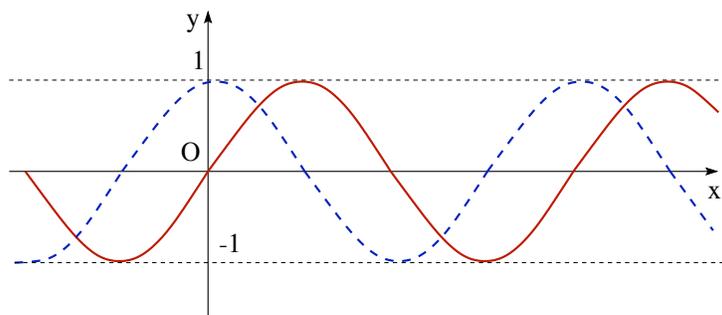


FIGURA 3. Il grafico della funzione $\sin x$ (linea continua) e della funzione $\cos x$ (linea tratteggiata).

Dato che $\sin^2 x + \cos^2 x = 1$, è sempre vero che $|\sin^2 x|, |\cos^2 x| \leq 1$ e quindi

$$|\sin x| \leq 1, \quad |\cos x| \leq 1 \quad \forall x \in \mathbb{R}.$$

ESERCIZIO 1.4. *Dimostrare che $|\sin x| \leq |x|$ per ogni $x \in \mathbb{R}$.*

ESERCIZIO 1.5. *Dedurre, dalle formule di somma e sottrazione, la formula*

$$\sin x - \sin y = 2 \cos \left(\frac{x+y}{2} \right) \sin \left(\frac{x-y}{2} \right).$$

Soluzione. Poniamo $\xi = \frac{x+y}{2}$ e $\eta = \frac{x-y}{2}$. Allora $x = \xi + \eta$ e $y = \xi - \eta$. Dunque $\sin x - \sin y = \sin(\xi + \eta) - \sin(\xi - \eta) = \sin \xi \cos \eta + \cos \xi \sin \eta - \sin \xi \cos \eta + \cos \xi \sin \eta = 2 \cos \xi \sin \eta$, e ricordando la definizione di ξ ed η si giunge alla conclusione.

Tramite le funzioni $\sin x$ e $\cos x$ si definiscono le funzioni **tangente** e **cotangente**:

$$\tan x := \frac{\sin x}{\cos x} \quad \text{e} \quad \cot x := \frac{\cos x}{\sin x}.$$

Dalla definizione e dalle proprietà di seno e coseno, discende che $\tan x$ è definita per $x \neq \frac{\pi}{2} + k\pi$ per $k \in \mathbb{Z}$ e $\cot x$ è definita per $x \neq k\pi$ per $k \in \mathbb{Z}$.

2. Operazioni elementari su grafici

Una volta noto il grafico di una funzione f è possibile, a partire da questo, ricostruire il grafico di altre funzioni g che si ottengano dalla prima per via elementare. Vediamo alcuni esempi significativi, tenendo conto che, qui, l'unica maniera per capire è *sperimentare* (anche usando un computer o una calcolatrice grafica, se possibile).

(i) *Traslazioni.* Il grafico di $g(x) = f(x) + c$ dove $c \in \mathbb{R}$ è dato da una traslazione in verticale del grafico di f della quantità c (la traslazione sarà quindi verso l'alto se $c > 0$ e verso il basso se $c < 0$).

Il grafico di $g(x) = f(x + c)$ dove $c \in \mathbb{R}$ è dato da una traslazione in orizzontale di $-c$ del grafico di f . Nota bene! La traslazione è di $-c$, quindi è verso sinistra se $c > 0$ e verso destra se $c < 0$.

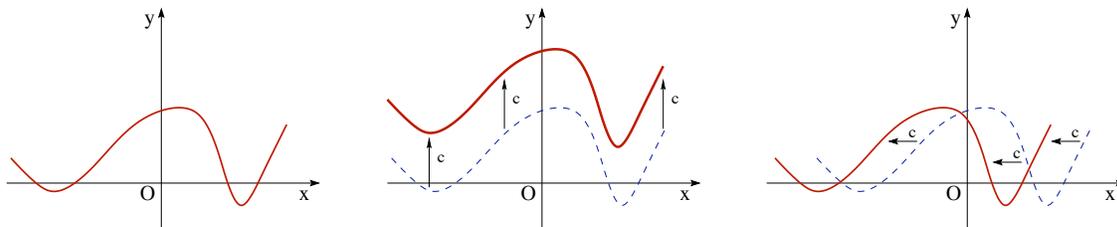


FIGURA 4. I grafici di (a) $y = f(x)$, (b) $y = f(x) + c$; e (c) $y = f(x + c)$.

(ii) *Dilatazioni/Compressioni.* Il grafico di $g(x) = kf(x)$ è ottenuto dilatando la variabile dipendente di un fattore k , il grafico è pertanto dilatato nella direzione verticale. Il grafico di $g(x) = f(kx)$ è ottenuto dilatando la variabile indipendente di un fattore

$1/k$, quello che per la funzione f accadeva in x ora per la funzione g accade in x/k . Questo vuol dire che se $k > 1$ il grafico risulta compresso in orizzontale verso l'asse y , mentre se $k < 1$ il grafico risulta dilatato. Un esempio? Fate il grafico di

$$f(x) = |x| - 1, \quad g(x) = |2x| - 1, \quad h(x) = \left| \frac{x}{2} \right| - 1.$$

Visto che ci siete, fate anche $l(x) = 2||x| - 1|$ e $m(x) = \frac{1}{2}||x| - 1|$.

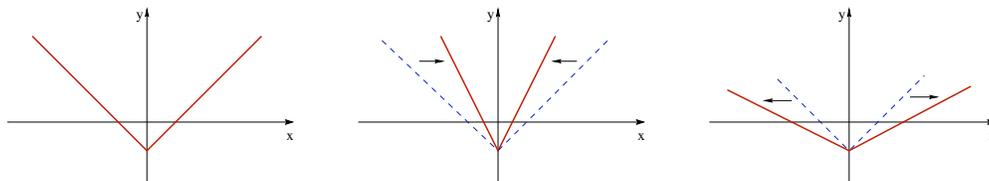


FIGURA 5. I grafici di (a) $y = f(x) = |x| - 1$, (b) $g(x) = |2x| - 1$, (c) $h(x) = \left| \frac{x}{2} \right| - 1$.

ESERCIZIO 2.1. *Disegnare i grafici delle funzioni*

$$f(x) = ||x| - 1|, \quad g(x) = ||3x| - 1|, \quad h(x) = \frac{1}{3}||x| - 1|.$$

(iii) *Somma/Sottrazione.* Dati i grafici di f e g è possibile stabilire un andamento qualitativo anche delle funzioni $h = f + g$ e $l = f - g$. Basta disegnare i due grafici di f e g sullo stesso piano (x, y) e poi calcolare punto per punto la somma e la differenza. Nel caso della differenza, il significato del grafico è di “distanza” con segno (cioè l è positiva se f è sopra g e negativa se f è sotto g) tra i punti, aventi stessa ascissa, dei grafici delle due funzioni. Quindi la “distanza” qui è calcolata in verticale (non è la distanza nel piano...).

(iv) *Passaggio al reciproco.* Dato il grafico della funzione f è possibile anche determinare i grafici delle funzioni $g(x) = \frac{1}{f(x)}$ e $h(x) = f(1/x)$. Come si dovrebbe essere capito dai casi precedenti, nel primo caso si ottiene una trasformazione “in verticale” (nel senso della variabile dipendente y), mentre nel secondo “in orizzontale” (nel senso della variabile dipendente x).

Il grafico di g si ottiene dalla f notando che i valori che vengono mandati da f vicino a zero sono trasformati per g in valori grandi, mentre i valori che la f trasforma in valori grandi, sono mandati da g in zero. I valori che vanno in ± 1 rimangono gli stessi. Un grafico di quel che fa la trasformazione $t \rightarrow s = 1/t$ dall'asse t all'asse s aiuta a capire cosa sta succedendo.

Per quanto riguarda il grafico della funzione h , questa volta l'inversione è compiuta sulla variabile indipendente x , quindi l'inversione è in orizzontale.

(v) *Modulo di una funzione.* Una classe significativa è quella delle funzioni della forma $g(x) = |f(x)|$ dove si suppone noto il grafico della funzione f . Dato che il modulo $|\cdot|$ trasforma un numero in sé stesso se è positivo, e nel suo opposto se è negativo, per fare il grafico di g basta lasciare invariata la parte del grafico di f che corrisponde a valori positivi della variabile dipendente, cioè la parte che è sopra l'asse delle x , e ribaltare attorno all'asse x la parte del grafico che si trova al di sotto.

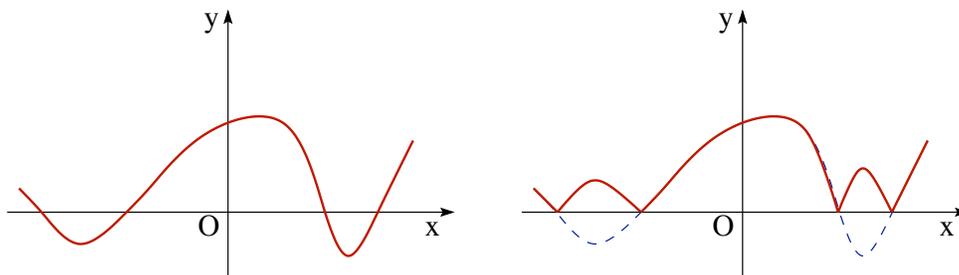


FIGURA 6. I grafici di (a) $y = f(x)$, (b) $y = |f(x)|$.

(vi) *Parte positiva e parte negativa di una funzione.* Data una funzione f , la funzione $\max\{f(x), 0\}$ è detta **parte positiva** di f e la funzione $\max\{-f(x), 0\}$ è detta **parte negativa** di f . Il grafico della prima delle due si ottiene molto facilmente a partire da quello della f : coincide con quest'ultimo dove $f(x) \geq 0$ e, in quel che resta, coincide con l'asse delle x . Il grafico della parte negativa richiede un piccolo sforzo in più: prima si ribalta il grafico della funzione f attorno all'asse x (cioè si disegna il grafico della funzione $-f$) e poi si procede come per la parte positiva. Si noti, in particolare, che sia il grafico della parte positiva che quello della parte negativa giacciono nel semipiano $y \geq 0$.

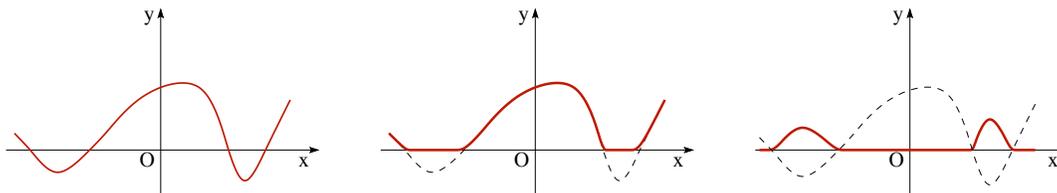


FIGURA 7. I grafici di (a) $y = f(x)$, (b) $y = \max\{f(x), 0\}$, (c) $y = \max\{-f(x), 0\}$.

Simmetrie di grafici. Capita spesso che le funzioni che si studiano abbiano della *simmetrie*, cioè abbiano la proprietà che il loro grafico rimane immutato quando vengono compiute certe trasformazioni. Ad esempio, il grafico di una funzione costante, dato che è una retta orizzontale, rimane invariato se viene traslato in orizzontale. Se si riconosce una simmetria di una funzione, lo studio è in genere semplificato, perché il

grafico può essere determinato studiandone semplicemente una parte e poi applicando una trasformazione opportuna. Vediamo rapidamente i principali tipi di simmetria.

Se il grafico di una funzione f è simmetrico rispetto all'asse y si dice che la funzione è **pari**. Analiticamente, questa proprietà corrisponde a

$$\text{funzione pari:} \quad f(-x) = f(x) \quad \forall x \in I.$$

Ad esempio le funzioni $y = x^2$, $y = |x|$ sono funzioni pari.

Se il grafico è simmetrico rispetto all'origine, la funzione è **dispari**

$$\text{funzione dispari:} \quad f(-x) = -f(x) \quad \forall x \in I.$$

Ad esempio, le funzioni $y = x^3$ e $y = 1/x$ sono dispari.

Le funzioni pari più semplici sono i polinomi che includano solo potenze pari di x . Le funzioni dispari più semplici sono i polinomi che includano solo potenze dispari di x . La funzione $\cos x$ è pari: sono pari quindi somme, differenze e prodotti di $\cos x$. La funzione $\sin x$ è dispari. E' vero che somme/prodotti di $\sin x$ sono dispari?

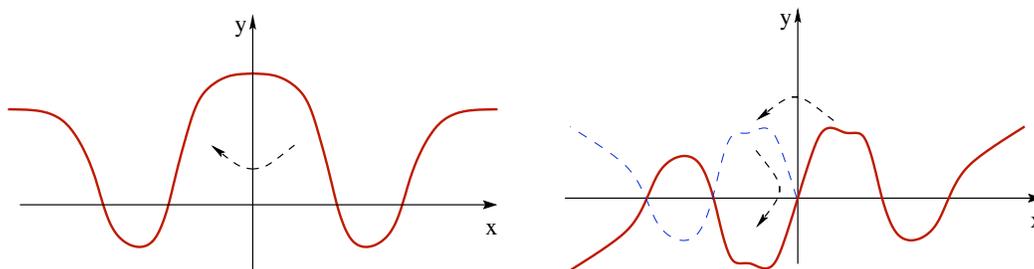


FIGURA 8. (a) una funzione pari, (b) una funzione dispari.

Una funzione $y = f(x)$ si dice **periodica** se

$$\text{funzione periodica:} \quad \exists T > 0 \text{ tale che } \forall x \quad f(x + T) = f(x).$$

Qualora esista, il più piccolo valore T per cui vale questa proprietà è il **periodo** della funzione f . Graficamente questa proprietà corrisponde al fatto che il grafico può essere ricostruito con “copia/incolla”: si determina il grafico della funzione in un intervallo di lunghezza T e poi lo si riproduce a destra e a sinistra dell'intervallo. Esempi di funzioni periodiche sono le funzioni trigonometriche: $\sin x$, $\cos x$, $\tan x$, $\cot x$,

$$\sin(x + 2\pi) = \sin x, \quad \cos(x + 2\pi) = \cos x,$$

$$\tan(x + \pi) = \tan x, \quad \cot(x + \pi) = \cot x.$$

Tanto per provare, verifichiamo la periodicità di $\tan x$:

$$\tan(x + \pi) = \frac{\sin(x + \pi)}{\cos(x + \pi)} = \frac{\sin x \cos \pi + \cos x \sin \pi}{\cos x \cos \pi - \sin x \sin \pi} = \frac{-\sin x}{-\cos x} = \frac{\sin x}{\cos x} = \tan x$$

ESERCIZIO 2.2. *Qualcuna delle funzioni seguenti è pari/dispari/periodica?*

$$\cos(x^2), \quad |\cos x|, \quad x|x|.$$

Soluzione. La prima funzione è pari, infatti $f(-x) = \cos((-x)^2) = \cos(x^2) = f(x)$. Anche la seconda è pari (verificare!). E' anche periodica di periodo π , infatti

$$|\cos(x + \pi)| = |\cos(x)\cos(\pi) - \sin(x)\sin(\pi)| = |-\cos(x)| = |\cos x|.$$

La terza funzione è dispari: $f(-x) = -x|-x| = -x|x| = -f(x)$.

Un altro esempio (meno frequente) di funzione periodica è la *parte frazionaria*. Sia

$$\text{parte intera di } x: \quad [x] := z \in \mathbb{Z},$$

dove z è l'unico intero per cui $z \leq x < z + 1$. Allora la funzione

$$\text{parte frazionaria (o mantissa) di } x: \quad \{x\} := x - [x]$$

è una funzione periodica, con periodo 1 (verificare!).

3. Funzioni invertibili e funzioni monotone

Data $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$ e dato $y \in \mathbb{R}$, un problema tipico è determinare se ci sono (e quante) soluzioni di $f(x) = y$. In altri termini, data y si vuole sapere quante sono le sue pre-immagini tramite f . Per definizione di insieme immagine (vd. il "Piccolo Glossario per Funzioni"), il problema ammette almeno una soluzione se e solo se $y \in f(I)$.

Problema: sia $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$ una funzione assegnata,

dato $y \in \mathbb{R}$, quante soluzioni esistono dell'equazione $f(x) = y$?

DEFINIZIONE 3.1. *Una funzione f è iniettiva se manda valori di x diversi in valori y diversi, ossia*

$$f \text{ è iniettiva se } f(x_1) = f(x_2) \Rightarrow x_1 = x_2.$$

In parole povere, se l'equazione $f(x) = y$ ammette sempre non più di una soluzione la funzione f si dice *iniettiva* (o *uno a uno*). Potrebbero però esserci dei valori y per cui il problema non ha soluzione.

Graficamente, l'iniettività corrisponde al fatto che *rette parallele all'asse delle x intersecano il grafico della funzione f al più una volta*. Ad esempio, la funzione x è iniettiva, mentre le funzioni x^2 e $|x|$ non lo sono.

Vediamo, in alcuni esempi concreti, come verificare se una funzione è iniettiva. La strategia pratica è supporre che valga l'uguaglianza $f(x_1) = f(x_2)$ per x_1, x_2 generici, e domandarsi se ne segue $x_1 = x_2$:

$$f(x_1) = f(x_2) \stackrel{?}{\Rightarrow} x_1 = x_2$$

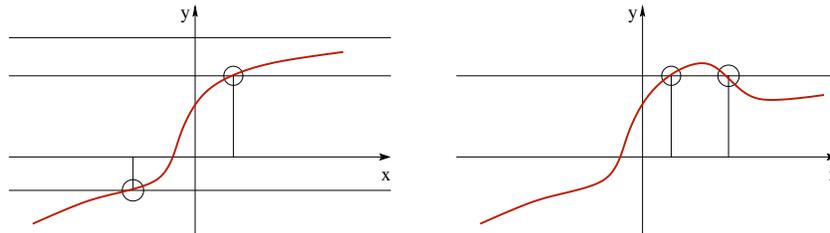


FIGURA 9. (a) una funzione iniettiva, (b) una funzione non iniettiva.

ESEMPIO 3.2. Consideriamo la funzione $f(x) = 3x - 2$ e supponiamo che esistano x_1, x_2 tale che $f(x_1) = f(x_2)$. Allora

$$3x_1 - 2 = 3x_2 - 2 \iff 3x_1 = 3x_2 \iff x_1 = x_2,$$

quindi la funzione è iniettiva. Alla stessa conclusione si giunge disegnando il grafico della retta $y = 3x - 2$ e osservando che, dato che la retta è obliqua, la proprietà geometrica dell'iniettività è soddisfatta.

ESEMPIO 3.3. Consideriamo

$$f(x) = \frac{x+1}{x+2} \quad I = \{x \neq -2\}.$$

Studiamone l'iniettività: siano $x_1, x_2 \in I$ tali che

$$\begin{aligned} \frac{x_1+1}{x_1+2} = \frac{x_2+1}{x_2+2} &\iff (x_2+2)(x_1+1) = (x_2+1)(x_1+2) \\ &\iff 2x_1 + x_2 = 2x_2 + x_1 \iff x_1 = x_2. \end{aligned}$$

Quindi la funzione è iniettiva.

ESEMPIO 3.4. Consideriamo la funzione $f(x) = x^2 + x$. Pensando al suo grafico, è evidente che non si tratta di una funzione iniettiva, ma come si può riconoscere questo fatto direttamente dai conti algebrici? Procediamo come in precedenza:

$$x_1^2 + x_1 = x_2^2 + x_2 \iff (x_1 - x_2)(x_1 + x_2 + 1) = 0.$$

Se supponiamo $x_1 \neq x_2$, allora otteniamo $x_1 + x_2 + 1 = 0$, da leggersi come una condizione su x_1 e x_2 che garantisce $f(x_1) = f(x_2)$. Ad esempio, scegliendo $x_1 = -1$ e $x_2 = 0$, otteniamo

$$x_1 = -1, \quad x_2 = 0 \implies f(x_1) = f(x_2) = 0,$$

quindi la funzione non è iniettiva.

Se una funzione è iniettiva, per tutti i valori y per cui l'equazione $y = f(x)$ ha soluzione è possibile associare un unico valore x dato proprio dalla soluzione dell'equazione $f(x) = y$. In altre parole, se una funzione è iniettiva è possibile definire

una nuova funzione che permette di “tornare indietro”, cioè che associa ad ogni y dell’insieme immagine, l’unico valore x di cui è immagine.

DEFINIZIONE 3.5. *Data una funzione f iniettiva, la funzione $f^{-1} : f(I) \subset \mathbb{R} \rightarrow \mathbb{R}$ che gode della proprietà*

$$f(f^{-1}(y)) = y \quad f^{-1}(f(x)) = x \quad \forall y \in f(I), x \in I,$$

si dice funzione inversa di f . La funzione f si dice invertibile.

Per determinarne l’inversa f^{-1} dobbiamo, sostanzialmente, esplicitare la funzione in termini della variabile y , cioè, a partire dall’espressione $y = f(x)$, arrivare ad una espressione *equivalente* della forma $x = f^{-1}(y)$.

ESEMPIO 3.6. Nel caso della funzione $f(x) = 3x - 2$, si ha

$$y = 3x - 2 \quad \iff \quad 3x = y + 2 \quad \iff \quad x = \frac{y + 2}{3}.$$

La funzione inversa è $f^{-1}(y) = \frac{y + 2}{3}$.

ESEMPIO 3.7. Per la funzione

$$f(x) = \frac{x + 1}{x + 2} \quad I = \{x \neq -2\}.$$

si ha

$$y = \frac{x + 1}{x + 2} \quad \iff \quad xy + 2y = x + 1 \quad \iff \quad x(y - 1) = 1 - 2y \quad \iff \quad x = \frac{1 - 2y}{y - 1}.$$

Tale espressione ha senso solo per $y \neq 1$. In effetti, nel caso $y = 1$, si trova la relazione $x + 1 = x + 2$ cioè $1 = 2$ che è falsa, quindi $1 \notin f(I)$. La funzione inversa è definita in $\{y \neq 1\}$ ed è data da

$$f^{-1}(y) = \frac{1 - 2y}{y - 1}.$$

Conoscendo il grafico di una funzione f , si può sempre ottenere il grafico dell’inversa f^{-1} . Infatti, dato che $y = f(x)$ se e solo se $x = f^{-1}(y)$, si ha $(x, f(x)) = (f^{-1}(y), y) \in \Gamma$; quindi il grafico della funzione inversa si ottiene scambiando il ruolo dell’asse x e dell’asse y , ossia ribaltando il grafico attorno alla retta $y = x$ (vedere Figura 10).

Funzioni monotone. Una funzione $y = f(x)$ il cui valore immagine cresce se cresce la variabile indipendente, cioè tale che, per ogni $x, x' \in I$,

$$x < x' \quad \iff \quad f(x) < f(x')$$

si dice **monotona strettamente crescente** in I o, più semplicemente, **strettamente crescente** in I . Analogamente, se, per ogni $x, x' \in I$,

$$x < x' \quad \iff \quad f(x) > f(x'),$$

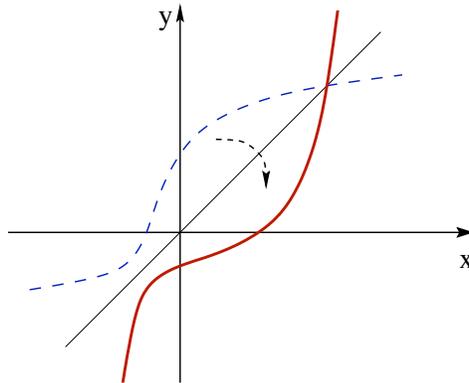


FIGURA 10. Il grafico della funzione inversa di una funzione assegnata.

la funzione è monotona strettamente decrescente o semplicemente strettamente decrescente. Equivalentemente, si può scrivere

$$\begin{aligned} f \text{ strettamente crescente} &\iff [f(x) - f(x')](x - x') > 0 & \forall x \neq x' \\ f \text{ strettamente decrescente} &\iff [f(x) - f(x')](x - x') < 0 & \forall x \neq x' \end{aligned}$$

Per ogni $n \in \mathbb{N}$, la funzione $y = x^n$ ristretta a $x \geq 0$ è una funzione monotona strettamente crescente (come si è visto nella sezione delle disequazioni). Più precisamente, per n dispari, la funzione x^n è strettamente crescente in \mathbb{R} (quindi anche per i negativi), mentre per n pari la funzione x^n non è monotona in \mathbb{R} .

ESERCIZIO 3.8. *Dimostrare che se f e g sono funzioni strettamente crescenti allora anche la funzione $f + g$ è strettamente crescente. E' vero che anche la funzione fg è strettamente crescente?*

Soluzione. Per ipotesi, se $x < y$, allora $f(x) < f(y)$ e $g(x) < g(y)$. Quindi, sommando i termini di destra e i termini di sinistra, si ottiene

$$f(x) + g(x) < f(y) + g(y) \quad \forall x, y \quad x < y.$$

La risposta alla domanda finale è "NO". Basta infatti considerare $f(x) = g(x) = x$, che è strettamente crescente, mentre $f(x)g(x) = x^2$ non lo è. Quale ipotesi aggiuntiva occorre per dedurre che il prodotto di funzioni strettamente crescenti è crescente?

Chiaramente valgono le implicazioni

$$f \text{ strettamente monotona} \implies f \text{ iniettiva} \iff f \text{ invertibile}$$

Ad esempio, le funzioni x^n per $x \geq 0$ e n pari e x^n per $x \in \mathbb{R}$ e n dispari sono funzioni invertibili.

OSSERVAZIONE 3.9. La monotonia di una funzione f ne garantisce l'invertibilità senza passare per la determinazione della formula per la funzione inversa: la funzione

f^{-1} c'è, ma non si vede! Ad esempio, per ogni n dispari, la funzione $f(x) = x + x^n$ è una funzione strettamente crescente, dato che è somma di funzioni strettamente crescenti. Quindi è anche una funzione invertibile.

Se la funzione f è strettamente crescente/decrescente, anche la sua inversa lo è. Infatti, siano $y = f(x)$ e $y' = f(x')$, allora $x = f^{-1}(y)$ e $x' = f^{-1}(y')$ e quindi

$$[f(x) - f(x')](x - x') = (y - y')[f^{-1}(y) - f^{-1}(y')].$$

Quindi il segno del secondo membro è lo stesso del primo, ossia le funzioni f e f^{-1} hanno lo stesso tipo di monotonia.

Nel caso in cui, nella definizione di monotonia, il simbolo di disequazione $<$ venga sostituito con la versione indebolita \leq e $>$ con \leq si parla di funzioni **non decrescenti** o **non crescenti**: Attenzione però al fatto che le funzioni non crescenti e quelle non decrescenti possono essere non iniettive, e quindi non invertibili (ad esempio, le funzioni costanti!).

4. Classi di funzioni più o meno comuni

Radici n -esime. Sia n un numero naturale $n \geq 2$. Dato che $y = x^n$ è strettamente crescente per $x \geq 0$, essa è iniettiva e quindi invertibile. La sua inversa si indica con

$$y = \sqrt[n]{x} = x^{1/n}.$$

Per definizione questa radice è sempre non negativa.

Per n dispari, però, la funzione x^n è strettamente crescente per tutti i valori $x \in \mathbb{R}$ (quindi anche per i negativi) e, di conseguenza, *per n dispari, $\sqrt[n]{x}$ è definita per tutti i valori della x* ; in questo caso $\sqrt[n]{x}$ è negativa per x negativa.

Più in generale, data una funzione g , possiamo considerare funzioni della forma

$$f(x) = \sqrt[n]{g(x)}.$$

Nel caso in cui n sia pari, una funzione di questo genere è definita solo per i valori della x per cui $g(x) \geq 0$. Ad esempio, dove è definita la funzione $f(x) = \sqrt{1 - x^2}$? Facile. Basta imporre la condizione $1 - x^2 \geq 0$, quindi per $x \in [-1, 1]$. Invece la funzione $h(x) = \sqrt[3]{1 - x^2}$ è definita per ogni valore x .

Inverse delle funzioni trigonometriche. Le funzioni trigonometriche sono periodiche, quindi non iniettive e non invertibili se considerate in tutto l'insieme in cui sono definite. Opportune restrizioni di queste funzioni sono però monotone e quindi invertibili. Vediamole in dettaglio.

Funzione arcoseno. La funzione $f(x) = \sin x$ è una funzione periodica su \mathbb{R} , quindi ad ogni elemento della sua immagine corrispondono infinite pre-immagini. Ad esempio,

$f^{-1}(1) = \{\frac{\pi}{2} + 2k\pi : k \in \mathbb{Z}\}$. Perciò la funzione $\sin x$ non è invertibile. Invece, la restrizione di $\sin x$ all'intervallo $[-\frac{\pi}{2}, \frac{\pi}{2}]$ è una funzione crescente e quindi anche invertibile. Questa inversa si chiama (**funzione**) **arcoseno** e si indica con $\arcsin x$. Il dominio dell'arcoseno è $[-1, 1]$ e l'insieme immagine è $[-\frac{\pi}{2}, \frac{\pi}{2}]$:

$$\arcsin : [-1, 1] \longrightarrow \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

Infine, dato che la funzione $\sin x$ è crescente nell'intervallo in cui la stiamo considerando, anche $\arcsin x$ è crescente.

Allo stesso modo si sarebbe potuto decidere di invertire la restrizione della funzione $\sin x$ ad un altro intervallo, ad esempio in $[\frac{3\pi}{2}, \frac{5\pi}{2}]$. La scelta dell'intervallo $[-\frac{\pi}{2}, \frac{\pi}{2}]$ è puramente convenzionale, per questo, qualche volta, si dice che $\arcsin x$ è il *valore principale dell'arcoseno*.

Funzione arcocoseno. In modo analogo, considerando la restrizione della funzione $\cos x$ all'intervallo $[0, \pi]$ e osservando che tale funzione è decrescente, è possibile definire la sua funzione inversa

$$\arccos : [-1, 1] \longrightarrow [0, \pi]$$

detta (**funzione**) **arcocoseno** (o anche *valore principale dell'arcocoseno*). Dalla monotonia di $\cos x$ in $[0, \pi]$ discende che la funzione $\arccos x$ è decrescente nel suo insieme di definizione.

Funzione arcotangente. La funzione $\tan x$ ristretta all'intervallo $(-\frac{\pi}{2}, \frac{\pi}{2})$ è crescente e quindi invertibile, con inversa crescente. La sua inversa si indica con $\arctan x$ ed è detta (**funzione**) **arcotangente**:

$$\arctan : \mathbb{R} \longrightarrow \left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Anche per l'arcotangente (come per arcoseno e arcocoseno) si sarebbe potuto decidere di invertire $\tan x$ in un altro intervallo.

Esponenziali e logaritmi (costruzione naïf). Oltre alle funzioni elementari sono importanti le funzioni esponenziali con base $a > 0$ e le loro inverse, i logaritmi in base $a > 0$:

$$y = a^x \quad \text{e} \quad y = \log_a x.$$

Diamo qui solo una definizione “leggera” (non rigorosa) delle funzioni esponenziali. Fissiamo $a > 0$, se $x = p/q \in \mathbb{Q}$ è possibile definire a^x

$$\text{fissato } a > 0, \quad a^x := \sqrt[q]{a^p} \quad \forall x = \frac{p}{q} \in \mathbb{Q},$$

dove la radice (come sempre) è scelta positiva. Per definire il valore a^x anche nel caso in cui x sia un numero irrazionale, è naturale approssimare $x \in \mathbb{R} \setminus \mathbb{Q}$ con numeri razionali sempre più vicini.

Per le funzioni esponenziali valgono le proprietà

- (i) $a^x > 0 \quad \forall x \in \mathbb{R}$,
- (ii) $a^0 = 1$,
- (iii) $a^x a^y = a^{x+y}$,
- (iv) $(a^x)^y = a^{xy}$
- (v) a^x è $\begin{cases} \text{decrescente} & \text{se } 0 < a < 1 \\ \text{crescente} & \text{se } a > 1. \end{cases}$

Dato che a^x è monotona per $a \neq 1$, essa è invertibile. La funzione inversa è $y = \log_a x$: è definita sull'insieme immagine dell'esponenziale, quindi solo per $x > 0$ ed associa ad x l'unico valore y che verifica $a^y = x$. Per i logaritmi valgono le proprietà

- (i) $\log_a x$ è definito per $x > 0$,
- (ii) $\log_a 1 = 0$,
- (iii) $\log_a x + \log_a y = \log_a(xy)$,
- (iv) $\log_a(x^\alpha) = \alpha \log_a x$
- (v) $\log_a x$ è $\begin{cases} \text{decrescente} & \text{se } 0 < a < 1 \\ \text{crescente} & \text{se } a > 1 \end{cases}$

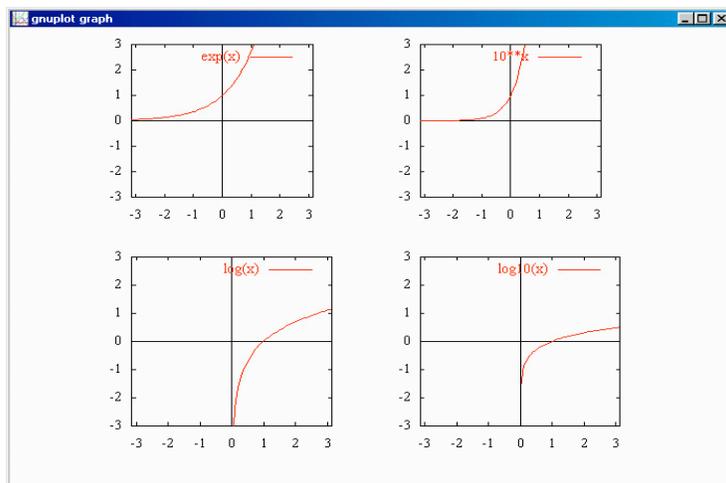


FIGURA 11. (a) $y = e^x$, (b) $y = 10^x$, (c) $y = \ln x$, (d) $y = \log_{10} x$

OSSERVAZIONE 4.1. Perché non si può definire “ a elevato ad x ” per a negativi? In effetti, se $a < 0$, l’espressione a^x ha senso per x del tipo p/q con p e q interi e q dispari. Quindi si potrebbe sperare di estendere questa strana funzione in modo da definire a^x anche per tutti gli altri valori di $x \in \mathbb{R}$. Ma non esiste una estensione che preservi le belle proprietà che abbiamo elencato per l’esponenziale. Ecco un esempio: supponiamo per un attimo che a^x sia ben definita per ogni a e per ogni x , allora $-1 = (-1)^1 = (-1)^{2 \cdot \frac{1}{2}} = [(-1)^2]^{\frac{1}{2}} = 1^{\frac{1}{2}} = 1$. Ops!

Funzioni composte. Si possono generare funzioni anche con la *composizione di funzioni*: se $\phi : \mathbb{R} \rightarrow \mathbb{R}$ e $g : \mathbb{R} \rightarrow \mathbb{R}$, allora la formula

$$f(x) := g(\phi(x))$$

definisce una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$. Ad esempio,

$$f(x) = \sin(1 + x^2) = g(\phi(x)) \quad \text{dove} \quad \begin{cases} \phi(x) = 1 + x^2, \\ g(u) = \sin u. \end{cases}$$

Analogamente

$$f(x) = 2^{\cos x} = g(\phi(x)) \quad \text{dove} \quad \begin{cases} \phi(x) = \cos x, \\ g(u) = 2^u. \end{cases}$$

Spesso la funzione composta $f(x) = g(\phi(x))$ si indica con $f = g \circ \phi$ (che si legge “ g composto ϕ ” o anche “ g dopo ϕ ”)².

La composizione di funzioni non è un’operazione commutativa: in generale, $g \circ \phi$ e $\phi \circ g$ non sono la stessa funzione. L’ordine con cui si fanno le operazioni è importante! Se, per esempio, l’operazione ϕ sta per “sommare 1” e g per “moltiplicare per 2”, allora

$$g(\phi(x)) = g(x + 1) = 2x + 2, \quad \phi(g(x)) = \phi(2x) = 2x + 1.$$

Nel contesto delle funzioni composte, la nozione di funzione inversa diviene ancora più chiara. La funzione $Id(x) = x$, si chiama *identità* (è una funzione affine che ha per grafico la bisettrice del primo e terzo quadrante). Se la funzione ϕ è iniettiva su \mathbb{R} e la sua funzione inversa ϕ^{-1} è anch’essa definita su tutto \mathbb{R} , ϕ^{-1} è l’unica funzione che gode delle proprietà $\phi^{-1} \circ \phi = Id$ e $\phi \circ \phi^{-1} = Id$, cioè

$$(\phi^{-1} \circ \phi)(x) = x \quad (\phi \circ \phi^{-1})(x) = x \quad \forall x \in \mathbb{R},$$

La composizione ha senso anche per funzioni non definite in tutto \mathbb{R} . Sia $\phi : I \subset \mathbb{R} \rightarrow \mathbb{R}$ e $g : J \subset \mathbb{R} \rightarrow \mathbb{R}$, allora $f(x) := g(\phi(x))$ è ben definita per ogni $x \in I$ per cui $\phi(x) \in J$. Ad esempio, la funzione $f(x) = \ln(1 + x)$ è la composizione della funzione $\phi(x) = 1 + x$, definita in $I = \mathbb{R}$, e della funzione $g(u) = \ln u$, definita in $J = (0, +\infty)$.

²Alle volte si omette il simbolo \circ e si scrive semplicemente $f = g\phi$, ma bisogna stare attenti a non fare confusione con la funzione prodotto.

La composizione f è definita per tutte i valori $x \in \mathbb{R}$ tali che $1 + x \in J$, cioè per $x \in (-1, +\infty)$.

Per poter comporre due funzioni ϕ e g e definire una nuova funzione $g \circ \phi$, il dominio di g deve includere almeno una parte dell'immagine di ϕ . Ad esempio, non possiamo formare la funzione $g \circ \phi$ quando $g(u) = \sqrt{u}$ e $\phi(x) = -1 - x^2$, dato che il dominio di g è $[0, +\infty)$ e l'immagine di ϕ è $(-\infty, -1)$.

Chiaramente è possibile comporre più di due funzioni. Ad esempio,

$$f(x) = \frac{1}{1 + \tan(x^2)}$$

può essere ottenuta componendo (nell'ordine) $\phi(x) = x^2$, $\psi(\phi) = 1 + \tan \phi$, $g(\psi) = \frac{1}{\psi}$. Quindi $f = g \circ \psi \circ \phi$.

Funzioni meno comuni. Esistono infiniti modi per definire funzioni. Una possibilità (vagamente esotica) è di decomporre l'insieme di definizione in un certo numero di sottoinsiemi disgiunti ed associare una opportuna regola di calcolo per ciascuno di tali sottoinsiemi.

(i) Sia $I \subset \mathbb{R}$, allora si definisce

$$\text{funzione caratteristica di } I: \quad \chi_I(x) = \begin{cases} 1 & \text{se } x \in I, \\ 0 & \text{se } x \notin I. \end{cases}$$

La funzione $\chi_{\mathbb{R}}$ vale identicamente 1, mentre χ_{\emptyset} vale sempre 0, la funzione $\chi_{[0,1]}$ vale 1 nell'intervallo $[0, 1]$ e 0 nel complementare.

(ii) E' possibile anche fare scelte più originali: ad esempio,

$$f(x) = \begin{cases} 3x + 1 & \text{se } x \leq 0, \\ x^2 & \text{se } x > 0. \end{cases}$$

(iii) Un altro modo per generare nuove funzioni è tramite i "comandi" max e min. Ad esempio, si è già visto che

$$\max\{-x, x\} = |x|.$$

Più in generale, date f e g ,

$$\max\{f, g\}(x) = \begin{cases} f(x) & \text{se } g(x) \leq f(x), \\ g(x) & \text{se } f(x) \leq g(x) \end{cases}$$

$$\min\{f, g\}(x) = \begin{cases} f(x) & \text{se } f(x) \leq g(x), \\ g(x) & \text{se } g(x) \leq f(x) \end{cases}$$

5. Problemi di massimo e minimo

Molti problemi pratici conducono a problemi di massimo e di minimo di funzioni: qual è il carico massimo sopportato da una trave? Qual è l'energia minima che occorre perchè un satellite sfugga dall'attrazione gravitazionale di un pianeta? Qual è il minimo sforzo che bisogna compiere per passare l'esame? Diamo perciò una definizione rigorosa di cosa si intenda per massimo e minimo di una funzione.

DEFINIZIONE 5.1. *Sia $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$. Se esiste $x_0 \in I$ tale che $f(x) \geq f(x_0)$ per ogni $x \in I$, la funzione ammette (valore) minimo in I e x_0 è un punto di minimo. Si scrive*

$$f(x_0) = \min_{x \in I} f(x) \quad (\text{valore) minimo di } f.$$

Analogamente se esiste un punto $x_1 \in I$ tale che $f(x) \leq f(x_1)$ per ogni $x \in I$, si dice che la funzione ammette (valore) massimo in I e x_1 è un punto di massimo. Si scrive

$$f(x_1) = \max_{x \in I} f(x) \quad (\text{valore) massimo di } f.$$

Il massimo ed il minimo dipendono dall'insieme di definizione I . In generale una restrizione di una funzione, se ammette massimo, ha un massimo minore o uguale a quello della funzione di partenza e, se ammette minimo, questo è maggiore o uguale di quello della funzione di partenza.

In effetti il massimo ed il minimo della funzione non sono altro che il massimo ed il minimo dell'insieme immagine $f(I)$:

$$\begin{aligned} \min_{x \in I} f(x) &= \min f(I) = \min\{f(x) : x \in I\}, \\ \max_{x \in I} f(x) &= \max f(I) = \max\{f(x) : x \in I\}. \end{aligned}$$

Come si fa a determinare il massimo o il minimo di una funzione di una variabile reale a partire dal grafico? Il significato geometrico di un punto di massimo è chiaro: il grafico della funzione f è al di sotto della retta di equazione $y = f(x_0) = \text{costante}$. Quindi per determinare il massimo a partire dal grafico, basta stabilire se esista una retta con tale proprietà.

Come si vede a partire da alcuni esempi, *non tutte le funzioni ammettono massimo e/o minimo nel loro insieme di definizione!* Per superare questo ostacolo si introducono i concetti di *estremo superiore* e di *estremo inferiore*. L'estremo superiore ed inferiore di una funzione f sono l'estremo superiore ed inferiore dell'insieme immagine $f(I)$

$$\begin{aligned} \inf_{x \in I} f(x) &= \inf f(I) = \inf\{f(x) : x \in I\}, \\ \sup_{x \in I} f(x) &= \sup f(I) = \sup\{f(x) : x \in I\}. \end{aligned}$$

Il risultato sull'esistenza di estremo superiore ed inferiore garantisce che se esiste almeno un maggiorante allora esiste l'estremo superiore. Quindi, dato il grafico della funzione f ci sono solo due possibilità: o esiste almeno una retta orizzontale di equazione $y = c \in \mathbb{R}$ che sia completamente sopra il grafico di f o non ne esiste nessuna. Nel primo caso, l'estremo superiore di f è il valore minimo che si può dare al valore c facendo sempre in modo che la retta $y = c$ sia sopra il grafico di f (tale retta può anche non intersecare il grafico). Nel secondo caso, la funzione f è **illimitata superiormente** e $\sup_I f = +\infty$. Analogamente per l'estremo inferiore.

DEFINIZIONE 5.2. Una funzione $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$ tale che $\inf_{x \in I} f(x) \in \mathbb{R}$ e $\sup_{x \in I} f(x) \in \mathbb{R}$ (quindi non sono $\pm\infty$) si dice **limitata**.

Il significato geometrico della limitatezza di una funzione è immediato: una funzione è limitata se e solo se il suo grafico è interamente contenuto in una striscia orizzontale $\{(x, y) \in \mathbb{R}^2 : c \leq y \leq d\}$ per qualche $c, d \in \mathbb{R}$.

La lattina più conveniente. Supponiamo che si voglia progettare una scatola di latta di forma cilindrica (altezza h e raggio di base r). Il problema è: fissato il volume della scatola, esiste una scelta di r e h che minimizzi il quantitativo di latta da utilizzare (cioè la superficie totale del cilindro)? Sia V il volume della scatola e S la sua superficie

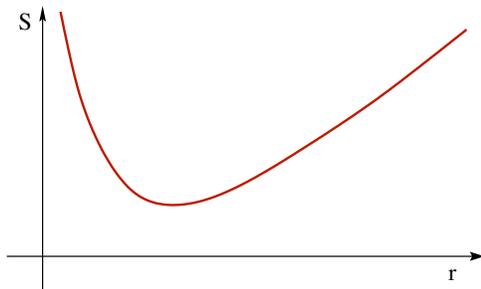


FIGURA 12. Il grafico di $S(r) = 2\pi \left(r^2 + \frac{V}{\pi r} \right)$.

totale. Allora $S = 2\pi r^2 + 2\pi hr = 2\pi (r^2 + hr)$ e $V = \pi r^2 h$. Fissare il volume V e minimizzare S , vuol dire scegliere $V = \text{costante}$. Quindi $h = V/\pi r^2$, e, sostituendo in S ,

$$h = \frac{V}{\pi r^2} \quad \Rightarrow \quad S = 2\pi \left(r^2 + \frac{V}{\pi r} \right).$$

Dal grafico della funzione S (che si può ottenere sommando i grafici $s = r^2$ e $s = V/\pi r$) si vede che tale minimo esiste. Il calcolo esplicito di quanto valga non è possibile per via elementare (ma con un minimo di cognizione di derivate, si può fare!).

CAPITOLO 3

Incontri ravvicinati con i limiti: le successioni

Una funzione $a : \mathbb{N} \rightarrow \mathbb{R}$ che associa ad ogni numero naturale n un valore $a(n)$ è una **successione (numerica)**. In genere, l' n -esimo elemento della successione si indica con a_n (invece di $a(n)$), questione di tradizione. Gli elementi della successione a_n possono essere pensati come una sequenza di valori ordinati in base al loro indice n

$$a_0, a_1, a_2, a_3, \dots$$

Un primo esempio è la successione dei numeri pari: $2, 4, 6, \dots$. In questo caso $a_n = 2n$. Un altro esempio semplice di funzione di n è l'espressione n -fattoriale, definita dal prodotto dei primi n numeri interi

$$a_n = n! = 1 \cdot 2 \cdot 3 \cdots n,$$

che dà luogo alla successione $1, 1, 2, 6, 24, 120, \dots$ (per definizione $0! := 1$).

Una successione può essere rappresentata disegnando nel piano cartesiano sopra ogni valore $n \in \mathbb{N}$ (dell'asse x) il valore definito da a_n , proprio come nel caso delle funzioni. Questo primo metodo è molto pratico nel caso di successioni definite da $a_n = f(n)$, dove si conosca il grafico della funzione f : basta prendere sul grafico di f solamente i punti con coordinata $x \in \mathbb{N}$. In alternativa, assegnata la successione a_n si può considerare come sua rappresentazione il grafico della funzione g definita da

$$g(x) = a_n \quad \text{per } x \in [n, n+1).$$

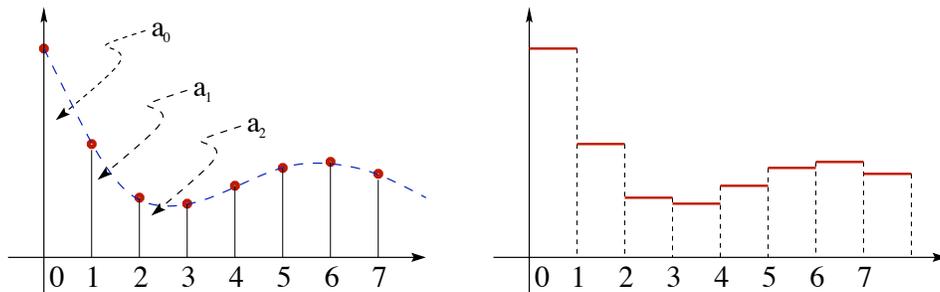


FIGURA 1.

Successioni definite per ricorrenza. In quel che segue utilizzeremo le successioni numeriche come una “cavia da laboratorio” per imparare, in una situazione particolarmente semplice, il concetto di limite e le procedure di base di calcolo. In realtà le successioni numeriche possono emergere anche da semplici modelli applicati. Supponiamo di voler studiare una popolazione di individui e di indicare con a_n il numero di abitanti all’anno n . Per controllare l’evoluzione della popolazione occorre conoscere il *tasso di incremento* R , definito da

$$R = \frac{a_{n+1} - a_n}{a_n},$$

che descrive quanto valga l’aumento di popolazione $a_{n+1} - a_n$ rispetto alla popolazione a_n all’anno n . Se si suppone che il tasso di crescita R sia costante ed uguale ad $r \in \mathbb{R}$, si ottiene, $a_{n+1} - a_n = ra_n$, cioè, esplicitando rispetto ad a_{n+1} ,

$$a_{n+1} = (1 + r)a_n.$$

Quindi, se si assegna la popolazione a_0 all’anno iniziale, si deduce che $a_1 = (1 + r)a_0$, $a_2 = (1 + r)a_1 = (1 + r)^2 a_0$. In generale questo semplice modello dà luogo ad una popolazione che cresce esponenzialmente in n , infatti

$$a_{n+1} = (1 + r)a_n = (1 + r)^2 a_{n-1} = \dots = (1 + r)^{n+1} a_0.$$

Ad esempio, se si parte da una popolazione di 100 abitanti e si suppone che il tasso di crescita annuale sia del 10%, cioè $r = 0,1$, dopo dieci anni la popolazione sarà di $a_{10} = 1,1^{10} \times 100$ abitanti (circa 259). Se si sceglie il tasso di incremento della forma $R = r(N - a_n)$ (questo vuol dire che c’è una popolazione critica, in questo caso pari a N , tale che se $a_n > N$ la popolazione decresce, mentre se $a_n < N$ la popolazione aumenta), si ottiene

$$a_{n+1} = (1 + Nr)a_n - ra_n^2 \quad (\text{equazione logistica}).$$

Già un oggetto così semplice e apparentemente innocuo è in grado di generare (per scelte opportune del parametro r) dinamiche particolarmente “stravaganti” e molto interessanti.

L’esempio precedente rientra nella classe delle successioni *definite per ricorrenza*: il termine $(n + 1)$ -esimo si ottiene in funzione dei termini precedenti. Nella forma più semplice, il termine $(n+1)$ -esimo è determinato dal solo termine n -esimo: assegnata la funzione f (dipende dal modello), si pone $a_{n+1} = f(a_n)$. Come nell’esempio precedente occorre anche assegnare una *condizione iniziale*, cioè deve essere dato il valore iniziale a_0 .

1. Limite di successioni

Il concetto fondamentale su cui si basa l'analisi matematica è quello di **limite**. L'idea che esprime il concetto di limite di una successione è semplice: assegnata la successione a_n , siamo in grado di “prevedere” quello che succederà per valori di n molto grandi? Più precisamente: è vero che la successione a_n “si stabilizza” per $n \rightarrow +\infty$, ovvero tende ad avvicinarsi ad un valore ℓ fissato? In caso affermativo, si dice che la successione ammette limite ℓ per $n \rightarrow +\infty$, altrimenti si dice che la successione non ha limite. Molte parole che abbiamo scritto nelle righe precedenti vanno precisate: che vuol dire “avvicinarsi”? E mandare n a $+\infty$ è da considerarsi un terribile insulto?

Partiamo da alcuni esempi. Sia

$$a_n = \frac{1}{n+1} \quad n \in \mathbb{N},$$

cioè consideriamo la successione $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$. Nessun numero di questa successione è zero, ma, per n che cresce, a_n si avvicina a zero. La frase “si avvicina a zero” va interpretata in questo senso: se decidiamo che l'essere vicino vuol dire che la distanza tra a_n e 0 deve essere minore di $1/10$, allora basta considerare gli elementi a_n della successione con indice $n \geq 10$; se rendiamo la condizione più stringente, ad esempio richiedendo che la distanza sia minore di $1/100$, basta considerare $n \geq 100$, e così via. In generale, comunque fissiamo una distanza $\varepsilon > 0$, da un certo indice n_ε in poi (n_ε dipende da ε) la distanza di a_n da 0 (che è data da $|a_n - 0|$) è minore di ε , cioè

$$(7) \quad \forall \varepsilon > 0 \quad \exists n_\varepsilon \in \mathbb{N} \quad \text{t.c.} \quad |a_n - 0| < \varepsilon \quad \forall n > n_\varepsilon.$$

In questo caso si dice che a_n **tende a 0** per $n \rightarrow +\infty$ (che si legge “ n tende a $+\infty$ ”). Per la successione $b_n = \frac{(-1)^n}{n+1}$ la situazione è esattamente la stessa, dato che

$$|b_n - 0| = \left| \frac{(-1)^n}{n+1} - 0 \right| = \left| \frac{1}{n+1} - 0 \right| = |a_n - 0|.$$

L'unica differenza è che i numeri b_n sono alternativamente più grandi e più piccoli di zero, cioè la successione *oscilla* attorno al valore limite 0, ma anche in questo caso vale la proprietà (7).

Consideriamo $a_n = \frac{n}{n+1}$. Scrivendo la successione nella forma

$$a_n = \frac{n}{n+1} = 1 - \frac{1}{n+1} \quad \Rightarrow \quad |a_n - 1| = \frac{1}{n+1},$$

vediamo che, per $n \rightarrow +\infty$, la distanza di a_n da 1 tende a zero, cioè il valore a_n si avvicina ad 1. Anche la successione $a_n = \frac{n^2 - 1}{n^2 + n + 1}$ si comporta in modo analogo,

infatti:

$$a_n = 1 - \frac{n+2}{n^2+n+1} \Rightarrow |a_n - 1| = \frac{n+2}{n^2+n+1} \leq \frac{2n}{n^2} = \frac{2}{n} \quad \forall n \geq 2,$$

da cui si deduce che la distanza di a_n da 1 tende a zero per $n \rightarrow \infty$.

DEFINIZIONE 1.1. Limite di successione. Si dice che la successione a_n converge ad $\ell \in \mathbb{R}$ per $n \rightarrow +\infty$ e si scrive $a_n \rightarrow \ell$ per $n \rightarrow +\infty$ o $\lim_{n \rightarrow +\infty} a_n = \ell$, se

$$(8) \quad \forall \varepsilon > 0 \quad \exists n_\varepsilon \in \mathbb{N} \quad \text{t.c.} \quad |a_n - \ell| < \varepsilon \quad \forall n > n_\varepsilon.$$

Il valore ℓ è il limite della successione a_n .

Se $a_n \rightarrow 0$ per $n \rightarrow +\infty$ (cioè se vale (7)), si dice che a_n è infinitesima.

La definizione esprime che, comunque si fissi una soglia di errore $\varepsilon > 0$, tutti gli elementi a_n della successione distano dal limite ℓ meno di ε , tranne al più un numero finito (quelli con indice da 1 ad n_ε). Quindi una maniera equivalente di dire che ℓ è il limite di a_n è affermare che *ogni intorno di ℓ contiene tutti i valori della successione a_n tranne al più un numero finito*.

Si faccia bene attenzione al fatto che la soglia ε vive sull'asse delle ordinate (e non su quello delle ascisse). In generale, scegliendo valori più piccoli per il margine di errore ε occorre scegliere valori più grandi di n_ε ; in altre parole, in generale, n_ε cresce quando ε tende a zero.

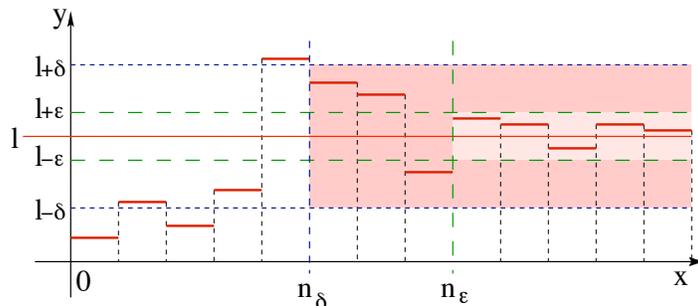


FIGURA 2.

OSSERVAZIONE 1.2. Per quale motivo si richiede che ε possa essere scelto arbitrariamente? Non basterebbe scegliere un fissato ε sufficientemente piccolo, ad esempio ε pari a un miliardesimo o a un miliardesimo di miliardesimo? Il problema è che i concetti di “grande” e “piccolo” sono soggettivi, mentre quello che si vuole definire qui è un criterio assoluto che vada bene sia per l’astronomo che usa la distanza Terra-Luna come parametro di “vicinanza”, sia per il fisico atomico per cui un millimetro è già una distanza abissale. La richiesta di una proprietà che valga *per ogni scelta di ε* rende la definizione “universale”, cioè indipendente dalla personale idea di piccolo o grande.

ESERCIZIO 1.3. Sia a_n una successione convergente ad ℓ per $n \rightarrow +\infty$ e sia b_n un'altra successione tale che $b_n = a_n$ per ogni $n > N_A$ dove N_A è il numero di Avogadro¹. Dimostrare che anche b_n converge ad ℓ per $n \rightarrow +\infty$.

Soluzione. Niente di più facile dato che la definizione di limite non dipende dal comportamento di un numero finito di elementi della successione. Per ipotesi,

$$\forall \varepsilon > 0 \quad \exists n_\varepsilon \in \mathbb{N} \quad \text{t.c.} \quad |a_n - \ell| < \varepsilon \quad \forall n > n_\varepsilon.$$

Ora vogliamo far vedere che vale una frase analoga anche per la successione b_n . Fissato $\varepsilon > 0$, scegliamo $n'_\varepsilon := \max\{n_\varepsilon, N_A\}$. Allora, se $n > n'_\varepsilon$, dato che $n > N_A$ si ha $b_n = a_n$ e dato che $n > n_\varepsilon$ vale anche $|a_n - \ell| < \varepsilon$. Dunque

$$|b_n - \ell| = |a_n - \ell| < \varepsilon \quad \forall n > n'_\varepsilon := \max\{n_\varepsilon, N_A\},$$

cioè la conclusione. E' essenziale che N_A sia proprio il numero di Avogadro o lo stesso ragionamento vale per N_A qualsiasi?

Calcolo diretto di un limite. Abbiamo una perfetta definizione di limite: logicamente ineccepibile. Ma come fare per verificarne la validità in un caso concreto? Proviamo a vedere un esempio. Tenete però conto che, nella pratica, non è questo il modo con cui si calcolano la maggior parte dei limiti! L'esempio che segue serve solo per acquisire maggiore familiarità con la definizione.

Consideriamo la successione

$$a_n = \frac{n^2}{n^2 + 1}.$$

Ammette limite? Ecco subito il primo problema: nella definizione di limite compare il valore ℓ del limite stesso, ma in generale ci si trova ad avere un'espressione per la successione, non per il suo (eventuale) limite. Questo va ottenuto per un'altra strada. Proviamo a ragionare in maniera casereccia. La domanda di fondo è: cosa succede dei valori a_n per n molto grande? Ad esempio, se $n = 1000$,

$$a_{1000} = \frac{1000^2}{1000^2 + 1} = \frac{1000000}{1000001}.$$

Bene... e se $n = 100000$? Allora

$$a_{100000} = \frac{100000^2}{100000^2 + 1} = \frac{10000000000}{10000000001}.$$

Come si vede, per valori di n molto grandi il termine $+1$ a denominatore diventa sempre più ridicolo perché va a sommarsi ad una quantità enormemente più grande. Allora è sensato aspettarsi che per $n \rightarrow +\infty$ valga un'approssimazione del tipo $n^2 + 1 \approx n^2$ e quindi $a_n = \frac{n^2}{n^2 + 1} \approx \frac{n^2}{n^2} = 1$. Questo per ora non dimostra un bel nulla, ma fa sospettare che la successione abbia limite e che il suo limite sia $\ell = 1$. Rimocchiamoci le maniche

¹Per chi non lo ricorda il numero di Avogadro è $N_A = 6,02214199 \times 10^{23}$.

e proviamo a dimostrarlo. Qual'è l'affermazione racchiusa nella definizione di limite? la distanza di a_n da ℓ è piccola se n è grande. Prima di tutto, quindi, scriviamo $|a_n - \ell|$:

$$|a_n - \ell| = \left| \frac{n^2}{n^2 + 1} - 1 \right| = \left| \frac{n^2 - n^2 - 1}{n^2 + 1} \right| = \frac{1}{n^2 + 1}.$$

Ora si tratta di far vedere che, fissato $\varepsilon > 0$, questa distanza è minore di ε se scegliamo $n > n_\varepsilon$ dove *abbiamo completa libertà di scelta per n_ε* . Imponiamo la disequazione a cui vogliamo arrivare e riscriviamola come condizione su n :

$$\frac{1}{n^2 + 1} < \varepsilon \quad \iff \quad n^2 > \frac{1}{\varepsilon} - 1 \quad \iff_{\varepsilon < 1} \quad n > \sqrt{\frac{1}{\varepsilon} - 1}.$$

Il gioco è fatto, basta scegliere $n_\varepsilon \geq \sqrt{\frac{1}{\varepsilon} - 1}$, ad esempio,

$$n_\varepsilon := \left[\sqrt{\frac{1}{\varepsilon} - 1} \right] + 1$$

dove $[\cdot]$ indica la funzione parte intera.

ESERCIZIO 1.4. *Dimostrare a partire dalla definizione la validità di*

$$\lim_{n \rightarrow +\infty} \frac{n}{2n + 1} = \frac{1}{2}, \quad \lim_{n \rightarrow +\infty} \frac{1}{n^k} = 0 \quad k \in \mathbb{N}, k \neq 0.$$

ESERCIZIO 1.5. *Dimostrare che, se $\lim_{n \rightarrow +\infty} a_n = a$, allora $\lim_{n \rightarrow +\infty} |a_n| = |a|$.*

Come dimostrare che una successione non ha limite? Data una successione a_n , una sottosuccessione a_{n_k} di a_n si ottiene scegliendo un sottoinsieme infinito degli elementi di a_n , scelti in modo che ogni elemento abbia indice strettamente maggiore di quello del precedente. Ad esempio, gli elementi di indice dispari $a_1, a_3, a_5, a_7, \dots$ costituiscono una sottosuccessione di $a_1, a_2, a_3, a_4, \dots$, così come gli elementi di indice pari $a_0, a_2, a_4, a_6, \dots$. Invece $a_5, a_3, a_{12}, a_{101}, \dots$ non è una sottosuccessione, perché il primo elemento ha indice maggiore del secondo.

Come individuare una sottosuccessione? Bisogna indicare il primo elemento, poi il secondo, quindi il terzo e così via. In definitiva bisogna scegliere un'applicazione da \mathbb{N} ad \mathbb{N} che ci dica al k -esimo posto qualè l'elemento n_k della successione scelto: al numero $k \in \mathbb{N}$ viene quindi associato un numero naturale n_k . Per fare in modo che l'ordine degli elementi venga preservato occorre che n_k sia crescente, cioè se $k_1 < k_2$ allora $n_{k_1} < n_{k_2}$. La sottosuccessione dei termini di indice dispari è espressa da $n_k = 2k + 1$: per $k = 0$ si prende $n_k = 1$, per $k = 2$ si prende $n_k = 3$ e così via...

ESERCIZIO 1.6. *Dire quale delle seguenti espressioni possono essere i primi elementi di una sottosuccessione di $a_n = n$*

$$\{1, 1, 3, 5, 7, \dots\}, \quad \{1, 4, 9, 16, 25, \dots\}, \quad \{1, 3, 5, 7, 6, 9, \dots\}.$$

ESERCIZIO 1.7. *Data la successione $a_n = n$, quali sono i primi termini della sottosuccessione a_{k^2} ? E se $k_n = 2n$? Ripetere l'esercizio nel caso in cui $a_n = n^2 + 1$.*

PROPOSIZIONE 1.8. *Sia a_n una successione convergente ad ℓ per $n \rightarrow +\infty$. Allora ogni sua sottosuccessione a_{n_k} converge allo stesso limite ℓ per $k \rightarrow +\infty$.*

La dimostrazione è lasciata per esercizio.

La Proposizione 1.8 può essere utilizzata “in negativo” per dimostrare che una assegnata successione non ammette limite. Consideriamo ad esempio la successione $a_n = (-1)^n$. La sottosuccessione dei termini di indice pari è $a_{2k} = (-1)^{2k} = 1$ per ogni k , quindi, essendo costantemente uguale ad 1, converge ad 1 per $k \rightarrow +\infty$. Invece, la sottosuccessione dei termini di indice dispari è $a_{2k+1} = (-1)^{2k+1} = -1$ per ogni k , quindi converge a -1 per $k \rightarrow +\infty$. Dato che due sottosuccessioni diverse convergono a limiti diversi, la conclusione della Proposizione 1.8 e quindi l'ipotesi non può essere vera: la successione $(-1)^n$ non è convergente. In generale, *se da una successione possono essere estratte due sottosuccessioni convergenti a limiti diversi, la successione non è convergente.*

Prime proprietà delle successioni convergenti. Prima di enunciare alcuni risultati che permettono di calcolare limiti in maniera più semplice di come si è fatto finora, dimostriamo alcune proprietà generali delle successioni convergenti.

PROPOSIZIONE 1.9. *Se una successione è convergente, allora il suo limite è unico.*

DIMOSTRAZIONE. Mostriamo che se la successione a_n tende sia ad ℓ che ad ℓ' , allora deve essere $\ell = \ell'$. Per definizione di limite, è vero che, per ogni $\varepsilon > 0$,

$$\exists n_\varepsilon \in \mathbb{N} \quad \text{t.c.} \quad |a_n - \ell| < \varepsilon \quad \forall n > n_\varepsilon \quad \text{e} \quad \exists n'_\varepsilon \in \mathbb{N} \quad \text{t.c.} \quad |a_n - \ell'| < \varepsilon \quad \forall n > n'_\varepsilon.$$

Allora, per $n > \max\{n_\varepsilon, n'_\varepsilon\}$, sono vere entrambe le affermazioni e quindi

$$0 \leq |\ell - \ell'| = |\ell - a_n + a_n - \ell'| \leq |\ell - a_n| + |a_n - \ell'| < 2\varepsilon.$$

In definitiva, abbiamo dimostrato che, per ogni $\varepsilon > 0$, vale $0 < |\ell - \ell'| < 2\varepsilon$. Quindi $|\ell - \ell'| = 0$, cioè $\ell = \ell'$. \square

DEFINIZIONE 1.10. *Una successione a_n tale che esista $M > 0$ per cui $|a_n| \leq M$ per ogni n , (cioè tutti gli a_n appartengono all'intervallo $[-M, M]$) si dice **limitata**.*

Se si ricorda che una successione a_n è una funzione da \mathbb{N} a \mathbb{R} , la condizione espressa nella Definizione 1.10 è equivalente alla frase “ $a(\mathbb{N})$ è un sottoinsieme limitato di \mathbb{R} ”, che è proprio la definizione di limitatezza data per funzioni di una variabile.

PROPOSIZIONE 1.11. *Se a_n è una successione convergente allora è anche limitata.*

DIMOSTRAZIONE. Fissato $\varepsilon = 1$, dato che a_n è convergente, esiste $N \in \mathbb{N}$ tale che $\ell - 1 < a_n < \ell + 1$ per ogni $n > N$. Quindi la “coda” della successione a_{N+1}, a_{N+2}, \dots vive in un insieme limitato. Dato che i termini mancanti sono in numero finito, tutta la successione “vive” in un insieme limitato (se necessario più grande del precedente). \square

Il viceversa non è vero: *esistono successioni limitate, non convergenti*. Ad esempio, la successione $a_n = (-1)^n$ non è convergente, ma è limitata dato che $|(-1)^n| = 1$ per ogni n (quindi si può scegliere $M = 1$ nella Definizione 1.10).

Successioni divergenti. Oltre alle successioni che tendono ad un limite, ci sono anche quelle il cui valore a_n diventa arbitrariamente grande: ad esempio la successione dei numeri pari $2n$, o la successione del fattoriale $n!$. La prossima definizione esprime cosa significhi in modo preciso l’affermazione “una successione tende a $+\infty$ ”.

DEFINIZIONE 1.12. *La successione a_n diverge a $+\infty$ (rispettivamente a $-\infty$) per $n \rightarrow +\infty$ se, comunque si scelga un valore M tutti i valori a_n sono più grandi di M (risp. più piccoli) tranne al più un numero finito. In tal caso si scrive*

$$\lim_{n \rightarrow +\infty} a_n = +\infty \quad (\text{risp. } -\infty).$$

In modo equivalente si può scrivere

$$(9) \quad \forall M \quad \exists n_M \in \mathbb{N} \quad \text{t.c.} \quad a_n \geq M \quad (\text{risp. } a_n \leq M) \quad \forall n > n_M.$$

Come nel caso delle successioni divergenti è utile vedere almeno un esempio di verifica diretta del fatto che una successione è divergente. Ad esempio, dimostriamo

$$\lim_{n \rightarrow +\infty} \frac{n^2}{n+1} = +\infty.$$

Fissato $M > 0$, dobbiamo far vedere che è possibile determinare n_M per cui valga $a_n > M$ per ogni $n > n_M$. Dato che

$$\frac{n^2}{n+1} > M \quad \iff \quad n^2 - Mn - M > 0,$$

la domanda da porsi è: per quali n è vera la disequazione finale? Le radici del polinomio di secondo grado $x^2 - Mx - M$ sono $x_{\pm} := (M \pm \sqrt{M^2 + 4M})/2$, quindi, se $n > x_+$, è vero che $n^2 - Mn - M > 0$. Ottimo, allora scegliamo

$$N_M := \left\lceil \frac{M + \sqrt{M^2 + 4M}}{2} \right\rceil + 1.$$

Strade diverse portano alla stessa conclusione. Lo stesso problema può essere risolto in un modo diverso, meno contoso. Prima di tutto si osserva che (verificate i

passaggi)

$$a_n = \frac{n^2}{n+1} = n - \frac{1}{n+1} \geq n - 1.$$

Quindi, se $n - 1 > M$, vale anche $a_n > M$. Perciò si può scegliere $N_M = [M + 1] + 1$ per dedurre la stessa conclusione. In effetti, il valore n_M della Definizione 1.12 (così come n_ε della Definizione 1.1), non è definito in maniera univoca, tutt'altro!

ESERCIZIO 1.13. *Dimostrare le implicazioni*

$$\begin{aligned} \lim_{n \rightarrow +\infty} a_n = 0 &\quad \Rightarrow \quad \lim_{n \rightarrow +\infty} \frac{1}{|a_n|} = +\infty. \\ \lim_{n \rightarrow +\infty} |a_n| = +\infty &\quad \Rightarrow \quad \lim_{n \rightarrow +\infty} \frac{1}{a_n} = 0. \end{aligned}$$

Criterio di convergenza di Cauchy. Ogni successione convergente definisce un numero ℓ , il suo limite, ma l'unico test di convergenza che emerge dalla definizione consiste nel dimostrare che la differenza $|a_n - \ell|$ è infinitesima, quindi è applicabile solo se il numero ℓ è già noto. Invece, è essenziale avere un test "intrinseco" di convergenza che non richieda la conoscenza *a priori* del valore del limite, ma che coinvolga solamente i termini stessi della successione.

TEOREMA 1.14. *Criterio di convergenza di Cauchy.* Una successione a_n è convergente se e solo se è una successione di Cauchy, cioè se vale la condizione

$$\forall \varepsilon > 0, \quad \exists n_\varepsilon \in \mathbb{N} \quad \text{t.c.} \quad |a_m - a_n| < \varepsilon \quad \forall m, n > n_\varepsilon.$$

La condizione di Cauchy descrive il fatto che gli elementi della successione distano tra loro meno di una soglia arbitraria $\varepsilon > 0$ a patto di considerare termini con un indice sufficientemente grande. Il fatto che i termini si avvicinino l'un l'altro per $n \rightarrow +\infty$ quando una successione è convergente è del tutto naturale: i termini si avvicinano al limite ℓ e quindi si avvicinano tra loro. La proprietà notevole è che vale anche il viceversa: se gli elementi della successione si avvicinano, allora la successione converge. Non è questa la sede per approfondire ulteriormente questo criterio, ma non si può mancare di dire che si tratta di una pietra miliare nella costruzione dei numeri reali a partire dai numeri razionali.

Piccolo glossario per le successioni

Se una successione $a_n \dots$

- \dots è convergente o divergente (a $\pm\infty$), allora è **regolare**;
- \dots non è convergente né divergente, allora è **non regolare**;
- \dots tende a 0 per $n \rightarrow +\infty$, allora è **infinitesima**;

... è tale che $|a_n| \leq M$ per qualche M e per ogni n , (cioè se i suoi elementi sono in un intervallo limitato) allora è **limitata**;

... è tale che $\lim_{n \rightarrow +\infty} |a_n| = +\infty$, allora **diverge in modulo**.

$$\begin{array}{ccccc}
 a_n \text{ è infinitesima} & \Rightarrow & a_n \text{ converge} & \Rightarrow & a_n \text{ è limitata} \\
 & & \downarrow & & \\
 & & a_n \text{ è regolare} & & \\
 & & \uparrow & & \\
 & & a_n \text{ diverge} & \Rightarrow & a_n \text{ diverge in modulo}
 \end{array}$$

2. Il limite entra in società

Fin qui abbiamo definito il senso della parola *limite* per successioni di numeri reali. Una successione in \mathbb{R} può avere una struttura complicata: ad esempio, può essere somma/prodotto di vari termini. Come si comporta l'operazione di "passaggio al limite" rispetto alle operazioni $+$ e \cdot definite in \mathbb{R} ? E rispetto ai segni \leq e $<$?

Operazioni razionali con i limiti. Per il calcolo dei limiti è possibile usare le operazioni elementari di somma, moltiplicazione, sottrazione e divisione.

(i) *Somma e sottrazione.* Il limite della somma/sottrazione di successioni convergenti è la somma/sottrazione dei limiti:

$$\lim_{n \rightarrow +\infty} a_n = a, \lim_{n \rightarrow +\infty} b_n = b \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} a_n \pm b_n = a \pm b.$$

(ii) *Prodotto.* Il limite del prodotto di successioni convergenti è il prodotto dei limiti:

$$\lim_{n \rightarrow +\infty} a_n = a, \lim_{n \rightarrow +\infty} b_n = b \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} a_n b_n = ab.$$

(iii) *Rapporto.* Il limite del rapporto di successioni convergenti è il rapporto dei limiti *a patto che la successione a denominatore non tenda a zero*:

$$\lim_{n \rightarrow +\infty} a_n = a, \lim_{n \rightarrow +\infty} b_n = b \neq 0 \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} \frac{a_n}{b_n} = \frac{a}{b}.$$

In altre parole: possiamo *invertire l'ordine di applicazione* delle operazioni razionali e del procedimento di limite ottenendo lo stesso risultato, o anche *operazioni razionali e limiti commutano*.

Dimostriamo la proprietà del prodotto. Supponiamo $a_n \rightarrow a$ e $b_n \rightarrow b$ per $n \rightarrow +\infty$. Allora²

$$\forall \varepsilon > 0 \quad \exists n_\varepsilon \quad : \quad |a - a_n| < \varepsilon, |b - b_n| < \varepsilon \quad \forall n > n_\varepsilon.$$

²Il valore di n_ε per la successione $\{a_n\}$ e quello per la successione $\{b_n\}$ potrebbero essere diversi, ma, in questo caso, potremmo scegliere il più grande dei due, per cui sono soddisfatte le relazioni sia per a_n che per b_n .

Scrivendo $ab - a_nb_n = b(a - a_n) + a_n(b - b_n)$, e ricordando che una successione convergente è sempre limitata (per cui esiste $M > 0$ tale che $|a_n| \leq M$ per ogni n), si ha che

$$|ab - a_nb_n| \leq |b||a - a_n| + |a_n||b - b_n| < (|b| + M)\varepsilon \quad \forall n > n_\varepsilon.$$

Dato che la quantità $(|b| + M)\varepsilon$ può essere resa arbitrariamente piccola scegliendo ε sufficientemente piccolo, la distanza tra ab e a_nb_n diviene arbitrariamente piccola scegliendo valori di n sufficientemente grandi e quindi vale la conclusione.

Tramite queste regole è possibile calcolare molti limiti. Ad esempio:

$$\lim_{n \rightarrow +\infty} \frac{2n^2 - 1}{3n^2 + n + 1} = \lim_{n \rightarrow +\infty} \frac{2 - \frac{1}{n^2}}{3 + \frac{1}{n} + \frac{1}{n^2}} = \frac{\lim_{n \rightarrow +\infty} (2 - \frac{1}{n^2})}{\lim_{n \rightarrow +\infty} (3 + \frac{1}{n} + \frac{1}{n^2})} = \frac{2}{3},$$

passando al limite sia nel numeratore che nel denominatore.

Per ora, non è chiaro cosa dire sul comportamento al limite della successione $\frac{a_n}{b_n}$ nel caso in cui la successione b_n sia infinitesima. Torneremo tra breve sulla questione.

ESERCIZIO 2.1. *Calcolare il limite per $n \rightarrow +\infty$ delle seguenti successioni*

$$\frac{n^4 + 1}{n^4 + n^2}, \quad \frac{3n^2 + 1}{n(2n^2 + 1)}, \quad \frac{(n+1)(2n+2)(3n+3)}{n^3}, \quad \frac{(n+1)^2}{(n^2+1)^2}.$$

Limiti e disequazioni. Un'altra questione importante è come si comporti l'operazione di limite rispetto all'ordinamento dei numeri reali.

TEOREMA 2.2. Monotonia del limite. *Sia $\lim_{n \rightarrow +\infty} a_n = a$ e $\lim_{n \rightarrow +\infty} b_n = b$. Se, per qualche $N \in \mathbb{N}$, vale $a_n < b_n$ (oppure $a_n \leq b_n$) per ogni $n \geq N$, allora $a \leq b$.*

DIMOSTRAZIONE. Dato che a_n tende a a e b_n tende a b , si ha che per ogni $\varepsilon > 0$, $a - \varepsilon < a_n \leq b_n < b + \varepsilon$ per ogni $n > N$ per un opportuno N . Guardando il primo e l'ultimo termine nella catena di disequazioni, si deduce che $b - a > -2\varepsilon$ per ogni $\varepsilon > 0$. Quindi, necessariamente, $b - a \geq 0$. \square

Il Teorema 2.2 afferma che l'operazione di limite è "monotona non decrescente", nel senso che vale l'implicazione

$$a_n < b_n \quad \Rightarrow \quad a := \lim_{n \rightarrow +\infty} a_n \leq \lim_{n \rightarrow +\infty} b_n =: b.$$

Si noti che, nel passaggio al limite, la disuguaglianza stretta diviene una disuguaglianza non stretta. Ad esempio, scegliendo $a_n = 1/2n$ e $b_n = 1/n$, si ha $a_n < b_n$ per ogni n , ma $\lim_{n \rightarrow +\infty} a_n = \lim_{n \rightarrow +\infty} b_n = 0$.

In maniera analoga, si dimostra il seguente (utilissimo) risultato.

TEOREMA 2.3. Teorema dei carabinieri. Siano a_n, b_n, c_n successioni tali che

(i) esiste $N \in \mathbb{N}$ tale che $a_n \leq b_n \leq c_n$ per ogni $n > N$,

(ii) $\lim_{n \rightarrow +\infty} a_n = \lim_{n \rightarrow +\infty} c_n = \ell \in \mathbb{R}$.

Allora la successione b_n è convergente e $\lim_{n \rightarrow +\infty} b_n = \ell$.

DIMOSTRAZIONE. L'obiettivo è dimostrare che per ogni $\varepsilon > 0$ c'è una scelta di $n_\varepsilon \in \mathbb{N}$ tale che $\ell - \varepsilon < b_n < \ell + \varepsilon$ per ogni $n > n_\varepsilon$. Quindi occorre stimare dall'alto e dal basso i termini b_n . Dalle ipotesi segue che comunque si fissi $\varepsilon > 0$ esiste $\tilde{n}_\varepsilon \in \mathbb{N}$ tale che, per ogni $n > \tilde{n}_\varepsilon$,

$$\ell - \varepsilon < a_n < \ell + \varepsilon \quad \text{e} \quad \ell - \varepsilon < c_n < \ell + \varepsilon.$$

Quindi, dato che per ipotesi esiste N tale che $a_n \leq b_n \leq c_n$ per ogni $n > N$, si ottiene

$$\ell - \varepsilon < a_n \leq b_n \leq c_n < \ell + \varepsilon \quad \forall n > N := \max\{N, \tilde{n}_\varepsilon\},$$

ossia la conclusione. □

ESEMPIO 2.4. Dato $x \in (0, 1)$, applichiamo il Teorema 2.3, alla successione

$$b_n = x^n.$$

Chiaramente esiste $h > 0$ tale che $x = 1/(1+h)$. Poiché $(1+h)^n \geq 1+nh$ per ogni $n \in \mathbb{N}$ (dimostrarlo!),

$$0 < b_n = \frac{1}{(1+h)^n} \leq \frac{1}{1+nh} \quad \forall n \in \mathbb{N}.$$

Dato che $\lim_{n \rightarrow +\infty} \frac{1}{1+nh} = 0$, scegliendo $a_n = 0$ e $c_n = \frac{1}{1+nh}$ e applicando il Teorema 2.3

$$\lim_{n \rightarrow +\infty} x^n = 0 \quad \text{per} \quad x \in (0, 1).$$

ESERCIZIO 2.5. Dati $\lambda \in [0, 1)$ e $a_0 \in \mathbb{R}$, sia a_n la successione definita per ricorrenza da $a_{n+1} = \lambda a_n$. Dimostrare che la successione a_n è infinitesima.

Conseguenza del Teorema 2.3 è questo piccolo criterio, che è una versione più generale dell'esercizio precedente: sostanzialmente si suppone che l'uguaglianza $a_{n+1} = \lambda a_n$ con $\lambda \in [0, 1)$ valga "all'infinito"...

COROLLARIO 2.6. Sia $a_n > 0$ per ogni n una successione tale che

$$(10) \quad \lim_{n \rightarrow +\infty} \frac{a_{n+1}}{a_n} = \lambda$$

Allora, se $\lambda \in [0, 1)$, la successione a_n è infinitesima.

DIMOSTRAZIONE. *Passo 1.* Dimostriamo che esistono $\sigma \in [0, 1)$ e $N \in \mathbb{N}$ tale che

$$(11) \quad a_{n+1} \leq \sigma a_n \quad \forall n \geq N.$$

Infatti, scegliamo $\sigma \in (\lambda, 1)$ e sia $\varepsilon := \sigma - \lambda$. Utilizzando la definizione di limite per la successione a_{n+1}/a_n , si deduce che esiste $n_\varepsilon \in \mathbb{N}$ tale che

$$\lambda - \varepsilon < \frac{a_{n+1}}{a_n} < \lambda + \varepsilon = \lambda + (\sigma - \lambda) = \sigma \quad \forall n > n_\varepsilon.$$

che porta alla (11) scegliendo $N = n_\varepsilon$.

Passo 2. Dimostriamo che a_n è infinitesima. Per semplicità, supponiamo $N = 0$ (cioè che (11) valga per ogni n), allora

$$0 < a_{n+1} \leq \sigma a_n \leq \sigma^2 a_{n-1} \leq \dots \leq \sigma^{n+1} a_0.$$

Dato che $\sigma \in [0, 1)$, per $n \rightarrow +\infty$ la quantità $\sigma^n a_1$ tende a zero e quindi si ha la conclusione. Nel caso generale, dato che

$$0 < a_{n+1} \leq \sigma^n \sigma^{1-N} a_N \quad \forall n \geq N,$$

la dimostrazione è analoga. □

È possibile dare criteri analoghi al Teorema 2.3 per dimostrare la divergenza di una successione. Ad esempio, una successione a_n che sia più grande di una successione b_n divergente a $+\infty$ è divergente a $+\infty$.

ESEMPIO 2.7. Consideriamo di nuovo la successione $b_n = x^n$, questa volta con $x > 1$. Allora $x = 1 + h$ con $h > 0$. Dalla disuguaglianza $(1 + h)^n \geq 1 + nh$, segue

$$b_n = (1 + h)^n \geq 1 + nh$$

Dato che $1 + nh \rightarrow +\infty$ per $n \rightarrow \infty$, anche la successione b_n diverge a $+\infty$.

Zeri a denominatore ed uso degli infiniti. Dall'analisi che abbiamo presentato fin qui restano fuori alcuni casi significativi:

- che succede della successione $\frac{a_n}{b_n}$ nel caso in cui $\lim_{n \rightarrow +\infty} b_n = 0$?
- che succede di somma/sottrazione/prodotto/rapporto quando qualcuno dei termini in gioco tende a $+\infty$ o a $-\infty$?

Partiamo dalla prima delle due questioni. Un numero molto piccolo a denominatore rende tutta la frazione molto grande. Quindi è ragionevole aspettarsi che, qualora il denominatore sia infinitesimo, il rapporto tenda a $+\infty$. Il banalissimo esempio:

$$a_n = 1, \quad b_n = \frac{1}{n} \quad \Rightarrow \quad \frac{a_n}{b_n} = \frac{1}{1/n} = n \rightarrow +\infty \quad \text{per } n \rightarrow +\infty,$$

dà conforto a questa prima ipotesi di lavoro. Arrischiamoci in una congettura:

$$\text{Congettura 1:} \quad \lim_{n \rightarrow +\infty} b_n = 0 \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} \frac{a_n}{b_n} = +\infty$$

Ma se consideriamo il caso

$$a_n = \frac{1}{n}, \quad b_n = \frac{1}{n} \quad \Rightarrow \quad \frac{a_n}{b_n} = \frac{1/n}{1/n} = 1 \rightarrow 1 \quad \text{per } n \rightarrow +\infty.$$

Cosa succede? Molto semplice, il denominatore è infinitesimo, ma lo è anche il numeratore. Quindi, può capitare che il tendere a zero del denominatore sia (in qualche modo) compensato dal tendere a zero del numeratore! Proviamo una nuova versione:

$$\text{Congettura 2:} \quad \lim_{n \rightarrow +\infty} a_n = a \neq 0, \quad \lim_{n \rightarrow +\infty} b_n = 0 \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} \frac{a_n}{b_n} = +\infty$$

Qui non siamo troppo lontani dal vero, ma ancora siamo stati un po' troppo leggeri nella questione dei segni: nel caso

$$a_n = 1, \quad b_n = -\frac{1}{n} \quad \Rightarrow \quad \frac{a_n}{b_n} = \frac{1}{-1/n} = -n \rightarrow -\infty \quad \text{per } n \rightarrow +\infty,$$

Proponiamo allora una versione (si spera finale) più precisa:

$$\text{Congettura 3:} \quad \lim_{n \rightarrow +\infty} a_n = a \neq 0, \quad \lim_{n \rightarrow +\infty} b_n = 0 \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} \left| \frac{a_n}{b_n} \right| = +\infty$$

La Congettura 3 è vera. Per dimostrarla, bisogna mostrare che per ogni $M > 0$, vale la disuguaglianza $|a_n/b_n| \geq M$ per n sufficientemente grande. Dato che la disequazione precedente è equivalente a $|a_n| - M|b_n| \geq 0$ bisogna mostrare che

$$\forall M \quad \exists n_M \in \mathbb{N} \quad \text{t.c.} \quad |a_n| - M|b_n| \geq 0 \quad \forall n > n_M.$$

Niente di più facile, dato che

$$\lim_{n \rightarrow +\infty} (|a_n| - M|b_n|) = |a| > 0.$$

ESERCIZIO 2.8. *Siamo tranquilli e soddisfatti del nostro risultato, quando, d'improvviso, giunge un tipo, sicuro del fatto suo, che afferma*

$$\text{Congettura 4:} \quad \left. \begin{array}{l} \exists \nu > 0 \quad \text{t.c.} \quad a_n \geq \nu \quad \forall n, \\ \lim_{n \rightarrow +\infty} b_n = 0 \end{array} \right\} \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} \left| \frac{a_n}{b_n} \right| = +\infty$$

Credergli o non credergli?

Resta fuori dalla nostra analisi il caso in cui sia il numeratore che il denominatore tendano a 0 per $n \rightarrow +\infty$. In questo caso può succedere di tutto! Ci sono casi in cui il rapporto tende a zero, casi in cui tende ad ∞ , casi in cui il limite del rapporto non esiste! Non esiste una regola generale e per questo si dice che si tratta di una forma

indeterminata (nel senso che non si può determinare subito se esista e quanto valga il limite ed occorre un'analisi più raffinata):

$$\text{forma indeterminata } \frac{0}{0} : \quad \lim_{n \rightarrow +\infty} a_n = \lim_{n \rightarrow +\infty} b_n = 0 \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} \left| \frac{a_n}{b_n} \right| = ?$$

ESERCIZIO 2.9. *Trovare due successioni a_n e b_n infinitesime tali che la successione dei rapporti a_n/b_n non abbia limite.*

Consideriamo il caso in cui uno dei termini sia divergente (per fissare le idee, divergente a $+\infty$) e l'altro convergente, cioè supponiamo

$$\lim_{n \rightarrow +\infty} a_n = +\infty \quad \text{e} \quad \lim_{n \rightarrow +\infty} b_n = b.$$

Cosa succede di somma e prodotto? La regola, detta in maniera molto poco ortodossa, è che “finché i termini in gioco non si contrastano tra loro tutto va bene...”. Ad esempio:

$$\lim_{n \rightarrow +\infty} a_n + b_n = +\infty, \quad \text{se } b \neq 0, \quad \lim_{n \rightarrow +\infty} |a_n b_n| = +\infty,$$

(dimostrate queste proprietà!). L'unica situazione di “contrasto” è quella in cui la successione da studiare sia prodotto di una successione divergente e di una infinitesima

$$\text{forma indeterminata } \infty \cdot 0 : \quad \left. \begin{array}{l} \lim_{n \rightarrow +\infty} a_n = \pm\infty, \\ \lim_{n \rightarrow +\infty} b_n = 0 \end{array} \right\} \Rightarrow \lim_{n \rightarrow +\infty} a_n \cdot b_n = ?$$

Infine, resta la situazione più drammatica di tutte: entrambe le successioni a_n e b_n sono divergenti. Nel caso in cui si sommino due successioni divergenti a $+\infty$, la conclusione è evidente: la successione somma è anch'essa divergente a $+\infty$:

$$\lim_{n \rightarrow +\infty} a_n = \lim_{n \rightarrow +\infty} b_n = +\infty \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} a_n + b_n = +\infty, \quad \lim_{n \rightarrow +\infty} a_n \cdot b_n = +\infty.$$

Nel caso in cui le due successioni divergono una a $+\infty$ e l'altra a $-\infty$, non si può dedurre nessuna conclusione generale:

$$\text{forma indeterminata } +\infty - \infty : \quad \left. \begin{array}{l} \lim_{n \rightarrow +\infty} a_n = +\infty, \\ \lim_{n \rightarrow +\infty} b_n = -\infty \end{array} \right\} \Rightarrow \lim_{n \rightarrow +\infty} a_n + b_n = ?$$

Il prodotto di successioni divergenti non crea nessun problema particolare (bisogna solo stare attenti al segno di ∞ , che si deduce con la buona vecchia regola del segno di un prodotto). Ad esempio,

$$\lim_{n \rightarrow +\infty} a_n = \lim_{n \rightarrow +\infty} b_n = +\infty \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} a_n \cdot b_n = +\infty.$$

Per quanto riguarda il rapporto, pochi minuti di riflessione portano alla conclusione che l'unica situazione problematica è la seguente:

$$\text{forma indeterminata } \frac{\infty}{\infty} : \left. \begin{array}{l} \lim_{n \rightarrow +\infty} a_n = \infty, \\ \lim_{n \rightarrow +\infty} b_n = \infty \end{array} \right\} \Rightarrow \lim_{n \rightarrow +\infty} \frac{a_n}{b_n} = ?$$

3. Calcolo di alcuni limiti

Utilizziamo ora le proprietà dei limiti per studiare alcune successioni specifiche.

ESEMPIO 3.1.

$$\lim_{n \rightarrow +\infty} \sqrt[n]{x} = 1 \quad \text{per ogni } x > 0.$$

Poniamo $a_n = \sqrt[n]{x}$ e usiamo la disequazione $(1+h)^n \geq 1+nh$. Nel caso $x > 1$, allora anche $\sqrt[n]{x} > 1$ e quindi $h_n := \sqrt[n]{x} - 1 > 0$ e $a_n = 1+h_n$. Esplicitando la disequazione rispetto a h_n ,

$$x = (a_n)^n = (1+h_n)^n \geq 1+nh_n \Rightarrow 0 < h_n \leq \frac{x-1}{n},$$

da cui segue $\lim_{n \rightarrow +\infty} h_n = 0$ e quindi otteniamo la conclusione nel caso $x > 1$.

Se $x = 1$ la successione è costante ed il risultato banale. Se $x < 1$, allora $1/x > 1$ e quindi $\sqrt[n]{1/x}$ converge ad 1 per quanto già visto. Dato che $\sqrt[n]{x} = \frac{1}{\sqrt[n]{1/x}}$, segue la conclusione.

ESEMPIO 3.2.

$$\lim_{n \rightarrow +\infty} \sqrt{n+1} - \sqrt{n} = 0.$$

In questo caso $a_n = \sqrt{n+1} - \sqrt{n}$ è la differenza di termini che tendono a $+\infty$. Passare al limite separatamente sui due termini dà l'espressione $+\infty - \infty$, che è *senza senso*.³ Ci troviamo di fronte ad una forma indeterminata e quindi per determinare l'esistenza o meno del limite bisogna lavorare un po' d'astuzia. Qui possiamo riscrivere a_n come

$$a_n = \frac{(\sqrt{n+1} - \sqrt{n})(\sqrt{n+1} + \sqrt{n})}{\sqrt{n+1} + \sqrt{n}} = \frac{(n+1) - n}{\sqrt{n+1} + \sqrt{n}} = \frac{1}{\sqrt{n+1} + \sqrt{n}},$$

che tende a 0 per $n \rightarrow +\infty$.

ESEMPIO 3.3. (Molto importante!)

$$\lim_{n \rightarrow +\infty} x^n = \begin{cases} 0 & x \in (-1, 1), \\ 1 & x = 1, \\ \text{non esiste} & x = -1, \\ +\infty & x > 1, \\ \text{diverge in modulo} & x < -1. \end{cases}$$

³In generale si pone la questione: si possono dare delle "regole algebriche" per l'uso dei simboli $\pm\infty$? Occorre ricordarsi che questi simboli sono definiti dall'operazione di limite e quindi, per definire tali regole, bisogna ricondursi alle proprietà dei limiti.

Abbiamo già visto che $\lim_{n \rightarrow +\infty} x^n = 0$ per $x \in (0, 1)$ e che $\lim_{n \rightarrow +\infty} x^n = +\infty$ per $x > 1$. Anche per $x \in (-1, 0]$, la successione x^n è infinitesima, dato che $\lim_{n \rightarrow +\infty} |x|^n = 0$.

Se $x = 1$ la successione è costantemente 1 (che quindi è il suo limite). Se $x = -1$, la successione oscilla tra i valori ± 1 ed è non regolare. Nel caso $x < -1$, la successione oscilla tra valori positivi e negativi e non è regolare, ma in valore assoluto diverge.

ESEMPIO 3.4.

$$\lim_{n \rightarrow +\infty} \frac{n^p}{x^n} = 0. \quad \forall x > 1, p \in \mathbb{N}.$$

Per dimostrare questo limite, applichiamo il Corollario 2.6 ad $a_n = \frac{n^p}{x^n}$:

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)^p x^n}{x^{n+1} n^p} = \frac{(n+1)^p}{x n^p} = \frac{1}{x} \left(1 + \frac{1}{n}\right)^p \longrightarrow \frac{1}{x} \in (0, 1) \quad \text{per } n \rightarrow +\infty.$$

Dato che il rapporto $\frac{a_{n+1}}{a_n}$ tende ad un numero in $(0, 1)$, sono verificate le ipotesi del Corollario e quindi vale la conclusione.

ESEMPIO 3.5.

$$(12) \quad \lim_{n \rightarrow +\infty} \frac{x^n}{n!} = 0 \quad \forall x > 1.$$

Applichiamo di nuovo il Corollario 2.6

$$\frac{a_{n+1}}{a_n} = \frac{x^{n+1}}{(n+1)!} \frac{n!}{x^n} = \frac{x}{n+1} \longrightarrow 0 \quad \text{per } n \rightarrow +\infty,$$

da cui segue la conclusione.

ESEMPIO 3.6.

$$\lim_{n \rightarrow +\infty} \frac{n!}{n^n} = 0.$$

Infatti

$$0 \leq a_n = \frac{n(n-1) \cdots 2 \cdot 1}{n \cdot n \cdots n} = \frac{n}{n} \frac{n-1}{n} \cdots \frac{1}{n} = 1 \left(1 - \frac{1}{n}\right) \cdots \frac{1}{n} \leq \frac{1}{n},$$

quindi, per il Teorema 2.3, vale la conclusione.

ESEMPIO 3.7. **La serie geometrica.** Fissato $q \in \mathbb{R}$, la successione

$$S_n := \sum_{k=0}^n q^k,$$

è convergente se e solo se $q \in (-1, 1)$. Inoltre

$$\lim_{n \rightarrow +\infty} S_n = \frac{1}{1-q} \quad \forall q \in (-1, 1).$$

Infatti, la successione S_n si può riscrivere nella forma

$$S_n = 1 + q + q^2 + \cdots + q^n = \begin{cases} \frac{1 - q^{n+1}}{1 - q} & q \neq 1, \\ n + 1 & q = 1 \end{cases}$$

Passando al limite per $n \rightarrow +\infty$ è utilizzando i risultati dell'Esempio 3.3 si ottiene la conclusione. Nel caso $q \geq 1$ la successione diverge, mentre per $q \leq -1$ la serie è non regolare. Per esprimere la convergenza di S_n a $1/(1-q)$ per $|q| < 1$, si usa la notazione

$$\text{serie geometrica :} \quad \sum_{k=0}^{+\infty} q^k := \lim_{n \rightarrow +\infty} \sum_{k=0}^n q^k = \frac{1}{1 - q} \quad \forall q \in (-1, 1).$$

Sul significato della parola “serie” in generale torneremo tra poco.

4. Successioni monotone

DEFINIZIONE 4.1. *Successioni monotone.* Una successione a_n è **strettamente crescente** se ogni termine è maggiore del precedente, cioè se $a_n < a_{n+1}$ per ogni $n \in \mathbb{N}$. Analogamente, è **strettamente decrescente** se $a_n > a_{n+1}$ per ogni $n \in \mathbb{N}$. Una successione a_n è **strettamente monotona** se è o strettamente crescente o strettamente decrescente.

Nel caso in cui valgano le disuguaglianze non strette valgono delle definizioni analoghe: a_n è **non decrescente** se $a_n \leq a_{n+1}$ per ogni n ed è **non crescente** se $a_n \geq a_{n+1}$. Una successione a_n è **monotona** se è o non decrescente o non crescente.

Per verificare se una successione è monotona occorre risolvere delle disequazioni. Ad esempio, la successione $a_n = \frac{1}{1+n^2}$ è strettamente decrescente, infatti

$$a_{n+1} < a_n \quad \iff \quad \frac{1}{1+(n+1)^2} < \frac{1}{1+n^2} \quad \iff \quad 1+n^2 < 1+(n+1)^2$$

che è verificata per ogni $n \in \mathbb{N}$.

ESERCIZIO 4.2. Dire se la successione $a_n = \frac{n}{1+n^2}$ è strettamente decrescente.

ESERCIZIO 4.3. Siano a_n e b_n due successioni positive e non decrescenti. Dimostrare che $a_n + b_n$ e $a_n b_n$ sono anch'esse successioni non decrescenti. E' ancora vera la conclusione se si rimuove l'ipotesi di positività?

Se una successione è monotona, allora è preclusa la possibilità che abbia delle oscillazioni: i termini o salgono sempre o scendono sempre, non possono fare “un po' su e un po' giù”. In termini di esistenza/non esistenza del limite questa proprietà semplifica molto la casistica.

TEOREMA 4.4. Regolarità delle successioni monotone. Una successione a_n monotona è sempre regolare (cioè o è convergente o è divergente) e

$$\begin{aligned} a_n \text{ non decrescente} &\Rightarrow \lim_{n \rightarrow +\infty} a_n = \sup_{n \in \mathbb{N}} a_n, \\ a_n \text{ non crescente} &\Rightarrow \lim_{n \rightarrow +\infty} a_n = \inf_{n \in \mathbb{N}} a_n. \end{aligned}$$

Se a_n è anche limitata, allora è convergente.

DIMOSTRAZIONE. Supponiamo a_n non decrescente e superiormente limitata,

$$a_n \leq a_m \leq \ell := \sup_{n \in \mathbb{N}} a_n < +\infty, \quad n \leq m.$$

Per dimostrare che tale successione converge ad ℓ bisogna mostrare che comunque scelto $\varepsilon > 0$ è possibile scegliere $N \in \mathbb{N}$ per cui si ha $\ell - \varepsilon < a_n < \ell + \varepsilon$ per ogni $n > N$. La seconda delle due disequazioni è sempre verificata per definizione di ℓ (è un maggiorante). Inoltre, dato che ℓ è il più piccolo dei maggioranti, per ogni $\varepsilon > 0$ esiste certamente un indice N tale che $\ell - \varepsilon < a_N$ (altrimenti $\ell - \varepsilon$ sarebbe un maggiorante e, quindi, ℓ non sarebbe il più piccolo!). Usando la monotonia si ha

$$\ell - \varepsilon < a_N \leq a_n < \ell \quad \forall n \geq N,$$

cioè la conclusione.

I casi rimanenti si dimostrano in maniera simile. □

ESERCIZIO 4.5. Sia a_n una successione non decrescente. Dimostrare che:

- (i) se da a_n si può estrarre una sottosuccessione a_{n_k} convergente, allora a_n è convergente;
- (ii) se da a_n si può estrarre una sottosuccessione a_{n_k} divergente, allora a_n diverge.

5. Serie numeriche

In generale, una **serie numerica** è definita da una successione a_n , con la richiesta di sommare i termini nell'ordine dato dall'indice n . In parole povere, si tratta di dare senso alla *somma di un numero infinito di termini*. Il procedimento più naturale (utilizzato nell'Esempio 3.7) è di considerare la successione S_n delle **somme parziali**

$$S_0 := a_0, \quad S_1 := a_0 + a_1, \quad \dots, \quad S_n := \sum_{k=0}^n a_k,$$

cioè la successione il cui termine n -esimo è la somma dei primi $n+1$ termini a_0, \dots, a_n . Se la successione delle somme parziali S_n è convergente, la serie si dice **semplicemente convergente**. La **somma della serie**, che si indica con $\sum_{n=0}^{\infty} a_n$, è definita da

$$\sum_{n=0}^{\infty} a_n := \lim_{n \rightarrow +\infty} S_n,$$

nel caso in cui tale limite esista. Si usa la stessa terminologia delle successioni: una serie è **convergente/divergente/regolare/non regolare** se la successione delle somme parziali è **convergente/divergente/regolare/non regolare**.

Grazie alla linearità del limite di successioni, anche le serie godono della proprietà di linearità: combinazioni lineari di serie convergenti danno luogo a serie convergenti.

TEOREMA 5.1. Linearità delle serie. *Siano a_k e b_k i termini generici di due serie convergenti. Allora, per ogni $\lambda, \mu \in \mathbb{R}$, la serie di termine generico $\lambda a_k + \mu b_k$ è convergente e vale la formula*

$$\sum_{k=0}^{\infty} (\lambda a_k + \mu b_k) = \lambda \sum_{k=0}^{\infty} a_k + \mu \sum_{k=0}^{\infty} b_k.$$

In generale non è facile stabilire se una serie sia o non sia convergente. Un primo ingrediente utile per questo studio è il seguente: *se una serie è convergente, allora il suo termine generico deve essere infinitesimo:*

$$\text{condizione necessaria} \quad : \quad \sum_{n=0}^{\infty} a_n \text{ convergente} \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} a_n = 0.$$

di convergenza

Infatti, sia $\sum_{n=0}^{\infty} a_n$ convergente, cioè esista finito $\lim_{n \rightarrow +\infty} S_n = S$. Dato che $a_n = S_n - S_{n-1}$ e, per ipotesi, la successione S_n converge a S per $n \rightarrow +\infty$, allora

$$\lim_{n \rightarrow +\infty} a_n = \lim_{n \rightarrow +\infty} (S_n - S_{n-1}) = \lim_{n \rightarrow +\infty} S_n - \lim_{n \rightarrow +\infty} S_{n-1} = S - S = 0.$$

Questa condizione necessaria rispecchia il fatto, suggerito dall'intuizione, secondo cui occorre sommare termini che tendono a zero per sperare che la somma di un numero infinito di termini dia un valore finito. E' importante notare che esistono serie il cui termine generico a_n è infinitesimo, ma che non sono convergenti. Ad esempio, possiamo considerare la serie definita dalla successione seguente: il primo termine vale 1, i due termini successivi valgono $\frac{1}{2}$, i tre termini successivi valgono $\frac{1}{3}$ e così via:

$$1, \frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \dots$$

E' chiaro che la successione è infinitesima, ma le corrispondenti somme ridotte verificano:

$$S_0 = 1, S_3 = 2, S_6 = 3, S_{10} = 4, S_{15} = 5, \dots$$

e quindi danno luogo ad una successione divergente.

Criteri di confronto per serie a termini positivi. Stabilire la convergenza o meno di una serie è più facile se la successione a_k è costituita da termini non negativi. Infatti, dato che $a_k \geq 0$, si ha $S_{n+1} - S_n = a_{n+1} \geq 0$, cioè $S_n \leq S_{n+1}$ per ogni n . Quindi la successione S_n è non decrescente e pertanto regolare: o converge o diverge,

$$a_k \geq 0 \quad \Rightarrow \quad \sum_{k=0}^{+\infty} a_k < +\infty \quad \text{oppure} \quad \sum_{k=0}^{+\infty} a_k = +\infty.$$

Dato che una serie a termini positivi può solo convergere o divergere a $+\infty$, condizione sufficiente di convergenza è che esista una serie maggiorante b_k convergente.

TEOREMA 5.2. Criterio di confronto per serie. *Sia $a_k \geq 0$ per ogni k .*

(i) *Se esiste b_k tali che $0 \leq a_k \leq b_k$ per ogni k , allora*

$$\sum_{k=0}^{\infty} b_k < +\infty \quad \Rightarrow \quad \sum_{k=0}^{\infty} a_k < +\infty$$

(ii) *se esiste $b_k \geq 0$ tali che $a_k \geq b_k$ per ogni k , allora*

$$\sum_{k=0}^{\infty} b_k = +\infty \quad \Rightarrow \quad \sum_{k=0}^{\infty} a_k = +\infty.$$

OSSERVAZIONE 5.3. Se in (i) e in (ii) la richiesta di stima con i termini b_k è soddisfatta definitivamente, cioè per ogni $k \geq \bar{N}$ per un opportuno \bar{N} , anziché per ogni $k \in \mathbb{N}$, la conclusione vale lo stesso, dato che *le proprietà di convergenza non cambiano quando si cambiano un numero finito di termini.*

DIMOSTRAZIONE. Dimostriamo solo la parte (i) lasciando la (ii) come esercizio. La successione delle somme parziali di a_n è non decrescente, quindi per dimostrarne la convergenza, è sufficiente mostrarne la limitatezza (grazie al Teorema 4.4). Dalle maggiorazioni

$$0 \leq \sum_{k=0}^n a_k \leq \sum_{k=0}^n b_k \leq \sum_{k=0}^{\infty} b_k < +\infty,$$

segue la conclusione. □

ESEMPIO 5.4. Consideriamo la serie (a termini non negativi)

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

Euristicamente, è ragionevole aspettarsi che, per valori di n grandi il termine generico $\frac{2^n - 1}{3^n + n}$ si possa approssimare come segue

$$\frac{2^n - 1}{3^n + n} \approx \frac{2^n}{3^n} = \left(\frac{2}{3}\right)^n,$$

quindi il sospetto è che sia convergente, vista la somiglianza con la serie geometrica. Poniamoci quindi l'obiettivo di stimarla dall'alto

$$\frac{2^n - 1}{3^n + n} \leq \frac{2^n}{3^n} = \left(\frac{2}{3}\right)^n.$$

Dato che il termine generico della serie è minore di quello della serie geometrica di ragione $\frac{2}{3}$, la serie è convergente.

OSSERVAZIONE 5.5. Una maniera facile per determinare una stima del tipo $a_k \leq Cb_k$ per k sufficientemente grande è tramite il limite del rapporto a_k/b_k . Supponiamo, ad esempio, di sapere che

$$\lim_{k \rightarrow +\infty} \frac{a_k}{b_k} = \ell \geq 0 \quad \text{e} \quad \sum_{k=1}^{\infty} b_k < +\infty,$$

Per definizione di limite, scelto $\varepsilon > 0$, esiste $N \in \mathbb{N}$ opportuno per cui, per ogni $n \geq N$,

$$\ell - \varepsilon < \frac{a_k}{b_k} < \ell + \varepsilon \quad \Rightarrow \quad a_k < (\ell + \varepsilon)b_k.$$

Quindi, applicando il precedente risultato, si dimostra che la serie $\sum_{k=0}^{\infty} a_k$ è convergente.

ESEMPIO 5.6. Fissato $x \geq 0$, studiamo la **serie esponenziale**

$$\sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Sappiamo già che il termine generico di questa serie è infinitesimo. Purtroppo questo non basta a stabilire la convergenza della serie. L'impressione è che, dato che il fattoriale cresce più rapidamente di qualsiasi esponenziale (vedi Esempio 3.5), il termine generico vada a zero molto rapidamente tanto da far convergere la serie. Scegliamo, allora, come serie b_k di confronto quella di termine generico $b_k = y^k$ con $y \in (0, 1)$. Dato che, per (12),

$$\lim_{k \rightarrow +\infty} \frac{x^k/k!}{y^k} = \lim_{k \rightarrow +\infty} \frac{(x/y)^k}{k!} = 0,$$

effettivamente, per k sufficientemente grande vale la stima $x^k/k! \leq y^k$ che mostra che la serie esponenziale è convergente per ogni scelta di $x > 0$.

Serie numeriche a segno qualsiasi. Come si è detto, verificare la convergenza semplice di una serie può essere molto complicato. Ben diverso è se la serie è a termini positivi, dato che in questo caso la successione delle somme parziali è non decrescente. Per questo motivo, per serie numeriche con termini a segno qualsiasi si introduce un nuovo tipo di convergenza che è più restrittivo di quella semplice, ma che è più facile da verificare, dato che richiede la verifica della convergenza di una serie a termini positivi.

DEFINIZIONE 5.7. La serie $\sum_{n=1}^{\infty} a_n$ converge assolutamente se è convergente la serie dei valori assoluti, cioè se $\sum_{n=1}^{\infty} |a_n|$ è convergente.

Nel caso in cui la serie sia a termini non negativi, cioè $a_n \geq 0$ per ogni n , la convergenza assoluta è equivalente alla convergenza semplice. In generale vale l'implicazione (mentre l'implicazione opposta è falsa!)

$$\text{convergenza assoluta} \quad \Rightarrow \quad \text{convergenza semplice}$$

Per dimostrarla, introduciamo due nuovi oggetti. Dato $a \in \mathbb{R}$, la parte positiva a^+ e la parte negativa a^- di a sono definite da

$$a^+ = \max(a, 0) \quad , \quad a^- = \max(-a, 0).$$

Quindi, se $a = 2$, $a^+ = 2$ e $a^- = 0$, mentre se $a = -3$, $a^+ = 0$ e $a^- = 3$. Dalla definizione discendono le proprietà seguenti:

$$0 \leq a^+ \leq |a|, \quad 0 \leq a^- \leq |a|, \quad a = a^+ - a^-, \quad |a| = a^+ + a^-.$$

ESERCIZIO 5.8. Disegnare il grafico delle funzioni parte positiva $f(x) = x^+$ e parte negativa $g(x) = x^-$.

Se a_k è il termine generico di una serie, consideriamo le altre due serie di termine generico a_k^+ e a_k^- . Ad esempio, se $a_k = (-1)^k / (k + 1)$,

$$\begin{array}{l} a_k : \quad 1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \frac{1}{5}, -\frac{1}{6}, \dots \\ a_k^+ : \quad 1, 0, \frac{1}{3}, 0, \frac{1}{5}, 0, \dots \\ a_k^- : \quad 0, \frac{1}{2}, 0, \frac{1}{4}, 0, \frac{1}{6}, \dots \end{array}$$

Che collegamento c'è tra la convergenza assoluta e la convergenza della serie delle parti positive e delle parti negative? Semplice... si tratta della stessa cosa!

PROPOSIZIONE 5.9. Una serie di termine generico a_k converge assolutamente se e solo se convergono (semplicemente) le serie di termine generico a_k^+ e a_k^- .

DIMOSTRAZIONE. Supponiamo che la serie $\sum a_k$ sia assolutamente convergente. Allora è possibile applicare il criterio di confronto per serie a termini positivi (Teorema 5.2): dato che $0 \leq a_k^\pm \leq |a_k|$ e la serie $\sum |a_k|$ è convergente, anche le serie $\sum a_k^+$ e $\sum a_k^-$ sono convergenti.

Viceversa, supponiamo che le serie $\sum a_k^+$ e $\sum a_k^-$ siano convergenti, allora per la linearità delle serie (Teorema 5.1) anche la loro somma è convergente. Dato che la loro somma è proprio la serie dei valori assoluti. \square

Da questa caratterizzazione della convergenza assoluta, discende rapidamente il risultato seguente.

TEOREMA 5.10. *Sia a_k il termine generico di un serie. Se $\sum_{k=0}^{\infty} a_k$ converge assolutamente, allora converge anche semplicemente.*

DIMOSTRAZIONE. Se la serie converge assolutamente, allora le serie $\sum a_k^+$ e $\sum a_k^-$ sono convergenti, quindi per il Teorema 5.1 anche la loro differenza è convergente: ma $a_k = a_k^+ - a_k^-$ e quindi la conclusione è a portata di mano. Allungatevi. \square

Se una serie converge assolutamente, allora vale la seguente versione (per somme infinite) della disuguaglianza triangolare:

$$\left| \sum_{k=0}^{\infty} a_k \right| \leq \sum_{k=0}^{\infty} |a_k|.$$

La dimostrazione (facile) è lasciata per esercizio.

ESEMPIO 5.11. Fissato $x \in \mathbb{R}$, consideriamo la serie

$$\sum_{k=1}^{\infty} \frac{\sin(kx)}{2^k}.$$

Questa serie è assolutamente convergente per ogni scelta di $x \in \mathbb{R}$. Infatti

$$\left| \frac{\sin(kx)}{2^k} \right| = \frac{|\sin(kx)|}{2^k} \leq \frac{1}{2^k},$$

e la serie $\sum \frac{1}{2^k}$ è convergente perché è la serie geometrica con ragione $\frac{1}{2}$. Quindi la serie $\sum \frac{\sin(kx)}{2^k}$ converge per ogni $x \in \mathbb{R}$.

La serie esponenziale. Dato $x \in \mathbb{R}$, consideriamo la serie

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} \quad x \in \mathbb{R}.$$

Abbiamo già visto che la serie è convergente per $x \geq 0$. Cosa succede per i valori di $x < 0$? Facile, basta ricorrere alla convergenza assoluta! Dato che

$$\sum_{k=0}^{\infty} \left| \frac{x^k}{k!} \right| = \sum_{k=0}^{\infty} \frac{|x|^k}{k!}$$

è convergente perché $|x| \geq 0$, la serie assegnata converge assolutamente e quindi converge anche semplicemente. Si definisce

$$(13) \quad \text{funzione esponenziale:} \quad e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad x \in \mathbb{R}.$$

Questa posizione è del tutto rigorosa dato che la serie è sempre convergente. A rigore, bisognerebbe verificare che questa definizione sia compatibile con la costruzione naïf e con le proprietà viste in precedenza per le funzioni esponenziali, ma non ci soffermeremo su tale aspetto.

Approssimazione del numero di Nepero. A partire dalla formula

$$e = \sum_{n=0}^{\infty} \frac{1}{n!},$$

è possibile ottenere approssimazioni del valore di e . Prima di tutto stimiamo l'errore che si commette sostituendo il numero e con la somma parziale $s_n = \sum_{k=0}^n \frac{1}{k!}$, cioè la differenza $e - s_n$

$$\begin{aligned} 0 < e - s_n &= \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \dots = \frac{1}{(n+1)!} \left(1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \dots \right) \\ &< \frac{1}{(n+1)!} \left(1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \dots \right) = \frac{1}{(n+1)!} \cdot \frac{1}{1 - \frac{1}{n+1}} = \frac{1}{n!n}, \end{aligned}$$

quindi vale

$$\left| e - \sum_{k=0}^n \frac{1}{k!} \right| \leq \frac{1}{n!n}.$$

Da questa stima si può determinare il numero di elementi che occorre sommare per ottenere una approssimazione di e con errore prescritto. Ad esempio, supponiamo di ammettere un errore massimo di 10^{-4} . Dato che

n	1	2	3	4	5	6	7
$n!n$	1	4	18	96	600	4320	35280
$1/(n!n)$	1	0,25	0,05	0,010412	0,0016	0,000231	0,000028

va bene scegliere $n = 7$. Quindi l'approssimazione richiesta è

$$e \approx 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} + \frac{1}{5040} = 2,718254.$$

La serie armonica e la serie armonica generalizzata. Una delle serie numeriche più famose è quella di termine generico $\frac{1}{n}$, che è detta **serie armonica**. Questa serie è divergente

$$(14) \quad \sum_{n=1}^{\infty} \frac{1}{n} = +\infty.$$

Infatti, sia $S_n := \sum_{k=1}^n \frac{1}{k}$ e consideriamo la sottosuccessione S_{2^k} di S_n . Allora si ha

$$S_1 = 1, \quad S_2 = 1 + \frac{1}{2}, \quad S_4 = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right), \dots$$

e così via. Manteniamo i termini S_1 e S_2 così come sono e preoccupiamoci dei successivi.

Il termine S_4 può essere stimato dal basso:

$$S_4 = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) \geq 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) = 1 + \frac{1}{2} + \frac{2}{4} = 1 + \frac{1}{2} + \frac{1}{2} = 1 + \frac{2}{2}.$$

Analogamente, vale

$$\begin{aligned} S_8 &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) = S_4 + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) \\ &\geq S_4 + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) = S_4 + \frac{1}{2} \geq 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1 + \frac{3}{2}. \end{aligned}$$

In generale, consideriamo il termine $S_{2^{k+1}}$:

$$\begin{aligned} S_{2^{k+1}} &= S_{2^k} + \left(\frac{1}{2^k+1} + \frac{1}{2^k+2} + \dots + \frac{1}{2^{k+1}}\right) \\ &\geq S_{2^k} + \left(\frac{1}{2^{k+1}} + \frac{1}{2^{k+1}} + \dots + \frac{1}{2^{k+1}}\right) = S_{2^k} + \frac{2^k}{2^{k+1}} = S_{2^k} + \frac{1}{2} \end{aligned}$$

dato che i termini in parentesi sono $2^{k+1} - 2^k = 2^k$. Quindi, ogni volta che l'indice k della sottosuccessione S_{2^k} aumenta di 1, la stima dal basso aumenta di $1/2$. Dato che $S_{2^0} = 1$, se ne deduce che

$$S_{2^{k+1}} \geq 1 + \frac{k+1}{2},$$

da cui segue che la sottosuccessione S_{2^k} è divergente. Dato che S_n è crescente ed ammette una sottosuccessione divergente, essa stessa è divergente.

In generale, la **serie armonica generalizzata**

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha} \quad \alpha > 0,$$

è convergente se e solo se $\alpha > 1$.

La divergenza nel caso $\alpha \in (0, 1)$ segue da $n^\alpha < n$ per ogni n (dimostrare!), che implica:

$$\sum_{k=1}^n \frac{1}{k^\alpha} > \sum_{k=1}^n \frac{1}{k}.$$

Nel caso $\alpha > 1$ tale serie è convergente. Dato che si tratta di una “somma” di termini positivi, la successione delle somme parziali è monotona crescente. Per la proprietà delle successioni monotone (Teorema 4.4), per dimostrarne la convergenza, basta verificare che la serie non diverga a $+\infty$. Perciò, con astuzia, consideriamo gli insiemi $I_k = \{n \in \mathbb{N} : 2^k \leq n < 2^{k+1}\}$,

notando che $\mathbb{N} = \bigcup_{k=0}^{\infty} I_k$. L'insieme I_k ha $2^{k+1} - 2^k = 2^k(2-1) = 2^k$ termini e, se $n \in I_k$, vale la maggiorazione $1/n^\alpha \leq 1/2^{\alpha k}$. Stimiamo la somma dei termini $1/n^\alpha$ per $n \in I_k$:

$$\sum_{n \in I_k} \frac{1}{n^\alpha} \leq \sum_{n \in I_k} \frac{1}{2^{\alpha k}} = 2^k \cdot \frac{1}{2^{\alpha k}} = \left(\frac{1}{2^{\alpha-1}} \right)^k.$$

Quindi, notando che $\frac{1}{2^{\alpha-1}} \in (-1, 1)$ per $\alpha > 1$,

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha} = \sum_{k=0}^{\infty} \sum_{n \in I_k} \frac{1}{n^\alpha} \leq \sum_{k=0}^{\infty} \left(\frac{1}{2^{\alpha-1}} \right)^k < +\infty$$

grazie alla convergenza della serie geometrica.

CAPITOLO 4

Le funzioni continue

Dopo l'*excursus* del Capitolo precedente sulle successioni numeriche, torniamo a parlare di funzioni reali di variabile reale in generale. Per fissare le idee, supponiamo di voler studiare funzioni f , definite in $I \subset \mathbb{R}$, dove I è un intervallo (limitato o illimitato) di \mathbb{R} . L'obiettivo principale del Capitolo è definire il significato della parola *continuità*.

1. Limite di funzioni

Tutto nasce dalla definizione di “limite”. Come abbiamo visto per le successioni, il limite formalizza l'idea di “previsione” del comportamento di un oggetto sotto osservazione per opportuni valori della variabile.

Limiti all'infinito. Partiamo prima di tutto dal concetto di funzione infinitesima per $x \rightarrow +\infty$. Una funzione $d : \mathbb{R} \rightarrow \mathbb{R}$ è **infinitesima** per $x \rightarrow +\infty$, se

$$(15) \quad \forall \varepsilon > 0 \quad \exists M \quad \text{tale che} \quad |d(x)| < \varepsilon \quad \forall x > M.$$

Il numero $M \in \mathbb{R}$ dipende dalla scelta di ε (come n_ε per le successioni): $M = M(\varepsilon)$.

La proprietà $|d(x)| < \varepsilon$ (equivalente a $-\varepsilon < d(x) < \varepsilon$) indica che il grafico della funzione d vive nella striscia infinita delimitata dalle rette $y = -\varepsilon$ e $y = \varepsilon$ per x sufficientemente grandi (Fig.1(a)), quindi la condizione (15) significa che il grafico della funzione d “tende a confondersi” con l'asse x per $x \rightarrow +\infty$.

Data una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$, questa tende ad un limite ℓ per $x \rightarrow +\infty$ se è la funzione $f(x) - \ell$ ad essere infinitesima.

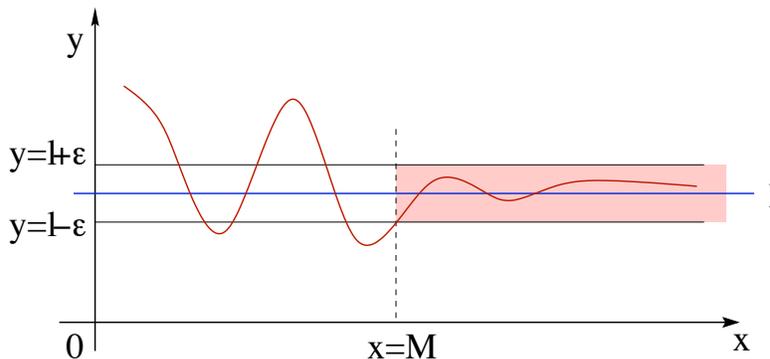
DEFINIZIONE 1.1. *Data $f : \mathbb{R} \rightarrow \mathbb{R}$, si dice che f converge ad $\ell \in \mathbb{R}$ per $x \rightarrow +\infty$*

$$(16) \quad \lim_{x \rightarrow +\infty} f(x) = \ell,$$

se $|f(x) - \ell|$ è infinitesima per $x \rightarrow +\infty$, cioè se

$$\forall \varepsilon > 0 \quad \exists M \quad \text{tale che} \quad |f(x) - \ell| < \varepsilon \quad \forall x > M.$$

La funzione $d(x) := |f(x) - \ell|$ rappresenta la lunghezza del segmento verticale di estremi $(x, f(x))$ e (x, ℓ) e “tendere ad ℓ ” indica che tale lunghezza tende a zero.

FIGURA 1. Una funzione che tende ad un limite per $x \rightarrow +\infty$.

Buona parte di quanto visto per le successioni si può ripetere. Ad esempio,

$$\lim_{x \rightarrow +\infty} \frac{x^2}{1+x^2} = 1.$$

Infatti, per ogni $\varepsilon \in (0, 1)$, si ha

$$\left| \frac{x^2}{1+x^2} - 1 \right| = \frac{1}{1+x^2} < \varepsilon \quad \forall x > M := \sqrt{\frac{1}{\varepsilon} - 1}.$$

OSSERVAZIONE 1.2. Nella definizione di limite di funzione per $x \rightarrow +\infty$, siamo partiti da una funzione f definita in tutto \mathbb{R} . Per definire il limite per $x \rightarrow +\infty$, basta anche di meno: l'unica cosa indispensabile è che l'insieme di definizione sia non limitato superiormente. Pensate al caso delle successioni: sono funzioni definite su \mathbb{N} (e quindi non su una semiretta) e il limite per $n \rightarrow +\infty$ ha perfettamente senso.

Analogamente si possono definire anche:

– limiti divergenti: $\lim_{x \rightarrow +\infty} f(x) = +\infty$ o $= -\infty$;

– limiti per x che tende a $-\infty$: $\lim_{x \rightarrow -\infty} f(x)$

Limiti in un punto. Per funzioni definite in intervalli, è possibile parlare di limite in un punto. Procediamo come in precedenza chiarendo prima il concetto di “funzione infinitesima in un punto” e poi il concetto di “limite di funzione in un punto”.

Sia $x_0 \in [a, b]$. Una funzione $d : (a, b) \rightarrow \mathbb{R}$ è **infinitesima** per $x \rightarrow x_0$ se

$$(17) \quad \forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \quad \text{t.c.} \quad |d(x)| < \varepsilon \quad \forall x \in (a, b), \quad 0 < |x - x_0| < \delta.$$

Rispetto alla definizione di funzione infinitesima per $x \rightarrow +\infty$, l'unica differenza sta nelle scelte di x per cui è soddisfatta la condizione $-\varepsilon < d(x) < \varepsilon$. In questo caso si tratta di tutti i valori x , diversi da x_0 , che distano da x_0 meno di $\delta > 0$.

DEFINIZIONE 1.3. Sia $f : (a, b) \rightarrow \mathbb{R}$ e sia $x_0 \in [a, b]$. La funzione f tende ad $\ell \in \mathbb{R}$ per $x \rightarrow x_0$ se $f(x) - \ell$ è infinitesima per $x \rightarrow x_0$, cioè se

$$\forall \varepsilon > 0 \quad \exists \delta = \delta(\varepsilon) > 0 \quad \text{t.c.} \quad |f(x) - \ell| < \varepsilon \quad \forall x \in (a, b), 0 < |x - x_0| < \delta.$$

In questo caso, si scrive

$$(18) \quad \lim_{x \rightarrow x_0} f(x) = \ell.$$

Il punto fondamentale è nella definizione di funzione infinitesima: per dimostrare che il limite della funzione è ℓ bisogna verificare che la distanza tra $f(x)$ e ℓ , cioè la quantità $d(x) := |f(x) - \ell|$, diventa piccola quando x è sufficientemente vicino a x_0 .

ESEMPIO 1.4. Proviamo a dimostrare il limite

$$\lim_{x \rightarrow 1} (3x - 5) = -2.$$

In questo caso $f(x) = 3x - 5$, $x_0 = 1$ e $\ell = -2$. Per definizione, basta mostrare che la quantità $|f(x) - \ell|$ è infinitesima per x che tende ad 1. Poniamoci quindi l'obiettivo di stimarla in termini di una funzione in cui compaia la distanza $|x - 1|$:

$$|f(x) - \ell| = |(3x - 5) - (-2)| = |3x - 3| = 3|x - 1|.$$

Perfetto! Da queste uguaglianze segue la conclusione. Se vogliamo conoscere esplicitamente il valore di δ in funzione di ε , così come richiesto dalla definizione, basta osservare che se $|x - 1| < \delta$, allora $|f(x) - \ell| < 3\delta$, quindi dato $\varepsilon > 0$, basta scegliere δ in modo che $3\delta = \varepsilon$, cioè $\delta = \varepsilon/3$.

ESEMPIO 1.5. Fissato $x_0 \in \mathbb{R}$, calcoliamo $\lim_{x \rightarrow x_0} \sin x$. E' ragionevole aspettarsi che tale limite esista e che valga $\ell = \sin x_0$, quindi proviamo a stimare $|\sin x - \sin x_0|$. Usando una delle (diaboliche) formule di prostaferesi,

$$|\sin x - \sin x_0| = \left| 2 \sin \left(\frac{x - x_0}{2} \right) \cos \left(\frac{x + x_0}{2} \right) \right| \leq 2 \left| \sin \left(\frac{x - x_0}{2} \right) \right|.$$

Dato che $|\sin x| \leq |x|$ per ogni $x \in \mathbb{R}$ (Cap. 2, Es. 1.4), si ottiene

$$|\sin x - \sin x_0| \leq 2 \left| \frac{x - x_0}{2} \right| = |x - x_0|,$$

da cui segue

$$\lim_{x \rightarrow x_0} \sin x = \sin x_0.$$

Volendo determinare esplicitamente δ , dato $\varepsilon > 0$, basta scegliere $\delta := \varepsilon$ per fare in modo che, se $|x - x_0| < \delta$, allora $|\sin x - \sin x_0| < \varepsilon$.

ESERCIZIO 1.6. Dimostrare che, per ogni $x_0 \in \mathbb{R}$, vale $\lim_{x \rightarrow x_0} \cos x = \cos x_0$.

ESEMPIO 1.7. Vediamo un limite più complicato:

$$\lim_{x \rightarrow 0} e^x = 1.$$

Euristicamente il risultato è più che ragionevole, dato che, dalla definizione dell'esponenziale data in (13), segue

$$e^x - 1 = \sum_{n=1}^{\infty} \frac{x^n}{n!} = x \sum_{n=0}^{\infty} \frac{x^n}{(n+1)!} = x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!} + \cdots,$$

e ciascuno dei termini sommati tende a zero per $x \rightarrow 0$. Il problema è che i termini sommati sono infiniti! Per dimostrare in modo rigoroso la validità del limite bisogna, come sempre, stimare il termine $|f(x) - \ell| = |e^x - 1|$:

$$|e^x - 1| = \left| x \sum_{n=0}^{\infty} \frac{x^n}{(n+1)!} \right| \leq |x| \sum_{n=0}^{\infty} \frac{|x|^n}{(n+1)!} \leq |x| \sum_{n=0}^{\infty} \frac{|x|^n}{n!} = |x|e^{|x|}$$

(nella riga precedente ci sono due disuguaglianze per serie... perché sono lecite?). Se scegliamo $|x| < 1$, si ha $e^{|x|} \leq e$ (si ricordi che la funzione e^x è crescente), quindi

$$|e^x - 1| \leq |x|e^{|x|} \leq e|x| \quad \forall x \in (-1, 1),$$

da cui si arriva alla conclusione.

OSSERVAZIONE 1.8. $x \neq x_0$. I punti x che intervengono nel limite per $x \rightarrow x_0$ sono, per definizione, *distinti da x_0* . In parole povere, il limite della funzione f per $x \rightarrow x_0$ è il comportamento che si prevede per la funzione f in x_0 , in base al grafico della funzione vicino a x_0 , ma *indipendentemente* da quello che succede nel punto limite.

A guardare bene, la Definizione 1.3 vale, così com'è per funzioni f definite in $(a, b) \setminus \{x_0\}$ (e $x_0 \in (a, b)$), cioè funzioni che non sono definite nel punto limite! Quello che conta è il punto limite x_0 sia "vicino" a punti in cui la funzione è definita: non ha senso calcolare il limite per $x \rightarrow 2$ di una funzione che è definita in $[0, 1]$!

Come si dimostra che un limite non esiste? Dalla definizione di limite, segue che, se $\lim_{x \rightarrow x_0} f(x) = \ell$, allora per ogni successione x_n , contenuta nell'insieme di definizione di f , e tale che x_n tende a x_0 per $n \rightarrow +\infty$, la successione $f(x_n)$ tende ad ℓ per $n \rightarrow +\infty$:

$$(19) \quad \lim_{n \rightarrow +\infty} x_n = x_0 \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} f(x_n) = \ell.$$

ESERCIZIO 1.9. *Dimostrare l'affermazione che avete appena letto.*

Dall'implicazione (19) discende il seguente

CRITERIO 1.10. Non esistenza del limite. *Se esistono due successioni x_n e ξ_n entrambe convergenti a x_0 e tali che*

$$\lim_{n \rightarrow +\infty} f(x_n) \neq \lim_{n \rightarrow +\infty} f(\xi_n),$$

la funzione f non può ammettere limite per $x \rightarrow x_0$.

Un esempio chiarirà meglio le idee. Consideriamo la funzione

$$\text{segno di } x: \quad \text{sgn } x := \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ +1 & x > 0 \end{cases}$$

e consideriamo $x_n = \frac{1}{n}$ e $\xi_n = -\frac{1}{n}$. È evidente che $\lim_{n \rightarrow +\infty} x_n = \lim_{n \rightarrow +\infty} \xi_n = 0$. Inoltre, per ogni n , $f(x_n) = 1$ e $f(\xi_n) = -1$, quindi

$$\lim_{n \rightarrow +\infty} f(x_n) = 1 \neq -1 = \lim_{n \rightarrow +\infty} f(\xi_n).$$

Pertanto la funzione sgn non ammette limite per $x \rightarrow 0$.

ESERCIZIO 1.11. *Dimostrare che $\sin(1/x)$ non ammette limite per $x \rightarrow 0$.*

Limite destro e limite sinistro. Quando si studia una funzione solo a destra o a sinistra del punto limite x_0 si parla di limite destro e di limite sinistro.

DEFINIZIONE 1.12. *La funzione f ha limite destro uguale ad ℓ per x che tende a x_0 , e si scrive $\lim_{x \rightarrow x_0^+} f(x) = \ell$, se*

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \text{t.c.} \quad |f(x) - \ell| < \varepsilon \quad \forall x \in (a, b), \quad x_0 < x < x_0 + \delta.$$

Analogamente per il limite sinistro, che si indica con $\lim_{x \rightarrow x_0^-} f(x) = \ell$.

ESERCIZIO 1.13. *Dimostrare che $\lim_{x \rightarrow 0^-} \text{sgn } x = -1$, $\lim_{x \rightarrow 0^+} \text{sgn } x = +1$.*

ESERCIZIO 1.14. *Calcolare i seguenti limiti*

$$\lim_{x \rightarrow 0^-} \arctan\left(\frac{1}{x}\right), \quad \lim_{x \rightarrow 0^+} \arctan\left(\frac{1}{x}\right).$$

Dalla definizione di limite destro e sinistro si deduce (con poca fatica) il seguente:

CRITERIO 1.15. Esistenza del limite. *Una funzione ammette limite in un punto se e solo se esistono sia il limite destro che quello sinistro e coincidono.*

Di conseguenza, se uno tra i limiti destro e sinistro non esiste, o se entrambi esistono, ma non coincidono, la funzione non ha limite per $x \rightarrow x_0$. Avendo risolto l'Esercizio 1.14, sapete dire se $\arctan(1/x)$ ammette limite per $x \rightarrow 0$?

Limiti e operazioni razionali. Limiti di somme, differenze, prodotti e rapporti di funzioni godono delle stesse proprietà viste per le successioni:

$$(20) \quad \left. \begin{array}{l} \lim_{x \rightarrow x_0} f(x) = \ell \\ \lim_{x \rightarrow x_0} g(x) = m \end{array} \right\} \Rightarrow \begin{cases} \lim_{x \rightarrow x_0} (f(x) \pm g(x)) = \ell \pm m, \\ \lim_{x \rightarrow x_0} f(x)g(x) = \ell m, \\ \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \frac{\ell}{m} \quad \text{se } m \neq 0. \end{cases}$$

ESEMPIO 1.16. Per calcolare

$$\lim_{x \rightarrow 1} \frac{3x^2 + x - 1}{x^5 + x^3}$$

non è una buona idea usare la definizione! Basta applicare le regole su descritte:

$$\begin{aligned} \lim_{x \rightarrow 1} \frac{3x^2 + x - 1}{x^5 + x^3} &= \frac{\lim_{x \rightarrow 1} (3x^2 + x - 1)}{\lim_{x \rightarrow 1} (x^5 + x^3)} = \frac{\lim_{x \rightarrow 1} 3x^2 + \lim_{x \rightarrow 1} x - \lim_{x \rightarrow 1} 1}{\lim_{x \rightarrow 1} x^5 + \lim_{x \rightarrow 1} x^3} \\ &= \frac{(\lim_{x \rightarrow 1} 3)(\lim_{x \rightarrow 1} x)^2 + \lim_{x \rightarrow 1} x - \lim_{x \rightarrow 1} 1}{(\lim_{x \rightarrow 1} x)^5 + (\lim_{x \rightarrow 1} x)^3} = \frac{3 + 1 - 1}{1 + 1} = \frac{3}{2}. \end{aligned}$$

Ora, se volete, provate a dimostrare il risultato usando solo la definizione di limite...

ESEMPIO 1.17. Fissato $x_0 \in \mathbb{R}$, calcoliamo

$$\lim_{x \rightarrow x_0} \frac{x^2 - x_0^2}{x - x_0}.$$

Qui non è possibile applicare direttamente le regole viste, perché il denominatore tende a zero per $x \rightarrow x_0$. Ma basta una riga di conto per risolvere il problema:

$$\lim_{x \rightarrow x_0} \frac{x^2 - x_0^2}{x - x_0} = \lim_{x \rightarrow x_0} \frac{(x - x_0)(x + x_0)}{x - x_0} = \lim_{x \rightarrow x_0} (x + x_0) = \lim_{x \rightarrow x_0} x + \lim_{x \rightarrow x_0} x_0 = 2x_0.$$

Analogamente, si dimostra che

$$\lim_{x \rightarrow x_0} \frac{x^3 - x_0^3}{x - x_0} = 3x_0^2,$$

utilizzando l'identità $x^3 - x_0^3 = (x^2 + xx_0 + x_0^2)(x - x_0)$. In generale, vale

$$\lim_{x \rightarrow x_0} \frac{x^n - x_0^n}{x - x_0} = nx_0^{n-1} \quad n \in \mathbb{N},$$

infatti

$$\frac{x^n - x_0^n}{x - x_0} = x^{n-1} + x^{n-2}x_0 + \cdots + xx_0^{n-2} + x_0^{n-1}.$$

Limiti e disequazioni. Anche per il rapporto tra limiti e disequazioni valgono le stesse regole già viste nel caso delle successioni: supponiamo che le funzioni f e g abbiano limite per $x \rightarrow x_0$, allora

$$f(x) < g(x) \quad (\text{o } f(x) \leq g(x)) \quad \Rightarrow \quad \lim_{x \rightarrow x_0} f(x) \leq \lim_{x \rightarrow x_0} g(x).$$

La disuguaglianza stretta diviene una disuguaglianza debole. Le dimostrazioni sono analoghe a quelle per le successioni.

Da queste proprietà discende la seguente proposizione (analogo al Teorema 2.3).

PROPOSIZIONE 1.18. *Siano f, g, h tre funzioni tali che $f(x) \leq g(x) \leq h(x)$ per tutti i valori x in un intorno di x_0 . Allora*

$$\lim_{x \rightarrow x_0} f(x) = \ell = \lim_{x \rightarrow x_0} h(x) \quad \Longrightarrow \quad \lim_{x \rightarrow x_0} g(x) = \ell.$$

Omettiamo la dimostrazione.

Zeri a denominatore ed uso degli infiniti. Anche per quanto riguarda quest'argomento, quello che c'è da capire è interamente contenuto nel caso delle successioni. In particolare, le forme indeterminate che si incontrano con più frequenza sono: $\frac{0}{0}$, $\infty \cdot 0$, $+\infty - \infty$, $\frac{\infty}{\infty}$.

Alcuni limiti notevoli. Il fatto che il limite sia compatibile con le operazioni di somma e prodotto fa in modo che nel calcolo effettivo dei limiti, nella maggior parte dei casi, non si debba utilizzare direttamente la definizione (con conseguente calcolo di ε e $\delta(\varepsilon)$, spesso tremendamente complicato), ma ci si possa ricondurre a limiti già noti. Il problema, a questo punto, è che di limiti noti ne abbiamo pochini... Corriamo al mercato ad acquistarne un po'.

ESEMPIO 1.19. Partiamo da un limite che non può mancare nella casa di nessuno:

$$(21) \quad \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1,$$

(il valore x , come sempre, è calcolato in radianti). Dal significato geometrico di $\sin x$ si deduce immediatamente che

$$\sin x < x < \tan x \quad \forall x \in (0, \pi/2).$$

Ne segue che, per ogni $x \in (0, \frac{\pi}{2})$,

$$1 < \frac{x}{\sin x} < \frac{1}{\cos x} \quad \Rightarrow \quad \cos x < \frac{\sin x}{x} < 1$$

Dato che $\cos x$ tende a $\cos 0 = 1$ per $x \rightarrow 0$, il rapporto $\frac{\sin x}{x}$ tende ad 1 per $x \rightarrow 0^+$. Lo stesso vale anche per $x \rightarrow 0^-$, dato che la funzione $\frac{\sin x}{x}$ è una funzione pari (verificare!). Quindi, per il Criterio 1.15, il gioco è fatto.

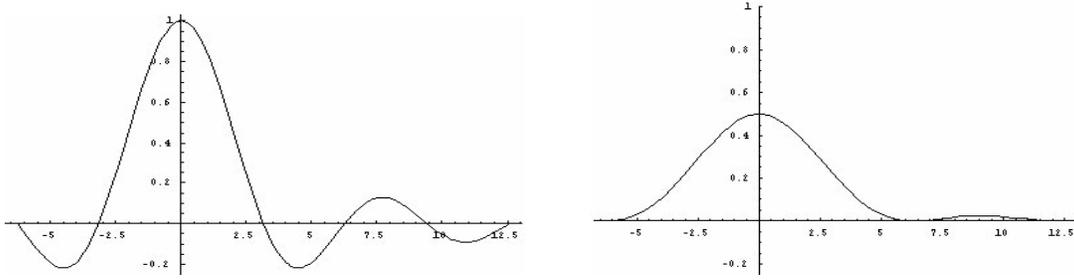


FIGURA 2. (a) $y = \frac{\sin x}{x}$, (b) $y = \frac{1-\cos x}{x^2}$

ESERCIZIO 1.20. Utilizzando (21), dimostrare

$$\lim_{x \rightarrow 0} \frac{\tan x}{x} = 1, \quad \lim_{x \rightarrow 0} \frac{1 - \cos x}{x} = 0, \quad \lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \frac{1}{2}.$$

Soluzione. Per il primo, basta ricordare la definizione di $\tan x$ e usare le proprietà dei limiti

$$\lim_{x \rightarrow 0} \frac{\tan x}{x} = \lim_{x \rightarrow 0} \frac{\sin x}{x} \frac{1}{\cos x} = \lim_{x \rightarrow 0} \frac{\sin x}{x} \frac{1}{\lim_{x \rightarrow 0} \cos x} = 1.$$

Per i restanti due, si può utilizzare l'uguaglianza

$$1 - \cos x = \frac{1 - \cos^2 x}{1 + \cos x} = \frac{\sin^2 x}{1 + \cos x},$$

da cui seguono

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{1 - \cos x}{x} &= \lim_{x \rightarrow 0} \frac{\sin x}{x} \frac{1}{1 + \cos x} \sin x = 0. \\ \lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} &= \lim_{x \rightarrow 0} \left(\frac{\sin x}{x} \right)^2 \frac{1}{1 + \cos x} = \frac{1}{2}. \end{aligned}$$

e quindi il risultato.

ESEMPIO 1.21. Una coppia di limiti molto importanti è

$$(22) \quad \lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1, \quad \lim_{x \rightarrow 0} \frac{\ln(1+x)}{x} = 1.$$

Per dimostrare il primo limite di (22), notiamo che

$$f(x) := \frac{e^x - 1}{x} - 1 = \sum_{n=1}^{\infty} \frac{x^n}{(n+1)!} = x \sum_{n=0}^{\infty} \frac{x^n}{(n+2)!},$$

dove si è usato che $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. Quindi

$$|f(x)| \leq |x| \sum_{n=0}^{\infty} \frac{|x|^n}{(n+2)!} \leq |x| \sum_{n=0}^{\infty} \frac{|x|^n}{n!} = |x|e^{|x|}.$$

Dato che $0 < e^{|x|} \leq e$ per tutti i valori $x \in [-1, 1]$, si ha $|f(x)| \leq e|x|$ che tende a zero per $x \rightarrow 0$.

Il secondo limite in (22) si può ottenere dal primo ponendo $y = e^x - 1$:

$$\lim_{y \rightarrow 0} \frac{y}{\ln(1+y)} = \lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1,$$

e passando agli inversi si ha la conclusione.

OSSERVAZIONE 1.22. Nel calcolo di quest'ultimo limite si è utilizzato il cambiamento di variabile per dedurre il valore del limite a partire dal precedente. La giustificazione rigorosa di questo procedimento può essere fatta, con un po' di attenzione, ma senza troppa difficoltà, a partire dalla definizione di limite.

I limiti appena presentati sono utili come esempi, ma allo stesso tempo, sono fondamentali per riuscire a calcolare altri limiti. Altri limiti importanti, di cui non diamo la dimostrazione, sono

$$\lim_{x \rightarrow 0} x^\alpha \ln x = 0, \quad \lim_{x \rightarrow +\infty} \frac{x^\alpha}{a^x} = 0, \quad \lim_{x \rightarrow +\infty} \frac{\log_a x}{x^\alpha} = 0 \quad \forall a > 1, \alpha > 0.$$

Il significato di ciascuno di questi è particolarmente interessante. Nel primo limite, la funzione x^α tende a 0 per $x \rightarrow 0$, mentre $\ln x$ tende a $-\infty$. Non è chiaro a priori quale sia il comportamento della funzione prodotto dato che sono presenti due termini contrastanti. Il fatto che il limite valga zero vuol dire che la funzione x^α tende a zero tanto rapidamente da riuscire a dominare la divergenza $-\infty$ del termine $\ln x$. Allo stesso modo, il secondo limite indica che l'esponenziale a^x , con $a > 1$ diverge più rapidamente di x^α , e il terzo esprime che, al contrario, il logaritmo $\log_a x$, con $a > 1$, diverge più lentamente di x^α . Sulle questioni di ordini di infinito e di infinitesimo ritorneremo più avanti.

2. Continuità

Il concetto di limite è collegato a quello di *continuità*. Intuitivamente la continuità significa che piccoli cambiamenti nella variabile indipendente x provocano piccoli cambiamenti nella variabile dipendente $y = f(x)$. Al contrario un grafico costituito da due parti separate da una "frattura" in corrispondenza dell'ascissa x_0 esibisce (in quel punto) una *discontinuità di salto* (ad esempio, la funzione $\operatorname{sgn} x$ ha una discontinuità di salto in $x_0 = 0$).

L'idea di continuità è implicita nell'uso quotidiano della matematica elementare. Quando una funzione $y = f(x)$ è descritta da tabelle (come nel caso dei logaritmi o delle funzioni trigonometriche), i valori di y possono essere dati solo per un insieme "discreto" di valori della variabile indipendente x , ad esempio in intervalli di lunghezza 10^{-3} (un millesimo) o 10^{-6} (un milionesimo). Però potrebbe essere utile conoscere il valore della funzione per valori intermedi. In questo caso, si assume tacitamente

che il valore $f(x_0)$ cercato, corrispondente ad un valore x_0 non presente nella tabella, sia approssimativamente lo stesso di $f(x)$ per un x che appaia nella tabella e che sia “vicino” ad x_0 .

DEFINIZIONE 2.1. *Sia I un intervallo di \mathbb{R} . La funzione $f : I \rightarrow \mathbb{R}$ è continua in $x_0 \in I$ se ha limite per $x \rightarrow x_0$ esiste e tale limite coincide con il valore di f in x_0 :*

$$\lim_{x \rightarrow x_0} f(x) = f(x_0),$$

cioè (ricordando la definizione di limite) se

$$\forall \varepsilon > 0, \quad \exists \delta > 0 \quad \text{t.c.} \quad |f(x) - f(x_0)| < \varepsilon \quad \forall x \in I, |x - x_0| < \delta.$$

Se una funzione f è continua in ogni punto $x_0 \in I$ allora f è continua in I .

Sia $P_0 = (x_0, y_0)$ un punto nel grafico. I punti (x, y) tali che $y_0 - \varepsilon < y < y_0 + \varepsilon$ costituiscono una striscia orizzontale J che contiene P_0 . La continuità di f in x_0 significa che per ogni striscia di questo genere J (di qualsiasi ampiezza) è possibile determinare una striscia verticale K data da $x_0 - \delta < x < x_0 + \delta$ sufficientemente piccola tale che tutti i punti del grafico di f che sono in K giacciono anche in J .

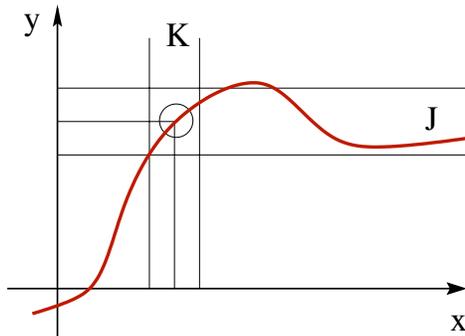


FIGURA 3. Significato geometrico della continuità

ESEMPIO 2.2. Per la funzione affine $f(x) = 5x + 3$ abbiamo

$$|f(x) - f(x_0)| = |(5x + 3) - (5x_0 + 3)| = 5|x - x_0|,$$

che esprime il fatto che la funzione $y = 5x + 3$ *dilata* le distanze di un fattore 5. In questo caso, ovviamente $|f(x) - f(x_0)| < \varepsilon$ per tutti valori x per cui $|x - x_0| < \varepsilon/5$. La condizione di continuità di f nel punto x_0 è soddisfatta scegliendo $\delta = \varepsilon/5$. Chiaramente è possibile scegliere un qualsiasi valore positivo tale che $\delta \leq \varepsilon/5$.

OSSERVAZIONE 2.3. Nella definizione di continuità, la condizione $|f(x) - f(x_0)| < \varepsilon$ è soddisfatta anche per x_0 , a differenza della definizione di limite dove si chiede $|f(x) - \ell|$ per valori x vicini a x_0 , ma diversi da x_0 stesso.

OSSERVAZIONE 2.4. *Scommettiamo che...* Per chiarire ulteriormente il significato di continuità, spieghiamo le regole di un gioco per due persone. Supponiamo assegnata una funzione f ed il punto x_0 nel suo insieme di definizione. Il giocatore B può scegliere un qualsiasi numero $\varepsilon > 0$ a suo gusto e piacimento. Per ogni scelta di ε compiuta da B , A deve essere in grado di determinare $\delta > 0$ in modo che tutti i valori immagine $f(x)$, per x che dista da x_0 meno di δ , distino da $f(x_0)$ meno di ε . Se il giocatore B trova un $\varepsilon > 0$ per cui A non possa rispondere, vince; viceversa, se per ogni ε , A è in grado di trovare δ opportuno, vince il giocatore A . Il giocatore A vince se e solo se la funzione f è continua in x_0 .

Se la funzione è $\sin(x^2)$ ed il punto $x_0 = 1$, quale giocatore vorreste essere: il giocatore A o il giocatore B ?

Ora che abbiamo una definizione chiara di continuità, vorremmo sapere quante e quali funzioni tra quelle che conosciamo sono continue. Dalle proprietà dei limiti di somma, prodotto, quoziente discende che

la somma, la differenza, il prodotto e il rapporto di funzioni continue danno luogo a funzioni continue (prudenza nel quoziente!¹).

Anche le operazioni di composizione e di inversione conservano la continuità:

*la composizione $f \circ g$ di funzioni f e g continue è continua
l'inversa f^{-1} di una funzione f continua è una funzione continua*

La prima delle due proprietà discende dalla catena di implicazioni

$$x \rightarrow x_0 \quad \Rightarrow \quad g(x) \rightarrow g(x_0) \quad \Rightarrow \quad f(g(x)) \rightarrow f(g(x_0)).$$

La continuità della funzione inversa è geometricamente evidente, una volta ricordato che il grafico di f^{-1} si può ottenere da quello di f tramite un ribaltamento attorno alla bisettrice del primo e terzo quadrante.

Ma tutte queste bellissime proprietà non servono a nulla fino a che non si conosca per lo meno una funzione continua. Passiamo quindi ad analizzare qualche esempio di base.

ESEMPIO 2.5. *Le funzioni costanti sono continue.* Banale! Infatti se $f(x) = c$ per ogni x , allora $|f(x) - f(x_0)| = |c - c| = 0$ sempre e comunque.

ESEMPIO 2.6. *La funzione $f(x) = x$ è continua.* Anche questo è facile, dato che, fissato $x_0 \in \mathbb{R}$, si ha $|f(x) - f(x_0)| = |x - x_0|$, quindi basta scegliere $\delta(\varepsilon) = \varepsilon$ nella definizione di continuità per giungere alla conclusione.

¹Come sempre nel caso della divisione, bisogna stare attenti al fatto che la divisione per zero non ha senso. Perciò se si hanno due funzione continue f e g , la funzione rapporto è una funzione continua dove è definito, cioè dove la funzione g non si azzerava.

ESEMPIO 2.7. *I polinomi sono funzioni continue.* Qui basta combinare le proprietà dei limiti (20), con la definizione di continuità e con i due esempi precedenti. Se $p(x) = a_0 + a_1x + \dots + a_nx^n$ per $a_0, a_1, \dots, a_n \in \mathbb{R}$ dati, allora

$$\begin{aligned} \lim_{x \rightarrow x_0} p(x) &= \lim_{x \rightarrow x_0} (a_0 + a_1x + \dots + a_nx^n) \\ &= \lim_{x \rightarrow x_0} a_0 + \lim_{x \rightarrow x_0} a_1 \lim_{x \rightarrow x_0} x + \dots + \lim_{x \rightarrow x_0} a_n (\lim_{x \rightarrow x_0} x)^n \\ &= a_0 + a_1x_0 + \dots + a_nx_0^n = p(x_0). \end{aligned}$$

ESEMPIO 2.8. *La funzione $f(x) = \sin x$ è una funzione continua.* Lo abbiamo già visto nell'Esempio 1.5. Stesso dicasi per $\cos x$ (avete risolto l'Esercizio 1.6?).

ESEMPIO 2.9. Cosa dire dell'esponenziale e^x ? L'Esempio 1.7 ne garantisce la continuità in $x_0 = 0$. Da questa è possibile dedurre la continuità anche negli altri punti, utilizzando la proprietà $e^{x+y} = e^x e^y$. Infatti:

$$\lim_{x \rightarrow x_0} e^x = \lim_{x \rightarrow x_0} e^{x-x_0+x_0} = \lim_{x \rightarrow x_0} e^{x-x_0} e^{x_0} = \lim_{x \rightarrow x_0} e^{x-x_0} \lim_{x \rightarrow x_0} e^{x_0} = \lim_{h \rightarrow 0} e^h \lim_{x \rightarrow x_0} e^{x_0} = e^{x_0}.$$

Una volta che abbiamo questi mattoni fondamentali, ecco a cascata una quantità impressionante di funzioni continue:

- le funzioni razionali,
- le funzioni trigonometriche,
- esponenziali e logaritmi,
- tutte le loro composizioni e inverse.

ESERCIZIO 2.10. *Perché le funzioni $f(x) = 10^x$ e $g(x) = \log_{10} x$ sono continue?*

Estensione per continuità. Quando una funzione f non è definita in x_0 , ma esiste il limite $\lim_{x \rightarrow x_0} f(x) = \ell$, è naturale definire una nuova funzione come segue

$$F(x) := \begin{cases} f(x) & x \neq x_0 \\ \ell & x = x_0. \end{cases}$$

La funzione F si chiama **estensione per continuità** di f , dato che, per costruzione, F è continua in x_0 . La domanda “è possibile estendere per continuità in x_0 una assegnata funzione f ?” equivale a “esiste il limite di f per $x \rightarrow x_0$?”

ESEMPIO 2.11. La funzione $\frac{\sin x}{x}$ non è definita in $x = 0$, ma ammette limite per $x \rightarrow 0$. Quindi può essere *estesa per continuità* in $x = 0$ attribuendole il valore 1. La nuova funzione (continua in \mathbb{R}) è

$$f(x) = \begin{cases} \frac{\sin x}{x} & x \neq 0, \\ 1 & x = 0. \end{cases}$$

ESERCIZIO 2.12. (a) Dire quale delle seguenti funzioni può essere estesa per continuità in $x = 0$

$$\sin\left(\frac{1}{x}\right), \quad x \sin\left(\frac{1}{x}\right), \quad (x+1) \sin\left(\frac{1}{x}\right), \quad x^2 \sin\left(\frac{1}{x}\right).$$

(b) Sia f una funzione continua in $x = 0$ e tale che $f(0) = 0$. E' vero che la funzione $f(x) \sin(1/x)$ può essere estesa per continuità in $x = 0$?

Funzioni lipschitziane. Una funzione $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$ è lipschitziana se esiste una costante $L > 0$ tale che

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2| \quad \forall x_1, x_2 \in I.$$

La lipschitzianità corrisponde al fatto che il *rapporto incrementale*, cioè il coefficiente della retta secante passante per i punti del grafico di f di coordinate $(x_1, f(x_1))$ e $(x_2, f(x_2))$

$$\frac{f(x_1) - f(x_2)}{x_1 - x_2},$$

è limitato in valore assoluto da un fissato valore finito L .

Esempi di funzioni lipschitziane sono le funzioni affini $f(x) = ax + b$. Un altro esempio è $f(x) = \sin x$, infatti, come già osservato in precedenza,

$$|\sin x - \sin x_0| \leq |x - x_0|.$$

Tutte le funzioni lipschitziane sono continue: dato $\varepsilon > 0$, per avere $|f(x) - f(x_0)| < \varepsilon$ basta scegliere $\delta = \varepsilon/L$,

$$|f(x) - f(x_0)| \leq L|x - x_0| < L\delta = \varepsilon.$$

ESERCIZIO 2.13. Dimostrare le seguenti affermazioni.

- (i) Se f, g sono funzioni lipschitziane, allora anche $f + g$ è lipschitziana.
- (ii) Se f, g sono lipschitziane e limitate, allora fg è lipschitziana.
- (iii) Se in (ii) si rimuove l'ipotesi di limitatezza, la conclusione non è vera.

Soluzione. (i) Indicate con L_f, L_g , due costanti per cui è soddisfatta la condizione di Lipschitz per f e g rispettivamente, allora

$$|(f(x) + g(x)) - (f(y) + g(y))| \leq |f(x) - f(y)| + |g(x) - g(y)| \leq (L_f + L_g)|x - y|.$$

(ii) Indichiamo con L_f, L_g , due costanti per cui è soddisfatta la condizione di Lipschitz per f e g rispettivamente, e sia $|f(x)| \leq M_f$ e $|g(x)| \leq M_g$, allora

$$\begin{aligned} |f(x)g(x) - f(y)g(y)| &= |f(x)g(x) - f(y)g(x) + f(y)g(x) - f(y)g(y)| \\ &\leq |f(x) - f(y)||g(x)| + |f(y)||g(x) - g(y)| \leq (L_f M_g + M_f L_g)|x - y|. \end{aligned}$$

(iii) Ad esempio, si può scegliere $f(x) = g(x) = x$: il prodotto è la funzione x^2 che non è lipschitziana dato che

$$\sup_{x \neq y} \frac{|x^2 - y^2|}{|x - y|} = \sup_{x \neq y} |x + y| = +\infty.$$

Chiaro, no?

ESERCIZIO 2.14. Una funzione f è *hölderiana* se esistono $L, \alpha > 0$ tali che

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2|^\alpha \quad \forall x_1, x_2.$$

Dimostrare che se una funzione è hölderiana allora è anche continua.

Soluzione. Infatti dato $\varepsilon > 0$, per avere $|f(x) - f(x_0)| < \varepsilon$ basta scegliere $\delta = L^{-1/\alpha} \varepsilon^{1/\alpha}$:

$$|f(x) - f(x_0)| \leq L|x - x_0|^\alpha < L\delta^\alpha = \varepsilon,$$

per giungere alla conclusione sani e salvi.

3. Esempi di discontinuità

Un modo per chiarire ulteriormente la definizione di continuità è “in negativo”, cioè dando esempi per cui non è soddisfatta.

ESEMPIO 3.1. Riprendiamo l'esempio $f(x) = \operatorname{sgn} x$. Chiaramente, in ogni punto $x_0 \neq 0$, questa funzione è continua (qual è la scelta di δ in funzione di ε dato?). In $x_0 = 0$ la funzione, invece, non è continua. Infatti non è possibile determinare nessun δ quando ε sia minore di 1, dato che $|f(x) - f(0)| = |f(x)| = 1$ per ogni $x \neq 0$.

La funzione $\operatorname{sgn} x$ è l'esempio più semplice di discontinuità in un punto x_0 detto **discontinuità di salto**: la funzione f si avvicina, per x che tende a x_0 da destra e da sinistra, a valori limite che non coincidono con il valore di f in x_0 .

ESEMPIO 3.2. Un esempio di discontinuità in cui non ci siano limiti né da destra né da sinistra è dato dalla funzione

$$f(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & x \neq 0, \\ 0 & x = 0. \end{cases}$$

Il grafico della funzione f può essere dedotto da quello della funzione $\sin x$ attraverso un “passaggio al reciproco” nella variabile indipendente. Grossolanamente parlando, tutte le oscillazioni (infinite!) della funzione $\sin x$ per $x > 1$ vengono compresse nell'intervallo limitato $(0, 1)$ e si accumulano sul segmento del piano (x, y) di estremi $(0, -1)$ e $(0, 1)$ e non c'è alcuna speranza che la funzione possa essere continua in $x = 0$. Una figura chiarisce più di mille parole (Fig.4(a)).

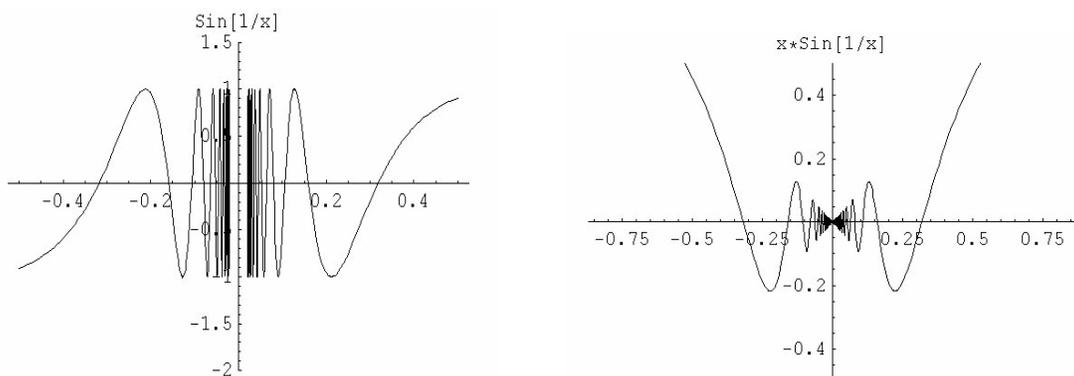


FIGURA 4. (a) Il grafico di $\sin(1/x)$; (b) Il grafico di $g(x)$.

Piccole varianti della funzione precedente possono condurre ad una funzione continua. Ad esempio consideriamo la funzione g seguente

$$g(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & x \neq 0, \\ 0 & x = 0. \end{cases}$$

Questa funzione (vedi Fig.4(b)) è continua in 0, infatti

$$|g(x) - 0| = \left| x \sin\left(\frac{1}{x}\right) \right| \leq |x| \rightarrow 0 \quad \text{per } x \rightarrow 0.$$

Sapete dire se è continua in 0 la funzione $(x^2 + 1)f(x)$, dove f è data nell'Esempio 3?

ESERCIZIO 3.3. Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione nondecreciente e discontinua in $x_0 \in (a, b)$. Che tipo di discontinuità ha la funzione f in x_0 ?

4. Teoremi sulle funzioni continue

Ora che abbiamo a disposizione un campionario vasto di funzioni continue e non, passiamo a stabilire alcune proprietà fondamentali che discendono dalla continuità: il *teorema dei valori intermedi* e il *teorema di Weierstrass* (che concerne il problema dell'esistenza di massimo e minimo). Entrambi discendono dal fatto che l'insieme dei numeri reali è *completo*, proprietà che traduce il fatto che la retta reale non ha buchi e che, rigorosamente, si basa sul postulato degli intervalli incapsulati e sull'assioma di Archimede. Nelle pagine che seguono ci dedichiamo prima a capire l'enunciato di questi due Teoremi fondamentali e solo successivamente ne vedremo le dimostrazioni.

Teorema del valore intermedio. Intuitivamente non c'è dubbio che se una funzione è continua, e quindi non ha salti, non può passare da un valore ad un altro senza

passare per tutti i valori intermedi. Pensiamo ad un esempio banale: se il signor Lafcadio fa una passeggiata in montagna e ci comunica che è partito da un rifugio che si trova a 2200 metri s.l.m. ed è arrivato in cima ad una montagna alta 3000 metri s.l.m., è vero che ad un certo punto si è trovato ad un'altitudine di 2800 metri? E più in generale, si è mai trovato ad una qualsiasi quota η compresa tra 2200 e 3000? La risposta (intuitiva) è "SÌ", a meno che non abbia utilizzato il teletrasporto...

TEOREMA 4.1. *Teorema del valore intermedio.* Sia $f : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}$ continua. Allora, per ogni $\eta \in [f(a), f(b)]$, esiste $x_0 \in [a, b]$ tale che $f(x_0) = \eta$.

Questo teorema dà condizioni sufficienti perchè l'equazione $f(x) = \eta$ abbia soluzione. Geometricamente, afferma che se i due punti $(a, f(a))$ e $(b, f(b))$ del grafico della funzione (continua) f giacciono su parti opposte rispetto alla retta $y = \eta$, allora il grafico di f interseca la retta in un punto intermedio.

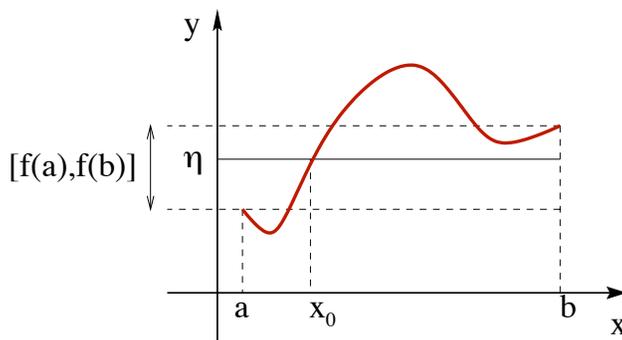


FIGURA 5. Il Teorema del valore intermedio

Controesempio 1. f non è continua. Nel caso di una funzione non continua la conclusione, in generale, è falsa. Ad esempio per la funzione

$$\text{segno di } x : \quad \text{sgn } x := \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ +1 & x > 0, \end{cases}$$

non esistono soluzioni di $\text{sgn } x = \eta$ per ogni $\eta \notin \{0, \pm 1\}$.

Controesempio 2. f definita in unione di intervalli disgiunti. Consideriamo la funzione $f : I = [-1, 0) \cup (0, 1] \rightarrow \mathbb{R}$ definita da $f(x) = \frac{1}{x}$. Allora, nonostante $0 \in [-1, 1] = [f(-1), f(1)]$, l'equazione $\frac{1}{x} = 0$ non ammette soluzioni! Analogamente per $g : I = [0, 1] \cup [2, 3] \rightarrow \mathbb{R}$ definita da $g(x) = x$, ci sono dei valori $\eta \in [g(0), g(3)] = [0, 3]$ tali che l'equazione $x = \eta$ non ammette soluzioni in I .

Qui si è persa una proprietà fondamentale degli intervalli: la *connessione*, cioè la garanzia che se x_1, x_2 appartengono all'intervallo I , allora $[x_1, x_2] \subset I$. In qualche

modo si può immaginare che una funzione continua non generi “strappi” o “buchi” nella trasformazione del dominio di partenza in quello di arrivo. E’ chiaro però che se il dominio di partenza è “già strappato”, cioè sconnesso (come nel caso di due intervalli chiusi disgiunti), è possibile che ci siano buchi anche nel dominio di arrivo.

Controesempio 3. *L’importanza di essere reale (razionale non basta!).* Consideriamo la funzione $f : \mathbb{Q} \cap [0, 2] \rightarrow \mathbb{Q}$ definita da $f(x) = x^2$. Allora $f(0) = 0$, $f(2) = 4$ ed è sensato domandarsi se ci siano soluzioni $x \in \mathbb{Q} \cap [0, 2]$ al problema $x^2 = 2 \in (0, 4)$. Come abbiamo già visto non c’è nessun valore razionale il cui quadrato sia 2. Quindi *il Teorema del valore intermedio non vale nei razionali!*

ESERCIZIO 4.2. Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua tale che

$$\lim_{x \rightarrow -\infty} f(x) = -\infty \quad \text{e} \quad \lim_{x \rightarrow +\infty} f(x) = +\infty.$$

Dimostrare, utilizzando il Teorema del valore intermedio, che $f(\mathbb{R}) = \mathbb{R}$. Cosa si può concludere se, invece, si suppone $\lim_{x \rightarrow -\infty} f(x) = \ell_- \in \mathbb{R}$ e $\lim_{x \rightarrow +\infty} f(x) = \ell_+ \in \mathbb{R}$?

Conseguenza del teorema del valore intermedio è il cosiddetto Teorema di esistenza degli zeri.

COROLLARIO 4.3. Teorema di esistenza degli zeri. Sia $f : [a, b] \rightarrow \mathbb{R}$ continua. Allora, se $f(a)f(b) < 0$ (cioè se $f(a)$ e $f(b)$ hanno segno discorde), esiste $x_0 \in (a, b)$ tale che $f(x_0) = 0$.

ESERCIZIO 4.4. Utilizzare il Teorema di esistenza degli zeri per dimostrare che ogni polinomio $p = p(x)$ di grado dispari ha sempre almeno uno zero.

Soluzione. Se p è un polinomio di grado dispari

$$\lim_{x \rightarrow \pm\infty} p(x) = \pm\infty \quad \text{oppure} \quad \lim_{x \rightarrow \pm\infty} p(x) = \mp\infty,$$

e quindi esiste certamente $L > 0$ per cui $p(-L)p(L) < 0$. Di conseguenza, per il Teorema di esistenza degli zeri, esiste $x_0 \in (-L, L)$ che azzera il polinomio.

Una funzione strettamente monotona, cioè tale che

$$x < x' \iff f(x) < f(x') \quad \text{oppure} \quad x < x' \iff f(x) > f(x'),$$

essendo iniettiva, è invertibile. In generale non è vero il viceversa: esistono funzioni invertibili che non sono monotone (sapete trovarne un esempio?). Invece, nel caso di funzioni continue definite in un intervallo, la stretta monotonia è una condizione necessaria e sufficiente di invertibilità. La dimostrazione è conseguenza del Teorema del valore intermedio.

COROLLARIO 4.5. *Sia f una funzione continua in $[a, b]$. Allora f è strettamente monotona se e solo se f è invertibile.*

DIMOSTRAZIONE. Basta dimostrare che se $f : [a, b] \rightarrow \mathbb{R}$ è continua ed invertibile, allora è anche strettamente monotona. Supponiamo per assurdo che non lo sia, allora esisterebbero $x_1, x_2, x_3 \in [a, b]$ tali che $x_1 < x_2 < x_3$ per cui o $f(x_1) < f(x_2)$ e $f(x_3) < f(x_2)$ oppure $f(x_1) > f(x_2)$ e $f(x_3) > f(x_2)$. Supponiamo di essere nel primo caso (l'altro si tratta in modo analogo) e scegliamo η tale che $\max\{f(x_1), f(x_3)\} < \eta < f(x_2)$. Applicando il teorema del valore intermedio agli intervalli $[x_1, x_2]$ e $[x_2, x_3]$ si ottiene che esistono $\xi_1 \in (x_1, x_2)$ e $\xi_2 \in (x_2, x_3)$ per cui $f(\xi_1) = f(\xi_2) = \eta$ che contraddice l'ipotesi di invertibilità di f . \square

Se si sostituisce l'ipotesi di “ f strettamente monotona” con “ f non strettamente monotona” la conclusione non è più vera. L'esempio più banale che si può pensare è quello di una funzione costante.

Teorema di Weierstrass. Un'altra proprietà fondamentale di una funzione continua f definita in un intervallo $[a, b]$ è l'esistenza del (valore) massimo e del (valore) minimo.

TEOREMA 4.6. Teorema di Weierstrass. *Sia $f : [a, b] \rightarrow \mathbb{R}$ continua. Allora esistono $x_0, x_1 \in [a, b]$ tali che $f(x_0) \leq f(x) \leq f(x_1)$ per ogni $x \in [a, b]$.*

Perseveriamo con la buona abitudine di cercare controesempi che mostrino il ruolo delle ipotesi del Teorema.

Controesempio 1. *f non è continua.* Se non è richiesta la continuità della funzione, è facile costruire casi di non esistenza di massimo/minimo. Ad esempio consideriamo $f : [-1, 1] \rightarrow \mathbb{R}$ definita da

$$f(x) = \begin{cases} x^2 & x \neq 0, \\ 1 & x = 0. \end{cases}$$

Chiaramente $\inf_{[-1,1]} f(x) = 0$, ma $f(x) \neq 0$ per ogni x . Analogamente si possono costruire casi in cui non c'è valore massimo.

Controesempio 2. $I = (a, b)$ (intervallo aperto). Anche in questo caso si possono trovare molti esempi che mostrano che le conclusioni del Teorema non sono vere. Ad esempio, $f(x) = x^2$ in $(-1, 1)$ non ammette massimo (l'estremo superiore è 1), oppure $g(x) = \sin x$ in $(0, \pi/2)$ non ammette né massimo né minimo (l'estremo superiore è 1 e quello inferiore è 0). Si noti che in entrambi questi esempi, quello che si vorrebbe essere punto di massimo/minimo è uno degli estremi dell'intervallo, che però non appartiene ad I visto che l'intervallo è considerato aperto.

Controesempio 3. *I illimitato.* L'esempio più facile è $f(x) = x$ per $x \in \mathbb{R}$ che non ammette né massimo né minimo. Esistono anche funzioni *limitate* in domini illimitati che non ammettono né massimo né minimo, ad esempio, $f(x) = \arctan x$.

Se si combinano insieme il Teorema del valore intermedio ed il Teorema di Weierstrass si può dimostrare la seguente affermazione.

COROLLARIO 4.7. *Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua. Allora l'insieme immagine $f([a, b])$ è un intervallo chiuso e limitato.*

5. Gli intervalli incapsulati: “divide et impera”

Come abbiamo già detto nella presentazione naïf dei numeri reali, i fatti fondamentali che accettiamo come assiomi sono

Postulato degli intervalli incapsulati. *Per ogni successione di intervalli $I_1, I_2, \dots, I_n, \dots$ chiusi e limitati che siano incapsulati, cioè tali che $I_{n+1} \subset I_n$ per ogni $n \in \mathbb{N}$, esiste sempre almeno un punto $x_0 \in \mathbb{R}$ tale che $x_0 \in I_n$ per ogni n .*

Assioma di Archimede. *Per ogni numero reale a , esiste un numero naturale n più grande di a : in simboli,*

$$\forall a \in \mathbb{R}, \quad \exists n \in \mathbb{N}, \quad \text{tale che } a \leq n.$$

Si ricordi che una delle conseguenze dell'assioma di Archimede è

$$x \leq \varepsilon \quad \forall \varepsilon > 0 \quad \Rightarrow \quad x \leq 0.$$

Daremo ora le dimostrazioni del Teorema del Valore Intermedio e del Teorema di Weierstrass a partire dal seguente risultato.

COROLLARIO 5.1. *Sia $I_n = [a_n, b_n] \subset \mathbb{R}$ una successione di intervalli tali che*

(i) $I_{n+1} \subset I_n$ (cioè $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$) per ogni $n \in \mathbb{N}$;

(ii) $\lim_{n \rightarrow +\infty} b_n - a_n = 0$.

Allora, esiste un unico $x_0 \in \mathbb{R}$ tale che $\bigcap_{n=1}^{\infty} I_n = \{x_0\}$, cioè $a_n \leq x_0 \leq b_n$ per ogni $n \in \mathbb{N}$. Inoltre $\lim_{n \rightarrow +\infty} a_n = \lim_{n \rightarrow +\infty} b_n = x_0$.

Il Corollario indica che se la successione degli intervalli incapsulati ha la proprietà aggiuntiva che la lunghezza $|I_n| = b_n - a_n$ è infinitesima per $n \rightarrow +\infty$, l'intersezione degli I_n (non vuota per il Postulato degli Intervalli Incapsulati) è costituita da un solo elemento.

DIMOSTRAZIONE DEL COROLLARIO 5.1. La proprietà (i) implica $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$ per il Postulato degli Intervalli Incapsulati; resta da dimostrare che tale intersezione è composta da un solo elemento. Siano $x_0, x_1 \in \bigcap_{n=1}^{\infty} I_n$ con $x_0 \leq x_1$. Allora $a_n \leq x_0 \leq x_1 \leq b_n$ per ogni n . Da questa relazione segue che $0 \leq x_1 - x_0 \leq b_n - a_n$. Dato che la successione $b_n - a_n$ è infinitesima, per ogni $\varepsilon > 0$, si ha $0 \leq x_1 - x_0 \leq b_n - a_n < \varepsilon$ per ogni n sufficientemente grande. In definitiva, per ogni $\varepsilon > 0$, si ha $0 \leq x_1 - x_0 < \varepsilon$, da cui segue $x_0 = x_1$, che conclude la prima parte del Teorema.

Inoltre, dato che $a_n \leq x_0 \leq b_n$, si ha anche $0 \leq x_0 - a_n \leq b_n - a_n \rightarrow 0$ per $n \rightarrow \infty$, quindi $a_n \rightarrow x_0$ per $n \rightarrow \infty$. Per b_n , basta notare che $b_n = a_n + (b_n - a_n) \rightarrow x_0 + 0 = x_0$. \square

Divide et impera. Per dimostrare i Teoremi useremo sempre la strategia del *Divide et Impera*. Il punto chiave è definire una successione di intervalli incapsulati $[a_n, b_n]$ di misura $b_n - a_n$ infinitesima per $n \rightarrow +\infty$. Tipicamente, costruiremo la successione di intervalli I_n , scegliendo un primo intervallo opportuno $I_0 = [a, b]$, poi prendendo il punto intermedio dell'intervallo $\frac{a+b}{2}$ e scegliendo (secondo un criterio che dipende da caso a caso) come intervallo I_1 una delle due metà di I_0 . Iterando il procedimento otterremo una successione di $I_n = [a_n, b_n]$ tale che

$$a_n \leq a_{n+1} \leq b_{n+1} \leq b_n, \quad \text{per ogni } n \in \mathbb{N}, \quad b_n - a_n = \frac{b-a}{2^n},$$

cioè soddisfacente le ipotesi del Corollario 5.1.

DIMOSTRAZIONE DEL TEOREMA DEL VALORE INTERMEDIO (TEOREMA 4.1). Supponiamo $f(a) < f(b)$ e $\eta \in (f(a), f(b))$. Se $f(b) < f(a)$, si può ragionare in modo simile. Come I_0 scegliamo l'intervallo di partenza $[a, b]$ e consideriamo il punto intermedio $\frac{a+b}{2}$. Procediamo come segue:

- (i) se $f\left(\frac{a+b}{2}\right) = \eta$, siamo arrivati alla conclusione;
- (i) se $f\left(\frac{a+b}{2}\right) > \eta$, poniamo $a_1 = a$ e $b_1 = \frac{a+b}{2}$;
- (ii) se $f\left(\frac{a+b}{2}\right) < \eta$, poniamo $a_1 = \frac{a+b}{2}$ e $b_1 = b$.

In questo modo o siamo giunti alla conclusione, o abbiamo costruito un intervallo $I_1 = [a_1, b_1]$ tale che $f(a_1) < \eta < f(b_1)$ e $b_1 - a_1 = \frac{b-a}{2}$. Iteriamo il procedimento: scegliamo il punto $\frac{a_1+b_1}{2}$, calcoliamo $f\left(\frac{a_1+b_1}{2}\right)$ e procediamo come sopra.

Così facendo, o si è dimostrata la conclusione dopo un numero finito di passi, o si è costruita una successione di intervalli incapsulati $I_n = [a_n, b_n]$ tale che $b_n - a_n = \frac{b-a}{2^n}$. Applicando il Corollario 5.1, deduciamo che esiste $x_0 \in [a, b]$ tale che

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = x_0.$$

Per la scelta di a_n, b_n , si ha $f(a_n) < \eta < f(b_n)$ per ogni n . Dato che f è continua in x_0 e $a_n, b_n \rightarrow x_0$, per $n \rightarrow \infty$ si deduce $f(x_0) \leq \eta \leq f(x_0)$, cioè $f(x_0) = \eta$. \square

DIMOSTRAZIONE DEL TEOREMA DI WEIERSTRASS (TEOREMA 4.6). Sia Λ l'estremo superiore di $\{f(x) : x \in [a, b]\}$. Per ora non sappiamo se Λ sia finito o no.

Poniamo $I_0 = [a, b]$. Dividiamo I_0 in due parti uguali, tramite il punto medio $\frac{a+b}{2}$. In almeno uno dei due sottointervalli $[a, \frac{a+b}{2}]$ e $[\frac{a+b}{2}, b]$ l'estremo superiore della funzione f è ancora uguale a Λ . Battezziamo il sottointervallo con questa proprietà I_1 e i suoi estremi con a_1 e b_1 . Nel caso in cui entrambi gli intervalli vadano bene ne scegliamo uno a nostro piacere. Iteriamo il procedimento, dividendo il sottointervallo tramite il suo punto medio. In questo modo, otteniamo una successione di intervalli incapsulati $I_n = [a_n, b_n]$ tali che $b_n - a_n = \frac{b-a}{2^n}$. Applicando il Teorema 5.1, deduciamo che esiste x_1 tale che $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = x_1$. Vogliamo a questo punto mostrare che Λ è finito e, inoltre, $f(x_1) = \Lambda$.

La funzione f è continua in x_1 , quindi per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che $f(x_1) - \varepsilon < f(x) < f(x_1) + \varepsilon$ per ogni $x \in (x_1 - \delta, x_1 + \delta)$. Fissato ε , e di conseguenza δ , scelgo $n \in \mathbb{N}$ tale che $I_n \subset (x_1 - \delta, x_1 + \delta)$. Allora

$$(23) \quad f(x) < f(x_1) + \varepsilon \quad \forall x \in I_n,$$

che esprime il fatto che $f(x_1) + \varepsilon$ è un maggiorante di $\{f(x) : x \in I_n\}$. Dato che $\sup\{f(x) : x \in I_n\} = \Lambda$ per costruzione, ne segue che Λ è finito, che $f(x_1) \leq \Lambda$ e che $\Lambda \leq f(x_1) + \varepsilon$. Ma, in quest'ultima relazione, ε può essere scelto arbitrariamente, quindi

$$0 \leq \Lambda - f(x_1) \leq \varepsilon \quad \forall \varepsilon > 0.$$

Perciò $\Lambda = f(x_1)$ e la dimostrazione è completa. \square

Si noti che, sebbene la costruzione porti a determinare un singolo punto di massimo, ce ne potrebbero essere anche molti altri! In effetti, ad ogni passo, nella scelta del sottointervallo ci può essere libertà di scelta, nel caso in cui in entrambi i sottointervalli l'estremo superiore della funzione f sia ancora uguale a Λ .

Per concludere, utilizziamo la strategia del *divide et impera* per dimostrare l'esistenza di estremo superiore ed inferiore (risultato concettualmente del tutto indipendente dai concetti di continuità di funzioni reali di variabile reale).

Ricordiamo che il valore $\Lambda \in \mathbb{R}$ è l'estremo superiore di $E \subset \mathbb{R}$ se

- (i) Λ è un maggiorante di E , cioè per ogni $y \in E$, si ha $y \leq \Lambda$;
- (ii) Λ è il più piccolo dei maggioranti, cioè se L è un maggiorante di E , allora $\Lambda \leq L$.

Teorema di esistenza dell'estremo superiore. *Se $E \neq \emptyset$ è limitato superiormente, allora esiste $\Lambda = \sup E \in \mathbb{R}$.*

DIMOSTRAZIONE. Sia $a \in E$ (E non è vuoto) e sia b un maggiorante di E (E è limitato superiormente). Indichiamo con I_0 l'intervallo chiuso e limitato $[a, b]$ e prendiamo il punto intermedio $\frac{a+b}{2}$. Allora $I_1 := [a_1, b_1]$, dove

- (i) se $\frac{a+b}{2}$ è un maggiorante di E , poniamo $a_1 = a$ e $b_1 = \frac{a+b}{2}$,
- (ii) se $\frac{a+b}{2}$ non è un maggiorante di E , poniamo $a_1 = \frac{a+b}{2}$ e $b_1 = b$.

In entrambi i casi in $I_1 = [a_1, b_1]$ c'è almeno un elemento di E e b_1 è un maggiorante di E . Iterando il procedimento, otteniamo la solita successione di intervalli incapsulati con $b_n - a_n = \frac{b-a}{2^n}$. Quindi, sempre per il Corollario 5.1, esiste Λ tale che $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = \Lambda$. Dato che $b_n \geq y$ per ogni $y \in E$, la stessa proprietà vale al limite: $\Lambda \geq y$ per ogni $y \in E$. Inoltre, per costruzione, ci sono elementi di E arbitrariamente vicini a Λ , quindi anche la condizione (ii) è soddisfatta. \square

CAPITOLO 5

Derivate, derivate e derivate

Prima di definire rigorosamente il concetto di derivabilità, prendiamoci il tempo di discutere un paio di punti di vista che indicano quale sia il significato di questo nuovo oggetto matematico: la derivata.

La derivata come velocità. Consideriamo un punto che si muova lungo l'asse y con posizione $y = f(t)$ all'istante t . Se la funzione f è affine, ossia $f(t) = At + B$, si parla di moto uniforme. La velocità A è il rapporto tra la distanza percorsa nell'intervallo di tempo $[t_0, t_1]$ e la durata di questo intervallo:

$$A = \frac{\Delta f}{\Delta t} = \frac{f(t) - f(t_0)}{t - t_0}.$$

Il moto è *uniforme* perchè la velocità è costante e, di conseguenza, in intervalli di tempo uguali, vengono percorse distanze uguali.

Se il moto non è uniforme, la quantità $\frac{f(t) - f(t_0)}{t - t_0}$ esprime la **velocità media** del punto nell'intervallo di tempo $[t_0, t]$. Se la velocità media tende ad un limite finito per $t \rightarrow t_0$, il valore del limite è detto **velocità (istantanea)**:

$$v(t_0) = \lim_{t \rightarrow t_0} \frac{f(t) - f(t_0)}{t - t_0}.$$

Se il limite non esiste, la velocità istantanea non è definita.

Un esempio semplice è il moto di un corpo in caduta libera, cioè sottoposto alla sola forza di gravità. Sperimentalmente, la distanza percorsa al tempo t da un corpo, lasciato cadere da fermo al tempo $t = 0$, è proporzionale a t^2 ; si rappresenta quindi con una funzione della forma

$$y = f(t) = at^2 \quad (a > 0).$$

La velocità v all'istante t si ottiene quindi calcolando

$$v(t_0) = \lim_{t \rightarrow t_0} \frac{f(t) - f(t_0)}{t - t_0} = \lim_{t \rightarrow t_0} \frac{at^2 - at_0^2}{t - t_0} = \lim_{t \rightarrow t_0} a(t + t_0) = 2at_0.$$

Quindi la velocità di un corpo in caduta libera cresce in modo proporzionale al tempo.

Nello studio del moto di un punto è utile osservare anche la variazione di velocità. Il procedimento è simile al precedente. L'**accelerazione media** è il rapporto tra la variazione di velocità nell'intervallo di tempo $[t_0, t]$ e la durata dell'intervallo, cioè è data da

$(v(t) - v(t_0))/(t - t_0)$. L'accelerazione (istantanea) a è il limite dell'accelerazione media per $t \rightarrow t_0$

$$a(t_0) = \lim_{t \rightarrow t_0} \frac{v(t) - v(t_0)}{t - t_0}.$$

Nel caso di moto uniforme $f(t) = At + B$,

$$v(t) = A \quad \forall t \quad \Rightarrow \quad a(t_0) = \lim_{t \rightarrow t_0} \frac{A - A}{t - t_0} = 0,$$

cioè l'accelerazione è nulla; nel caso del corpo in caduta libera $f(t) = at^2$,

$$v(t) = 2at \quad \forall t \quad \Rightarrow \quad a(t_0) = \lim_{t \rightarrow t_0} \frac{2at - 2at_0}{t - t_0} = 2a,$$

cioè il moto è *uniformemente accelerato*.

La derivata come approssimazione lineare. In generale, supponiamo di esaminare l'evoluzione di una quantità (la posizione di un punto in movimento, la temperatura dell'acqua sul fuoco, o altro...) descritta all'istante t , dal numero reale $y = f(t)$. Fissiamo un istante iniziale t_0 e misuriamo il valore di $y = f(t_0)$. Per controllare quello che succederà da t_0 in poi, dobbiamo studiare la **variazione** di f , cioè la quantità

$$\Delta f(t; t_0) := f(t) - f(t_0).$$

Se la funzione f è costante, non c'è evoluzione: $\Delta f = 0$ per ogni scelta di t_0 e t . Se f è una funzione affine, cioè se $f(t) = At + B$ per qualche $A, B \in \mathbb{R}$, allora

$$\Delta f(t; t_0) = (At + B) - (At_0 + B) = A(t - t_0) = A\Delta t,$$

dove $\Delta t = t - t_0$ rappresenta l'intervallo di tempo trascorso dall'istante iniziale t_0 a quello finale t . Come si vede, se f è affine, la funzione Δf è lineare nell'incremento Δt della variabile indipendente t , cioè Δf è proporzionale a Δt . La costante di proporzionalità è A ed è data da $A = \Delta f / \Delta t$.

Proviamo con un polinomio di secondo grado in t : $f(t) = at^2 + bt + c$ con $a, b, c \in \mathbb{R}$. In questo caso, scrivendo $t = t_0 + \Delta t$

$$\begin{aligned} \Delta f(t_0 + \Delta t; t_0) &= [a(t_0 + \Delta t)^2 + b(t_0 + \Delta t) + c] - [at_0^2 + bt_0 + c] \\ &= (2at_0 + b)\Delta t + a(\Delta t)^2. \end{aligned}$$

Questa volta l'incremento Δf non è lineare in Δt , dato che compare il termine quadratico $A(\Delta t)^2$. Però Δf ha la gentilezza di decomporre in due parti:

$$\Delta f(t_0 + \Delta t; t_0) = (\text{termine lineare in } \Delta t) + (\text{resto}).$$

Quanto è grande il resto $a(\Delta t)^2$? Consideriamo un caso semplice: $a = 1, b = 0, c$ qualsiasi, $t_0 = 1$. In questo caso:

$$f(t) = t^2 + c, \quad \Delta f(1 + \Delta t; 1) = 2\Delta t + (\Delta t)^2,$$

la parte lineare è $2\Delta t$ ed il resto è $(\Delta t)^2$. Vediamo i valori di questi due termini per diverse scelte di Δt :

Δt	1	0,1	0,01	0,001	0,0001	0,00001
parte lineare: $2\Delta t$	2	0,2	0,02	0,002	0,0002	0,00002
resto: $(\Delta t)^2$	1	0,01	0,0001	0,000001	0,00000001	0,0000000001

Come si vede dalla tabella, sia il termine lineare che il resto diminuiscono per $\Delta t \rightarrow 0$ (sono infinitesimi). Ma c'è una differenza fondamentale: il resto $(\Delta t)^2$ diviene piccolo molto più rapidamente del termine lineare. Dunque è ragionevole approssimare, per $\Delta t \rightarrow 0$, l'incremento Δf tramite una funzione lineare in Δt :

$$\Delta f(t_0 + \Delta t; t_0) \approx (2at_0 + b) \Delta t \quad \text{per } \Delta t \rightarrow 0.$$

In generale, data f qualsiasi, se è possibile scrivere l'incremento Δf nella forma $\Delta f(t_0 + \Delta t; t_0) = A\Delta t + R$ con R che tende a zero più rapidamente del termine lineare $A\Delta t$ ha senso utilizzare l'approssimazione

$$\Delta f(t_0 + \Delta t; t_0) \approx A\Delta t \quad \text{per } \Delta t \rightarrow 0.$$

Tutte le volte che questa operazione è possibile, la funzione f si dice *derivabile* e il valore A è la *derivata prima* di f in t_0 . La derivata, dunque, dà un'informazione sulla variazione Δf della funzione f quando la variabile indipendente t subisca una variazione Δt piccola.

Restano un paio di perplessità: che vuol dire la frase “il resto tende a zero più rapidamente del termine lineare”? E, in concreto, data una funzione f come stabilire se esiste e come calcolare il valore A ? La risposta alla prima domanda permette magicamente di risolvere anche il secondo angoscioso quesito. Dire che il resto R tende a zero più rapidamente del termine lineare vuol dire richiedere che valga

$$\lim_{\Delta t \rightarrow 0} \frac{R}{\Delta t} = 0.$$

In questo limite ci si trova di fronte ad una forma indeterminata del tipo $\frac{0}{0}$, e la richiesta è che il resto R tende a zero tanto rapidamente da dominare l'effetto del termine infinitesimo a denominatore.

Dalla condizione sul resto si deduce un modo per calcolare A : se esiste A tale che $\Delta f = A\Delta t + R$ con R che soddisfa $\lim_{\Delta t \rightarrow 0} R/\Delta t = 0$, allora, dividendo per Δt ,

$$A = \frac{\Delta f}{\Delta t} - \frac{R}{\Delta t},$$

e passando al limite per $\Delta t \rightarrow 0$, si ha

$$A = \lim_{\Delta t \rightarrow 0} \frac{\Delta f}{\Delta t},$$

che dà una maniera rigorosa di definire la derivata di f e, allo stesso tempo, una maniera per calcolarne il valore.

1. Definizione di derivata

Riprendiamo il discorso da capo e mettiamo ordine nel *brainstorming* fatto fin qui.

DEFINIZIONE 1.1. Derivabilità. Una funzione $f : [a, b] \rightarrow \mathbb{R}$ è derivabile in $x_0 \in (a, b)$ se esiste finito il limite

$$(24) \quad \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Se esiste, il limite si indica con $f'(x_0)$ e si dice **derivata (prima)** della funzione f in x_0 . Se f è derivabile in tutti i punti di $[a, b]$, si dice che f è derivabile in $[a, b]$.

Per la derivata si usano anche altri simboli:

$$f' = \frac{df}{dx} = Df = \frac{dy}{dx} = \dot{y} = \dots,$$

e il limite (24) può essere scritto in maniere equivalenti

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x} = \dots$$

Dato che la derivata f' dipende dal punto di derivazione, f' è essa stessa una funzione, il cui insieme di definizione è contenuto nell'insieme di definizione della funzione f (non è detto che i due domini di definizione coincidano).

Significato geometrico. Data una funzione $y = f(x)$, consideriamo il problema di determinare la retta tangente al grafico della funzione nel punto $P_0 = (x_0, f(x_0))$. L'idea è la seguente: dato un secondo punto $P = (x, f(x))$ sul grafico di f , per P_0 e P passa un'unica retta, detta **retta secante**. Se, muovendo P verso P_0 , la retta secante

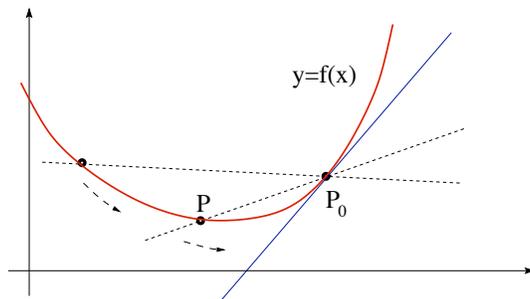


FIGURA 1. Il grafico di una funzione con tangente e secanti.

tende ad una posizione limite, tale retta limite è la **retta tangente**. Formuliamo ora, in

maniera rigorosa, il processo geometrico di limite che abbiamo appena raccontato. Il coefficiente angolare della retta secante per P_0 e P è

$$\text{rapporto incrementale: } \frac{f(x) - f(x_0)}{x - x_0}.$$

Se la funzione f è derivabile in x_0 , esiste il limite del rapporto incrementale e vale $f'(x_0)$, quindi il valore della derivata prima in x_0 rappresenta il coefficiente angolare della retta tangente al grafico della funzione nel punto $P_0 = (x_0, f(x_0))$.

OSSERVAZIONE 1.2. Determinare una derivata vuol dire fare (con successo) un limite: i limiti si fanno nei punti interni ad un intervallo di definizione. Negli estremi si fanno al più limiti sinistri o limiti destri. In punti isolati non si fanno neanche i limiti... Chi penserebbe di fare la tangente in un singolo punto?

La derivata è dunque il limite di una funzione opportuna, il rapporto incrementale. Vediamo come calcolare *esplicitamente* tale funzione derivata. Partiamo da alcuni casi semplici:

– se $f(x) = c \in \mathbb{R}$ per ogni x , si ha

$$\frac{f(x+h) - f(x)}{h} = \frac{c - c}{h} = 0 \quad \implies \quad f'(x) = \lim_{h \rightarrow 0} 0 = 0;$$

– se $f(x) = x$, vale

$$\frac{f(x+h) - f(x)}{h} = \frac{(x+h) - x}{h} = 1 \quad \implies \quad f'(x) = \lim_{h \rightarrow 0} 1 = 1;$$

– infine, se $f(x) = x^2$, si ha

$$\frac{f(x+h) - f(x)}{h} = \frac{2xh + h^2}{h} = 2x + h \quad \implies \quad f'(x) = \lim_{h \rightarrow 0} (2x + h) = 2x.$$

Passiamo ora ad un esempio meno facile: sia $f(x) = \sqrt{x}$ per $x \geq 0$. Il rapporto incrementale in $x \neq 0$ è

$$\frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{\sqrt{x+h} - \sqrt{x}}{h} \frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}} = \frac{1}{\sqrt{x+h} + \sqrt{x}}$$

Passando al limite per $h \rightarrow 0$ si ottiene

$$\lim_{h \rightarrow 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{1}{2\sqrt{x}} \quad x > 0.$$

Nel punto $x = 0$ la funzione ha una *singularità* (come è fatto il grafico di \sqrt{x} ?): pur essendo definita e continua, si ha

$$\lim_{h \rightarrow 0^+} \frac{\sqrt{x+h} - \sqrt{x}}{h} = +\infty.$$

Ecco altri due esempi di funzioni continue, ma non derivabili in $x = 0$:

$$f(x) = |x| \quad \text{e} \quad g(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & x \neq 0, \\ 0 & x = 0 \end{cases}$$

Per la funzione f , la non derivabilità in 0 è dovuta al fatto che i limiti destro e sinistro del rapporto incrementale esistono finiti ma non coincidono (il rapporto incrementale ha una discontinuità di salto in 0)

$$\lim_{x \rightarrow 0^+} \frac{|h|}{h} = 1 \neq -1 = \lim_{x \rightarrow 0^-} \frac{|h|}{h}.$$

Nel grafico, un comportamento di questo genere si traduce nella presenza di un *punto angoloso*. Nel caso della funzione g , il rapporto incrementale ha l'espressione

$$\frac{g(h) - g(0)}{h} = \frac{h \sin(1/h) - 0}{h} = \sin\left(\frac{1}{h}\right).$$

Come si è già visto, questa funzione non ha limite (né destro né sinistro) per $h \rightarrow 0$. In termini di grafico (controllare di persona!), questa funzione ha delle variazioni sempre più rapide di pendenza man mano che ci sia avvicina ad $x = 0$.

Due conseguenze. Vediamo cosa si può dedurre in un soffio dalla derivabilità.

1. Derivabilità \Rightarrow Continuità. Se una funzione f è derivabile in x_0 , allora è anche continua in x_0 . Infatti la continuità della funzione f nel punto x_0 è equivalente all'affermazione $\lim_{x \rightarrow x_0} (f(x) - f(x_0)) = 0$, e, dato che

$$f(x) - f(x_0) = \frac{f(x) - f(x_0)}{x - x_0} (x - x_0),$$

passando al limite per $x \rightarrow x_0$ si ottiene la conclusione.

2. Equazione della retta tangente. Data $f : [a, b] \rightarrow \mathbb{R}$, sia $x_0 \in [a, b]$ un punto in cui f è derivabile, la retta tangente è, per definizione, la retta passante per il punto $(x_0, f(x_0))$, il cui coefficiente angolare è pari a $f'(x_0)$

$$\text{equazione della retta tangente:} \quad y = f(x_0) + f'(x_0)(x - x_0).$$

Fissato il punto x_0 , il polinomio di primo grado in x a secondo membro può essere visto come un'approssimazione della funzione f vicino al punto x_0 .

Nel sostituire la funzione con la sua retta tangente l'errore R_{x_0} , è pari a

$$R(x; x_0) = f(x) - f(x_0) - f'(x_0)(x - x_0).$$

Per $x \rightarrow x_0$, l'errore che si commette tende a zero, cioè

$$(25) \quad \lim_{x \rightarrow x_0} R(x; x_0) = \lim_{x \rightarrow x_0} (f(x) - f(x_0) - f'(x_0)(x - x_0)) = 0.$$

Ma (attenzione!) lo stesso è vero per qualsiasi altra retta per il punto $(x_0, f(x_0))$, infatti

$$\lim_{x \rightarrow x_0} (f(x) - f(x_0) - m(x - x_0)) = 0 \quad \forall m \in \mathbb{R}.$$

Quindi la proprietà (25) non è indicativa! Il fatto *fondamentale* è che per $R(x; x_0)$ vale

$$(26) \quad \lim_{x \rightarrow x_0} \frac{R(x; x_0)}{x - x_0} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = 0.$$

Questa condizione è più restrittiva della precedente e, tra le funzioni affini, è verificata *solo* da quella che rappresenta la retta tangente ad f in x_0 . In maniera equivalente, avremmo potuto dire che una funzione è derivabile in x_0 se esiste un valore $\ell \in \mathbb{R}$ per cui

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - \ell(x - x_0)}{x - x_0} = 0.$$

Il valore ℓ è pari a $f'(x_0)$.

Prime formule di derivazione. Applichiamo ora la definizione per calcolare esplicitamente le derivate di alcune funzioni semplici.

Polinomi e potenze. Si è già visto che valgono le regole di derivazione

$$(c)' = 0, \quad (x)' = 1, \quad (x^2)' = 2x.$$

Per un generico polinomio di grado 2, $f(x) = ax^2 + bx + c$ si può procedere in modo analogo. Il rapporto incrementale è

$$\frac{f(x+h) - f(x)}{h} = \frac{a(x+h)^2 + b(x+h) + c - ax^2 - bx - c}{h} = 2ax + b + h.$$

Quindi, passando al limite per $h \rightarrow 0$, si ottiene

$$(ax^2 + bx + c)' = \lim_{h \rightarrow 0} (2ax + b + h) = 2ax + b.$$

In modo simile è possibile derivare un qualsiasi polinomio. Calcoliamo prima di tutto la derivata di $f(x) = x^n$ dove $n \in \mathbb{N}$. Il rapporto incrementale si può scrivere come

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{x_1^n - x^n}{x_1 - x} = x_1^{n-1} + x_1^{n-2}x + \dots + x^{n-1},$$

dato che $x_1^n - x^n = (x_1 - x)(x_1^{n-1} + x_1^{n-2}x + \dots + x^{n-1})$ per ogni $x_1, x \in \mathbb{R}$. Passando al limite per $x_1 \rightarrow x$, ciascuno dei termini tende a x^{n-1} e quindi, dato che si tratta di n termini, si ottiene

$$(27) \quad (x^n)' = nx^{n-1} \quad \forall n \in \mathbb{N},$$

(per $n = 1, 2$ si ottengono le relazioni già note per x e x^2).

Una volta noto che è possibile calcolare esplicitamente la derivata di un qualsiasi polinomio, è naturale chiedersi se sia possibile fare lo stesso per funzioni razionali. Partiamo dal caso più semplice:

$$f(x) = \frac{1}{x} \quad (x \neq 0) \quad \Rightarrow \quad \frac{f(x_1) - f(x)}{x_1 - x} = \frac{\frac{1}{x_1} - \frac{1}{x}}{x_1 - x} = \frac{x - x_1}{x_1 x (x_1 - x)} = -\frac{1}{x_1 x}.$$

Quindi passando al limite $x_1 \rightarrow x$, si ottiene la formula

$$\left(\frac{1}{x}\right)' = -\frac{1}{x^2} \quad (x \neq 0).$$

Allo stesso modo è possibile trattare funzioni $f(x) = \frac{1}{x^\beta}$ con $\beta \in \mathbb{N}$ ($x \neq 0$):

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{\frac{1}{x_1^\beta} - \frac{1}{x^\beta}}{x_1 - x} = \frac{x^\beta - x_1^\beta}{x_1^\beta x^\beta (x_1 - x)} = -\frac{x_1^{\beta-1} + x_1^{\beta-2}x + \dots + x^{\beta-1}}{x_1^\beta x^\beta}.$$

Passando al limite per $x_1 \rightarrow x$, si ottiene

$$(28) \quad (x^{-\beta})' \equiv \left(\frac{1}{x^\beta}\right)' = -\frac{\beta}{x^{\beta+1}} \equiv -\beta x^{-\beta-1} \quad \forall \beta \in \mathbb{N}, \quad \forall x \neq 0.$$

Vedremo più avanti come si possa calcolare la derivata di una generica funzione razionale.

Le formule (27) e (28) si possono sintetizzare nell'unica formula

$$(29) \quad (x^\alpha)' = \alpha x^{\alpha-1} \quad \forall \alpha \in \mathbb{Z}.$$

Dimostriamo che è possibile scegliere $\alpha \in \mathbb{Q}$ ottenendo ancora la formula (29). Supponiamo la funzione $f(x) = x^\alpha$ con $\alpha = p/q$ con p e q interi ($q \neq 0$). Consideriamo, per semplicità, il caso $p, q > 0$. Il rapporto incrementale è

$$\frac{x_1^\alpha - x^\alpha}{x_1 - x} = \frac{x_1^{p/q} - x^{p/q}}{x_1 - x}.$$

Ponendo $x_1^{1/q} = \xi_1$ e $x^{1/q} = \xi$, otteniamo

$$\frac{x_1^\alpha - x^\alpha}{x_1 - x} = \frac{\xi_1^p - \xi^p}{\xi_1^q - \xi^q} = \frac{\xi_1^{p-1} + \xi_1^{p-2}\xi + \dots + \xi^{p-1}}{\xi_1^{q-1} + \xi_1^{q-2}\xi + \dots + \xi^{q-1}}.$$

Passando al limite per $x_1 \rightarrow x$, cioè per $\xi_1 \rightarrow \xi$, si ottiene

$$\lim_{x_1 \rightarrow x} \frac{x_1^\alpha - x^\alpha}{x_1 - x} = \lim_{\xi_1 \rightarrow \xi} \frac{\xi_1^{p-1} + \xi_1^{p-2}\xi + \dots + \xi^{p-1}}{\xi_1^{q-1} + \xi_1^{q-2}\xi + \dots + \xi^{q-1}} = \frac{p \xi^{p-1}}{q \xi^{q-1}} = \frac{p}{q} \xi^{p-q} = \frac{p}{q} x^{\frac{p}{q}-1},$$

cioè la formula (29) per α razionale positivo.

In generale si può dimostrare che (29) vale *per ogni* $\alpha \in \mathbb{R}$, cioè

$$(30) \quad (x^\alpha)' = \alpha x^{\alpha-1} \quad \forall \alpha \in \mathbb{R} \quad \forall x \neq 0.$$

Funzioni trigonometriche. Grazie alle formule di addizione è possibile scrivere i rapporti incrementali di $\sin x$ e $\cos x$ come

$$\frac{\sin(x+h) - \sin x}{h} = \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} = \sin x \frac{\cos h - 1}{h} + \cos x \frac{\sin h}{h},$$

$$\frac{\cos(x+h) - \cos x}{h} = \frac{\cos x \cos h - \sin x \sin h - \cos x}{h} = \cos x \frac{\cos h - 1}{h} - \sin x \frac{\sin h}{h}.$$

Passando al limite per $h \rightarrow 0$ e ricordando che $\lim_{h \rightarrow 0} \frac{\cos h - 1}{h} = 0$ e $\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1$,

$$(\sin x)' = \cos x \quad \text{e} \quad (\cos x)' = -\sin x.$$

Esponenziale e logaritmo. Come ultimo esempio, consideriamo le funzioni e^x e $\ln x$. Nel caso dell'esponenziale, il rapporto incrementale è

$$\frac{f(x+h) - f(x)}{h} = \frac{e^{x+h} - e^x}{h} = e^x \frac{e^h - 1}{h}.$$

Passando al limite per $h \rightarrow 0$ e usando il limite notevole $\lim_{h \rightarrow 0} \frac{e^h - 1}{h} = 1$,

$$(e^x)' = e^x,$$

che esprime una proprietà notevole dell'esponenziale: *la derivata di e^x è e^x* . In effetti, la funzione e^x è l'unica funzione f che verifica l'equazione (differenziale) $f' = f$ e la condizione $f(0) = 1$.

Il rapporto incrementale del logaritmo naturale si riscrive come

$$\frac{f(x+h) - f(x)}{h} = \frac{\ln(x+h) - \ln x}{h} = \frac{1}{h} \ln \left(\frac{x+h}{x} \right) = \frac{1}{h} \ln \left(1 + \frac{h}{x} \right).$$

Quindi, ponendo $t = h/x$ (x è fissato) e usando il limite notevole $\lim_{t \rightarrow 0} \frac{\ln(1+t)}{t} = 1$,

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{1}{h} \ln \left(1 + \frac{h}{x} \right) = \lim_{t \rightarrow 0} \frac{1}{x} \frac{\ln(1+t)}{t} = \frac{1}{x}.$$

2. Regole fondamentali di derivazione

Dalla definizione dell'operazione di derivazione, discendono alcune regole basilari che permettono di derivare una classe ampia di funzioni, a partire da una classe più ristretta di derivate note.

Linearità. Dati $\alpha, \beta \in \mathbb{R}$ e f, g derivabili, allora anche $\alpha f + \beta g$ è derivabile e

$$\phi(x) = \alpha f(x) + \beta g(x) \quad \implies \quad \phi'(x) = \alpha f'(x) + \beta g'(x).$$

Basta infatti osservare che il rapporto incrementale di ϕ si può riscrivere come

$$\frac{\phi(x+h) - \phi(x)}{h} = \alpha \frac{f(x+h) - f(x)}{h} + \beta \frac{g(x+h) - g(x)}{h},$$

e passare al limite per $h \rightarrow 0$, applicando le proprietà note dei limiti.

Ad esempio, la derivata di un polinomio $p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0$ si può calcolare senza bisogno di passare per il limite del rapporto incrementale, ma semplicemente usando la linearità della derivazione e la formula $(x^k)' = kx^{k-1}$:

$$\begin{aligned}(p(x))' &= a_n (x^n)' + a_{n-1} (x^{n-1})' + \dots + a_1 (x)' + (a_0)' \\ &= n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \dots + a_1.\end{aligned}$$

Derivata di un prodotto. Date f, g derivabili, allora anche fg è derivabile e

$$\phi(x) = f(x)g(x) \implies \phi'(x) = f(x)g'(x) + f'(x)g(x).$$

Per dimostrare la formula, scriviamo il rapporto incrementale

$$\begin{aligned}\frac{\phi(x+h) - \phi(x)}{h} &= \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ &= f(x+h) \frac{g(x+h) - g(x)}{h} + \frac{f(x+h) - f(x)}{h} g(x),\end{aligned}$$

(si è aggiunto e sottratto a numeratore la quantità $f(x+h)g(x)$). Per $h \rightarrow 0$, la conclusione.

Ad esempio, per calcolare la derivata della funzione $\phi(x) = x \sin x$,

$$(x \sin x)' = x(\sin x)' + (x)' \sin x = x \cos x + \sin x,$$

avendo usato le formule di derivazione per x e $\sin x$.

Derivata di un rapporto. Se f e g sono derivabili ($g(x) \neq 0$ per ogni x), allora anche il rapporto f/g è derivabile e vale la formula

$$\phi(x) = \frac{f(x)}{g(x)} \implies \phi'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}.$$

La dimostrazione discende dalla struttura del rapporto incrementale per la funzione rapporto. Niente di sorprendente. Si ha:

$$\begin{aligned}\frac{\phi(x+h) - \phi(x)}{h} &= \frac{1}{h} \left[\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)} \right] = \frac{f(x+h)g(x) - f(x)g(x+h)}{hg(x)g(x+h)} \\ &= \frac{f(x+h)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(x+h)}{hg(x)g(x+h)} \\ &= \frac{1}{g(x)g(x+h)} \left[\frac{f(x+h) - f(x)}{h} g(x) - f(x) \frac{g(x+h) - g(x)}{h} \right].\end{aligned}$$

Per $h \rightarrow 0$, si ottiene la conclusione.

OSSERVAZIONE 2.1. La formula di derivazione del rapporto è stata scritta nel caso in cui $g(x) \neq 0$ per ogni x . Ripercorrendo la dimostrazione ci si convince che basta supporre $g(x) \neq 0$ nel punto considerato. Infatti, se g è derivabile in x è anche continua nel punto, e quindi, c'è tutto un intorno I di x in cui g non si azzera.

Ad esempio, la derivata di $f(x) = \tan x$ è data da

$$\begin{aligned} (\tan x)' &= \left(\frac{\sin x}{\cos x} \right)' = \frac{(\sin x)' \cos x - \sin x (\cos x)'}{\cos^2 x} \\ &= \frac{\cos x \cos x - \sin x (-\sin x)}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}. \end{aligned}$$

Anche per derivare funzioni razionali basta applicare la formula di derivazione del rapporto. Ad esempio,

$$\left(\frac{x^2}{x+1} \right)' = \frac{2x(x+1) - x^2 \cdot 1}{(x+1)^2} = \frac{x(x+2)}{(x+1)^2}.$$

Derivata di una funzione composta. Siano g, h derivabili, allora la funzione composta $f = h \circ g$ è derivabile e vale la formula (in inglese, nota come chain rule)

$$(31) \quad f(x) = h(g(x)) \implies f'(x) = h'(g(x)) g'(x).$$

Usare concretamente questa regola è molto più semplice di quel che possa sembrare. Vediamo, ad esempio, come calcolare la derivata di $f(x) = e^{x^2}$.

i. Riconosciamo la struttura di funzione composta:

$$f(x) = h(g(x)) \quad \text{dove} \quad g(x) = x^2, \quad h(s) = e^s.$$

ii. Dato che $g(x) = x^2$ e $h(s) = e^s$, si ha $g'(x) = 2x$ e $h'(s) = e^s$.

iii. Ora occorre fare il prodotto delle derivate, *calcolando h' in $s = g(x) = x^2$* :

$$D(e^{x^2}) = 2xe^{x^2}.$$

Analogamente, dato che $D(\sin x) = \cos x$ e $D(\sqrt{s}) = 1/(2\sqrt{s})$,

$$D\left(\sqrt{1+\sin x}\right) = \frac{\cos x}{2\sqrt{1+\sin x}}.$$

Se la funzione è composta da più di due funzioni, si itera il procedimento:

$$D(h(g(f(x)))) = h'(g(f(x))) \cdot g'(f(x)) \cdot f'(x).$$

Ad esempio,

$$D\left(\sqrt{1+\sin^2 x}\right) = \frac{1}{2\sqrt{1+\sin^2 x}} \cdot 2 \sin x \cdot \cos x = \frac{\sin x \cos x}{\sqrt{1+\sin^2 x}}.$$

Per dimostrare la formula (31), scriviamo il rapporto incrementale

$$(32) \quad \frac{\Delta f}{\Delta x} = \frac{\Delta h}{\Delta x} = \begin{cases} 0 & \text{se } \Delta g = 0, \\ \frac{\Delta h}{\Delta g} \frac{\Delta g}{\Delta x} & \text{se } \Delta g \neq 0, \end{cases}$$

dove

$$\begin{aligned} \Delta x &= x_2 - x_1 & \Delta f &= f(x_2) - f(x_1) \\ \Delta h &= h(g(x_2)) - h(g(x_1)) & \Delta g &= g(x_2) - g(x_1). \end{aligned}$$

Se, per x_2 vicino ad x_1 , si ha $\Delta g \neq 0$, la conclusione segue da

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Delta h}{\Delta g} \frac{\Delta g}{\Delta x} = \lim_{\Delta g \rightarrow 0} \frac{\Delta h}{\Delta g} \lim_{\Delta x \rightarrow 0} \frac{\Delta g}{\Delta x} = h'(g(x_1)) g'(x_1),$$

dato che $\Delta g \rightarrow 0$ quando $\Delta x \rightarrow 0$. Se in ogni intorno di x_1 ci sono punti per cui $\Delta g = 0$, la derivata di g in x_1 deve essere nulla (come si dimostra?), e quindi vale la conclusione, dato che entrambe le rappresentazioni di $\Delta f/\Delta x$ in (32) tendono a zero per $\Delta x \rightarrow 0$.

Applicando la formula (31) è possibile ottenere le formule per le derivate di

$$x^\alpha \quad (\alpha \in \mathbb{R}) \quad \text{e} \quad a^x \quad (a > 0).$$

Per entrambe è utile osservare utilizzare la formula

$$(33) \quad a^b = e^{b \ln a} \quad \forall a > 0, b \in \mathbb{R}.$$

Usando la formula (33),

$$D(x^\alpha) = D(e^{\alpha \ln x}) = \frac{\alpha}{x} e^{\alpha \ln x} = \frac{\alpha x^\alpha}{x} = \alpha x^{\alpha-1} \quad \forall \alpha \in \mathbb{R},$$

$$D(a^x) = D(e^{x \ln a}) = e^{x \ln a} \ln a = a^x \ln a \quad \forall a > 0.$$

Derivata di una funzione inversa. Una conseguenza della formula di derivazione di funzione composta è la formula della derivata dell'inversa di una funzione. La prima domanda naturale da porsi è: *se la funzione f è invertibile e derivabile, lo è anche la funzione inversa?* La risposta è immediata se si pensa a come si ottiene il grafico della funzione inversa a partire da quello della funzione originale e se si ricorda il significato geometrico della derivabilità: la funzione f è derivabile in x se in tale punto il grafico ammette tangente e tale retta tangente *non è verticale* (quando la tangente al grafico è verticale, il rapporto incrementale tende ad ∞). Il grafico di f^{-1} si può ottenere da quello della f tramite un ribaltamento attorno alla bisettrice del primo e terzo quadrante. In questa operazione di ribaltamento, rette orizzontali diventano verticali e viceversa. Quindi un punto in cui la tangente al grafico di f è orizzontale (cioè $f'(x) = 0$), corrisponde, nel grafico di f^{-1} , ad un punto in cui la tangente è verticale e viceversa (Fig.2). Questo significa che:

Se $f'(x) \neq 0$, la funzione inversa f^{-1} è derivabile nel punto $y = f(x)$.

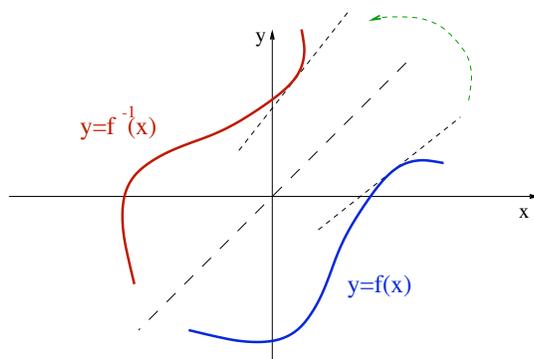


FIGURA 2. Una funzione e la sua inversa, con le relative tangenti.

Come calcolare la derivata della inversa f^{-1} ? Dato che $f(f^{-1}(x)) = x$ per ogni x , derivando membro a membro tramite la formula di derivazione di funzione composta,

$$f(f^{-1}(x)) = x \quad \implies \quad f'(f^{-1}(x))(f^{-1})'(x) = 1.$$

Esplicitando $(f^{-1})'(x)$, si ottiene la formula di derivazione della funzione inversa

$$(34) \quad (f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}.$$

Verifichiamo questa formula, calcolando di nuovo la derivata della funzione $f(x) = \ln x$ (in precedenza la formula si è ottenuta in modo diverso). In questo caso

$$\begin{cases} f(t) = e^t \\ f^{-1}(x) = \ln x \end{cases} \implies \begin{cases} f'(t) = e^t \\ f'(f^{-1}(x)) = e^{\ln x} = x \end{cases} \implies (\ln x)' = \frac{1}{x}.$$

Consideriamo le inverse delle funzioni trigonometriche e calcoliamone le derivate. Dato che $\cos t = \sqrt{1 - \sin^2 t}$ per $t \in [-\pi/2, \pi/2]$ e $\sin t = \sqrt{1 - \cos^2 t}$ per $t \in [0, \pi]$, si ha

$$\begin{aligned} \begin{cases} f(t) = \sin t \\ f^{-1}(x) = \arcsin x \end{cases} &\implies \begin{cases} f'(t) = \cos t \\ f'(f^{-1}(x)) = \cos(\arcsin x) = \sqrt{1 - x^2} \end{cases} \\ &\implies (\arcsin x)' = \frac{1}{\sqrt{1 - x^2}} \quad \forall x \in (-1, 1), \end{aligned}$$

$$\begin{aligned} \begin{cases} f(t) = \cos t \\ f^{-1}(x) = \arccos x \end{cases} &\implies \begin{cases} f'(t) = -\sin t \\ f'(f^{-1}(x)) = -\sin(\arccos x) = -\sqrt{1 - x^2} \end{cases} \\ &\implies (\arccos x)' = -\frac{1}{\sqrt{1 - x^2}} \quad \forall x \in (-1, 1). \end{aligned}$$

Per quanto riguarda la funzione $\arctan x$, è utile ricordare che

$$D(\tan t) = \frac{\cos^2 t + \sin^2 t}{\cos^2 t} = \frac{1}{\cos^2 t} = 1 + \tan^2 t.$$

Quindi

$$\begin{aligned} \begin{cases} f(t) = \tan t \\ f^{-1}(x) = \arctan x \end{cases} &\Rightarrow \begin{cases} f'(t) = 1 + \tan^2 t \\ f'(f^{-1}(x)) = 1 + \tan^2(\arctan x) = 1 + x^2 \end{cases} \\ &\Rightarrow (\arctan x)' = \frac{1}{1 + x^2}. \end{aligned}$$

Ultima, ma non ultima, la formula della derivata di $f^{-1}(x) = \log_a x$ con $a > 0$ qualsiasi:

$$\begin{cases} f(x) = a^x \\ f^{-1}(x) = \log_a x \end{cases} \Rightarrow \begin{cases} f'(t) = a^t \ln a \\ f'(f^{-1}(x)) = a^{\log_a x} \ln a = x \ln a \end{cases} \Rightarrow (\log_a x)' = \frac{1}{x \ln a}.$$

3. Derivate successive

L'operazione di derivazione porta da una funzione f ad una nuova funzione f' . E' naturale chiedersi se la funzione derivata f' sia a sua volta derivabile.

DEFINIZIONE 3.1. Sia $f : [a, b] \rightarrow \mathbb{R}$ derivabile e sia $x \in [a, b]$. Se esiste finito

$$(35) \quad \lim_{h \rightarrow 0} \frac{f'(x+h) - f'(x)}{h},$$

la funzione f è derivabile due volte in x , il limite si indica con $f''(x)$ e si chiama **derivata seconda** di f in x . Come sempre, se f è derivabile due volte in tutti i punti dell'intervallo $[a, b]$, si dice che f è derivabile due volte in $[a, b]$.

Per la derivata seconda si usano anche le notazioni

$$f'' = \frac{d^2 f}{dx^2} = D^2 f = \frac{d^2 y}{dx^2} = \dots,$$

Analogamente, nel caso di una funzione derivabile due volte, è possibile domandarsi se esista la derivata terza f''' . Iterando il procedimento si può parlare di derivata n -esima, che si indica con $f^{(n)}$. Simboli equivalenti sono

$$f^{(n)} \equiv D^n f \equiv \frac{d^n f}{dx^n} = \frac{d^n y}{dx^n}.$$

Qualche volta si indica la funzione f come la sua derivata 0-esima: $f^{(0)} \equiv f$.

La maniera *operativa* di calcolare derivate successive è semplicemente di iterare le formule note per la derivazione. Ad esempio,

$$f(x) = x^3 + x \Rightarrow f'(x) = 3x^2 + 1 \Rightarrow f''(x) = 6x \Rightarrow f'''(x) = 6.$$

Le derivate di ordine superiore al terzo della funzione $f(x) = x^3 + x$ esistono e sono tutte nulle. In generale, un polinomio p di grado n è infinitamente derivabile (cioè

ammette derivate di qualsiasi ordine), e le sue derivate di ordine maggiore o uguale ad $n + 1$ sono tutte nulle. Anche le funzioni $\sin x$ e $\cos x$ sono infinitamente derivabili:

$$\begin{aligned} D(\sin x) &= \cos x, & D^2(\sin x) &= -\sin x, & D^3(\sin x) &= -\cos x, & D^4(\sin x) &= \sin x \\ D(\cos x) &= -\sin x, & D^2(\cos x) &= -\cos x, & D^3(\cos x) &= \sin x, & D^4(\cos x) &= \cos x. \end{aligned}$$

Le derivate successive ripetono lo stesso schema in modo periodico, ossia

$$\begin{aligned} D^{2n-1}(\sin x) &= (-1)^{n+1} \cos x, & D^{2n}(\sin x) &= (-1)^n \sin x, \\ D^{2n-1}(\cos x) &= (-1)^n \sin x, & D^{2n}(\cos x) &= (-1)^n \cos x, \end{aligned} \quad \forall n \in \mathbb{N}.$$

Pensando al caso di polinomi e funzioni trigonometriche, si potrebbe essere indotti a credere che tutte le funzioni siano infinitamente derivabili. Un esempio di funzione che sia derivabile due volte in un punto, ma non tre volte è $f(x) = x^{5/2}$. Infatti $f''(x) = \frac{15}{4}\sqrt{x}$ che, come sappiamo, non è derivabile in zero.

ESERCIZIO 3.2. Se $f(x) = \cos x$, quanto vale l'espressione $f''(x) + f(x)$? E se, dato $\lambda \in \mathbb{R}$, $g(x) = e^{\lambda x}$, quanto vale $g''(x) - \lambda^2 g(x)$?

Notazioni. Comunemente si usano le notazioni (qui I è un intervallo aperto e $k \in \mathbb{N}$)

$$\begin{aligned} C(I) &\equiv C^0(I) := \{\text{funzioni continue in } I\} \\ C^1(I) &:= \{\text{funzioni derivabili in } I \text{ e con } f' \in C(I)\} \\ C^k(I) &:= \{\text{funzioni derivabili } k \text{ volte in } I \text{ e con } f^{(k)} \in C(I)\} \\ C^\infty(I) &:= \{\text{funzioni infinitamente derivabili in } I\}. \end{aligned}$$

4. Il Teorema di Lagrange

Dato che il rapporto incrementale è determinato dai valori della funzione in due punti distinti, esso riflette proprietà della funzione “in grande”. Invece, la derivata, che si ottiene con un procedimento di limite, riflette solo proprietà “in piccolo”. E' molto utile poter dedurre proprietà *globali* della funzione (cioè “in grande”) a partire da proprietà *locali* (cioè “in piccolo”) date dalla derivata prima della funzione. Lo strumento più utile per questa operazione è il *teorema di Lagrange* (o *teorema del valor medio del calcolo differenziale*).¹

Graficamente il Teorema di Lagrange afferma che data una funzione f continua nell'intervallo chiuso $[x_1, x_2]$ e derivabile nell'intervallo aperto (x_1, x_2) , la retta passante per i punti $(x_1, f(x_1))$ e $(x_2, f(x_2))$ (detta retta secante) è parallela alla retta tangente

¹Nei testi americani, spesso il Teorema di Lagrange è denominato “mean value theorem of differential calculus” o “intermediate value theorem”.

al grafico nel punto $(\xi, f(\xi))$ per almeno un valore $\xi \in (x_1, x_2)$. Dato che il coefficiente angolare della secante è

$$\frac{\Delta f}{\Delta x} = \frac{f(x_2) - f(x_1)}{x_2 - x_1},$$

per questo valore intermedio ξ vale la relazione $f'(\xi) = [f(x_2) - f(x_1)]/[x_2 - x_1]$.

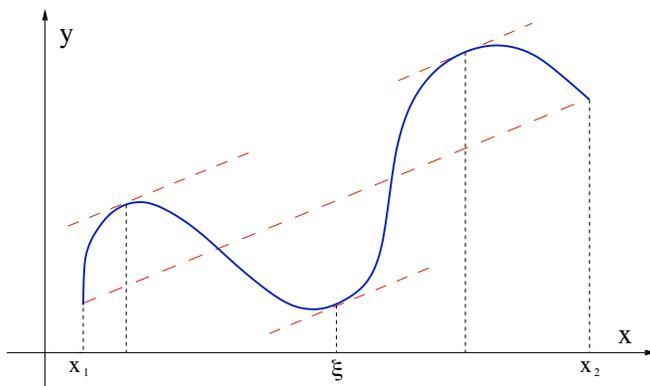


FIGURA 3. Il teorema di Lagrange

TEOREMA 4.1. Teorema di Lagrange. *Sia f continua in $[x_1, x_2]$ e derivabile in (x_1, x_2) . Allora esiste $\xi \in (x_1, x_2)$ tale che*

$$f'(\xi) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

La tesi del Teorema equivale ad affermare che esiste $\theta \in (0, 1)$ per cui

$$f'(x_1 + \theta(x_2 - x_1)) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

Le due formulazioni sono equivalenti dato che il punto intermedio ξ può sempre essere scritto nella forma $\xi = x_1 + \theta(x_2 - x_1)$ per $\theta \in (0, 1)$ opportuno. Oppure, sostituendo x_1 con x e x_2 con $x + h$, possiamo scrivere

$$\frac{f(x+h) - f(x)}{h} = f'(x + \theta h), \quad \theta \in (0, 1).$$

OSSERVAZIONE 4.2. Il punto di partenza nella definizione di derivabilità è dare solidità ad approssimazioni del tipo

$$\Delta f \approx f'(x_0)\Delta x \quad \text{per } x \approx x_0.$$

dove $\Delta f = f(x) - f(x_0)$ e $\Delta x = x - x_0$. Il Teorema di Lagrange garantisce che $\Delta f = f'(\xi)\Delta x$ per qualche ξ compreso tra x e x_0 . Quindi se si è disposti a pagare il prezzo di calcolare la derivata di f in un misterioso punto ξ , anziché in x_0 , l'errore commesso è nullo (ma non si dimentichi che ξ dipende da x e da x_0).

CONTROESEMPIO 4.3. “Datemi un punto (interno) di non derivabilità, e vi darò un controesempio.” Se la funzione f non è derivabile in tutti i punti dell’intervallo aperto (x_1, x_2) , non è detto che valga la conclusione del Teorema di Lagrange: può capitare che nessuna parallela della secante che congiunge gli estremi del grafico sia tangente al grafico stesso. Consideriamo la funzione $f(x) = |x|$ nell’intervallo $[-1, 1]$. Questa funzione è derivabile per ogni $x \neq 0$ e si ha

$$D(|x|) = \begin{cases} -1 & -1 \leq x < 0, \\ +1 & 0 < x \leq 1, \end{cases}$$

ma, dato che

$$\frac{f(1) - f(-1)}{1 - (-1)} = \frac{1 - 1}{2} = 0 \neq D(|x|) \quad \forall x,$$

la conclusione del Teorema non vale.

Il Teorema di Lagrange è conseguenza del seguente risultato.

TEOREMA 4.4. *Teorema di Rolle.* Sia ϕ continua in $[x_1, x_2]$ e derivabile in (x_1, x_2) . Se $\phi(x_1) = \phi(x_2)$, allora esiste $\xi \in (x_1, x_2)$ tale che $\phi'(\xi) = 0$.

Geometricamente, il Teorema di Rolle afferma che, se $\phi(x_1)$ e $\phi(x_2)$ coincidono allora il grafico di ϕ ha tangente orizzontale in un punto interno dell’intervallo (x_1, x_2) .

DIMOSTRAZIONE DEL TEOREMA 4.4. Sia $\ell = \phi(x_1) = \phi(x_2)$. Dato che la funzione ϕ è continua in $[x_1, x_2]$, per il Teorema di Weierstrass, esistono sia il massimo M che il minimo m di ϕ in $[x_1, x_2]$. Chiaramente, $m \leq \ell \leq M$.

Se $M = m$, deve essere $\phi(x) = M$ in tutto l’intervallo $[x_1, x_2]$, quindi $\phi'(x) = 0$ in tutti i punti dell’intervallo.

Se $M \neq m$, almeno uno dei due valori deve essere diverso da ℓ . Supponiamo che sia $M \neq \ell$ (l’altro caso si tratta in modo simile). Allora $M > \ell$ ed esiste $\xi \in [x_1, x_2]$ tale che $\phi(\xi) = M$. Inoltre visto che $\phi(x_1) = \phi(x_2) = \ell \neq M$, $\xi \neq x_1, x_2$, ossia $\xi \in (x_1, x_2)$. Dato che $\phi(x) \leq M = \phi(\xi)$ per ogni $x \in [x_1, x_2]$,

$$\frac{\phi(x) - \phi(\xi)}{x - \xi} \begin{cases} \leq 0 & \forall x > \xi, \\ \geq 0 & \forall x < \xi, \end{cases}$$

Passando al limite per $x \rightarrow \xi$ da destra e da sinistra e, sapendo che i limiti destro e sinistro esistono e coincidono, si ha

$$\phi'(\xi) = \lim_{x \rightarrow \xi^+} \frac{\phi(x) - \phi(\xi)}{x - \xi} \leq 0 \quad \text{e} \quad \phi'(\xi) = \lim_{x \rightarrow \xi^-} \frac{\phi(x) - \phi(\xi)}{x - \xi} \geq 0$$

da cui $0 \leq \phi'(\xi) \leq 0$, e quindi $\phi'(\xi) = 0$. □

DIMOSTRAZIONE DEL TEOREMA 4.1. Definiamo la funzione (ausiliaria) ϕ

$$\phi(x) := f(x) - f(x_1) - \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1),$$

che rappresenta la distanza verticale tra il punto $(x, f(x))$ del grafico della funzione e la retta secante passante per i suoi estremi. La funzione ϕ soddisfa le ipotesi di regolarità del Teorema di Rolle (cioè è continua in $[x_1, x_2]$ e derivabile in (x_1, x_2)). Inoltre

$$\phi(x_1) = f(x_1) - f(x_1) - \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x_1 - x_1) = 0,$$

$$\phi(x_2) = f(x_2) - f(x_1) - \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x_2 - x_1) = 0.$$

Quindi esiste un valore $\xi \in (x_1, x_2)$ tale che $\phi'(\xi) = 0$. Dato che

$$\phi'(x) = f'(x) - \frac{f(x_2) - f(x_1)}{x_2 - x_1} \quad x \in (x_1, x_2),$$

si deduce che

$$\phi'(\xi) = f'(\xi) - \frac{f(x_2) - f(x_1)}{x_2 - x_1} = 0,$$

cioè la conclusione. □

Conseguenze del Teorema di Lagrange. L'apparentemente innocuo Teorema di Lagrange è un'arma estremamente potente. Vediamo perché.

a. Funzioni monotone. Sia f derivabile in (a, b) . Allora

$$f'(x) > 0 \quad \forall x \in (a, b) \quad \implies \quad f \quad \text{strettamente crescente in } (a, b).$$

Infatti, supponiamo $f'(x) > 0$ per ogni $x \in (a, b)$ e siano x_1, x_2 in (a, b) tali che $x_1 < x_2$. Per il Teorema di Lagrange, esiste $\xi \in (x_1, x_2) \subset (a, b)$ tale che

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1).$$

Dato che $f'(\xi) > 0$ per ipotesi, ne segue $f(x_2) > f(x_1)$.

Analogamente si dimostra che

$$f'(x) < 0 \quad \forall x \in (a, b) \quad \implies \quad f \quad \text{strettamente decrescente in } (a, b).$$

Se invece dell'informazione $f'(x) > 0$ o $f'(x) < 0$, si ha l'informazione più debole $f'(x) \geq 0$ o $f'(x) \leq 0$, la conclusione va sostituita con le analoghe proprietà di monotonía deboli (nondecrescente/noncrescente).

Consideriamo, come esempio, la funzione

$$f(x) = \frac{1}{1 + x^2},$$

e studiamone la monotonía. Da quanto si è appena detto, basta studiare il segno della derivata prima:

$$\left(\frac{1}{1+x^2}\right)' = -\frac{2x}{(1+x^2)^2},$$

Dato che $f'(x)$ è strettamente positiva per $x < 0$ e strettamente negativa per $x > 0$, la funzione è crescente in $(-\infty, 0]$ ed è decrescente $[0, +\infty)$.

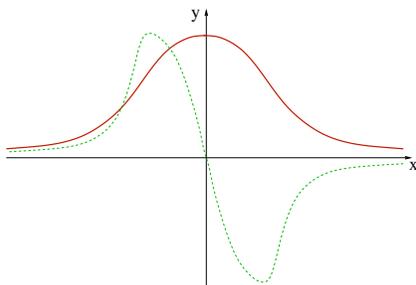


FIGURA 4. Grafico (qualitativo) di $1/(1+x^2)$ e della sua derivata

Vediamo un secondo esempio. Consideriamo la funzione $f(x) = \frac{1}{x}$. Dato che la derivata di questa funzione è

$$f'(x) = -\frac{1}{x^2} < 0 \quad \forall x \neq 0,$$

concludiamo che la funzione f è decrescente... Se però calcoliamo la differenza del valore della funzione in 1 e in -1 , otteniamo una contraddizione: $f(1) - f(-1) = 1 + 1 > 0$. Cosa sta succedendo? Bisogna stare attenti al fatto che le conclusioni sulla monotonía delle funzioni seguono dal Teorema di Lagrange che vale su intervalli, cioè su insiemi “senza buchi” (si dicono *insiemi connessi*). Se togliamo dall’enunciato del Teorema l’ipotesi di “assenza di buchi”, la conclusione non è più vera.² Nel caso della funzione $1/x$ stiamo applicando il Teorema all’insieme $(-\infty, 0) \cup (0, +\infty)$ che invece ha un buco: non contiene il punto 0. Ecco l’errore. Quindi la funzione $f(x) = 1/x$ NON è decrescente in $\mathbb{R} \setminus \{0\}$! Possiamo invece correttamente applicare i risultati sulla monotonía alle semirette $(-\infty, 0)$ e $(0, \infty)$ *separatamente* e concludere che $\frac{1}{x}$ è decrescente in $(-\infty, 0)$ ed è decrescente in $(0, +\infty)$.

OSSERVAZIONE 4.5. Cogliamo l’occasione per far notare una sottigliezza. Se $f'(x) > 0$ in un intervallo, necessariamente la funzione f è crescente. Cosa succede se $f'(x_0) > 0$ *nel solo punto* x_0 ? La possibilità di tracciare la retta tangente (che è crescente) suggerirebbe il fatto che la funzione f sia crescente, per lo meno in un intorno di x_0 . Invece

²Da cui il noto modo di dire, attribuito a N. Barbecue, “Non tutti i Teoremi riescono col buco”...

questa affermazione è falsa! Consideriamo la funzione

$$f(x) = \begin{cases} \frac{1}{2}x + x^2 \left(2 - \sin\left(\frac{1}{x}\right)\right) & x \neq 0, \\ 0 & x = 0, \end{cases}$$

Questa funzione è derivabile dappertutto e (c'è da dirlo? ...verificare!)

$$f'(x) = \begin{cases} 1/2 + \cos\left(\frac{1}{x}\right) + 2x \left(2 - \sin\left(\frac{1}{x}\right)\right) & x \neq 0, \\ 1/2 & x = 0, \end{cases}$$

Quindi $f'(0) = \frac{1}{2} > 0$, ma in ogni intorno di $x = 0$ cadono punti in cui la derivata è negativa: si tratta dei punti in cui $\cos\left(\frac{1}{x}\right)$ è uguale a -1 . Quindi *non è vero che f è crescente in un intorno dell'origine*. Il problema sta nel fatto che f' non è continua in 0 . Se fosse stata continua, $f'(x_0) > 0$ avrebbe implicato $f'(x) > 0$ in un intorno di x_0 e quindi la monotonia in tale intorno.

b. Funzioni a derivata nulla. Una seconda conseguenza del Teorema di Lagrange:

$$f'(x) = 0 \quad \forall x \in (a, b) \quad \implies \quad f \quad \text{è costante in } (a, b).$$

Infatti, per ogni coppia di valori $x_1, x_2 \in (a, b)$, esiste un valore ξ , compreso tra i due, per cui $f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1)$. Dato che $f'(x) = 0$ per ogni $x \in (a, b)$, si avrà, in particolare, $f'(\xi) = 0$, cioè

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) = 0 \quad \implies \quad f(x_2) = f(x_1).$$

Si noti che, anche qui, ha un ruolo fondamentale il fatto che si lavori su intervalli. Ad esempio, la funzione f definita da

$$f(x) = \begin{cases} 0 & x \in [0, 1], \\ 1 & x \in [2, 3], \end{cases}$$

è derivabile nel suo insieme di definizione $[0, 1] \cup [2, 3]$ e la sua derivata è ovunque nulla, ma la funzione si guarda bene dall'essere costante.

c. Lipschitzianità di funzioni a derivata limitata. Siano I l'intervallo aperto (α, β) e $f : I \rightarrow \mathbb{R}$ una funzione derivabile in I . Se f' è limitata in I , cioè se

$$\exists L > 0 \quad \text{tale che} \quad |f'(x)| \leq L \quad \forall x \in I.$$

allora dal Teorema di Lagrange segue

$$|f(x_2) - f(x_1)| = |f'(\xi)(x_2 - x_1)| \leq L|x_2 - x_1| \quad \forall x_1, x_2 \in (\alpha, \beta).$$

Quindi *una funzione derivabile con derivata limitata è lipschitziana*.

In particolare, se $f \in C^1(I)$ (sempre con $I = (\alpha, \beta)$), la derivata prima è continua in $[a, b]$ per ogni $[a, b] \subset I$ e quindi, per il Teorema di Weierstrass, è limitata in $[a, b]$. Se ne deduce che le funzioni in $C^1(I)$ sono **localmente lipschitziane**, cioè lipschitziane in ogni intervallo chiuso e limitato contenuto in I .

d. Approssimazione lineare. Un'ulteriore applicazione interessante del Teorema di Lagrange è la stima dell'errore che si commette approssimando una funzione con la sua tangente in un punto. Sia f derivabile in $[a, b]$. Supponiamo conoscere il valore della funzione f e della sua derivata prima f' in un punto assegnato $x_0 \in [a, b]$. Si può pensare che il valore della funzione f in un qualsiasi altro punto sia dato *approssimativamente* dal valore della funzione lineare che definisce la tangente al grafico di f in x_0 , cioè

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

Questo corrisponde ad approssimare il grafico della funzione f con quello della sua tangente. *E' possibile stimare l'errore che commettiamo facendo questa approssimazione?* Consideriamo un esempio concreto. Vogliamo calcolare, in modo approssimato, il valore di $\sin(1/10)$. Dato che $1/10$ è ragionevolmente vicino a 0, possiamo pensare di approssimare la funzione $\sin x$ con la sua tangente in $x = 0$, cioè $\sin(x) \approx x$. Calcolando in $x = 0,1$ otteniamo l'approssimazione richiesta

$$\sin(0,1) \approx 0,1.$$

Il problema fondamentale è: quant'è grande l'errore commesso? In altri termini, è possibile stimare $|\sin(0,1) - 0,1|$?

Torniamo al caso generale. Supponiamo di lavorare con una funzione f che sia *derivabile due volte* nell'intervallo (a, b) e supponiamo che la derivata seconda f'' sia limitata, cioè esista $M > 0$ tale che $|f''(x)| \leq M$ per ogni $x \in [a, b]$. Dato $x_0 \in [a, b]$, vogliamo stimare *il valore assoluto* della quantità

$$R(x; x_0) = f(x) - f(x_0) - f'(x_0)(x - x_0).$$

Applicando il Teorema di Lagrange otteniamo l'espressione

$$R(x; x_0) = f'(\xi)(x - x_0) - f'(x_0)(x - x_0) = (f'(\xi) - f'(x_0))(x - x_0),$$

dove ξ è compreso tra x e x_0 . Applicando il Teorema di Lagrange a $f'(\xi) - f'(x_0)$

$$R(x; x_0) = f''(\eta)(\xi - x_0)(x - x_0),$$

dove η è tra ξ e x_0 . Quindi il valore assoluto dell'errore $R(x; x_0)$ è stimato da

$$(36) \quad |R(x; x_0)| = |f''(\eta)||\xi - x_0||x - x_0| \leq M|x - x_0|^2,$$

dove si è usata la limitatezza della derivata seconda f'' e il fatto che $|\xi - x_0| \leq |x - x_0|$. Nel caso-modello di $f(x) = \sin x$, $x_0 = 0$ e $x = 1/10$, si ha $M = 1$, pertanto

$$|R(10^{-1}; 0)| \leq \frac{1}{100},$$

dove si è usato che $|f''(x)| = |-\sin x| \leq 1$ e $|x - x_0| = 1/10$. Quindi

$$0.09 < \sin(0.1) < 0.11$$

e. Derivabilità tramite il limite della derivata. In alcune situazioni, capita di lavorare con funzioni definite tramite formule diverse in diversi intervalli. Consideriamo come caso modello una funzione della forma

$$f(x) = \begin{cases} f_1(x) & x < x_0, \\ \ell & x = x_0, \\ f_2(x) & x > x_0, \end{cases}$$

dove $\ell \in \mathbb{R}$, e f_1, f_2 sono funzioni note. La domanda naturale è se la funzione f sia derivabile nel punto x_0 oppure no. Come abbiamo già visto, la derivabilità implica la continuità, quindi, prima di tutto, deve essere verificata la condizione

$$\lim_{x \rightarrow x_0^-} f_1(x) = \ell = \lim_{x \rightarrow x_0^+} f_2(x).$$

Se questa condizione non è verificata, la funzione non è continua in x_0 e quindi, a maggior ragione, non è neanche derivabile in x_0 . Nel caso in cui la funzione sia continua in x_0 , per stabilirne la derivabilità occorre calcolare il limite del rapporto incrementale in x_0 . Dato che la funzione f è definita da espressioni diverse a seconda che ci si trovi a destra o a sinistra di x_0 , è sensato calcolare il limite del rapporto incrementale da destra e da sinistra.³ Per definizione, *la funzione f è derivabile in x_0 se e solo se questi limiti esistono e coincidono*, ossia se e solo se

$$\lim_{x \rightarrow x_0^-} \frac{f_1(x) - \ell}{x - x_0} = \lim_{x \rightarrow x_0^+} \frac{f_2(x) - \ell}{x - x_0}.$$

La derivata in x_0 è il valore comune di questi due limiti.

In molte situazioni le f_1 e f_2 sono funzioni derivabili in tutto il loro insieme di definizione ed è possibile calcolare esplicitamente la funzione derivata. Invece di calcolare il limite del rapporto incrementale, può essere più semplice calcolare le derivate f_1' e f_2' nei rispettivi domini e calcolare il limite di queste funzioni derivate. Quale informazione dà questa procedura?

PROPOSIZIONE 4.6. *Dato $x_0 \in \mathbb{R}$ e $r > 0$, sia f continua in x_0 e derivabile in $(x_0 - r, x_0 + r) \setminus \{x_0\}$ e supponiamo che esistano finiti il limite destro e sinistro $\lim_{x \rightarrow x_0^\pm} f'(x) = \ell^\pm$. Allora f è derivabile in x_0 se e solo se $\ell^+ = \ell^-$.*

³Se esiste il limite destro del rapporto incrementale di una funzione f in x_0 , si dice che f è *derivabile da destra in x_0* . Analogamente per il limite sinistro. Per indicare il limite destro/sinistro del rapporto incrementale (qualora esistano), cioè per indicare la *derivata destra/sinistra* si usa il simbolo $D_\pm f(x_0)$, o varianti.

DIMOSTRAZIONE. Grazie al Teorema di Lagrange,

$$\frac{f(x) - f(x_0)}{x - x_0} = \begin{cases} f'(\xi_-) & \text{se } x < x_0, & (x < \xi_- < x_0) \\ f'(\xi_+) & \text{se } x > x_0, & (x_0 < \xi_+ < x) \end{cases}$$

dove ξ_{\pm} sono punti opportuni tra x e x_0 . Passando al limite per $x \rightarrow x_0$ da sinistra, dato che il limite sinistro della derivata f' esiste ed è uguale ad ℓ^- ,

$$\lim_{x \rightarrow x_0^-} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \rightarrow x_0^-} f'(\xi_-) = \ell^-.$$

Analogamente per il limite destro. Quindi, nelle ipotesi della Proposizione 4.6, i limiti destro e sinistro del rapporto incrementale esistono e sono uguali, rispettivamente, a ℓ^+ e ℓ^- . A questo punto, la conclusione è evidente. \square

Ad esempio, studiamo la derivabilità in 0 della funzione $f(x) = x|x|$. Dato che

$$x|x| = \begin{cases} -x^2 & x < 0, \\ x^2 & x \geq 0, \end{cases}$$

la funzione è certamente derivabile per $x \neq 0$ e

$$D(x|x|) = \begin{cases} -2x & x < 0, \\ 2x & x > 0. \end{cases}$$

Dato che $\lim_{x \rightarrow 0^-} -2x = \lim_{x \rightarrow 0^+} 2x = 0$, la funzione è derivabile in 0.

Consideriamo, invece, la funzione $e^{-|x|}$. In questo caso

$$e^{-|x|} = \begin{cases} e^x & x < 0, \\ e^{-x} & x \geq 0. \end{cases}$$

La funzione è derivabile per $x \neq 0$ e

$$D(e^{-|x|}) = \begin{cases} e^x & x < 0, \\ -e^{-x} & x > 0. \end{cases}$$

Dato che $\lim_{x \rightarrow 0^-} e^x = 1 \neq -1 = \lim_{x \rightarrow 0^+} -e^{-x}$, la funzione non è derivabile in 0.

La verifica della derivabilità in x_0 tramite il calcolo del limite della derivata a destra e a sinistra di x_0 è lecita *solo quando la derivata ammetta limiti destro e sinistro in x_0* . Quando questi limiti non esistano, il criterio non è più valido. La funzione può essere derivabile o può non esserlo. Ad esempio, consideriamo la funzione

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & x \neq 0, \\ 0 & x = 0. \end{cases}$$

Per $x \neq 0$, la derivata prima f' di questa funzione è

$$f'(x) = 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right).$$

Per $x \rightarrow 0$, il primo dei due termini è infinitesimo, mentre il secondo non ammette limite, quindi non esiste $\lim_{x \rightarrow 0^\pm} f'(x)$. La Proposizione 4.6 non è applicabile. Per studiare la derivabilità in zero, calcoliamo direttamente il limite del rapporto incrementale

$$\lim_{h \rightarrow 0} \frac{h^2 \sin(1/h) - 0}{h - 0} = \lim_{h \rightarrow 0} h \sin\left(\frac{1}{h}\right) = 0.$$

Quindi la funzione è derivabile in 0 e $f'(0) = 0$.

Tabella delle derivate

funzione f	derivata prima f'	funzione f	derivata prima f'
costante	0	x^α	$\alpha x^{\alpha-1}$
$\sin x$	$\cos x$	$\cos x$	$-\sin x$
e^x	e^x	$\ln x$	$\frac{1}{x}$
$\tan x$	$\frac{1}{\cos^2 x} = 1 + \tan^2 x$	$\cot x$	$-\frac{1}{\sin^2 x} = -1 - \cot^2 x$
a^x	$a^x \ln a$	$\arctan x$	$\frac{1}{1+x^2}$
$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$	$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$
$\sinh x$	$\cosh x$	$\cosh x$	$\sinh x$

CAPITOLO 6

Analisi locale e analisi globale

Una proprietà di una funzione f è *locale* se dipende dal comportamento della funzione nell'intorno di un punto x . Continuità e derivabilità in x sono proprietà locali. Una proprietà di una funzione f è *globale* se vale in tutto l'insieme di definizione della f . Ad esempio le funzioni e^x , $\arctan x$, x^3 , ... sono funzioni globalmente monotone crescenti e pertanto (globalmente) invertibili. Nella prima parte di questo capitolo approfondiamo l'uso della derivazione per determinare proprietà locali di funzioni: massimi/minimi relativi, punti di singolarità, ... Torneremo più avanti sulle proprietà globali, concentrandoci sul problema di determinare massimi e minimi (assoluti) di una funzione assegnata.

1. Punti stazionari

Abbiamo già definito massimo e minimo di una funzione: data $f : D \rightarrow \mathbb{R}$, un punto $x_0 \in D$ è **punto di massimo** di f se $f(x) \leq f(x_0)$ per ogni $x \in D$. Il valore $f(x_0) = \max_{x \in D} f(x)$ è il **massimo della funzione f in D** . Analogamente per i minimi.

L'esistenza di massimo e/o minimo è una proprietà globale della funzione. E' utile introdurre un analogo locale del concetto di massimo e di minimo.

DEFINIZIONE 1.1. *Massimo e minimo locale.* Il punto $x_0 \in D$ è un **punto di massimo locale (o relativo)** e il valore $f(x_0)$ è un **massimo locale (o relativo)** di f se esiste un intorno $(x_0 - \delta, x_0 + \delta)$ del punto x_0 tale che $f(x_0)$ è il massimo di f in $(x_0 - \delta, x_0 + \delta)$:

$$\exists \delta > 0 \quad \text{tale che} \quad f(x) \leq f(x_0) \quad \forall x \in D \cap (x_0 - \delta, x_0 + \delta).$$

Analogamente per il **minimo locale**.

Un punto x_0 che sia o di massimo o minimo locale è un **punto di estremo locale**.

Per distinguere in modo più chiaro il massimo e il minimo dagli analoghi concetti locali, si parla di **massimo globale (o assoluto)** e di **minimo globale (o assoluto)**. Dalla definizione segue immediatamente che se x_0 è punto di massimo globale, allora è anche punto di massimo locale. Il viceversa invece non è vero, come nel caso del grafico rappresentato in Figura 1. Per un esempio analitico, si può considerare la funzione

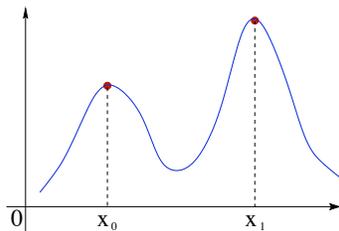


FIGURA 1. Il punto x_0 è di massimo locale, ma non globale; il punto x_1 è di massimo globale.

$f(x) = x^4 - x^2$. Dato che $f(0) = 0$ e $f(x) \leq 0$ per $x \in (-1, 1)$, il punto $x = 0$ è un punto di massimo locale, ma non è di massimo globale dato che $\lim_{x \rightarrow \pm\infty} f(x) = +\infty$.

DEFINIZIONE 1.2. Sia $D \subset \mathbb{R}$. Un punto $x_0 \in \mathbb{R}$ è **interno a D** se esiste un intorno di x_0 interamente contenuto in D , cioè se esiste $\delta > 0$ per cui $(x_0 - \delta, x_0 + \delta) \subset D$.

Se una funzione f ha un massimo o un minimo locale in corrispondenza di un punto x_0 interno all'insieme di definizione e in cui la funzione è derivabile, necessariamente

$$f'(x_0) = 0.$$

Basta infatti pensare alla necessaria posizione orizzontale della retta tangente (Fig.2(a)). Per una dimostrazione analitica, sia x_0 un punto di massimo locale interno e supponiamo f derivabile in x_0 . Dato che $f(x) \leq f(x_0)$ per ogni $x \in (x_0 - \delta, x_0 + \delta)$

$$\frac{f(x) - f(x_0)}{x - x_0} \begin{cases} \leq 0 & \forall x_0 < x < x_0 + \delta, \\ \geq 0 & \forall x_0 - \delta < x < x_0, \end{cases}$$

Passando al limite per $x \rightarrow x_0$ si deduce $f'(x_0) = 0$.

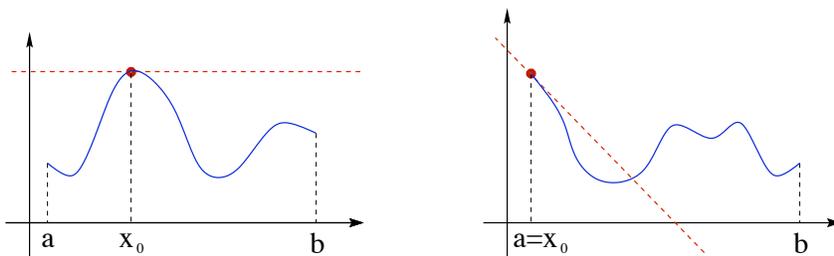


FIGURA 2. Se il punto di massimo locale è interno, la tangente è orizzontale. Se, invece, si trova sul bordo... non è detto!

Nel caso in cui il punto x_0 non sia interno al dominio, non è detto che la retta tangente sia orizzontale (Fig.2(b)). Conclusioni analoghe per i punti di minimo.

Pertanto, risolvere l'equazione $f'(x) = 0$ permette di determinare i possibili candidati a punti di minimo o massimo locale *interno* in cui f è *derivabile*. È comunque

possibile che un estremo locale cada in un punto in cui la funzione non è derivabile; ad esempio $f(x) = |x|$ ha un minimo (globale) in $x = 0$ dove non ha retta tangente.

DEFINIZIONE 1.3. Punto stazionario. Sia $f : D \rightarrow \mathbb{R}$. Un punto x_0 , interno a D , tale che $f'(x_0) = 0$ si dice **punto stazionario**¹ (o **punto critico**) della funzione f .

Equivalentemente, si può affermare che i punti critici di f sono i punti x per cui la tangente al grafico di f in $(x, f(x))$ è orizzontale.

ESERCIZIO 1.4. Determinare i punti critici di $f(x) = x^7 + 14x^4 + 1$.

Classificazione dei punti stazionari. Se x_0 è un punto di massimo o di minimo locale interno e f è derivabile in x_0 , necessariamente x_0 è un punto critico, cioè $f'(x_0) = 0$. *Il viceversa non è vero:* esistono punti x_0 tali che $f'(x_0) = 0$, ma che non sono né punti di massimo locale, né punti di minimo locale. Ad esempio, la funzione $f(x) = x^3$ è strettamente crescente (quindi non ha né punti di massimo né punti di minimo in \mathbb{R}), ma $f'(x) = 3x^2$ si azzera nel punto $x = 0$.

Conoscendo il segno della derivata prima alla destra e alla sinistra del punto in questione, grazie al legame tra monotonia e segno di f' , si può individuare quando un punto stazionario sia di massimo o di minimo. Supponiamo f derivabile in $(x_0 - \delta, x_0 + \delta)$ con $\delta > 0$, allora

$$f'(x) \begin{cases} \geq 0 & x_0 - \delta < x < x_0, \\ \leq 0 & x_0 < x < x_0 + \delta, \end{cases} \implies x_0 \text{ punto di massimo locale.}$$

Analogamente, per il minimo, vale

$$f'(x) \begin{cases} \leq 0 & x_0 - \delta < x < x_0, \\ \geq 0 & x_0 < x < x_0 + \delta, \end{cases} \implies x_0 \text{ punto di minimo locale.}$$

ESERCIZIO 1.5. Determinare i punti critici della funzione $f(x) = x^2(3x^2 - 8x + 6)$ e dire quali di essi sono punti di massimo o di minimo.

Soluzione. La derivata prima della funzione è

$$f'(x) = 2x(3x^2 - 8x + 6) + x^2(6x - 8) = 12x(x^2 - 2x + 1) = 12x(x - 1)^2.$$

I punti critici sono $x = 0$ e $x = 1$; il punto $x = 0$ è punto di minimo, mentre il punto $x = 1$ non è né di massimo né di minimo. Il grafico qualitativo della funzione f è in Figura 3.

¹Il termine *stazionario* è ereditato dalla cinematica. Se x è la posizione di un punto su cui agisce una forza conservativa con potenziale dato dalla funzione $f = f(x)$, allora l'accelerazione del punto è proporzionale a f' . Se il punto viene collocato a riposo nella posizione x_0 e $f'(x_0) = 0$, allora, dato che l'accelerazione (cioè la variazione di velocità) è nulla, il punto *stazionerà* nella posizione x_0 per ogni tempo successivo

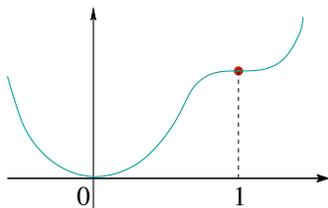


FIGURA 3. Il grafico di $f(x) = x^2(3x^2 - 8x + 6)$ dell'Esercizio 1.5.

Se x_0 è un punto critico di f , per riconoscere se f' cambia segno attraversando x_0 , basta considerare il segno di f'' , qualora esista:

$$f'(x_0) = 0, \quad f''(x_0) > 0 \quad \Longrightarrow \quad x_0 \text{ punto di minimo locale;}$$

$$f'(x_0) = 0, \quad f''(x_0) < 0 \quad \Longrightarrow \quad x_0 \text{ punto di massimo locale.}$$

Si noti che si tratta solo di condizioni sufficienti: ad esempio, la funzione $f(x) = x^4$ ha un punto di minimo in 0, ma $f''(0) = 0$.

ESEMPIO 1.6. *Una raffinatezza per buongustai.* In genere, si immagina il grafico di una funzione vicino al punto di minimo x_0 con $f'(x) \leq 0$ per $x_0 - \delta < x < x_0$ e $f'(x) \geq 0$ per $x_0 < x < x_0 + \delta$. Esistono però anche situazioni in cui una funzione alla sinistra del punto di minimo non è decrescente e alla destra non è crescente. Scetticismo? Ecco un esempio:

$$f(x) = \begin{cases} x^2 \left(2 - \sin \left(\frac{1}{x} \right) \right) & x \neq 0, \\ 0 & x = 0. \end{cases}$$

La funzione f è derivabile in tutto \mathbb{R} e

$$f(x) \geq 0 \quad \forall x \in \mathbb{R}, \quad \text{e} \quad f(x) = 0 \quad \Longleftrightarrow \quad x = 0.$$

Il punto $x = 0$ è punto di minimo globale, e quindi di minimo locale. Necessariamente $f'(0) = 0$ (come si può ottenere anche tramite il calcolo del limite del rapporto incrementale). La derivata prima di f nei punti $x \neq 0$ è

$$f'(x) = 2x \left(2 - \sin \left(\frac{1}{x} \right) \right) + \cos \left(\frac{1}{x} \right),$$

quindi, per $x \approx 0$, si ha $f'(x) \approx \cos \left(\frac{1}{x} \right)$, che assume valori sia positivi che negativi.

2. Analisi al microscopio

Nello studio dell'andamento qualitativo del grafico di una funzione, è interessante approfondire quello che succede in prossimità di certi punti significativi. Qui consideriamo come punti "significativi" quelli che corrispondono ad una delle seguenti situazioni:

- (a) punti x_0 che non sono nell'insieme di definizione di f , ma che sono sul “bordo” (ad esempio, se $f : (a, b] \rightarrow \mathbb{R}$, il punto $x_0 = a$, oppure se $f : [a, b] \setminus \{c\} \rightarrow \mathbb{R}$, $x_0 = c$);
- (b) punti x_0 dell'insieme di definizione in cui f non è continua;
- (c) punti x_0 in cui f è continua, ma non derivabile.

Nei casi (b) e (c) si parla talvolta di *punti di singolarità*.

Asintoti verticali. Sia nel caso (a) che nel caso (b), si calcola il limite

$$\lim_{x \rightarrow x_0} f(x).$$

Se il limite è $+\infty$ o $-\infty$, la funzione ha in $x = x_0$ un **asintoto verticale**. Lo stesso è vero

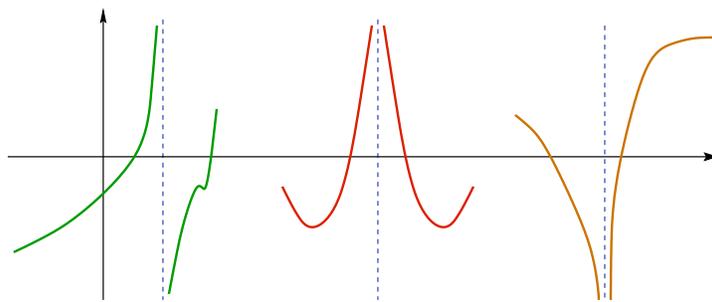


FIGURA 4. Alcuni esempi di asintoti verticali.

nel caso in cui sia il limite destro che il limite sinistro tendano a $+\infty$ o $-\infty$, ma con segni opposti. In generale, zeri del denominatore di una funzione razionale (che non siano anche zeri del numeratore), corrispondono a punti di asintoto verticale.

Ci sono situazioni più esotiche: il limite potrebbe non esistere oppure potrebbero esistere i limiti destro e sinistro, ma con valori diversi, ... A voi individuare possibili esempi e corrispondenti grafici.

ESERCIZIO 2.1. Studiare la funzione $f(x) = \arctan(1/x)$ vicino al punto $x = 0$.

Punti angolosi e cuspidi. Consideriamo il caso (c), quindi supponiamo x_0 tale che la funzione f sia continua in x_0 , ma non derivabile. Se esistono finiti i limiti destro e sinistro della derivata prima

$$\lim_{x \rightarrow x_0^\pm} f'(x) = \ell^\pm,$$

dato che f non è derivabile in x_0 , deve essere $\ell^+ \neq \ell^-$. Un punto di questo genere si chiama **punto angoloso** (o **spigolo**). Per disegnarlo correttamente è possibile tracciare le rette tangenti destra e sinistra, cioè le rette di equazione $y = f(x_0) + \ell^\pm(x - x_0)$.

Nel caso in cui i limiti destro e sinistro siano $\pm\infty$ si possono avere due situazioni differenti. Se entrambi sono $+\infty$ (o $-\infty$), cioè se

$$\lim_{x \rightarrow x_0^\pm} f'(x) = +\infty \quad (-\infty),$$

il punto x_0 è un punto a tangente verticale. Se invece i limiti destro e sinistro sono $\pm\infty$, ma con segni opposti, il punto x_0 è una **cuspid**e del grafico di f . Per un esempio di cuspid, si consideri la funzione $f(x) = \sqrt{|x|}$. In questo caso

$$\lim_{x \rightarrow 0^+} f'(x) = \lim_{x \rightarrow 0^+} \frac{1}{2\sqrt{x}} = +\infty, \quad \lim_{x \rightarrow 0^-} f'(x) = \lim_{x \rightarrow 0^-} -\frac{1}{2\sqrt{-x}} = -\infty.$$

Ovviamente sono possibili comportamenti analoghi a quelli descritti, ma misti: ad

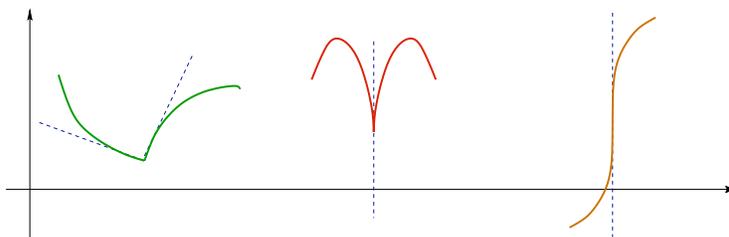


FIGURA 5. Da sinistra: un punto angoloso, una cuspid e un punto a tangente verticale.

esempio, una funzione può avere derivata prima che tende ad un valore dato da destra e che diverge da sinistra, o tutte le varianti che la mente è in grado di inventare.

3. Comportamento asintotico

Se la funzione f è definita in insiemi illimitati, è interessante studiarne il comportamento per $x \rightarrow \pm\infty$. Per fissare le idee, consideriamo una funzione f definita su una semiretta $[a, +\infty)$ con $a \in \mathbb{R}$. In questo caso si vuole stabilire cosa succeda per $x \rightarrow +\infty$, cioè determinare il *comportamento asintotico* per $x \rightarrow +\infty$. Considerazioni analoghe valgono per il caso di semirette del tipo $(-\infty, a]$, per \mathbb{R} , e, in generale, per domini illimitati.

La prima operazione sensata è il calcolo del limite per $x \rightarrow +\infty$. Se

$$\exists \lim_{x \rightarrow +\infty} f(x) = \ell \in \mathbb{R},$$

si dice che la funzione f tende asintoticamente ad ℓ , oppure che f ha un asintoto orizzontale (di equazione $y = \ell$) per $x \rightarrow +\infty$. Il grafico della funzione f si avvicina alla retta di equazione $y = \ell$ per x sempre più grandi. Per un disegno più preciso, si può studiare il segno della funzione $f(x) - \ell$, che indica se il grafico della funzione f sia al di sopra o al di sotto dell'asintoto.

ESEMPIO 3.1. Consideriamo la funzione

$$f(x) = \frac{2x^2}{x^2 + 1} \quad x \in [1, +\infty).$$

In questo caso,

$$\lim_{x \rightarrow +\infty} \frac{2x^2}{x^2 + 1} = 2,$$

quindi la funzione ha l'asintoto orizzontale di equazione $y = 2$. Dato che

$$f(x) - \ell = \frac{2x^2}{x^2 + 1} - 2 = -\frac{2}{x^2 + 1} < 0,$$

quindi f tende a $y = 2$ dal basso. A voi il gusto di tracciare il grafico di questa funzione.

Invece, la funzione

$$f(x) = \frac{2x^2 \sin x}{x^2 + 1} \quad x \in [1, +\infty),$$

non ha limite per $x \rightarrow +\infty$ e quindi non ha asintoto orizzontale.

Se il limite della funzione f esiste, ma è $+\infty$ o $-\infty$, evidentemente non c'è asintoto orizzontale. Che cosa si può dire in questo caso? È possibile che la funzione tenda ad un **asintoto obliquo**, ossia è possibile che esistano $a, b \in \mathbb{R}$ tali che

$$(37) \quad \lim_{x \rightarrow +\infty} [f(x) - (ax + b)] = 0.$$

Questa proprietà indica che il grafico della funzione f si avvicina al grafico della retta $y = ax + b$ per $x \rightarrow +\infty$. Il problema è: *come determinare (qualora esistano) le costanti a e b ?* Supponiamo che valga (37), allora

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{x} = \lim_{x \rightarrow +\infty} \frac{f(x) - ax}{x} + a = a.$$

Una volta noto a , è possibile determinare b (qualora esista) calcolando

$$\lim_{x \rightarrow +\infty} [f(x) - ax] = b.$$

Ecco, quindi, le istruzioni per determinare la presenza di un asintoto obliquo:

i. calcolare $\lim_{x \rightarrow +\infty} f(x)$: se il limite esiste finito, c'è un asintoto orizzontale (fine dello studio a $+\infty$), se il limite non esiste, non c'è né asintoto obliquo, né asintoto orizzontale (fine dello studio a $+\infty$), se il limite è $+\infty$ o $-\infty$ si va al punto (ii);

ii. calcolare $\lim_{x \rightarrow +\infty} f(x)/x$: se il limite esiste finito, il suo valore è a e si va al punto (iii), se il limite non esiste o se vale $\pm\infty$, non c'è asintoto obliquo (fine dello studio a $+\infty$);

iii. calcolare $\lim_{x \rightarrow +\infty} [f(x) - ax]$: se il limite esiste finito, il suo valore è b , la funzione

ha asintoto obliquo di equazione $y = ax + b$, se il limite non esiste o se vale $\pm\infty$, non c'è asintoto obliquo (fine dello studio a $+\infty$).

ESEMPIO 3.2. Consideriamo la funzione

$$f(x) = \frac{x^2 - 1}{3x + 1} \quad x \in [0, +\infty).$$

Si ha

$$\begin{aligned} \lim_{x \rightarrow +\infty} f(x) &= \lim_{x \rightarrow +\infty} \frac{x^2 - 1}{3x + 1} = +\infty, \\ \lim_{x \rightarrow +\infty} \frac{f(x)}{x} &= \lim_{x \rightarrow +\infty} \frac{x^2 - 1}{3x(x + 1)} = \frac{1}{3} =: a, \\ \lim_{x \rightarrow +\infty} f(x) - \frac{x}{3} &= \lim_{x \rightarrow +\infty} \frac{x^2 - 1}{3x + 1} - \frac{x}{3} = \lim_{x \rightarrow +\infty} \frac{-3 - x}{3(3x + 1)} = -\frac{1}{9} =: b. \end{aligned}$$

Quindi la funzione ha un asintoto obliquo di equazione $y = \frac{1}{3}x - \frac{1}{9}$. Anche in questo caso, per disegnare un grafico più preciso, si può studiare il segno della funzione

$$f(x) - (ax + b) = \frac{x^2 - 1}{3x + 1} - \left(\frac{1}{3}x - \frac{1}{9}\right) = -\frac{8}{9(3x + 1)} < 0 \quad \forall x > -\frac{1}{3}.$$

La differenza è negativa, quindi la funzione tende all'asintoto dal basso.

Dopo il punto (i), se esiste finito $\lim_{x \rightarrow +\infty} f'(x)$, allora a è uguale al valore di questo limite e si può proseguire direttamente dal punto (iii). Se invece il limite di f' non esiste, bisogna necessariamente seguire il procedimento esposto sopra. Ad esempio, per $f(x) = x + \frac{\sin(x^2)}{x}$, si ha

$$f'(x) = 1 + 2 \cos(x^2) - \frac{\sin(x^2)}{x^2},$$

che non ammette limite per $x \rightarrow +\infty$, ma è facile vedere che la funzione ha un asintoto obliquo per $x \rightarrow +\infty$ di equazione $y = x$.

Esercizio 3.3. Sia $f(x) = \frac{p(x)}{q(x)}$ con p polinomio di grado k e q polinomio di grado h . Dimostrare che:

- (i) f ha un asintoto orizzontale se e solo se $k \leq h$;
- (ii) f ha un asintoto obliquo, non orizzontale, se e solo se $k = h + 1$.

Altri profili asintotici. Il caso dell'asintoto obliquo è solo una situazione molto particolare: può capitare che una funzione tenda asintoticamente ad una funzione, che non sia un polinomio di primo grado. Consideriamo, ad esempio,

$$f(x) = \frac{x^3 + 1}{x - 2}.$$

Grazie all'algoritmo di divisione di polinomi, possiamo riscrivere questa funzione come

$$f(x) = x^2 + 2x + 4 + \frac{9}{x-2}.$$

Da questa espressione è immediato vedere che

$$\lim_{x \rightarrow \pm\infty} [f(x) - (x^2 + 2x + 4)] = 0,$$

e quindi il grafico di f tende asintoticamente alla parabola $y = x^2 + 2x + 4$.

Riconsideriamo la funzione

$$f(x) = \frac{2x^2 \sin x}{x^2 + 1} \quad x \in [1, +\infty).$$

Dato che $\frac{2x^2}{x^2+1} \rightarrow 2$ per $x \rightarrow +\infty$, è sensato immaginare che questa funzione “assomigli” alla funzione $f(x) = 2 \sin x$ per $x \rightarrow +\infty$. Calcoliamo la differenza tra $f(x)$ e $2 \sin x$ e vediamo se è infinitesima:

$$\left| \frac{2x^2 \sin x}{x^2 + 1} - 2 \sin x \right| = \frac{2 |\sin x|}{1 + x^2} \leq \frac{2}{1 + x^2} \rightarrow 0 \quad \text{per } x \rightarrow +\infty.$$

Quindi

$$f(x) = \frac{2x^2 \sin x}{x^2 + 1} = 2 \sin x + h(x) \quad \text{con} \quad \lim_{x \rightarrow +\infty} h(x) = 0.$$

In generale se siamo in grado di riscrivere la funzione f nella forma

$$f(x) = g(x) + h(x)$$

con g funzione di cui si conosce il grafico e $h \rightarrow 0$ per $x \rightarrow \infty$, il grafico della funzione f tende verso quello della funzione g . Non esiste alcuna strategia generale per determinare

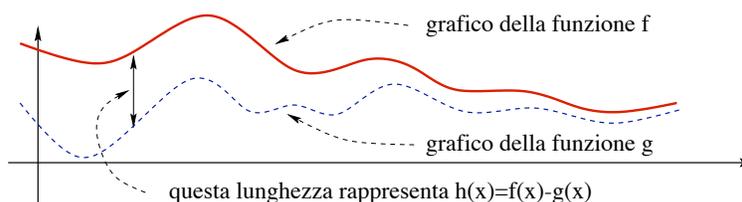


FIGURA 6. Il grafico di f con $f(x) = g(x) + h(x)$ e h infinitesima per $x \rightarrow +\infty$.

una decomposizione di questo genere.

4. Funzioni convesse

Come già detto, ripetuto ed usato ampiamente, il segno della derivata prima dà informazioni relative alla monotonia della funzione f . Quale ruolo gioca, invece, il segno della derivata seconda? Partiamo da una definizione.

DEFINIZIONE 4.1. Una funzione $f : [a, b] \rightarrow \mathbb{R}$ è **convessa** in $[a, b]$ se

$$(38) \quad f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \quad \forall x, y \in [a, b] \quad \forall t \in (0, 1).$$

Una funzione per cui valga la disuguaglianza opposta si dice **concava**.

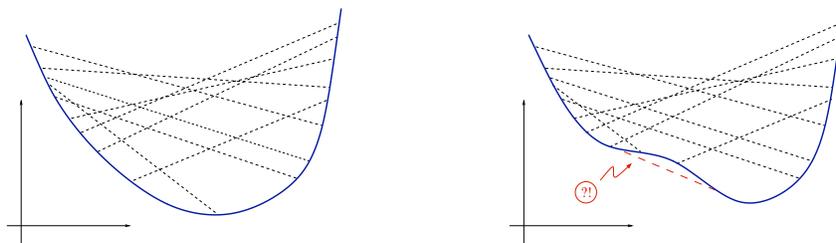


FIGURA 7. Una funzione convessa ed una non convessa.

Dalla definizione segue che se f è concava, allora $-f$ è convessa, e viceversa. Quindi studiare la convessità è sufficiente per comprendere anche la concavità.

OSSERVAZIONE 4.2. Esiste una maniera diversa di introdurre il concetto di convessità. Un sottoinsieme E del piano è **convesso** se scelta una qualsiasi coppia di punti P e Q appartenenti ad E , il segmento che li congiunge è interamente contenuto in E . Una funzione $f : [a, b] \rightarrow \mathbb{R}$ è **convessa** in $[a, b]$ se l'insieme $E := \{(x, y) : y \geq f(x)\}$, composto dai punti che si trovano sopra il suo grafico (detto **epigrafico**) è un insieme convesso. Le due definizioni sono comunque equivalenti (sapete dimostrarlo?).

Cerchiamo di capire il significato geometrico della condizione (38). Fissiamo $x = \bar{x}$ e $y = \bar{y}$ con $\bar{x} < \bar{y}$. Per $t \in (0, 1)$, definiamo $z(t) := t\bar{x} + (1-t)\bar{y} \in (\bar{x}, \bar{y})$. Scriviamo la retta che passa per $(\bar{x}, f(\bar{x}))$ e $(\bar{y}, f(\bar{y}))$:

$$\Phi(x) = f(\bar{x}) + \frac{f(\bar{y}) - f(\bar{x})}{\bar{y} - \bar{x}}(x - \bar{x}),$$

e calcoliamo questa funzione in $z(t)$. Dato che

$$\begin{aligned} \Phi(z(t)) &= f(\bar{x}) + \frac{f(\bar{y}) - f(\bar{x})}{\bar{y} - \bar{x}}(t\bar{x} + (1-t)\bar{y} - \bar{x}) \\ &= f(\bar{x}) + \frac{f(\bar{y}) - f(\bar{x})}{\bar{y} - \bar{x}}(1-t)(\bar{y} - \bar{x}) \\ &= f(\bar{x}) + (f(\bar{y}) - f(\bar{x}))(1-t) = tf(\bar{x}) + (1-t)f(\bar{y}), \end{aligned}$$

la condizione (38), si può riscrivere come

$$f(z(t)) \leq \Phi(z(t)) \quad \forall x, y \in [a, b] \quad \forall t \in (0, 1).$$

Questa scrittura ha un'interpretazione in termini di grafico immediata: *una funzione f è convessa, se per ogni scelta di x e y , il grafico di f giace al di sotto della retta secante che congiunge i punti $(x, f(x))$ e $(y, f(y))$ nell'intervallo di estremi x e y .*

ESERCIZIO 4.3. *Se f e g sono due funzioni convesse in $[a, b]$, allora una tra $\max\{f, g\}$ e $\min\{f, g\}$ è convessa. Sapete dire quale?*

PROPOSIZIONE 4.4. *Una funzione $f : [a, b] \rightarrow \mathbb{R}$ è convessa in $[a, b]$ se e solo se*

$$(39) \quad \frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(y) - f(z)}{y - z}$$

per ogni x, y, z tali che $a \leq x < z < y \leq b$.

La dimostrazione della formula (39) si ottiene riscrivendo in termini di rapporti incrementali la formula (38). I dettagli sono lasciati alla buona volontà del lettore.

La proprietà (39) può essere interpretata graficamente in termini di monotonia delle pendenze delle secanti: fissato y , la funzione

$$\phi(x) := \frac{f(x) - f(y)}{x - y}$$

è crescente in x . Quando la funzione è derivabile, questa proprietà diviene una richiesta di monotonia della derivata prima f' . Nel caso in cui la funzione f sia derivabile due volte, la monotonia della funzione f' può essere tradotta in termini di segno della derivata seconda f'' .

TEOREMA 4.5. *Sia $f : [a, b] \rightarrow \mathbb{R}$. Allora*

(i) se f è derivabile una volta, la funzione f è convessa in $[a, b]$ se e solo se f' è non decrescente in $[a, b]$;

(ii) se f è derivabile due volte, la funzione f è convessa in $[a, b]$ se e solo se $f''(x) \geq 0$ per ogni $x \in [a, b]$.

Per le funzioni concave, vale un risultato analogo sostituendo a “ f' crescente” la frase “ f' non crescente” e a “ $f'' \geq 0$ ” la frase “ $f'' \leq 0$ ”.

Dal Teorema 4.5(i) discende un'altra interpretazione geometrica della convessità: *se la funzione f è derivabile e convessa, il suo grafico è interamente al di sopra di ogni retta tangente ad esso.* Infatti, scriviamo la differenza tra f e la retta tangente in $(x_0, f(x_0))$

$$R(x; x_0) = f(x) - f(x_0) - f'(x_0)(x - x_0),$$

con l'obiettivo di dimostrare che se f è convessa, la funzione $R(x; x_0)$ è positiva. Appliciamo il Teorema di Lagrange e riscriviamo $R(x; x_0)$ come

$$R(x; x_0) = f'(\xi)(x - x_0) - f'(x_0)(x - x_0) = (f'(\xi) - f'(x_0))(x - x_0).$$

Se $x > x_0$ allora $\xi > x_0$ e quindi, essendo f' crescente, $f'(\xi) > f'(x_0)$. Ne segue che il termine a destra è positivo perché prodotto di termini positivi. Se $x < x_0$ allora $\xi < x_0$ e, sempre per la monotonía di f' , $f'(\xi) < f'(x_0)$. Questa volta i due termini sono entrambi negativi, ma comunque il loro prodotto è positivo.

DIMOSTRAZIONE DEL TEOREMA 4.5. (i) Supponiamo che f sia convessa, allora vale la (39). Quindi, passando al limite per $z \rightarrow x^+$ si ottiene

$$f'(x) \leq \frac{f(y) - f(x)}{y - x}.$$

Analogamente, passando al limite nella (39) per $z \rightarrow y^-$,

$$\frac{f(y) - f(x)}{y - x} \leq f'(y).$$

Ne segue che $f'(x) \leq f'(y)$ per ogni $x \leq y$.

Viceversa, supponiamo che la funzione f' sia non decrescente e dimostriamo la (38) studiando la funzione differenza

$$F(t) := tf(x) + (1 - t)f(y) - f(tx + (1 - t)y), \quad t \in [0, 1],$$

con x, y fissati. Consideriamo il caso $y < x$ (l'altro è analogo). Calcolando la derivata di F e applicando il Teorema di Lagrange, si deduce che esiste $\xi \in (y, x)$ tale che

$$F'(t) = f(x) - f(y) - f'(tx + (1 - t)y)(x - y) = [f'(\xi) - f'(tx + (1 - t)y)](x - y).$$

Dato che, per $t \in [0, 1]$, il punto $tx + (1 - t)y$ descrive l'intervallo $[y, x]$, esiste t^* tale che $t^*x + (1 - t^*)y = \xi$. Inoltre, dato che f' è non decrescente, $f'(tx + (1 - t)y) \leq f'(\xi)$ per $t \in [0, t^*]$ e $f'(tx + (1 - t)y) \geq f'(\xi)$ per $t \in [t^*, 1]$. Perciò:

$$F'(t) \geq 0 \quad \text{per } t \in [0, t^*] \quad \text{e} \quad F'(t) \leq 0 \quad \text{per } t \in [t^*, 1].$$

Perciò il minimo globale della funzione F è assunto in uno degli estremi $t = 0$ o $t = 1$ e, dato che $F(0) = F(1) = 0$, $F(t) \geq 0$ per ogni $t \in [0, 1]$, cioè la formula (38). \square

ESERCIZIO 4.6. Sia f una funzione convessa e derivabile in (a, b) . Se esistono $x_0, x_1 \in (a, b)$ con $x_0 \neq x_1$ tali che $f'(x_0) = f'(x_1) = 0$, che cosa si può dedurre sulla funzione f ?

DEFINIZIONE 4.7. Sia $f : (a, b) \rightarrow \mathbb{R}$ derivabile due volte. Se x_0 è tale che f'' cambia segno in x_0 (cioè è negativa da una parte e positiva dall'altra), allora x_0 si chiama punto di flesso della funzione f .

Grazie al Teorema 4.5, se f'' ha segno opposto alla destra e alla sinistra di x_0 , necessariamente x_0 è un punto di flesso.

ESEMPIO 4.8. Consideriamo la funzione $f(x) = \sin x$. La sua derivata seconda è $f''(x) = -\sin x$, quindi tutti punti della forma $x = k\pi$ per $k \in \mathbb{Z}$ sono punti di flesso.

Per la funzione $f(x) = \frac{1}{1+x^2}$, si ha

$$f'(x) = -\frac{2x}{(1+x^2)^2} \quad \Longrightarrow \quad f''(x) = \frac{2(3x^2-1)}{(1+x^2)^3},$$

e quindi i punti di flesso di f sono $x = \pm 1/\sqrt{3}$. La funzione f è convessa in $(-\infty, -1/\sqrt{3})$ e in $(1/\sqrt{3}, +\infty)$ e concava in $(-1/\sqrt{3}, 1/\sqrt{3})$.

La convessità è utile per determinare l'esistenza di minimi di una funzione. Infatti, vale la seguente implicazione

$$f \text{ convessa, } f'(x_0) = 0 \quad \Longrightarrow \quad x_0 \text{ punto di minimo globale.}$$

La dimostrazione è lasciata per esercizio. Analogamente, per le funzioni concave ed i punti di massimo. Chiaramente se la convessità è solo locale (cioè in un intorno del punto x_0), x_0 è punto di minimo locale.

ESERCIZIO 4.9. *Dimostrare la seguente implicazione*

$$f : [a, b] \rightarrow \mathbb{R} \text{ convessa} \quad \Rightarrow \quad f \text{ continua in } (a, b).$$

5. A caccia di massimi e minimi assoluti

Problema 1. Abbiamo già considerato il problema di determinare il cilindro di volume $V = k = \text{costante}$ con superficie totale S minima, con l'obiettivo (malcelato) di diventare ricchi grazie all'uso della matematica, applicando il risultato alla costruzione di scatole di fagioli, o, più in generale, di confezioni cilindriche con minima spesa di materiali. La speranza si era presto infranta quando ci siamo resi conto che per via elementare non riuscivamo a determinare il minimo della funzione S , cioè a risolvere il problema ($r = \text{raggio della base del cilindro}$)

$$\text{determinare il minimo di } S(r) = 2\pi \left(r^2 + \frac{k}{\pi r} \right) \quad r > 0.$$

Torniamo al problema con la conoscenza delle derivate e studiamo la monotonìa di S :

$$\frac{dS}{dr} = 2\pi \left(2r - \frac{k}{\pi r^2} \right) = \frac{4\pi}{r^2} \left(r^3 - \frac{k}{2\pi} \right).$$

Perciò $S'(r) \geq 0$ se e solo se $r \geq r^*$ dove $r^* = (k/2\pi)^{1/3}$. Quindi la funzione S è decrescente in $(0, r^*)$ ed è crescente in (r^*, ∞) . Ne segue che il punto di minimo richiesto esiste ed è proprio $r = r^*$ (Fig.8(a)). Problema risolto, corriamo in fabbrica!

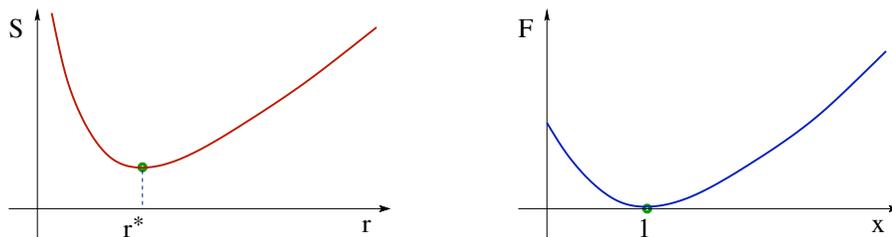


FIGURA 8. (a) Il grafico della funzione $S(r) = 2\pi \left(r^2 + \frac{k}{\pi r}\right)$; (b) il grafico della funzione $F(x) = x^p - 1 - p(x - 1)$, $p > 1$.

Problema 2. Vogliamo dimostrare la disequazione

$$x^p - 1 \geq p(x - 1) \quad \forall p > 1, \quad \forall x \geq 0.$$

Fissiamo $p > 1$ e consideriamo la funzione

$$F(x) = x^p - 1 - p(x - 1) \quad x \geq 0.$$

Dato che $F'(x) = p(x^{p-1} - 1)$, $F'(x) < 0$ per $x \in (0, 1)$ e $F'(x) > 0$ per $x \in (1, +\infty)$. Quindi la funzione F è decrescente in $[0, 1)$ e crescente in $(1, +\infty)$ e $x = 1$ è un punto di minimo. Ne segue che $F(x) \geq F(1) = 0$, da cui la conclusione (Fig.8(b)).

Problema 3. Siano $a_1, a_2, \dots, a_n \in \mathbb{R}$ assegnati. Supponiamo di voler determinare $x \in \mathbb{R}$ tale che sia minima la quantità

$$(40) \quad \sum_{i=1}^n (a_i - x)^2.$$

Possiamo immaginare che i valori a_i provengano da misurazioni di un fenomeno sotto osservazione e che si stia cercando un valore medio per questi numeri, che minimizzi l'errore commesso misurato dal valore in (40). Consideriamo la funzione $F(x) = \sum_{i=1}^n (a_i - x)^2$ e calcoliamone la derivata:

$$F'(x) = -2 \sum_{i=1}^n (a_i - x) = -2 \left[\sum_{i=1}^n a_i - \sum_{i=1}^n x \right] = 2n \left[x - \frac{1}{n} \sum_{i=1}^n a_i \right].$$

La funzione F è decrescente a sinistra di $\frac{1}{n} \sum_{i=1}^n a_i$ e crescente a destra. Il valore

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

è il punto di minimo (e coincide con la media aritmetica di a_1, \dots, a_n).

ESERCIZIO 5.1. Dati $a_1, a_2, \dots, a_n \in \mathbb{R}$, determinare x che minimizzi $\sum_{i=1}^n \lambda_i (a_i - x)^2$ dove $\lambda_1, \dots, \lambda_n > 0$ sono pesi (positivi) assegnati.

I problemi che abbiamo appena presentato mostrano alcune tra le miriadi di situazioni in cui si pone il problema: data una funzione f , come determinarne massimo e minimo globali (qualora esistano)? Proviamo ad affrontare il problema in generale.

Supponiamo di lavorare con una funzione f definita nell'intervallo $[a, b]$ e continua. Grazie al teorema di Weierstrass, l'ipotesi di continuità garantisce l'esistenza del massimo e del minimo assoluti. Abbiamo già visto che le soluzioni di $f'(x) = 0$ (cioè i punti critici di f) permettono di determinare i possibili candidati a punti di minimo o massimo locale *interno derivabile*. Chiaramente, è possibile che un estremo locale cada in un punto in cui la funzione non è derivabile. Quindi, la strategia per individuare il massimo ed il minimo di una funzione continua in $[a, b]$ è la seguente:

- ★ determinare l'insieme \mathcal{S} dei punti stazionari in (a, b) ;
- ★ determinare l'eventuale insieme \mathcal{N} dei punti in cui f non è derivabile;
- ★ calcolare la funzione in \mathcal{S} , in \mathcal{N} e negli estremi dell'intervallo a e b .
- ★ individuare il più grande e il più piccolo tra i valori calcolati.

ESERCIZIO 5.2. Determinare il massimo ed il minimo assoluti di

$$f(x) = (x^2 - 5x + 7)e^x \quad x \in [0, 2].$$

Soluzione. La funzione f è derivabile dappertutto. Per determinare i punti singolari:

$$f'(x) = (2x - 5)e^x + (x^2 - 5x + 7)e^x = (x^2 - 3x + 2)e^x = (x - 2)(x - 1)e^x.$$

Quindi $f'(x) = 0$ se e solo se $x = 1$ o $x = 2$. L'insieme dei punti critici *interni* è $\mathcal{S} = \{1\}$. Dato che $f(0) = 7 < f(2) = e^2 < f(1) = 3e$, si ha

$$\min_{x \in [0, 2]} f(x) = f(0) = 7, \quad \max_{x \in [0, 2]} f(x) = f(1) = 3e,$$

che è quanto richiesto dall'esercizio.

Spesso è utile conoscere il massimo del modulo di una funzione assegnata f , cioè risolvere il problema

$$\text{data } f : [a, b] \rightarrow \mathbb{R} \text{ continua, calcolare } \max_{x \in [a, b]} |f(x)|.$$

In questo caso, si può procedere come detto sopra, o, alternativamente, determinare il massimo ed il minimo della funzione f in $[a, b]$ e poi sfruttare la relazione (evidente?)

$$\max_{x \in [a, b]} |f(x)| = \max \left\{ \left| \max_{x \in [a, b]} f(x) \right|, \left| \min_{x \in [a, b]} f(x) \right| \right\}.$$

ESERCIZIO 5.3. Calcolare $\max\{|x^2 - 1| : x \in [-1, 2]\}$.

Nel caso in cui si studi una funzione f continua, ma definita su un dominio illimitato (ad esempio, $f : [a, +\infty) \rightarrow \mathbb{R}$), le ipotesi del Teorema di Weierstrass non sono soddisfatte e quindi non è detto che esistano il massimo ed il minimo della funzione. Comunque ha senso domandarsi: *quanto valgono l'estremo superiore e l'estremo inferiore? Nel caso in cui siano finiti, si tratta di massimo o di minimo?* La strategia per risolvere questo problema è simile a quanto appena visto. Il punto che bisogna modificare è quello relativo al calcolo della funzione negli estremi dell'intervallo. In questo caso almeno uno degli estremi dell'intervallo sarà $+\infty$ o $-\infty$ e le espressioni $f(+\infty)$ e $f(-\infty)$, in generale, non hanno senso, ma vanno sostituite con $\lim_{x \rightarrow \pm\infty} f(x)$. Vediamo la procedura negli esercizi che seguono.

ESERCIZIO 5.4. *Determinare l'estremo superiore e l'estremo inferiore della funzione $f(x) = e^{x^2}$ in \mathbb{R} e dire se si tratta di massimo e minimo.*

Soluzione. L'estremo superiore è presto detto: dato che $\lim_{x \rightarrow +\infty} f(x) = +\infty$, è chiaro che $\sup f = +\infty$. Che possiamo dire sull'estremo inferiore? Visto che la funzione f è continua su \mathbb{R} , esistono massimo e minimo di f in $[-M, M]$ per ogni scelta di $M > 0$. Quindi

$$\inf_{x \in \mathbb{R}} f(x) = \min \left\{ \min_{|x| \leq M} f(x), \inf_{|x| > M} f(x) \right\}.$$

Inoltre, dato che $f(x) \rightarrow +\infty$ per $x \rightarrow \pm\infty$, per M grande, $\min_{|x| \leq M} f(x) \leq \inf_{|x| > M} f(x)$. Non resta che cercare i punti di minimo relativo in $[-M, M]$. Derivando, $f'(x) = 2xe^{x^2}$ che si azzerava se e solo se $x = 0$, quindi l'unico punto di minimo relativo è $x = 0$ che, per quanto detto, è anche minimo assoluto: $\min_{x \in \mathbb{R}} e^{x^2} = e^0 = 1$.

ESERCIZIO 5.5. *Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua tale che*

$$\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow -\infty} f(x) = +\infty.$$

Dimostrare che la funzione f ammette minimo assoluto su \mathbb{R} .

ESERCIZIO 5.6. *Determinare l'estremo superiore e l'estremo inferiore della funzione $f(x) = e^{-x^2}$ in \mathbb{R} e dire se si tratta di massimo e minimo.*

Soluzione. La funzione è derivabile su tutto \mathbb{R} e la derivata vale $f'(x) = -2xe^{-x^2}$. Quindi c'è un unico punto critico $x = 0$ in cui la funzione vale $f(0) = 1$. Inoltre

$$\lim_{x \rightarrow -\infty} e^{-x^2} = \lim_{x \rightarrow +\infty} e^{-x^2} = 0.$$

Confrontando i valori deduciamo che

$$\inf_{x \in \mathbb{R}} e^{-x^2} = 0 \quad \sup_{x \in \mathbb{R}} e^{-x^2} = f(0) = 1.$$

Dato che l'estremo superiore fa parte dell'insieme immagine, l'estremo superiore è massimo. Invece l'estremo inferiore non è minimo, perchè la funzione f è strettamente positiva.

Analogamente nel caso di funzioni continue definite in insiemi non chiusi, cioè $f : (a, b) \rightarrow \mathbb{R}$ oppure $f : [a, b) \rightarrow \mathbb{R}$, o varianti, non si applica il Teorema di Weierstrass. Anche in questi casi, per determinare l'estremo superiore/inferiore bisogna considerare i limiti agli estremi.

Concludiamo la Sezione, analizzando altri due problemi di massimo e minimo.

Un problema di statica: la puleggia di De L'Hôpital. Consideriamo gli assi cartesiani (x, y) posti in modo che l'asse y sia in verticale rispetto al suolo e che la forza di gravità sia direzionata nel verso delle y negative. Indichiamo con $O = (0, 0)$ e con $A = (a, 0)$ dove $a > 0$ è una lunghezza fissata. Nel punto O , fissiamo una corda di lunghezza b e all'estremità B di questa corda fissiamo una puleggia. Lasciamo pendere per fatti suoi la puleggia e fissiamo una seconda corda di lunghezza ℓ al punto A . Dopo aver fatto passare la seconda corda attraverso la puleggia (quindi facendola passare per il punto B) fissiamo un peso M all'altra estremità. La configurazione finale è disegnata in Figura 9. Il problema è: in quale punto (x, y) si posizionerà il peso M ?

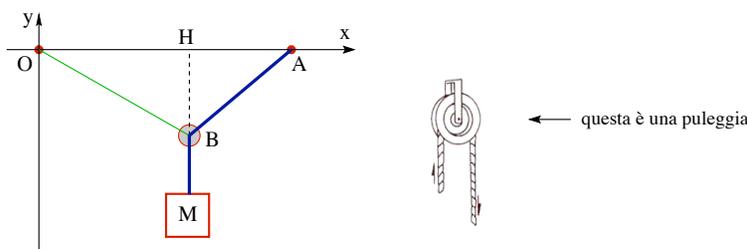


FIGURA 9. La puleggia di De L'Hôpital.

Qui stiamo supponendo che l'unica forza esterna che agisce sul sistema è la forza di gravità. Se, inoltre, supponiamo che il peso delle corde e della puleggia sia trascurabile rispetto al peso di M , e che non sia presente nessun tipo di attrito, il peso M si collocherà nella posizione più bassa possibile, cioè minimizzerà il valore y . Quest'affermazione discende dal principio seguente: *la posizione d'equilibrio (stabile) del sistema minimizza l'energia potenziale di M* . Dato che l'energia potenziale di M è della forma $Ay + B$ con $A > 0$, minimizzare l'energia potenziale equivale a minimizzare l'altezza y .

Riscriviamo per chiarezza i dati del problema:

$$\overline{OA} = a, \quad \overline{OB} = b, \quad \overline{AB} + \overline{BM} = \ell.$$

L'incognita è la posizione di $M = (x, y)$. Supponiamo inoltre $0 < b < a < \ell$. La soluzione del problema può essere divisa in due passi:

passo 1. assegnata la coordinata x di M , determinare la coordinata $y = f(x)$;

passo 2. calcolare (se esiste) il minimo della funzione f .

Da brave persone ordinate, partiamo dal primo passo. Indichiamo con H la proiezione ortogonale di M sull'asse x , cioè $H = (x, 0)$. Allora

$$y = -(\overline{HB} + \overline{BM}) = -(\overline{HB} + \ell - \overline{AB}) = \overline{AB} - \overline{HB} - \ell.$$

Si tratta ora di scrivere le lunghezze \overline{AB} e \overline{HB} in funzione di x . Basta usare il Teorema di Pitagora per ottenere:

$$\overline{HB} = \sqrt{\overline{OB}^2 - \overline{OH}^2} = \sqrt{b^2 - x^2}, \quad \overline{AB} = \sqrt{\overline{HB}^2 + \overline{HA}^2} = \sqrt{b^2 - x^2 + (a - x)^2}$$

Quindi la funzione da studiare è

$$y = f(x) = \sqrt{b^2 - x^2 + (a - x)^2} - \sqrt{b^2 - x^2} - \ell \quad x \in [0, b].$$

L'intervallo di variazione di x si deduce direttamente dal problema considerato.

Secondo passo: qual è la scelta di x che minimizza y ? Dato che la funzione è continua in $[0, b]$ e l'intervallo è chiuso e limitato, il Teorema di Weierstrass ci assicura che il problema ha soluzione, ma non ci dà nessuna informazione su quale sia il punto di minimo. Implementiamo, quindi, la strategia proposta poche pagine fa.

Prima di tutto, notiamo che la funzione f non è derivabile nei punti in cui si azzerava l'argomento di una delle due radici. Dato che $b < a$, il termine $b^2 - x^2 + (a - x)^2$ è sempre non nullo. La seconda radice $\sqrt{b^2 - x^2}$ si azzerava per $x = \pm b$. Di questi due punti, $-b$ va scartato perché è fuori da $[0, b]$ e b è uno degli estremi dell'intervallo.

Determiniamo l'insieme \mathcal{S} dei punti stazionari di f in $(0, b)$. Dato che

$$f'(x) = -\frac{a}{\sqrt{b^2 - x^2 + (a - x)^2}} + \frac{x}{\sqrt{b^2 - x^2}}$$

i punti stazionari $x \in (0, b)$ verificano

$$\frac{a}{\sqrt{b^2 - x^2 + (a - x)^2}} = \frac{x}{\sqrt{b^2 - x^2}},$$

e, passando ai quadrati,

$$\frac{a^2}{b^2 + a^2 - 2ax} = \frac{x^2}{b^2 - x^2}.$$

Ora ci vogliono un po' di conti: l'equazione precedente è equivalente a

$$2ax^3 - 2a^2x^2 - b^2x^2 + a^2b^2 = 0 \quad \iff \quad 2ax^2(x - a) - b^2(x + a)(x - a) = 0.$$

Dividendo per $x - a \neq 0$,

$$2ax^2 - b^2x - ab^2 = 0 \quad \iff \quad x = x_{\pm} := \frac{b^2 \pm \sqrt{b^4 + 8a^2b^2}}{4a}$$

Dato che $x_- < 0 < x_+ < b$ (verificare!), c'è un unico punto critico in $(0, b)$: $\mathcal{S} = \{x_+\}$.

Ricapitolando, la nostra strategia propone tre punti di minimo assoluto possibili: $0, x_+, b$. Per determinare chi di questi sia il punto di minimo bisognerebbe confrontare i tre valori $f(0), f(x_+)$ e $f(b)$. Fattibile, ma non particolarmente semplice. Seguiamo una strada diversa. Dato che

$$f'(0) = -\frac{a}{a-b} < 0$$

il punto 0 non può essere di minimo relativo e dunque nemmeno di minimo assoluto! La lotta rimane tra x_+ e b . In b non possiamo ragionare come in 0 dato che la funzione f non è derivabile in b , quindi $f'(b)$ non ha senso. Non perdiamoci d'animo: dato che

$$\lim_{x \rightarrow b^-} f'(x) = \lim_{x \rightarrow b^-} -\frac{a}{\sqrt{b^2 - x^2} + (a-x)} + \frac{x}{\sqrt{b^2 - x^2}} = +\infty$$

la funzione f è crescente in un intorno (sinistro) di b , quindi nemmeno b può essere il punto di minimo richiesto. Perfetto: resta un unico sopravvissuto x_+ che è il punto di

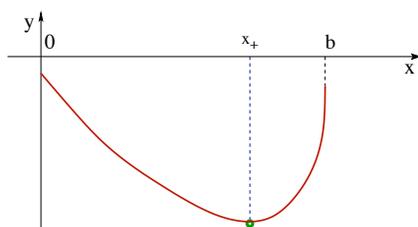


FIGURA 10. Il grafico della funzione f .

minimo cercato. Il peso M si collocherà nella posizione di coordinate $(x_+, f(x_+))$. In alternativa, per stabilire che x_+ è il punto di minimo assoluto, si sarebbe potuto anche notare che $f'' \geq 0$, quindi la funzione f è convessa e, necessariamente, il suo punto critico x_+ è di minimo assoluto.

Il principio di Fermat. Passiamo ora a considerare due problemi di *ottica geometrica* che si traducono nella ricerca del punto di minimo assoluto di certe funzioni. In quel che segue, considereremo un raggio di luce in maniera naïf: come un qualcosa che viaggia da un punto ad un altro, si riflette sugli oggetti, entra nell'occhio...

Il problema fondamentale è stabilire quale sia il tragitto percorso dal raggio per passare da un punto A ad un punto B . Il principio, proposto da Fermat, è il seguente: *il tragitto prescelto è quello che minimizza il tempo di percorrenza*. Se il raggio viaggia sempre nello stesso mezzo, la sua velocità v è costante, quindi minimizzare il tempo di percorrenza T equivale a minimizzare la lunghezza del percorso. Perciò il raggio percorre linee rette. Che succede in situazioni un po' più complicate?

Riflessione. Consideriamo un raggio di luce che parta dal punto $A = (0, a)$ e che si diriga verso l'asse delle x dove immaginiamo collocato uno specchio. Il raggio viene riflesso nel punto $P = (x, 0)$ e da lì arriva nel punto $B = (b, c)$. I tragitti da A a P e da P a B sono percorsi lungo segmenti. Supponendo assegnati $a, b, c > 0$ e quindi assegnati i punti A e B , qual'è il punto P prescelto dal raggio luminoso?

Ad ogni scelta di $P = (x, 0)$, corrisponde un certo tempo di percorrenza $T = T(x)$. Il principio di Fermat afferma che il punto di riflessione $(x_0, 0)$ è tale che $T(x_0) = \min T(x)$. Bene, non resta che determinare l'espressione esplicita di $T = T(x)$ e trovare in quale punto sia assunto il minimo. Indicando con T_{AP} e T_{PB} , il tempo impiegato dal raggio per andare da A a P e da P a B rispettivamente, e con v la velocità della luce nel mezzo in considerazione,

$$T(x) = T_{AP}(x) + T_{PB}(x) = \frac{\overline{AP}}{v} + \frac{\overline{PB}}{v}.$$

Quindi, grazie al teorema di Pitagora,

$$T(x) = \frac{1}{v} \left(\sqrt{x^2 + a^2} + \sqrt{(b-x)^2 + c^2} \right) \quad x \in [0, b].$$

Dato che la funzione T è derivabile infinite volte, applicando la strategia per il calcolo di massimi e minimi assoluti, il punto di minimo x_0 sarà o 0 , o b , o tale che $T'(x_0) = 0$. Cerchiamo quindi i punti critici di T . La derivata prima T' è esplicitamente data da

$$T'(x) = \frac{1}{v} \left(\frac{x}{\sqrt{x^2 + a^2}} - \frac{b-x}{\sqrt{(b-x)^2 + c^2}} \right).$$

Essa si azzerava se e solo se $x = x_* := ab/(a+c)$.

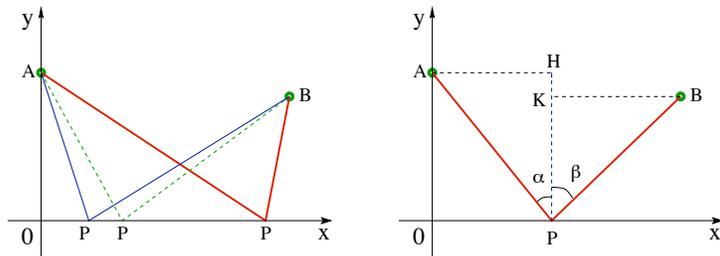


FIGURA 11. Quale sarà il punto P prescelto da un raggio riflesso in uno specchio? Quello che fa in modo che gli angoli di incidenza α e di riflessione β coincidano.

Invece di confrontare i valori di T per $x = 0, x_*, b$, si può notare che

$$T'(0) = -\frac{b}{v\sqrt{b^2 + c^2}} < 0 \quad \text{e} \quad T'(b) = \frac{b}{v\sqrt{b^2 + a^2}} > 0,$$

quindi nessuno dei due estremi dell'intervallo è di minimo. Pertanto il punto x_* è il punto di minimo assoluto.

Il punto di riflessione P è individuato da una condizione geometrica semplice. Sia α l'angolo, detto di *incidenza*, determinato dal segmento AP e dalla semiretta da P , perpendicolare all'asse x , contenuta nel semipiano $y > 0$, e sia β l'angolo, detto di *riflessione*, determinato dal segmento PB e dalla stessa semiretta di prima. Allora, indicando con H il punto di coordinate (x_*, a) e con K il punto di coordinate (x_*, c) ,

$$\tan \alpha = \frac{\overline{AH}}{\overline{PH}} = \frac{x_*}{a} = \frac{b}{a+c} \quad \tan \beta = \frac{\overline{BK}}{\overline{PK}} = \frac{b-x_*}{c} = \frac{ab+bc-ab}{a+c} \frac{1}{c} = \frac{b}{a+c}$$

Quindi $\tan \alpha = \tan \beta$ e, dato che $\alpha, \beta \in [0, \pi/2]$, ne segue che $\alpha = \beta$. In definitiva, il principio di Fermat implica che *l'angolo di riflessione coincide con l'angolo di incidenza*.

Rifrazione. Cambiamo tipo di esperimento. Consideriamo un raggio luminoso che viaggi in due mezzi differenti in cui la sua velocità è v_+ e v_- . Per semplicità, supponiamo che il mezzo in cui la velocità è v_+ corrisponda alla regione di piano con $y > 0$ e quello in cui la velocità è v_- corrisponda a $y < 0$. Se un raggio parte da $A = (0, a)$ con $a > 0$ ed arriva a $B = (b, c)$ con $c < 0 < b$, che tragitto sceglie?

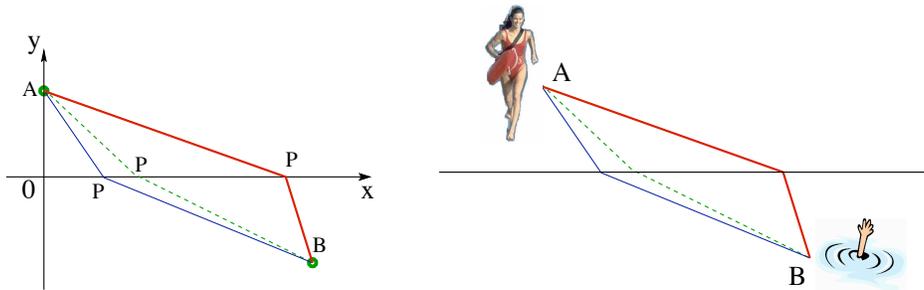


FIGURA 12. Quale percorso scegliere? Supponiamo che nel semipiano $y > 0$ ci sia la terra ferma e il mare nel semipiano $y < 0$. Nel punto A c'è la prestante bagnina di Baywatch pronta ad intervenire per salvare la vita di un affogando sito nel punto B . Sapendo che la soccorritrice quando corre sulla spiaggia va a velocità v_+ e quando nuota in mare va a velocità v_- , qual'è il percorso che le permette di soccorrere il malcapito nel minor tempo possibile?

Esattamente come prima, utilizziamo il principio di Fermat. Indicando con T_{AP} e T_{PB} , il tempo impiegato dal raggio per andare da A a P e da P a B rispettivamente, il tempo impiegato per andare da A a B è

$$T(x) = T_{AP}(x) + T_{PB}(x) = \frac{\overline{AP}}{v_+} + \frac{\overline{PB}}{v_-}.$$

e, di nuovo per il teorema di Pitagora,

$$T(x) = \frac{\sqrt{x^2 + a^2}}{v_+} + \frac{\sqrt{(b-x)^2 + c^2}}{v_-} \quad x \in [0, b].$$

Anche questa funzione è derivabile infinite volte in $[0, b]$. La sua derivata prima è

$$T'(x) = \frac{x}{v_+ \sqrt{x^2 + a^2}} - \frac{b-x}{v_- \sqrt{(b-x)^2 + c^2}}.$$

Quindi $T'(0) < 0 < T'(b)$ e anche in questo caso gli estremi non sono punti di minimo. Perciò il minimo è in $(0, b)$. Inoltre, con un po' di pazienza, si ottiene

$$T''(x) = \frac{a^2}{v_+ (x^2 + a^2)^{3/2}} + \frac{c^2}{v_- [(b-x)^2 + c^2]^{3/2}} > 0$$

Dato che la derivata seconda è (strettamente) positiva, la derivata prima T' è strettamente crescente, quindi esiste un unico x_* tale che $T'(x_*) = 0$ (si ricordi che $T'(0) < 0 < T'(b)$). Necessariamente x_* è il punto di minimo che andiamo cercando.

Come individuarlo? Per ora sappiamo solo che x_* è individuato in maniera univoca dalla relazione $T'(x_*) = 0$, cioè x_* è l'unico valore per cui

$$(41) \quad \frac{1}{v_+} \frac{x_*}{\sqrt{x_*^2 + a^2}} = \frac{1}{v_-} \frac{b-x_*}{\sqrt{(b-x_*)^2 + c^2}}.$$

Come nel caso della riflessione, ragioniamo in termini di angoli. Sia α , *angolo di*

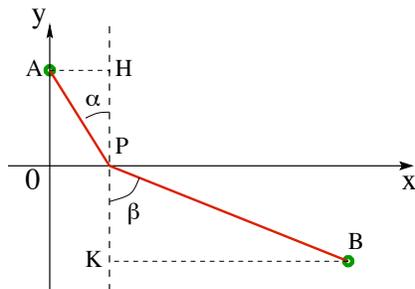


FIGURA 13. Gli angoli di incidenza e di rifrazione.

incidenza, l'angolo formato dal segmento AP con la semiretta verticale per P contenuta in $y > 0$, e sia β , *angolo di rifrazione*, l'angolo formato dal segmento PB con la semiretta verticale per P contenuta in $y < 0$. Se H e K sono le proiezioni di A e B sulla retta $x = x_*$, i triangoli APH e BPK sono rettangoli e quindi

$$\sin \alpha = \frac{\overline{AH}}{\overline{AP}} = \frac{x_*}{\sqrt{x_*^2 + a^2}} \quad \sin \beta = \frac{\overline{BK}}{\overline{BP}} = \frac{b-x_*}{\sqrt{(b-x_*)^2 + c^2}}.$$

Sostituendo nella formula (41), si deduce che il punto di rifrazione P è scelto in modo che valga

$$\frac{\sin \alpha}{v_+} = \frac{\sin \beta}{v_-}.$$

Tale relazione è nota come *legge di rifrazione di Snell*. .

Ordini di grandezza e la formula di Taylor

1. Verso lo zero e ad un passo dall'infinito

Le funzioni con cui ci si trova a dover lavorare possono avere una struttura complicata, e anche molto. Se si è interessati al comportamento della funzione solo per determinati regimi, cioè per valori dell'incognita in opportune regioni, può bastare conoscere quali siano i termini dominanti all'interno della funzione. Ad esempio, se il valore $f(t)$ rappresenta la posizione di una particella all'istante t , si potrebbe essere interessati solo al comportamento della funzione f per valori grandi di t . Se $f(t) = e^t + \sin t$, è chiaro che saremo soddisfatti di un'approssimazione del tipo $f(t) \approx e^t$ per $t \rightarrow +\infty$, dato che questo termine diverge a $+\infty$, mentre l'altro rimane limitato. Ma se invece $f(t) = e^t + t$? Anche qui l'approssimazione sensata, per $t \rightarrow +\infty$, è $f(t) \approx e^t$, dato che l'esponenziale cresce più rapidamente del termine di primo grado t . Come formalizzare in modo preciso la frase “cresce ben più rapidamente”?

Lo stesso tipo di problema sorge nel caso di quantità infinitesime. Come confrontare termini che diventano molto piccoli (tendenti a zero)?

Ordine di infinito per $x \rightarrow +\infty$. Consideriamo qui funzioni f tali che

$$\lim_{x \rightarrow +\infty} |f(x)| = +\infty$$

(il caso $x \rightarrow -\infty$ è analogo). Come distinguere tra funzioni di questo genere quelle che divergono “più rapidamente” e quelle che divergono “meno rapidamente”? Ad esempio le funzioni x^α , $\ln x$, e^x , a^x (con $\alpha > 0$ e $a > 1$) divergono per $x \rightarrow +\infty$ in modi essenzialmente differenti. Quale di queste funzioni cresce più rapidamente delle altre?

Dato che

x	1	10	100	1000
x^2	1	100	10000	1000000
x^3	1	1000	1000000000	1000000000000

ci aspettiamo che x^3 tenda all'infinito più rapidamente di x^2 . La maniera rigorosa per esprimere questo concetto è studiare il rapporto delle due quantità. Dato che

$$\lim_{x \rightarrow +\infty} \frac{x^3}{x^2} = +\infty,$$

la grandezza di x^3 relativamente a quella di x^2 è maggiore. Questa proprietà si esprime dicendo che x^3 *tende a $+\infty$ più rapidamente di x^2 per $x \rightarrow +\infty$.*

Allo stesso modo, x^α cresce più rapidamente di x^β per $\alpha > \beta$

$$\lim_{x \rightarrow +\infty} \frac{x^\alpha}{x^\beta} = +\infty \quad \forall \alpha > \beta.$$

DEFINIZIONE 1.1. Ordine di infinito I. *Siano f e g tali che*

$$(42) \quad \lim_{x \rightarrow +\infty} |f(x)| = \lim_{x \rightarrow +\infty} |g(x)| = +\infty.$$

Si dice che: f ha ordine di infinito superiore rispetto a g per $x \rightarrow +\infty$ se

$$\lim_{x \rightarrow +\infty} \frac{|f(x)|}{|g(x)|} = +\infty.$$

Analogamente, f ha ordine di infinito inferiore rispetto a g per $x \rightarrow +\infty$ se

$$\lim_{x \rightarrow +\infty} \frac{|f(x)|}{|g(x)|} = 0.$$

Può capitare che due funzioni abbiano lo stesso tipo di andamento all'infinito. Ad esempio: x e $2x$ sono entrambi polinomi di grado 1 e vale:

$$\lim_{x \rightarrow +\infty} \frac{x}{2x} = \frac{1}{2}.$$

Il fatto che il limite del rapporto sia una costante non zero, suggerisce che le due funzioni hanno lo stesso ordine di grandezza. Si sarebbe tentati di dire che due funzioni hanno lo stesso ordine di grandezza se e solo se

$$\lim_{x \rightarrow +\infty} \frac{|f(x)|}{|g(x)|} = \ell > 0.$$

Questa definizione, seppure ragionevole, è troppo restrittiva: non è in grado di coprire casi con termini oscillanti. Un esempio: per $f(x) = x(1 + \sin^2 x)$ e $g(x) = x$, si ha

$$\frac{|f(x)|}{|g(x)|} = \frac{|x||1 + \sin^2 x|}{|x|} = 1 + \sin^2 x \in [1, 2],$$

che non ammette limite per $x \rightarrow +\infty$. Ma, dato che $x \leq f(x) \leq 2x$, è ragionevole affermare che f tende all'infinito con la stessa velocità di x .

DEFINIZIONE 1.2. Ordine di infinito II. *Si dice che le funzioni f e g , soddisfacenti (42), hanno lo stesso ordine di infinito per $x \rightarrow +\infty$ se esistono $C_1, C_2 > 0$ tali che*

$$(43) \quad 0 < C_1 \leq \frac{|f(x)|}{|g(x)|} \leq C_2 \quad \text{per } x \text{ sufficientemente grande.}$$

In generale, per verificare la condizione (43) occorre stimare il rapporto $|f|/|g|$ e non sempre questa operazione è facile. Però, se esiste ed è diverso da zero il limite

$$\lim_{x \rightarrow +\infty} \frac{|f(x)|}{|g(x)|} = \ell \neq 0,$$

la condizione (43) è automaticamente soddisfatta. Ad esempio, consideriamo le funzioni $f(x) = 2x^2 - 1$ e $g(x) = x^2 + x + 3$:

$$\lim_{x \rightarrow +\infty} \frac{2x^2 - 1}{x^2 + x + 3} = 2 \neq 0,$$

quindi hanno lo stesso ordine di infinito. In generale, se f è un polinomio di grado m e g un polinomio di grado p , allora: se $m > p$, f è di ordine superiore a g ; se $m = p$, f e g sono dello stesso ordine; se $m < p$, f è di ordine inferiore a g .

OSSERVAZIONE 1.3. Se f è di ordine superiore rispetto a g , allora la funzione somma $f + g$ ha lo stesso ordine di f , infatti

$$\lim_{x \rightarrow +\infty} \frac{f(x) + g(x)}{f(x)} = \lim_{x \rightarrow +\infty} \left(1 + \frac{g(x)}{f(x)} \right) = 1.$$

Ad esempio, la funzione $x + \ln x$ ha lo stesso ordine di infinito di x per $x \rightarrow +\infty$

$$\lim_{x \rightarrow +\infty} \frac{x + \ln x}{x} = \lim_{x \rightarrow +\infty} \left(1 + \frac{\ln x}{x} \right) = 1.$$

DEFINIZIONE 1.4. Se per una funzione f esiste un valore $\alpha > 0$ tale che f è dello stesso ordine di infinito di $|x|^\alpha$ per $x \rightarrow +\infty$ si dice che f è un infinito di ordine α . La funzione $|x|$ è detta infinito campione per $x \rightarrow +\infty$.

Ad esempio, la funzione $\sqrt{1+x^2}$ è tale che

$$\lim_{x \rightarrow +\infty} \frac{\sqrt{1+x^2}}{x} = \lim_{x \rightarrow +\infty} \sqrt{\frac{1}{x^2} + 1} = 1,$$

quindi ha ordine di infinito uguale ad 1 per $x \rightarrow +\infty$. La funzione $\frac{4x^3+1}{x-5}$ è tale che

$$\lim_{x \rightarrow +\infty} \frac{(4x^3+1)/(x-5)}{x^2} = \lim_{x \rightarrow +\infty} \frac{4x^3+1}{x^3-5x^2} = 4$$

quindi ha ordine di infinito 2. In generale, l'ordine di infinito di una funzione razionale con numeratore di grado m e denominatore di grado p (con $m > p$) è $m - p$ (dimostratelo!).

OSSERVAZIONE 1.5. Se una funzione f ha ordine di infinito α , allora

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{x^{\alpha+\varepsilon}} = \lim_{x \rightarrow +\infty} \frac{f(x)}{x^\alpha} \frac{1}{x^\varepsilon} = 0 \quad \forall \varepsilon > 0;$$

cioè ha ordine di infinito inferiore rispetto ad $x^{\alpha+\varepsilon}$ per ogni $\varepsilon > 0$. Analogamente

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{x^{\alpha-\varepsilon}} = \lim_{x \rightarrow +\infty} \frac{f(x)}{x^\alpha} x^\varepsilon = \infty \quad \forall \varepsilon > 0,$$

quindi ha ordine di infinito superiore rispetto ad $x^{\alpha-\varepsilon}$ per ogni $\varepsilon > 0$.

Si potrebbe pensare di introdurre una “scala assoluta” di ordine di grandezza delle funzioni, attribuendo a ciascuna funzione divergente la corrispondente potenza che la rappresenta. Questa scala, però, non adempie il compito richiesto: ci sono funzioni la cui “velocità di divergenza” non corrisponde a quella di nessuna potenza e quindi che, in questo senso, non possono essere classificate. I due casi più rilevanti sono dati dalla funzione $\ln x$ e da e^x per cui vale

$$(44) \quad \lim_{x \rightarrow +\infty} \frac{\ln x}{x^\alpha} = 0, \quad \lim_{x \rightarrow +\infty} \frac{e^x}{x^\alpha} = +\infty \quad \forall \alpha > 0.$$

Per dimostrare (44), osserviamo preliminarmente che vale la disequazione¹

$$(45) \quad \ln t \leq t \quad \forall t > 0.$$

Ora, dato $\alpha > 0$, scegliamo $\varepsilon \in (0, \alpha)$. Applicando la disequazione (45) per $t = x^{\alpha-\varepsilon}$ e usando le proprietà del logaritmo:

$$\ln x \leq \frac{1}{\alpha - \varepsilon} x^{\alpha-\varepsilon} \quad \forall x > 0, 0 < \varepsilon < \alpha.$$

Dividendo entrambi i membri per x^α e passando al limite,

$$0 \leq \frac{\ln x}{x^\alpha} \leq \frac{1}{(\alpha - \varepsilon)x^\varepsilon} \quad \forall x > 0 \quad \Rightarrow \quad \lim_{x \rightarrow +\infty} \frac{\ln x}{x^\alpha} = 0.$$

Per il secondo limite in (44), ponendo $y = e^x$,

$$\lim_{x \rightarrow +\infty} \frac{e^x}{x^\alpha} = \lim_{y \rightarrow +\infty} \frac{y}{(\ln y)^\alpha} = \lim_{y \rightarrow +\infty} \left(\frac{y^{1/\alpha}}{\ln y} \right)^\alpha = +\infty.$$

Le formule in (44) affermano che $\ln x$ diverge per $x \rightarrow +\infty$ più lentamente di qualsiasi potenza x^α e che e^x diverge più rapidamente di qualsiasi potenza x^α .

OSSERVAZIONE 1.6. Con le funzioni esponenziali e con i logaritmi, è possibile costruire funzioni che divergono a velocità sempre più grandi o a velocità sempre più piccole. Ad esempio,

$$\lim_{x \rightarrow +\infty} \frac{\ln(\ln x)}{\ln x} = \lim_{y \rightarrow +\infty} \frac{\ln y}{y} = 0, \quad \lim_{x \rightarrow +\infty} \frac{e^{(e^x)}}{e^x} = \lim_{x \rightarrow +\infty} \frac{e^y}{y} = +\infty,$$

e quindi $\ln(\ln x)$ è un infinito di ordine inferiore a $\ln x$ e $e^{(e^x)}$ è di ordine superiore a e^x .

¹Infatti, detta $F(t) = t - \ln t$, allora $F'(t) = 1 - 1/t$ e quindi F ha un punto di minimo assoluto per $t = 1$. Perciò $F(t) \geq F(1) = 1 > 0$.

OSSERVAZIONE 1.7. Cosa succede per a^x e $\log_a x$ con $a > 1$ (e diverso da e)? La funzione $\log_a x$ si può scrivere in termini della funzione $\ln x$:

$$\log_a x = \frac{\ln x}{\ln a}$$

(sapete giustificare questa formula?). Quindi

$$\lim_{x \rightarrow +\infty} \frac{\log_a x}{x^\alpha} = \lim_{x \rightarrow +\infty} \frac{\ln x}{x^\alpha \ln a} = 0 \quad \forall \alpha > 0.$$

Procedendo come nel passaggio dal limite riguardante il logaritmo naturale al limite per l'esponenziale con base e , deduciamo che

$$\lim_{x \rightarrow +\infty} \frac{a^x}{x^\alpha} = +\infty \quad \forall \alpha > 0, a > 1.$$

E' interessante confrontare tra loro esponenziali e logaritmi con basi diverse:

$$\lim_{x \rightarrow +\infty} \frac{\log_a x}{\log_b x} = \lim_{x \rightarrow +\infty} \frac{\ln x \ln b}{\ln a \ln x} = \frac{\ln b}{\ln a} \quad \forall a, b > 1,$$

quindi logaritmi con basi diverse hanno lo stesso ordine di infinito. Per gli esponenziali, invece, dati $a, b > 1$

$$\lim_{x \rightarrow +\infty} \frac{a^x}{b^x} = \lim_{x \rightarrow +\infty} \left(\frac{a}{b}\right)^x = \begin{cases} 0 & a < b, \\ 1 & a = b, \\ +\infty & b < a, \end{cases}$$

che mostra che esponenziali con base maggiore hanno ordine di infinito maggiore.

Ordine di infinito e comportamento asintotico. È possibile che per una funzione f valga una decomposizione del tipo

$$(46) \quad f(x) = g(x) + h(x) \quad \text{con} \quad \lim_{x \rightarrow +\infty} h(x) = 0,$$

dove la funzione g è una funzione "nota" (ad esempio, una funzione con un asintoto obliquo). Se la funzione $|g|$ diverge a $+\infty$ per $x \rightarrow +\infty$, allora

$$\lim_{x \rightarrow +\infty} |f(x)| = \lim_{x \rightarrow +\infty} |g(x) + h(x)| = +\infty.$$

Come sono collegati gli ordini di infinito di f e g ? La risposta è semplice: *le funzioni f e g hanno lo stesso ordine di infinito*, infatti

$$(47) \quad \lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow +\infty} \frac{g(x) + h(x)}{g(x)} = \lim_{x \rightarrow +\infty} \left(1 + \frac{h(x)}{g(x)}\right) = 1.$$

Ad esempio, tutte le funzioni che possiedono un asintoto obliquo (non orizzontale) per $x \rightarrow +\infty$ hanno ordine di infinito 1.

E' vero il viceversa? Supponiamo di sapere che una funzione f abbia lo stesso ordine di infinito di una funzione g :

$$\lim_{x \rightarrow +\infty} |f(x)| = \lim_{x \rightarrow +\infty} |g(x)| = +\infty \quad \text{e} \quad 0 < C_1 \leq \frac{|f(x)|}{|g(x)|} \leq C_2.$$

E' vero che vale una rappresentazione come quella data in (46)? La risposta, in generale, è negativa. Ad esempio, per le funzioni $f(x) = x + \ln x$ e $g(x) = x$

$$\lim_{x \rightarrow +\infty} \frac{x + \ln x}{x} = \lim_{x \rightarrow +\infty} 1 + \frac{\ln x}{x} = 1,$$

ma la differenza tra f e g se ne guarda bene dal tendere a zero per $x \rightarrow +\infty$

$$h(x) = f(x) - g(x) = (x + \ln x) - x = \ln x \quad \Rightarrow \quad \lim_{x \rightarrow +\infty} h(x) = +\infty.$$

Ordine di infinito per $x \rightarrow x_0$. Così come si confrontano i comportamenti delle funzioni per $x \rightarrow +\infty$, è possibile confrontare funzioni che divergono in un punto x_0 . La terminologia è analoga alla precedente.

DEFINIZIONE 1.8. Ordine di infinito III. Date due funzioni f e g tali che

$$\lim_{x \rightarrow x_0} |f(x)| = \lim_{x \rightarrow x_0} |g(x)| = +\infty,$$

si dice che f è un infinito di ordine superiore rispetto a g per $x \rightarrow x_0$ se

$$\lim_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} = +\infty.$$

Se il limite è 0, f è un infinito di ordine inferiore rispetto a g per $x \rightarrow x_0$. Infine, se esistono C_1, C_2 e $\delta > 0$ per cui

$$(48) \quad 0 < C_1 \leq \frac{|f(x)|}{|g(x)|} \leq C_2 \quad \text{per} \quad 0 < |x - x_0| < \delta$$

si dice che f e g hanno stesso ordine di infinito per $x \rightarrow x_0$.

Come nel caso di $x \rightarrow +\infty$, se $\lim_{x \rightarrow x_0} f(x)/g(x) = \ell \neq 0$, allora è automaticamente verificata la condizione (48) e, quindi, f e g sono dello stesso ordine. Ad esempio

$$\lim_{x \rightarrow 0} \left| \frac{1}{x} \right| = \lim_{x \rightarrow 0} \left| \frac{1}{\sin x} \right| = +\infty \quad \text{e} \quad \lim_{x \rightarrow 0} \frac{1/x}{1/\sin x} = \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1,$$

quindi $f(x) = 1/x$ e $g(x) = 1/\sin x$ sono "infiniti" dello stesso ordine per $x \rightarrow 0$.

DEFINIZIONE 1.9. Se la funzione f ha lo stesso ordine di infinito di $1/|x - x_0|^\alpha$ per qualche $\alpha > 0$, si dice che f è un infinito di ordine α per $x \rightarrow x_0$. La funzione $1/|x - x_0|$ è detta infinito campione per $x \rightarrow x_0$.

Anche qui si può ripetere quanto detto in precedenza: esistono funzioni che non hanno un ordine di infinito per $x \rightarrow x_0$. Un esempio? Con il cambio di variabile $y = -\ln x$, si deduce che

$$\lim_{x \rightarrow 0^+} \frac{\ln x}{1/x^\alpha} = \lim_{y \rightarrow +\infty} \frac{y}{e^{\alpha y}} = 0,$$

che mostra che $\ln x$ ha un ordine di infinito in 0 inferiore a qualsiasi potenza.

Ordine di infinitesimo. Così come si confrontano infiniti, è possibile confrontare funzioni infinitesime per $x \rightarrow x_0$ con $x_0 \in \mathbb{R}$ oppure per $x \rightarrow \pm\infty$. Qui, per abbreviare l'esposizione, scriviamo x_0 per indicare o un numero reale, o uno dei due simboli $\pm\infty$.

DEFINIZIONE 1.10. Ordine di infinitesimo. Siano f e g infinitesime per $x \rightarrow x_0$. Si dice che f è un infinitesimo di ordine superiore a g per $x \rightarrow x_0$ se

$$(49) \quad \lim_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} = 0.$$

Se il limite è $+\infty$, f è un infinitesimo di ordine inferiore a g . Infine, f e g hanno lo stesso ordine di infinitesimo se in un intorno di x_0 (nel caso di $x_0 = \pm\infty$ si intende per valori sufficientemente grandi),

$$0 < C_1 \leq \left| \frac{f(x)}{g(x)} \right| \leq C_2 \quad \text{per qualche } C_1, C_2 > 0.$$

Come nel caso degli infiniti, si introducono infinitesimi campione.

- Se $x_0 \in \mathbb{R}$, si dice che f è un infinitesimo di ordine α , se è dello stesso ordine di $|x - x_0|^\alpha$. La funzione $|x - x_0|^\alpha$ è l'infinitesimo campione per $x \rightarrow x_0$.
- Se $x_0 = \pm\infty$, si dice che f è un infinitesimo di ordine α , se è dello stesso ordine di $1/|x|^\alpha$. La funzione $1/|x|^\alpha$ è l'infinitesimo campione per $x \rightarrow \pm\infty$.

Qualche esempio (tanto per gradire):

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1 \quad \Rightarrow \quad \sin x \text{ è infinitesimo di ordine 1 per } x \rightarrow 0$$

$$\lim_{x \rightarrow \pm\infty} \frac{1/(1+x^2)}{1/x^2} = 1 \quad \Rightarrow \quad \frac{1}{1+x^2} \text{ è infinitesimo di ordine 2 per } x \rightarrow \pm\infty$$

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \frac{1}{2} \quad \Rightarrow \quad 1 - \cos x \text{ è infinitesimo di ordine 2 per } x \rightarrow 0$$

$$\lim_{x \rightarrow \pm\infty} \frac{\arctan(1/x)}{1/x} = 1 \quad \Rightarrow \quad \arctan\left(\frac{1}{x}\right) \text{ è infinitesimo di ordine 1 per } x \rightarrow \pm\infty$$

I simboli di Landau: “O” e “o”. Per indicare che una funzione f ha un ordine di grandezza inferiore a quello di un'altra funzione g si usa la notazione $f = o(g)$ per $x \rightarrow x_0$ (si legge f è un “o piccolo” di g). Il significato di questa affermazione è che f/g tende a zero per $x \rightarrow x_0$. Ad esempio,

$$x^\alpha = o(x^\beta) \quad \forall \alpha < \beta \quad \text{per } x \rightarrow +\infty.$$

Abbiamo anche visto che

$$\begin{aligned} \ln x &= o(x^\alpha) & \forall \alpha > 0 & \quad \text{per } x \rightarrow +\infty, \\ x^\alpha &= o(e^x) & \forall \alpha > 0 & \quad \text{per } x \rightarrow +\infty, \\ 1 - \cos x &= o(x) & & \quad \text{per } x \rightarrow 0. \end{aligned}$$

ESERCIZIO 1.11. Verificare la validità di

$$\frac{1}{\sqrt{1+4x^2}} = \frac{1}{2x} + o\left(\frac{1}{x}\right) \quad \text{per } x \rightarrow +\infty.$$

Analogamente si introduce la notazione $f = O(g)$ (si legge f è un “O grande” di g) per indicare che la funzione f ha al più l'ordine di grandezza di g , ossia se il rapporto f/g è limitato in un intorno di x_0

$$\left| \frac{f(x)}{g(x)} \right| \leq C,$$

per qualche $C > 0$, in un intorno di x_0 . Ad esempio,

$$\sqrt{10x-1} = O(\sqrt{x}) \quad \text{per } x \rightarrow +\infty.$$

Infatti, dato che

$$\lim_{x \rightarrow +\infty} \frac{\sqrt{10x-1}}{\sqrt{x}} = \lim_{x \rightarrow +\infty} \sqrt{10 - \frac{1}{x}} = \sqrt{10},$$

il rapporto di $\frac{\sqrt{10x-1}}{\sqrt{x}}$ è limitato per valori della x sufficientemente grandi.

Dai limiti notevoli

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1, \quad \lim_{x \rightarrow 0} \frac{\ln(1+x)}{x} = 1, \quad \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1, \quad \lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \frac{1}{2},$$

segue che, per $x \rightarrow 0$,

$$e^x = 1 + O(x), \quad \ln(1+x) = O(x), \quad \sin x = O(x), \quad \cos x = 1 + O(x^2).$$

Derivabilità con i simboli di Landau. Tramite i simboli di Landau si può riscrivere la derivabilità di una funzione in modo diverso. La derivabilità di una funzione f in x può essere scritta nella forma

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x) - f'(x)h}{h} = 0.$$

Quindi, se una funzione è derivabile in x , vale la relazione (particolarmente significativa)

$$(50) \quad f(x+h) - f(x) = f'(x)h + o(h) \quad \text{per } x \rightarrow 0.$$

L'interpretazione di questa formula è che l'incremento di f si può rappresentare come un termine $f'(x)h$, lineare nell'incremento h , più un resto che ha un ordine di grandezza inferiore ad h per $h \rightarrow 0$. Si usa la terminologia:

$$\text{differenziale di } f: \quad df(x; h) := f'(x)h.$$

Fissato il valore di x , il differenziale $df(x; h)$ rappresenta un'approssimazione (valida a meno di $o(h)$) dell'incremento $\Delta f(x; h) := f(x+h) - f(x)$:

$$\Delta f(x; h) \approx df(x; h) \quad \text{per } h \rightarrow 0.$$

Come si precisa il senso del simbolo \approx ? Proprio tramite i simboli di Landau:

$$|\Delta f(x; h) - df(x; h)| = o(h) \quad \text{per } h \rightarrow 0,$$

che esprime il fatto che l'errore che si commette sostituendo all'incremento Δf , il differenziale df è un infinitesimo di ordine superiore al primo.

2. Il Teorema di de l'Hôpital

Supponiamo f e g continue nell'intervallo (a, b) e sia $x_0 \in (a, b)$ tale che

$$f(x_0) = g(x_0) = 0.$$

In questa situazione non è evidente se esista e quanto valga il limite

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)}.$$

Sostituendo, formalmente, a f e g il valore nel punto limite si ottiene l'espressione $\frac{0}{0}$ che non ha senso. L'esistenza o meno del limite è legata alla rapidità con cui le funzioni f e g tendono a 0 per $x \rightarrow x_0$, ossia alla relazione che c'è tra i loro ordini di infinitesimo. Come risolvere un limite del genere? Se le funzioni f e g sono derivabili in x_0 , si può pensare di approssimare le funzioni f e g con la loro retta tangente nel punto x_0 :

$$\frac{f(x)}{g(x)} \approx \frac{f(x_0) + f'(x_0)(x - x_0)}{g(x_0) + g'(x_0)(x - x_0)} = \frac{f'(x_0)(x - x_0)}{g'(x_0)(x - x_0)} = \frac{f'(x_0)}{g'(x_0)}.$$

Come rendere rigorosa tale affermazione? Utilizzando (50), si ha

$$(51) \quad \frac{f(x)}{g(x)} = \frac{f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)}{g(x_0) + g'(x_0)(x - x_0) + o(x - x_0)} = \frac{f'(x_0) + \frac{o(x-x_0)}{x-x_0}}{g'(x_0) + \frac{o(x-x_0)}{x-x_0}}$$

avendo utilizzato la proprietà $f(x_0) = g(x_0) = 0$. Passando al limite per $x \rightarrow x_0$, nel caso in cui $g'(x_0) \neq 0$, si ottiene

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \frac{f'(x_0)}{g'(x_0)} \quad (g'(x_0) \neq 0).$$

Questa formula è nota come **regola di de l'Hôpital**². Lavorando in maniera più raffinata, si dimostra una variante della precedente formula di de l'Hôpital, che non richiede la derivabilità delle funzioni f e g nel punto limite, ma solo l'esistenza del limite del rapporto delle derivate. Anche tale variante è nota sotto lo stesso nome.

TEOREMA 2.1. Regola di de l'Hôpital. *Siano f e g due funzioni derivabili tali che $f(x_0) = g(x_0) = 0$. Se esiste finito*

$$\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)} = \ell \in \mathbb{R},$$

allora esiste anche il limite $\lim_{x \rightarrow x_0} f(x)/g(x)$ ed ha lo stesso valore ℓ .

La conclusione è valida anche nel caso in cui il rapporto delle derivate abbia limite $+\infty$ o $-\infty$. Non si può dedurre nessuna conclusione nel caso in cui il rapporto delle derivate non abbia limite.

Esistono in commercio anche altre versioni del Teorema di de l'Hôpital che si applicano a casi differenti: forme indeterminate del tipo $0/0$ per $x \rightarrow \pm\infty$ o del tipo ∞/∞ per $x \rightarrow x_0$ e per $x \rightarrow \pm\infty$

$$\lim_{x \rightarrow \pm\infty} f(x) = \lim_{x \rightarrow \pm\infty} g(x) = 0, \quad \lim_{x \rightarrow \pm\infty} \frac{f'(x)}{g'(x)} = \ell \quad \Rightarrow \quad \lim_{x \rightarrow \pm\infty} \frac{f(x)}{g(x)} = \ell$$

$$\lim_{x \rightarrow x_0} |f(x)| = \lim_{x \rightarrow x_0} |g(x)| = \infty, \quad \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)} = \ell \quad \Rightarrow \quad \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \ell$$

$$\lim_{x \rightarrow \pm\infty} |f(x)| = \lim_{x \rightarrow \pm\infty} |g(x)| = \infty, \quad \lim_{x \rightarrow \pm\infty} \frac{f'(x)}{g'(x)} = \ell \quad \Rightarrow \quad \lim_{x \rightarrow \pm\infty} \frac{f(x)}{g(x)} = \ell.$$

Il simbolo ℓ può essere sia un numero reale sia $+\infty$ o $-\infty$.

Il principio è sempre lo stesso: nel caso di una forma indeterminata $0/0$ o ∞/∞ , si può calcolare il limite del rapporto delle derivate. Se tale limite esiste, allora dà anche il valore del limite iniziale. Se, invece, il rapporto delle derivate non esiste, non si può

²Questa regola porta il nome il nome del matematico francese Guillaume Francois Antoine, marchese de L'Hôpital (1661 – 1704), che pubblicò la formula nel suo libro *Analyse des infiniment petits pour l'intelligence des lignes courbes* (1692). La regola è in realtà dovuta a Jean Bernoulli, a cui L'Hôpital pagava una pensione di 300 franchi annui in cambio delle informazioni relative ai suoi progressi nel calcolo infinitesimale e della risoluzione di alcuni problemi posti dal de L'Hôpital (tra cui quello di determinare il limite di forme indeterminate). L'Hôpital, riconoscendo che parte del contenuto del suo trattato era dovuta a Bernoulli, preferì pubblicarlo in forma anonima. Ciò nonostante, una volta scoperto l'autore del libro, la formula fu associata al suo nome.

concludere nulla. Nel caso in cui il limite del rapporto delle derivate dia luogo, esso stesso, ad una forma indeterminata $0/0$ o ∞/∞ , si può applicare di nuovo il Teorema di de l'Hôpital e (provare a) calcolare il limite del rapporto delle derivate seconde.

ESEMPIO 2.2. Per calcolare il limite

$$(52) \quad \lim_{x \rightarrow 0} \frac{x - \arctan x}{x - \sin x},$$

studiamo il limite del rapporto delle derivate

$$\lim_{x \rightarrow 0} \frac{1 - \frac{1}{1+x^2}}{1 - \cos x} = \lim_{x \rightarrow 0} \frac{x^2}{(1+x^2)(1-\cos x)} = 2,$$

dato che $\lim_{x \rightarrow 0} (1 - \cos x)/x^2 = 1/2$. Quindi il limite (52) esiste e vale 2.

ESEMPIO 2.3. Calcoliamo

$$\lim_{x \rightarrow +\infty} \left(\frac{\pi}{2} - \arctan x \right) x.$$

Questo limite è della forma $0 \cdot \infty$, ma si può ricondurre alla tipologia trattabile con il teorema di de l'Hôpital riscrivendolo come

$$\lim_{x \rightarrow +\infty} \frac{\pi/2 - \arctan x}{1/x}.$$

Il rapporto delle derivate ha limite:

$$\lim_{x \rightarrow +\infty} \frac{-1/(1+x^2)}{-1/x^2} = \lim_{x \rightarrow +\infty} \frac{x^2}{1+x^2} = 1.$$

Quindi, il limite richiesto esiste e vale 1.

ESERCIZIO 2.4. *Calcolare il limite*

$$\lim_{x \rightarrow 0} \frac{e^x - 1 - x}{\cos x - 1}.$$

La dimostrazione del Teorema di de L'Hôpital. Per cominciare, enunciamo (e dimostriamo) una variante del Teorema di Lagrange, nota come Teorema di Cauchy.

TEOREMA 2.5. Teorema di Cauchy. *Siano f e g due funzioni continue in $[a, b]$ e derivabili in (a, b) . Allora esiste $\xi \in (a, b)$ tale che*

$$(53) \quad \det \begin{pmatrix} f(b) - f(a) & f'(\xi) \\ g(b) - g(a) & g'(\xi) \end{pmatrix} = 0,$$

cioè $(f(b) - f(a))g'(\xi) = f'(\xi)(g(b) - g(a))$.

Interpretazione geometrica. Date le funzioni f e g , consideriamo la funzione *vetto-riale* (f, g) che associa ad un punto dell'intervallo $[a, b]$ il punto del piano di coordinate (f, g) . Il Teorema di Cauchy asserisce che esiste sempre un valore $\xi \in (a, b)$ tale che il vettore incremento $(f(b) - f(a), g(b) - g(a))$ e il vettore “derivata” $(f'(x), g'(x))$ calcolato in $x = \xi$ sono paralleli.

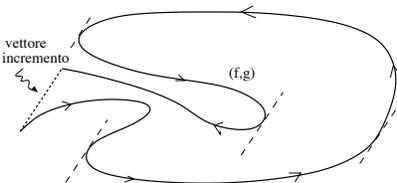


FIGURA 1. L'interpretazione geometrica del Teorema di Cauchy.

DIMOSTRAZIONE DEL TEOREMA 2.5. Consideriamo la funzione

$$\Phi(x) := \det \begin{pmatrix} f(b) - f(a) & f(x) \\ g(b) - g(a) & g(x) \end{pmatrix} \equiv [f(b) - f(a)]g(x) - f(x)[g(b) - g(a)].$$

La funzione Φ è continua in $[a, b]$ e derivabile in (a, b) . Inoltre, si ha

$$\Phi(a) = \det \begin{pmatrix} f(b) - f(a) & f(a) \\ g(b) - g(a) & g(a) \end{pmatrix} = \det \begin{pmatrix} f(b) & 0 \\ g(b) & 0 \end{pmatrix} = 0.$$

$$\Phi(b) = \det \begin{pmatrix} f(b) - f(a) & f(b) \\ g(b) - g(a) & g(b) \end{pmatrix} = \det \begin{pmatrix} -f(a) & 0 \\ -g(a) & 0 \end{pmatrix} = 0.$$

Quindi, per il Teorema di Rolle, esiste $\xi \in (a, b)$ tale che $\Phi'(\xi) = 0$. Dato che $\Phi'(x) = (f(b) - f(a))g'(x) - f'(x)(g(b) - g(a))$, segue la conclusione. \square

Armati del precedente risultato, si può dimostrare il Teorema di de l'Hôpital.

DIMOSTRAZIONE DEL TEOREMA 2.1. Scegliamo, nella formula (53), $a = x_0$ e $b = x$. Dato che, per ipotesi, $f(x_0) = g(x_0) = 0$, si ha

$$(54) \quad \frac{f(x)}{g(x)} = \frac{f'(\xi)}{g'(\xi)},$$

dove ξ è compreso tra x_0 e x . Per $x \rightarrow x_0$, necessariamente $\xi \rightarrow x_0$, e il termine $f'(\xi)/g'(\xi)$ tende ad ℓ per ipotesi. Dato che questo termine è uguale a $f(x)/g(x)$, ne segue la conclusione \square

Approssimazioni polinomiali. Utilizziamo adesso il Teorema di de l'Hôpital per dedurre delle approssimazioni polinomiali di funzioni con un errore che sia infinitesimo di ordine sempre più alto. Scegliamo come cavia la funzione $\sin x$. Dato che è derivabile in 0 e la sua derivata è 1,

$$(55) \quad \sin x = x + o(x) \quad \text{per } x \rightarrow 0$$

o, equivalentemente,

$$\lim_{x \rightarrow 0} \frac{\sin x - x}{x} = 0.$$

Per dedurre un'approssimazione per $\sin x$ con un errore che sia $o(x^2)$, calcoliamo

$$\lim_{x \rightarrow 0} \frac{\sin x - x}{x^2}.$$

Per applicare il Teorema di de l'Hôpital, studiamo il limite del rapporto delle derivate

$$\lim_{x \rightarrow 0} \frac{\cos x - 1}{2x} = -\frac{1}{2} \lim_{x \rightarrow 0} \frac{1 - \cos x}{x} = 0.$$

Dato che tale limite esiste finito, anche il limite di partenza esiste e vale 0. Quindi

$$(56) \quad \sin x = x + o(x^2) \quad \text{per } x \rightarrow 0$$

La formula (56) dice che l'errore che si commette approssimando $\sin x$ con x è un infinitesimo di ordine superiore al secondo per $x \rightarrow 0$. Questa informazione è migliore di quella data da (55), che ci garantiva solamente un errore di ordine superiore al primo.

Per ottenere un'approssimazione con errore di ordine superiore al terzo, ragioniamo come in precedenza e calcoliamo

$$\lim_{x \rightarrow 0} \frac{\sin x - x}{x^3} = \lim_{x \rightarrow 0} \frac{\cos x - 1}{3x^2} = -\frac{1}{3} \lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = -\frac{1}{6},$$

che implica $\sin x = x + O(x^3)$. Portando il termine $-1/6$ a primo membro, otteniamo

$$\lim_{x \rightarrow 0} \frac{\sin x - x + \frac{1}{6}x^3}{x^3} = 0$$

cioè la funzione $\sin x$ è pari a $x - \frac{1}{6}x^3$ più un errore superiore a x^3

$$(57) \quad \sin x = x - \frac{1}{6}x^3 + o(x^3) \quad \text{per } x \rightarrow 0.$$

Possiamo iterare il procedimento e calcolare

$$\lim_{x \rightarrow 0} \frac{\sin x - x + \frac{1}{6}x^3}{x^4} = \lim_{x \rightarrow 0} \frac{\cos x - 1 + \frac{1}{2}x^2}{4x^3} = \lim_{x \rightarrow 0} \frac{-\sin x + x}{12x^2} = 0,$$

quindi

$$(58) \quad \sin x = x - \frac{1}{6}x^3 + o(x^4) \quad \text{per } x \rightarrow 0.$$

Ripetiamo l'esperimento su una cavia diversa: e^x . Il fatto che e^x sia derivabile in $x = 0$ e la derivata valga 1 si traduce nella formula

$$(59) \quad e^x = 1 + x + o(x) \quad \text{per } x \rightarrow 0.$$

Per migliorare l'espressione, calcoliamo

$$\lim_{x \rightarrow 0} \frac{e^x - 1 - x}{x^2} = \lim_{x \rightarrow 0} \frac{e^x - 1}{2x} = \frac{1}{2},$$

cioè $e^x = 1 + x + O(x^2)$. Il limite può essere riscritto come

$$\lim_{x \rightarrow 0} \frac{e^x - 1 - x}{x^2} - \frac{1}{2} = 0 \quad \implies \quad \lim_{x \rightarrow 0} \frac{e^x - 1 - x - \frac{1}{2}x^2}{x^2} = 0,$$

cioè $e^x - 1 - x - \frac{1}{2}x^2 = o(x^2)$, o anche

$$e^x = 1 + x + \frac{1}{2}x^2 + o(x^2).$$

Allo stesso modo

$$(60) \quad \lim_{x \rightarrow 0} \frac{e^x - 1 - x - \frac{1}{2}x^2}{x^3} = \lim_{x \rightarrow 0} \frac{e^x - 1 - x}{3x^2} = \frac{1}{6},$$

La relazione (60) si può riscrivere come

$$\lim_{x \rightarrow 0} \frac{1}{x^3} \left(e^x - 1 - x - \frac{1}{2}x^2 - \frac{1}{6}x^3 \right) = 0 \quad \iff \quad e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + o(x^3)$$

Cosa stiamo facendo iterando questo procedimento? Stiamo ottenendo delle approssimazioni ad ordini sempre più alti di una funzione data. La relazione $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + o(x^3)$ esprime il fatto che la funzione e^x si può approssimare, per $x \rightarrow 0$ con il polinomio $1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$ commettendo un errore (la differenza tra e^x e il polinomio) che tende a zero per $x \rightarrow 0$ con ordine superiore a 3 (cioè più rapidamente di x^3).

L'iterazione dell'algorithmo che abbiamo visto conduce direttamente a quello che si chiama *polinomio di Taylor*.

3. La formula di Taylor

Replichiamo, in generale, l'esperimento fatto su $\sin x$ e e^x alla fine del paragrafo precedente. Se f è una funzione derivabile in x_0 , si ha

$$\lim_{x \rightarrow x_0} \frac{f(x) - [f(x_0) + f'(x_0)(x - x_0)]}{x - x_0} = 0,$$

o, equivalentemente,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(|x - x_0|) \quad \text{per } x \rightarrow x_0,$$

che esprime che la funzione f , vicino ad x_0 , si approssima con la funzione $f(x_0) + f'(x_0)(x - x_0)$, con un errore che è un infinitesimo di ordine superiore ad 1.

Per ottenere un'approssimazione più precisa, supponendo che la funzione f si derivabile due volte, possiamo calcolare il limite

$$\lim_{x \rightarrow x_0} \frac{f(x) - [f(x_0) + f'(x_0)(x - x_0)]}{(x - x_0)^2} = \lim_{x \rightarrow x_0} \frac{f'(x) - f'(x_0)}{2(x - x_0)} = \frac{1}{2}f''(x_0),$$

avendo applicato il Teorema di de l'Hôpital. Il precedente limite si può scrivere come

$$\lim_{x \rightarrow x_0} \frac{f(x) - [f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2]}{(x - x_0)^2} = 0$$

ossia

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + o(|x - x_0|^2).$$

Così abbiamo scoperto che la funzione $f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$ approssima f , vicino ad x_0 , con un errore di ordine superiore a 2. Il grafico della funzione $p(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$ rappresenta la parabola che meglio approssima la funzione f per $x \rightarrow x_0$.

Iterando ancora una volta il procedimento e supponendo che la funzione f sia derivabile tre volte in x_0 , si ottiene

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \frac{1}{3!}f'''(x_0)(x - x_0)^3 + o(|x - x_0|^3).$$

E in generale?

TEOREMA 3.1. Formula di Taylor. *Sia $f : (a, b) \rightarrow \mathbb{R}$ derivabile n volte in (a, b) e sia $x_0 \in (a, b)$. Dato $n \in \mathbb{N}$, posto*

$$T_n(x; x_0) := f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{1}{n!}f^{(n)}(x_0)(x - x_0)^n,$$

si ha

$$\lim_{x \rightarrow x_0} \frac{f(x) - T_n(x; x_0)}{(x - x_0)^n} = 0,$$

cioè vale la decomposizione $f(x) = T_n(x; x_0) + o(|x - x_0|^n)$.

DEFINIZIONE 3.2. *Il polinomio $T_n(x; x_0)$ si chiama polinomio di Taylor³ di grado n della funzione f nel punto x_0 e rappresenta un'approssimazione di f vicino ad x_0 .*

La peculiarità della formula di Taylor sta nel fatto che il resto R_n , definito da

$$R_n(x; x_0) := f(x) - T_n(x; x_0),$$

è un infinitesimo di ordine superiore ad $|x - x_0|^n$ per $x \rightarrow x_0$.

³Se $x_0 = 0$, il polinomio p_n viene, a volte, chiamato polinomio di McLaurin.

DIMOSTRAZIONE. Applichiamo il Teorema di de l'Hôpital al limite

$$\begin{aligned} & \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0) - \dots - \frac{1}{(n-1)!} f^{(n-1)}(x_0)(x - x_0)^{n-1}}{(x - x_0)^n} \\ &= \lim_{x \rightarrow x_0} \frac{f'(x) - f'(x_0) - f''(x_0)(x - x_0) - \dots - \frac{1}{(n-2)!} f^{(n-1)}(x_0)(x - x_0)^{n-2}}{n(x - x_0)^{n-1}}. \end{aligned}$$

Dato che sia il numeratore che il denominatore sono infinitesimi, possiamo applicare nuovamente il Teorema di de l'Hôpital, ottenendo

$$\lim_{x \rightarrow x_0} \frac{f''(x) - f''(x_0) - f'''(x_0)(x - x_0) \dots - \frac{1}{(n-3)!} f^{(n-1)}(x_0)(x - x_0)^{n-3}}{n(n-1)(x - x_0)^{n-2}}.$$

Iterando $n - 1$ volte il procedimento si ottiene

$$\lim_{x \rightarrow x_0} \frac{f^{(n-1)}(x) - f^{(n-1)}(x_0)}{n!(x - x_0)} = \frac{1}{n!} f^{(n)}(x_0).$$

Quindi vale

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - \dots - \frac{1}{(n-1)!} f^{(n-1)}(x_0)(x - x_0)^{n-1}}{(x - x_0)^n} = \frac{1}{n!} f^{(n)}(x_0),$$

da cui segue

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - \dots - \frac{1}{(n-1)!} f^{(n-1)}(x_0)(x - x_0)^{n-1}}{(x - x_0)^n} - \frac{1}{n!} f^{(n)}(x_0) = 0,$$

che porta alla conclusione. □

ESEMPIO 3.3. Polinomi. Assegnati a_0, \dots, a_p , sia

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_px^p.$$

Consideriamo, prima di tutto, lo sviluppo in $x_0 = 0$. Dato che

$$\begin{aligned} f'(x) &= a_1 + 2a_2x + \dots + pa_px^{p-1} \\ f''(x) &= 2a_2 + 3 \cdot 2a_3x + \dots + p(p-1)a_px^{p-2} \\ &\vdots \\ f^{(p)}(x) &= p(p-1) \dots 2 \cdot 1 a_p \\ f^{(k)}(x) &= 0 \qquad \qquad \qquad k > p \end{aligned}$$

si ha

$$f'(0) = a_1, \quad f''(0) = 2a_2, \quad \dots \quad f^{(p)}(0) = p! a_p, \quad f^{(k)}(0) = 0 \quad k > p.$$

Quindi,

$$T_n(x; 0) = \begin{cases} a_0 + a_1x + \dots + a_nx^n & n < p, \\ a_0 + a_1x + \dots + a_px^p & n \geq p. \end{cases}$$

Come era naturale aspettarsi, il polinomio di Taylor di f in $x_0 = 0$ e di grado n si ottiene considerando i termini del polinomio con grado minore o uguale ad n .

Per il polinomio di Taylor in $x_0 \neq 0$, occorre riscrivere il polinomio in termini di potenze di $h := x - x_0$. In questo modo si otterrà un'espressione del tipo

$$f(x) = b_0 + b_1(x - x_0) + \cdots + b_p(x - x_0)^p$$

con b_0, b_1, \dots, b_p opportuni. Il polinomio di Taylor è dato da

$$T_n(x; x_0) = \begin{cases} b_0 + b_1(x - x_0) + \cdots + b_n(x - x_0)^n & n < p, \\ b_0 + b_1(x - x_0) + \cdots + b_p(x - x_0)^p & n \geq p. \end{cases}$$

Consideriamo, ad esempio, la funzione $f(x) = x + x^3$. Fissato $x_0 \in \mathbb{R}$, per scriverla in termini di potenze di $h = x - x_0$ calcoliamo

$$f(x_0 + h) = (x_0 + h) + (x_0 + h)^3 = x_0 + x_0^3 + (1 + 3x_0^2)h + 3x_0h^2 + h^3.$$

Quindi vale l'identità

$$x + x^3 = x_0 + x_0^3 + (1 + 3x_0^2)(x - x_0) + 3x_0(x - x_0)^2 + (x - x_0)^3.$$

Ad esempio, il polinomio di Taylor di grado 2 in x_0 è

$$T_2(x; x_0) = x_0 + x_0^3 + (1 + 3x_0^2)(x - x_0) + 3x_0(x - x_0)^2.$$

Lo stesso, evidentemente, si ottiene applicando direttamente la formula: da

$$f(x) = x + x^3, \quad f'(x) = 1 + 3x^2, \quad f''(x) = 6x,$$

segue

$$T_2(x; x_0) = x_0 + x_0^3 + (1 + 3x_0^2)(x - x_0) + 3x_0(x - x_0)^2.$$

ESEMPIO 3.4. Esponenziale in $x_0 = 0$. Siano

$$f(x) = e^x \quad \text{e} \quad x_0 = 0.$$

Dato che $f^{(k)}(x) = e^x$ per ogni $k \in \mathbb{N}$, si ha $f^{(k)}(0) = e^0 = 1$ per ogni k , quindi il polinomio di Taylor di grado n è

$$T_n(x; 0) = \sum_{k=0}^n \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!}.$$

Questa formula è coerente con la definizione di esponenziale data in precedenza.

Che succede se $x_0 \neq 0$? I conti non sono molto diversi:

$$T_n(x; 0) = e^{x_0} \sum_{k=0}^n \frac{(x - x_0)^k}{k!} = e^{x_0} \left[1 + (x - x_0) + \frac{(x - x_0)^2}{2} + \cdots + \frac{(x - x_0)^n}{n!} \right].$$

ESEMPIO 3.5. Seno e coseno in $x_0 = 0$. Sia $f(x) = \sin x$, allora

$$f^{(2k)}(x) = (-1)^k \sin x, \quad f^{(2k+1)}(x) = (-1)^k \cos x \quad \forall k = 0, 1, 2, \dots$$

Calcolando in $x_0 = 0$, otteniamo

$$f^{(2k)}(0) = 0, \quad f^{(2k+1)}(0) = (-1)^k \quad \forall k = 0, 1, 2, \dots$$

Ne segue che

$$T_n(x; 0) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots + (-1)^k \frac{x^{2k+1}}{(2k+1)!}$$

è il polinomio di Taylor di grado n con $n = 2k+1$ o $2k+2$. In effetti si può dimostrare che vale l'uguaglianza

$$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \quad \forall x \in \mathbb{R}.$$

Analogamente se consideriamo la funzione $f(x) = \cos x$ abbiamo

$$f^{(2k)}(x) = (-1)^k \cos x, \quad f^{(2k+1)}(x) = (-1)^{k+1} \sin x \quad \forall k = 0, 1, 2, \dots$$

Calcolando in $x_0 = 0$, otteniamo

$$f^{(2k)}(0) = (-1)^k, \quad f^{(2k+1)}(0) = 0 \quad \forall k.$$

Ne segue che

$$T_n(x; 0) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots + (-1)^k \frac{x^{2k}}{(2k)!}$$

è il polinomio di Taylor di $\cos x$ centrato in 0 di grado n con $n = 2k$ o $2k+1$. Anche per il coseno vale un'uguaglianza analoga alla precedente:

$$\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} \quad \forall x \in \mathbb{R}.$$

ESERCIZIO 3.6. Calcolare il polinomio di Taylor di grado 4 della funzione $\sin x$ centrato in $x_0 = \pi/2$ e quello centrato in $x_0 = \pi/4$.

ESEMPIO 3.7. Siano $f(x) = \frac{1}{1-x}$ e $x_0 = 0$. Le derivate di f sono

$$f'(x) = \frac{1}{(1-x)^2}, \quad f''(x) = \frac{2}{(1-x)^3}, \quad \dots, \quad f^{(k)}(x) = \frac{k!}{(1-x)^{k+1}}.$$

Perciò $f(0) = 1$, $f'(0) = 1$, $f''(0) = 2, \dots$, $f^{(k)}(0) = k!$. Quindi il polinomio di Taylor è

$$T_n(x; 0) = 1 + x + x^2 + \dots + x^n = \sum_{k=0}^n x^k$$

coerente con l'espressione nota per la serie geometrica.

ESEMPIO 3.8. Come ultimo esempio, consideriamo

$$f(x) = \ln(1+x) \quad \text{e} \quad x_0 = 0.$$

Le derivate di f sono

$$f'(x) = \frac{1}{1+x}, \quad f''(x) = -\frac{1}{(1+x)^2}, \quad \dots, \quad f^{(k)}(x) = (-1)^{k+1} \frac{(k-1)!}{(1+x)^k},$$

e quindi $f(0) = 0$, $f'(0) = 1$, $f''(0) = -1$, $f^{(k)}(0) = (-1)^{k+1}(k-1)!$. Il polinomio di Taylor di grado n in $x_0 = 0$ è

$$T_n(x; 0) = x - \frac{x^2}{2} + \dots + (-1)^{n+1} \frac{x^n}{n} = \sum_{k=1}^n (-1)^{k+1} \frac{x^k}{k}.$$

4. Espressioni del resto

Data una funzione f , con un buon numero di derivate, sappiamo determinare un polinomio che la approssimi vicino ad un punto assegnato x_0 . In questa approssimazione viene commesso un errore pari a

$$R_n(x; x_0) := f(x) - \sum_{k=0}^n \frac{1}{k!} f^{(k)}(x_0) (x - x_0)^k.$$

Quali proprietà conosciamo su R_n ? Per ora sappiamo solo che

$$R_n(x; x_0) = o(|x - x_0|^n) \quad \text{cioè} \quad \lim_{x \rightarrow x_0} \frac{R_n(x; x_0)}{(x - x_0)^n} = 0.$$

Questa è solo un'informazione sul comportamento al limite, quindi non dice nulla di preciso sulla grandezza della quantità R_n in punti $x \neq x_0$. Per poter stimare l'errore occorre una rappresentazione migliore di R_n . Ecco il nostro nuovo obiettivo.

TEOREMA 4.1. Resto in forma di Lagrange. *Se la funzione f è derivabile $n+1$ volte, esiste ξ , tra x_0 e x , tale che*

$$f(x) = \sum_{k=0}^n \frac{1}{k!} f^{(k)}(x_0) (x - x_0)^k + \frac{1}{(n+1)!} f^{(n+1)}(\xi) (x - x_0)^{n+1},$$

DIMOSTRAZIONE. Sia f derivabile $n+1$ volte e consideriamo le funzioni

$$F(x) := R_n(x; x_0) = f(x) - \sum_{k=0}^n \frac{1}{k!} f^{(k)}(x_0) (x - x_0)^k \quad \text{e} \quad G(x) := (x - x_0)^{n+1}.$$

Dato che $F(x_0) = G(x_0) = 0$, applicando il Teorema di Cauchy a F e G , segue

$$\frac{F(x) - F(x_0)}{G(x) - G(x_0)} = \frac{F(x)}{G(x)} = \frac{F'(\xi_1)}{G'(\xi_1)},$$

per qualche ξ_1 , compreso tra x_0 e x . Derivando le espressioni di F e G , otteniamo

$$F'(x) = f'(x) - \sum_{k=0}^{n-1} \frac{1}{k!} f^{(k+1)}(x_0)(x-x_0)^k \quad \text{e} \quad G'(x) = (n+1)(x-x_0)^n.$$

Se $n \geq 1$ è possibile riapplicare il Teorema di Cauchy, trovando ξ_2 , compreso tra x_0 e ξ_1 e quindi anche tra x_0 e x , per cui

$$\frac{F(x) - F(x_0)}{G(x) - G(x_0)} = \frac{F'(\xi_1)}{G'(\xi_1)} = \frac{F''(\xi_2)}{G''(\xi_2)}.$$

Iterando $n+1$ volte il procedimento, si dimostra l'esistenza di ξ_{n+1} tra x_0 e x tale che

$$\frac{F(x) - F(x_0)}{G(x) - G(x_0)} = \frac{F^{(n+1)}(\xi_{n+1})}{G^{(n+1)}(\xi_{n+1})}.$$

Dato che $F^{(n+1)}(x) = f^{(n+1)}(x)$ e $G^{(n+1)}(x) = (n+1)!$, si deduce (qui $\xi = \xi_{n+1}$)

$$F(x) - F(x_0) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) (G(x) - G(x_0))$$

da cui, ricordando le definizioni di F e G ,

$$R_n(x; x_0) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) (x-x_0)^{n+1},$$

cioè la conclusione. □

A partire da questa espressione del resto, possiamo stimare l'errore commesso quando si approssimi una funzione f con il suo polinomio di Taylor: se $M > 0$ è tale che $|f^{(n+1)}(t)| \leq M$ per ogni t tra x e x_0 , allora

$$|R_n(x; x_0)| = \left| \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-x_0)^{n+1} \right| = \frac{|f^{(n+1)}(\xi)|}{(n+1)!} |x-x_0|^{n+1} \leq \frac{M}{(n+1)!} |x-x_0|^{n+1}.$$

Calcolo approssimato di $\sin(1/10)$ con stima dell'errore. Abbiamo già considerato questo problema proponendo come "candidato" per l'approssimazione il valore $1/10$. In quell'occasione avevamo stimato l'errore commesso con $1/100$. Il procedimento era basato sul Teorema di Lagrange e sull'approssimazione della funzione $\sin x$ con la sua tangente nell'origine:

$$\sin x \approx x \quad \text{per } x \rightarrow 0.$$

Detta $f(x) = \sin x$, la stima dell'errore discendeva da

$$\begin{aligned} |f(x) - f(x_0) - f'(x_0)(x-x_0)| &= |(f'(\xi) - f'(x_0))(x-x_0)| \\ &= |f''(\eta)(\xi-x_0)(x-x_0)| \leq |f''(\eta)||x-x_0|^2. \end{aligned}$$

dove $x = 1/10$ e $x_0 = 0$. Dato che $f''(x) = -\sin x$, la stima è immediata.

Come ottenere stime migliori? La scelta naturale è approssimare la funzione $\sin x$ con un suo polinomio di Taylor di grado opportuno. Ad esempio,

$$\sin x \approx x - \frac{x^3}{6} \quad \text{per } x \rightarrow 0.$$

Quindi un'approssimazione migliore della precedente è

$$\sin\left(\frac{1}{10}\right) \approx \frac{1}{10} - \frac{1}{6000} = 0,0998\bar{3}.$$

Scriviamo l'errore con la forma di Lagrange $R_3(x; x_0) = \frac{1}{4!} f^{(4)}(\xi)(x - x_0)^4$, cioè

$$\sin \frac{1}{10} - \left(\frac{1}{10} - \frac{1}{6000} \right) = R_3(1/10; 0) = \frac{\sin \xi}{4!} \frac{1}{10^4},$$

quindi

$$|R_3(1/10; 0)| \leq \frac{1}{24 \cdot 10^4} = 4,1\bar{6} \times 10^{-6}.$$

In realtà il polinomio $x - \frac{x^3}{6}$ è anche il polinomio di Taylor di $\sin x$ in 0 di grado 4, quindi il resto può anche essere scritto come

$$\sin \frac{1}{10} - \frac{1}{10} + \frac{1}{6000} = R_4(1/10; 0) = \frac{\cos \xi}{5!} \frac{1}{10^5},$$

quindi

$$|R_4(1/10; 0)| \leq \frac{1}{120 \cdot 10^5} = 8,3 \times 10^{-8}.$$

In definitiva

$$\sin\left(\frac{1}{10}\right) = 0,0998\bar{3} \pm 8,3 \times 10^{-8}.$$

CAPITOLO 8

L'integrale

Il problema della misurazione delle lunghezze di segmenti ci ha condotti per un sentiero particolarmente interessante: dai numeri naturali, ai relativi, ai razionali ed infine ai numeri reali. E se ci viene la voglia di misurare aree di regioni del piano? Nel caso di un rettangolo, l'area è nota: è il prodotto delle lunghezze dei lati. Nel caso di una regione che sia unione finita di rettangoli che si toccano al più lungo il perimetro, basta sommare le aree dei singoli rettangoli. Ma per regioni più generali?

1. L'area di un sottografico e la definizione di integrale

Data una funzione $f : [a, b] \rightarrow [0, +\infty)$, l'insieme

$$S_f = \{(x, y) \mid a \leq x \leq b, \quad 0 \leq y \leq f(x)\}$$

si dice **sottografico** di f nell'intervallo $[a, b]$. Come definire e/o calcolare l'area di un sottografico di funzione? L'area del sottografico delle funzioni costanti è elementare: se $f(x) = C \geq 0$, il sottografico S_f di f in $[a, b]$ è un rettangolo, la cui area è, da sempre, $\mathcal{A}(S_f) = (b - a)C$. Con poco impegno, possiamo definire una classe di funzioni il cui sottografico ha un'area facile da calcolare.

DEFINIZIONE 1.1. *Dato l'intervallo $[a, b]$, un insieme $P = \{x_0, x_1, \dots, x_n\}$ tale che $a = x_0 < x_1 < \dots < x_n = b$ è una **partizione** di $[a, b]$. L'ampiezza della partizione P è il numero $|P| := \max\{x_i - x_{i-1} : i = 1, \dots, n\}$.*

*Una funzione $f : [a, b] \rightarrow \mathbb{R}$ è una **funzione costante a tratti** (o **funzione a scala**) se esiste una partizione $P = \{x_0, x_1, \dots, x_n\}$ di $[a, b]$ tale che f è costante su ogni intervallo $[x_{i-1}, x_i)$ per $i = 1, \dots, n - 1$ e su $[x_{n-1}, x_n]$, ossia*

$$f(x) = \begin{cases} \alpha_i & x \in [x_{i-1}, x_i) \\ \alpha_n & x \in [x_{n-1}, x_n] \end{cases} \quad i = 1, \dots, n - 1$$

dove $\alpha_1, \dots, \alpha_n$ sono numeri reali.

Dato che per una funzione f costante a tratti e non negativa il sottografico è un'unione finita di rettangoli, l'area di S_f è data dalla somma delle aree di questi

rettangoli:

$$\mathcal{A}(S_f) = \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) = \sum_{i=1}^n \alpha_i(x_i - x_{i-1})$$

dove ξ_i è un qualsiasi punto di $[x_{i-1}, x_i]$.

Data una funzione $f : [a, b] \rightarrow \mathbb{R}$, non negativa, possiamo considerare approssimazioni del sottografico S_f date da sottografici di funzioni costanti a tratti g ed h , con $g \leq f \leq h$ in $[a, b]$ (vedi Figura 1(a)) e utilizzare le aree dei sottografici approssimanti come approssimazioni dell'area di S_f . Per poter stimare per eccesso la funzione f con una funzione costante a tratti non negativa occorre che la funzione f sia *superiormente limitata* (vedi Figura 1(b)). Per la stima dal basso con funzioni costanti a tratti, serve anche che f sia inferiormente limitata, ipotesi che nel nostro presente caso è automaticamente soddisfatta dato che f è non negativa. L'ipotesi di limitatezza della funzione f è fondamentale qui, così come in tutto il Capitolo e in tutta la definizione dell'integrale definito.

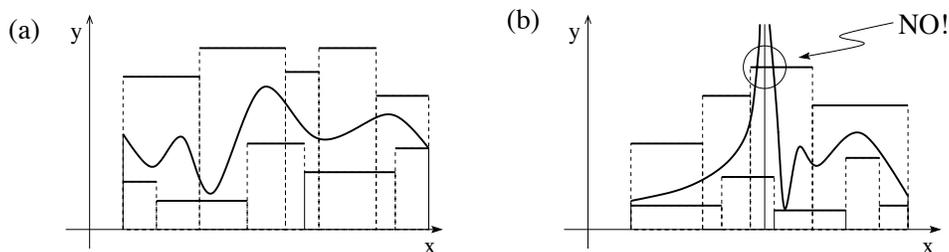


FIGURA 1. (a) Una funzione approssimata per eccesso e per difetto con due funzioni costanti a tratti. (b) Un tentativo (fallito) di approssimazione per eccesso di una funzione non limitata superiormente.

Data una partizione $P = \{x_0, x_1, \dots, x_n\}$ ci sono due funzioni \underline{f} e \bar{f} , costanti a tratti negli intervalli definiti da P e tali che $\underline{f} \leq f \leq \bar{f}$, che danno la migliore stima per difetto e la migliore per eccesso. Tali funzioni sono definite da

$$\begin{aligned} \underline{f}(x) &= \alpha_i := \inf_{[x_{i-1}, x_i]} f(x) & \forall x \in [x_{i-1}, x_i], \\ \bar{f}(x) &= \beta_i := \sup_{[x_{i-1}, x_i]} f(x) & \forall x \in [x_{i-1}, x_i]. \end{aligned}$$

In definitiva, data una partizione P dell'intervallo $[a, b]$, (tutte le volte che $\mathcal{A}(S_f)$ ha senso) si ha

$$\underline{\mathcal{A}}(S_f; P) \leq \mathcal{A}(S_f) \leq \bar{\mathcal{A}}(S_f; P).$$

dove

$$\underline{\mathcal{A}}(S_f; P) := \mathcal{A}(S_{\underline{f}}) = \sum_{i=1}^n \alpha_i(x_i - x_{i-1}) \quad \bar{\mathcal{A}}(S_f; P) := \mathcal{A}(S_{\bar{f}}) = \sum_{i=1}^n \beta_i(x_i - x_{i-1}).$$

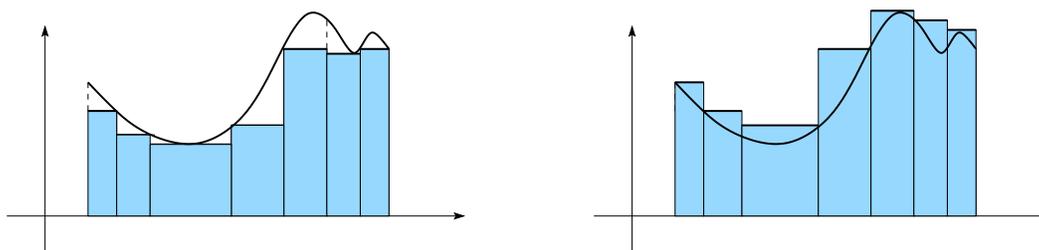


FIGURA 2. Due approssimazioni per S_f tramite le funzioni \underline{f} e \bar{f} .

Dopo, si procede a migliorare l'approssimazione tramite la scelta di una partizione con un'ampiezza più piccola. Una possibilità è scegliere una successione di partizioni P_n con ampiezza $|P_n|$ che tenda a zero per $n \rightarrow \infty$. In questo modo, passando al limite nelle due successioni numeriche $\underline{A}(S_f; P_n)$ e $\bar{A}(S_f; P_n)$ si dovrebbe (incrociando le dita) ottenere proprio l'area richiesta.

ESEMPIO 1.2. Sia $f(x) = x$ per $x \in [a, b]$ con $0 \leq a < b$. Il sottografico della funzione f nell'intervallo $[a, b]$ ha una forma familiare: nel caso in cui $a = 0$, si tratta di un triangolo, e nel caso di $a > 0$ si tratta di un trapezio. In entrambi casi, la geometria elementare fornisce una formula per il calcolo dell'area, che è data da:

$$\frac{(b - a)(b + a)}{2} = \frac{b^2}{2} - \frac{a^2}{2}.$$

Cosa succede se si calcola l'area attraverso il procedimento di approssimazione per eccesso e per difetto proposto in precedenza?

Dividiamo l'intervallo $[a, b]$ in n parti di uguale lunghezza tramite la partizione $P_n = \{x_k = a + kh : k = 0, \dots, n\}$ dove $h = (b - a)/n$. Dato che $f(x) = x$ è crescente,

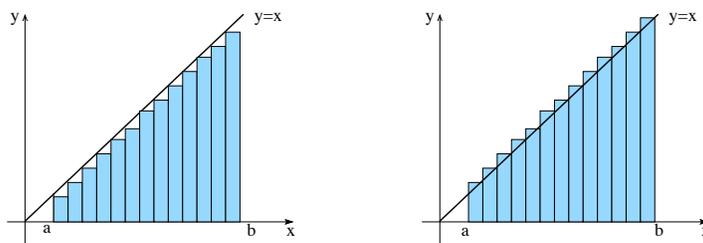


FIGURA 3. La funzione $f(x) = x$ e le approssimazioni determinate dai punti $a + \frac{k(b-a)}{n}$ con $k = 0, \dots, n$.

$$\alpha_k = \inf_{[x_{k-1}, x_k]} x = x_{k-1} = a + (k - 1)h, \quad \beta_k = \sup_{[x_{k-1}, x_k]} x = x_k = a + kh,$$

e $x_k - x_{k-1} = h$, valgono

$$\underline{A}(S_f; P_n) = ah + \cdots + (a + (n-1)h)h = h \sum_{k=0}^{n-1} (a + kh) = hna + h^2 \sum_{k=0}^{n-1} k,$$

$$\overline{A}(S_f; P_n) = (a+h)h + \cdots + (a+nh)h = h \sum_{k=1}^n (a + kh) = hna + h^2 \sum_{k=1}^n k$$

Tenendo conto della formula $\sum_{k=1}^n k = \frac{n(n+1)}{2}$ (dimostratela!) e di $h = \frac{b-a}{n}$

$$\underline{A}(S_f; P_n) = hna + h^2 \frac{n(n-1)}{2} = (b-a)a + (b-a)^2 \frac{n(n-1)}{2n^2}$$

$$\overline{A}(S_f; P_n) = hna + h^2 \frac{n(n+1)}{2} = (b-a)a + (b-a)^2 \frac{n(n+1)}{2n^2}.$$

Per $n \rightarrow +\infty$, le due quantità tendono allo stesso limite:

$$\lim_{n \rightarrow +\infty} \underline{A}(S_f; P_n) = \lim_{n \rightarrow +\infty} \overline{A}(S_f; P_n) = (b-a)a + \frac{(b-a)^2}{2} = \frac{b^2}{2} - \frac{a^2}{2},$$

che concorda con le formule note dalla geometria elementare.

ESEMPIO 1.3. Passiamo a considerare $f(x) = x^2$ in $[a, b]$ con $0 \leq a < b$. Qual è l'espressione dell'area per il sottografico di x^2 ?

Con la stessa partizione P_n di prima. Dato che x^2 è crescente su $[a, b]$ per $a \geq 0$,

$$\alpha_i = \inf_{[x_{k-1}, x_k]} x^2 = (a + (k-1)h)^2, \quad \beta_i = \sup_{[x_{k-1}, x_k]} x^2 = (a + kh)^2,$$

da cui segue

$$\underline{A}(S_f; P_n) = a^2h + (a+h)^2h + (a+2h)^2h + \cdots + (a+(n-1)h)^2h$$

$$\overline{A}(S_f; P_n) = (a+h)^2h + (a+2h)^2h + \cdots + (a+nh)^2h$$

Svolti i quadrati e tenuto presente che $h = \frac{b-a}{n}$, si ottiene

$$\underline{A}(S_f; P_n) = h \left\{ na^2 + 2ah \sum_{k=0}^{n-1} k + h^2 \sum_{k=0}^{n-1} k^2 \right\}$$

$$= (b-a) \left\{ a^2 + \frac{2a(b-a)}{n^2} \sum_{k=0}^{n-1} k + \frac{(b-a)^2}{n^3} \sum_{k=0}^{n-1} k^2 \right\}$$

$$\overline{A}(S_f; P_n) = h \left\{ na^2 + 2ah \sum_{k=1}^n k + h^2 \sum_{k=1}^n k^2 \right\}$$

$$= (b-a) \left\{ a^2 + \frac{2a(b-a)}{n^2} \sum_{k=1}^n k + \frac{(b-a)^2}{n^3} \sum_{k=1}^n k^2 \right\}$$

Utilizzando le formule $\sum_{k=1}^n k = \frac{1}{2}n(n+1)$ e $\sum_{k=1}^n k^2 = \frac{1}{6}n(n+1)(2n+1)$ (dimostratele!), si ottiene

$$\begin{aligned}\underline{\mathcal{A}}(S_f; P_n) &= (b-a) \left\{ a^2 + a(b-a) \frac{n-1}{n} + (b-a)^2 \frac{(n-1)(2n-1)}{6n^2} \right\} \\ \overline{\mathcal{A}}(S_f; P_n) &= (b-a) \left\{ a^2 + a(b-a) \frac{n+1}{n} + (b-a)^2 \frac{(n+1)(2n+1)}{6n^2} \right\}\end{aligned}$$

Quindi, passando al limite per $n \rightarrow \infty$,

$$\lim_{n \rightarrow +\infty} \underline{\mathcal{A}}(S_f; P_n) = \lim_{n \rightarrow +\infty} \overline{\mathcal{A}}(S_f; P_n) = \frac{b^3}{3} - \frac{a^3}{3} =: \mathcal{A}(S_f).$$

che è il valore dell'area cercato.

ESEMPIO 1.4. Sperimentiamo la tecnica per una funzione più bizzarra:

$$\text{funzione di Dirichlet: } D(x) := \begin{cases} 0 & x \in [0, 1] \setminus \mathbb{Q}, \\ 1 & x \in [0, 1] \cap \mathbb{Q}, \end{cases}$$

Data una qualsiasi partizione $P = \{a = x_0 < x_1 < \dots < x_n = b\}$, in ognuno dei sottointervalli $[x_{k-1}, x_k]$ cadono sia numeri razionali che numeri irrazionali, quindi

$$\alpha_i = \inf_{[x_{k-1}, x_k]} D(x) = 0, \quad \beta_i = \sup_{[x_{k-1}, x_k]} D(x) = 1.$$

Pertanto, per ogni partizione P , $\underline{\mathcal{A}}(S_D; P) = 0$ e $\overline{\mathcal{A}}(S_D; P) = 1$ e quindi

$$\sup_P \underline{\mathcal{A}}(S_D; P) = 0 < 1 = \inf_P \overline{\mathcal{A}}(S_D; P).$$

Anche scegliendo partizioni P con ampiezza sempre più piccola, le stime per difetto e quelle per eccesso restano sempre ben lontane le une dalle altre. L'interpretazione che diamo di questa situazione è che *esistono sotto insiemi del piano a cui non è possibile associare un'area*, ossia, per alcuni insiemi S , l'espressione $\mathcal{A}(S)$ non ha senso!

ESEMPIO 1.5. Ancora un esempio: $f(x) = e^x$ in $[a, b]$ con $a < b$. Scegliamo ancora una volta la partizione $P_n = \{x_k = a + kh : k = 0, \dots, n\}$ dove $h = (b-a)/n$. La funzione e^x è crescente su $[a, b]$ quindi

$$\alpha_i = \inf_{[x_{k-1}, x_k]} e^x = e^{x_{k-1}} = e^a e^{(k-1)h}, \quad \beta_i = \sup_{[x_{k-1}, x_k]} e^x = e^{x_k} = e^a e^{kh}.$$

Perciò

$$\begin{aligned}\underline{\mathcal{A}}(S_f; P_n) &= e^a h [1 + e^h + e^{2h} + \dots + e^{(n-1)h}] = e^a h [1 + e^h + (e^h)^2 + \dots + (e^h)^{n-1}] \\ \overline{\mathcal{A}}(S_f; P_n) &= e^a h [e^h + e^{2h} + e^{3h} + \dots + e^{nh}] = e^{a+h} h [1 + e^h + (e^h)^2 + \dots + (e^h)^{n-1}]\end{aligned}$$

Dato che $1 + e^h + (e^h)^2 + \dots + (e^h)^{n-1} = (e^{nh} - 1)/(e^h - 1)$ e $nh = b - a$,

$$\underline{A}(S_f; P_n) = \frac{e^a h (e^{b-a} - 1)}{e^h - 1} = \frac{h}{e^h - 1} (e^b - e^a)$$

$$\overline{A}(S_f; P_n) = \frac{e^{a+h} h (e^{b-a} - 1)}{e^h - 1} = \frac{h}{e^h - 1} e^h (e^b - e^a)$$

Quindi, passando al limite per $n \rightarrow \infty$,

$$\lim_{n \rightarrow +\infty} \underline{A}(S_f; P_n) = \lim_{n \rightarrow +\infty} \overline{A}(S_f; P_n) = e^b - e^a,$$

che è il valore cercato.

Tiriamo le fila di quello che abbiamo fatto fin qui. Considerando l'“area di una regione del piano” un concetto intuitivo, abbiamo considerato il caso di sottografici di funzioni non negative, proponendo un algoritmo per il calcolo dell'area: approssimare per difetto e per eccesso l'area richiesta tramite aree di sottografici di funzioni costanti a tratti e, facendo tendere l'ampiezza della partizione a 0, ottenere il valore dell'area cercata. Tramite un certo numero di esempi, ci siamo resi conto che il procedimento è ragionevole, ma che in alcune situazioni non porta a nessuna conclusione (come per la funzione di Dirichlet).

Mettendo da parte, per il momento, il problema dell'area e battezziamo le funzioni per cui il procedimento di sopra descritto converge *funzioni integrabili* e il valore limite delle approssimazioni *integrale (definito) della funzione*. Rispetto a quanto fatto in precedenza, c'è una differenza essenziale: *le funzioni che consideriamo possono avere segno qualsiasi*.

Sia $f : [a, b] \rightarrow \mathbb{R}$ limitata e $P = \{x_0, x_1, \dots, x_n\}$ una partizione di $[a, b]$. Poniamo

$$\alpha_i := \inf_{[x_{i-1}, x_i]} f(x) \quad \text{e} \quad \beta_i := \sup_{[x_{i-1}, x_i]} f(x).$$

DEFINIZIONE 1.6. Somme integrali. Si chiamano *somma integrale per difetto/per eccesso di f relativamente alla partizione P , i valori delle somme*

$$\underline{S}(f; P) := \sum_{i=1}^n \alpha_i (x_i - x_{i-1}) \quad \text{e} \quad \overline{S}(f; P) := \sum_{i=1}^n \beta_i (x_i - x_{i-1}).$$

La condizione di limitatezza della funzione f garantisce che, per ogni partizione P , le somme integrali per difetto e per eccesso sono sempre ben definite.

DEFINIZIONE 1.7. Integrale definito. Una funzione $f : [a, b] \rightarrow \mathbb{R}$ limitata si dice *integrabile (secondo Riemann) se vale l'uguaglianza*

$$(61) \quad \sup_P \underline{S}(f; P) = \inf_P \overline{S}(f; P),$$

dove l'estremo superiore e l'estremo inferiore sono presi nell'insieme di tutte le partizioni P dell'intervallo $[a, b]$.

Il valore comune è l'integrale definito di f in $[a, b]$ e si indica¹ con

$$\int_a^b f(x) dx.$$

La lettera usata per indicare la variabile di integrazione è indifferente: al posto di $\int_a^b f(x) dx$, si può scrivere $\int_a^b f(t) dt$ o $\int_a^b f(u) du$ (come per le sommatorie, in cui il nome dato all'indice è ininfluente).

OSSERVAZIONE 1.8. Nella definizione di integrale, non viene fatta nessuna richiesta di positività della funzione. Se la *funzione integranda* f è positiva in $[a, b]$ ed è integrabile, l'integrale da la definizione di area del sottografico di f in $[a, b]$

$$f \geq 0 \quad \Rightarrow \quad \mathcal{A}(\{(x, y) : a \leq x \leq b, 0 \leq y \leq f(x)\}) := \int_a^b f(x) dx.$$

Se f è negativa in tutto o in parte dell'intervallo, il significato dell'integrale non è più quello di un'area: *l'integrale è somma di termini positivi e negativi, gli uni e gli altri in corrispondenza delle zone in cui il grafico è sopra o sotto l'asse x* (vedi Figura 4).

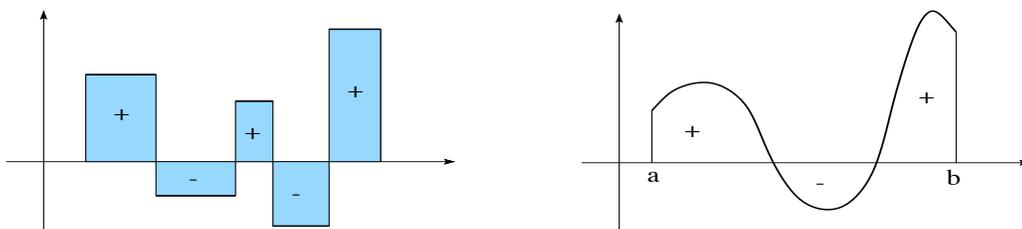


FIGURA 4. L'integrale non è l'area: (a) una funzione costante a tratti, (b) una funzione qualsiasi.

La condizione di integrabilità data dalla definizione è chiara e limpida da un punto di vista di rigore matematico. Meno chiaro è come rispondere concretamente alla domanda: quali classi di funzioni sono integrabili?

PROPOSIZIONE 1.9. Una funzione $f : [a, b] \rightarrow \mathbb{R}$ è integrabile in $[a, b]$ se e solo se

$$(62) \quad \forall \varepsilon > 0 \quad \exists P_\varepsilon \quad t.c. \quad \overline{S}(f; P_\varepsilon) - \underline{S}(f; P_\varepsilon) < \varepsilon.$$

¹Il simbolo dell'integrale è una variante del simbolo di somma indicato da una lunga S come si usava al tempo di Leibnitz. Il simbolo dx è l'erede della lunghezza dell'intervallo $x_i - x_{i-1}$. L'uso della "d" minuscola ricorda che le approssimazioni migliori si ottengono considerando partizioni con ampiezza piccola: $\Delta x := x_i - x_{i-1} \approx dx$.

La condizione espressa in (62) può essere riscritta come

$$\overline{S}(f; P_\varepsilon) - \underline{S}(f; P_\varepsilon) = \sum_{i=1}^n (\beta_i - \alpha_i) (x_i - x_{i-1}) < \varepsilon$$

con α_i e β_i definiti in precedenza. Dato che tale condizione è sufficiente, per dimostrare l'integrabilità di una funzione occorre stimare la differenza $\beta_i - \alpha_i$, cioè la variazione $|f(y) - f(x)|$ per $x, y \in [x_{i-1}, x_i]$ per partizioni con ampiezza piccola.

DIMOSTRAZIONE DELLA PROPOSIZIONE 1.9. La definizione di integrabilità e le proprietà dell'estremo superiore e dell'estremo inferiore indicano che per ogni $\varepsilon > 0$, esistono partizioni P'_ε e P''_ε tali che $\overline{S}(f; P'_\varepsilon) - \underline{S}(f; P''_\varepsilon) < \varepsilon$. Occorre ora dimostrare che la stessa stima vale per una scelta opportuna di una stessa partizione P_ε .

Passo 1. Dimostriamo prima di tutto che

$$(63) \quad \underline{S}(f; P_1) \leq \underline{S}(f; P_2) \quad \text{e} \quad \overline{S}(f; P_2) \leq \overline{S}(f; P_1) \quad \forall P_2 \subset P_1.$$

Dato che $P_2 \subset P_1$, si può costruire P_2 a partire da P_1 aggiungendo un numero finito di punti, basta studiare il caso in cui $P_2 = P_1 \cup \{\xi\}$ (il caso generale si ottiene iterando il procedimento). Inoltre consideriamo solo le somme per difetto, l'altra parte è analoga.

Supponiamo $P_1 = \{x_0, x_1, \dots, x_n\}$ e $\xi \in (x_{k-1}, x_k)$ per un opportuno $k \in \{1, \dots, n\}$. Le espressioni di $\underline{S}(f; P_1)$ e di $\underline{S}(f; P_2)$ coincidono in tutti i termini tranne in quelli relativi all'intervallo $[x_{k-1}, x_k]$ e quindi, indicando con

$$\alpha = \inf_{[x_{k-1}, x_k]} f(x), \quad \alpha' = \inf_{[x_{k-1}, \xi]} f(x), \quad \alpha'' = \inf_{[\xi, x_k]} f(x),$$

vale, dato che $\alpha \leq \alpha'$ e $\alpha \leq \alpha''$,

$$\begin{aligned} \underline{S}(f; P_2) - \underline{S}(f; P_1) &= \left[\alpha'(x_k - \xi) + \alpha''(\xi - x_{k-1}) \right] - \alpha(x_k - x_{k-1}) \\ &= \left[\alpha'(x_k - \xi) + \alpha''(\xi - x_{k-1}) \right] - \left[\alpha(x_k - \xi) + \alpha(\xi - x_{k-1}) \right] \\ &= (\alpha' - \alpha)(x_k - \xi) + (\alpha'' - \alpha)(\xi - x_{k-1}) \geq 0, \end{aligned}$$

Passo 2. Ora dimostriamo che

$$\underline{S}(f; P') \leq \overline{S}(f; P'') \quad \forall P', P'' \text{ partizioni.}$$

Infatti, per costruzione, $\underline{S}(f; P) \leq \overline{S}(f; P)$ per ogni partizione P . Quindi, se consideriamo la partizione $P = P' \cup P''$ e utilizziamo (63):

$$\underline{S}(f; P') \leq \underline{S}(f; P) \leq \overline{S}(f; P) \leq \overline{S}(f; P'').$$

Passo 3. Infine, dimostriamo la Proposizione. Per ogni partizione P , vale

$$0 \leq \inf_P \overline{S}(f; P) - \sup_P \underline{S}(f; P) \leq \overline{S}(f; P) - \underline{S}(f; P)$$

Se per ogni $\varepsilon > 0$ esiste una partizione P_ε che verifica (62), allora

$$0 \leq \inf_P \overline{S}(f; P) - \sup_P \underline{S}(f; P) < \varepsilon \quad \forall \varepsilon > 0,$$

cioè $\inf_P \overline{S}(f; P) = \sup_P \underline{S}(f; P)$, e f è integrabile.

Se invece supponiamo che la funzione f sia integrabile, per ogni $\varepsilon > 0$ esistono due partizioni P'_ε e P''_ε per cui vale $\overline{S}(f; P'_\varepsilon) - \underline{S}(f; P''_\varepsilon) < \varepsilon$. Scegliendo $P_\varepsilon = P'_\varepsilon \cup P''_\varepsilon$, grazie a $\underline{S}(f; P''_\varepsilon) \leq \underline{S}(f; P_\varepsilon)$ e $\overline{S}(f; P_\varepsilon) \leq \overline{S}(f; P'_\varepsilon)$,

$$\overline{S}(f; P_\varepsilon) - \underline{S}(f; P_\varepsilon) \leq \overline{S}(f; P'_\varepsilon) - \underline{S}(f; P''_\varepsilon) < \varepsilon,$$

cioè la condizione (62). □

Se si costruisce una successione di partizioni P_n per cui

$$\lim_{n \rightarrow +\infty} \underline{S}(f; P_n) = \lim_{n \rightarrow +\infty} \overline{S}(f; P_n) = \ell,$$

a maggior ragione si avrà che: per ogni $\varepsilon > 0$ esiste una partizione P_n per cui vale

$$0 \leq \overline{S}(f; P_n) - \underline{S}(f; P_n) = [\overline{S}(f; P_n) - \ell] + [\ell - \underline{S}(f; P_n)] < 2\varepsilon$$

cioè la funzione f è integrabile e l'integrale è il valore ℓ . In particolare, gli Esempi 1.2, 1.3, 1.5 indicano che le funzioni x , x^2 e e^x sono integrabili in $[a, b]$ e che valgono

$$\int_a^b x \, dx = \frac{b^2}{2} - \frac{a^2}{2}, \quad \int_a^b x^2 \, dx = \frac{b^3}{3} - \frac{a^3}{3}, \quad \int_a^b e^x \, dx = e^b - e^a.$$

L'Esempio 1.4 mostra che *esistono anche funzioni non integrabili!*

Il volume di solidi di rotazione. L'integrale è utile per il calcolo delle aree, ma anche, in alcune situazioni speciali, per il calcolo di volumi di solidi. Consideriamo una funzione $f : [a, b] \rightarrow [0, +\infty)$ continua e disegniamone il grafico Γ_f . Se si fa ruotare il grafico Γ_f attorno all'asse x , si ottiene la superficie laterale di un solido Σ . Qual'è il suo volume? Come per il calcolo delle aree, anche nel calcolo dei volumi bisogna partire da una formula nota. Qui, diamo per buona la regola che ci è stata insegnata da bambini: il volume di un cilindro di raggio di base r e altezza h è $\pi r^2 h$.

Ragionando in maniera simile a quanto fatto per la determinazione dell'area di un sottografo, approssimiamo il solido con l'unione di oggetti di cui conosciamo il volume. La scelta più ragionevole è l'unione di cilindri ottenuti tramite una rotazione di rettangoli con lati paralleli agli assi x e y come in Figura 5(b) Un'approssimazione di questo genere di Σ , si ottiene tramite la scelta di una partizione $P = \{a \equiv x_0 <$

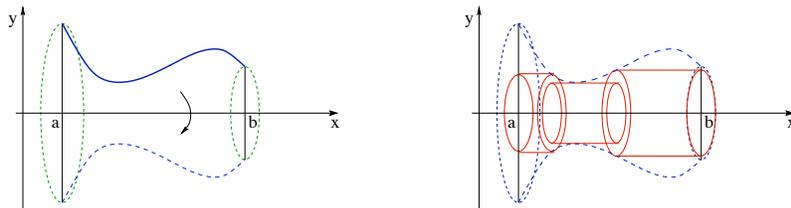


FIGURA 5. (a) Un solido di rotazione; (b) Una sua approssimazione con cilindri.

$x_1 < \dots < x_n = b$ dell'intervallo $[a, b]$ e di n punti ξ_1, \dots, ξ_n tali che $\xi_i \in [x_{i-1}, x_i]$. Quindi l'approssimazione del volume $\mathcal{V}(\Sigma)$ di Σ è data da

$$\mathcal{V}(\Sigma) \approx \sum_{i=1}^n \pi f^2(\xi_i)(x_i - x_{i-1}).$$

Quando l'ampiezza della partizione tende a zero, l'errore di approssimazione tende a zero e si ottiene la formula desiderata

$$\mathcal{V}(\Sigma) = \pi \int_a^b f^2(x) dx.$$

Collaudiamo la formula con un paio di esempi.

ESEMPIO 1.10. Il volume del cono. Un cono di altezza h e raggio di base r , si può ottenere tramite una rotazione del grafico della funzione

$$f(x) = \frac{rx}{h} \quad x \in [0, h].$$

Quindi

$$\mathcal{V}(\Sigma) = \pi \int_0^h \left(\frac{rx}{h}\right)^2 dx = \frac{\pi r^2}{h^2} \int_0^h x^2 dx = \frac{\pi r^2}{h^2} \frac{h^3}{3} = \frac{1}{3} \pi r^2 h,$$

che è la tradizionale formula del volume del cono.

ESEMPIO 1.11. Il volume della sfera. Una sfera di raggio r si ottiene con la rotazione del grafico di

$$f(x) = \sqrt{r^2 - x^2} \quad x \in [-r, r].$$

Applicando la formula del volume

$$\mathcal{V}(\Sigma) = \pi \int_{-r}^r \left(\sqrt{r^2 - x^2}\right)^2 dx = \pi \int_{-r}^r (r^2 - x^2) dx = \pi \left\{ 2r^3 - \frac{2}{3}r^3 \right\} = \frac{4}{3} \pi r^3.$$

Anche questa volta, giustamente, i conti tornano!

ESERCIZIO 1.12. Qual'è il volume del solido ottenuto facendo ruotare il grafico di e^x per $x \in [0, 1]$ attorno all'asse x ?

2. Istruzioni per l'uso

L'integrale di una funzione limitata è l'estremo superiore/inferiore di opportune sommatorie. Per le sommatorie valgono tre proprietà fondamentali: dati $a_1, \dots, a_n, b_1, \dots, b_n, \alpha, \beta \in \mathbb{R}$, si hanno

$$\text{linearità:} \quad \sum_{k=1}^n (\alpha a_k + \beta b_k) = \alpha \sum_{k=1}^n a_k + \beta \sum_{k=1}^n b_k$$

$$\text{additività:} \quad \sum_{k=1}^n a_k = \sum_{k=1}^m a_k + \sum_{k=m+1}^n a_k$$

$$\text{monotonìa:} \quad a_k \leq b_k \quad \Rightarrow \quad \sum_{k=1}^n a_k \leq \sum_{k=1}^n b_k.$$

Di conseguenza, analoghe proprietà vengono ereditate dagli integrali definiti.

Linearità. Una combinazione lineare di funzioni integrabile è integrabile: dati $c_1, c_2 \in \mathbb{R}$ e f, g integrabili in $[a, b]$, la funzione $c_1 f + c_2 g$ è integrabile e vale

$$(64) \quad \int_a^b [c_1 f(x) + c_2 g(x)] dx = c_1 \int_a^b f(x) dx + c_2 \int_a^b g(x) dx$$

Additività. Sia f integrabile in $[a, b]$ e sia $c \in (a, b)$. Allora f è integrabile in $[a, c]$ e in $[c, b]$ e vale

$$(65) \quad \int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx \quad \forall a < c < b.$$

In particolare, allora f è integrabile in ogni sottointervallo di $[a, b]$.

Monotonìa. Per ogni coppia di funzioni f, g integrabili in $[a, b]$,

$$(66) \quad f(x) \leq g(x) \quad \forall x \in [a, b] \quad \Rightarrow \quad \int_a^b f(x) dx \leq \int_a^b g(x) dx.$$

Per ora abbiamo definito $\int_a^b f(x) dx$ solo nel caso $a < b$. E' usanza diffusa definire l'integrale nel caso di $a = b$ o $a > b$, in modo che sia preservata la regola dell'additività. Scrivendo (65) con $c = a$, si ottiene

$$\int_a^b f(x) dx = \int_a^a f(x) dx + \int_a^b f(x) dx,$$

da cui segue

$$\int_a^a f(x) dx := 0,$$

coerente con l'eventuale interpretazione in termini di aree.

Scriviamo (65) per $b = a$, allora

$$0 = \int_a^a f(x) dx = \int_a^c f(x) dx + \int_c^a f(x) dx$$

da cui segue la definizione

$$\int_c^a f(x) dx := - \int_a^c f(x) dx \quad a < c,$$

dove il membro destro ha il significato definito nel paragrafo precedente.

Dalla proprietà di monotonia dell'integrale discende una proprietà che è, sostanzialmente, una "disuguaglianza triangolare per integrali":

PROPOSIZIONE 2.1. *Se la funzione limitata $f : [a, b] \rightarrow \mathbb{R}$ è integrabile in $[a, b]$, allora anche $|f|$ è integrabile in $[a, b]$ e vale*

$$(67) \quad \left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx.$$

Dedichiamoci ora a dimostrare le proprietà dell'integrale. Le proprietà di linearità e di additività contengono due parti. Una prima parte concerne il fatto che dall'integrabilità di alcune funzioni se ne deduce l'integrabilità di certe altre. La seconda parte mette in relazione tra loro gli integrali definiti delle varie funzioni.

Linearità. Siano f e g integrabili in $[a, b]$ e $c_1, c_2 \in \mathbb{R}$. Data $P = \{x_0, x_1, \dots, x_n\}$, partizione di $[a, b]$, battezziamo

$$\alpha_i = \inf_{[x_{i-1}, x_i]} [c_1 f(x) + c_2 g(x)] \quad \text{e} \quad \beta_i = \sup_{[x_{i-1}, x_i]} [c_1 f(x) + c_2 g(x)],$$

e, analogamente,

$$\alpha_i^f = \inf_{[x_{i-1}, x_i]} f(x), \quad \beta_i^f = \sup_{[x_{i-1}, x_i]} f(x), \quad \alpha_i^g = \inf_{[x_{i-1}, x_i]} g(x), \quad \beta_i^g = \sup_{[x_{i-1}, x_i]} g(x).$$

Per ogni $x, y \in (x_{i-1}, x_i)$, si ha

$$\begin{aligned} [c_1 f(x) + c_2 g(x)] - [c_1 f(y) + c_2 g(y)] &= c_1 [f(x) - f(y)] + c_2 [g(x) - g(y)] \\ &\leq |c_1| (\beta_i^f - \alpha_i^f) + |c_2| (\beta_i^g - \alpha_i^g), \end{aligned}$$

quindi, passando all'estremo superiore in x e all'estremo inferiore in y , si deduce che

$$\beta_i - \alpha_i \leq |c_1| (\beta_i^f - \alpha_i^f) + |c_2| (\beta_i^g - \alpha_i^g).$$

Perciò, per ogni partizione P di $[a, b]$,

$$\begin{aligned} \overline{S}(c_1f + c_2g; P) - \underline{S}(c_1f + c_2g; P) &= \sum_{i=1}^n [\beta_i - \alpha_i] (x_i - x_{i-1}) \\ &\leq |c_1| \sum_{i=1}^n [\beta_i^f - \alpha_i^f] (x_i - x_{i-1}) + |c_2| \sum_{i=1}^n [\alpha_i^g - \beta_i^g] (x_i - x_{i-1}) \\ &= |c_1| [\overline{S}(f; P) - \underline{S}(f; P)] + |c_2| [\overline{S}(g; P) - \underline{S}(g; P)]. \end{aligned}$$

Dato che la funzione f è integrabile, grazie alla Proposizione 1.9, esiste una partizione P_1 tale che $\overline{S}(f; P_1) - \underline{S}(f; P_1) < \varepsilon$. Analogamente, esiste una partizione P_2 per cui $\overline{S}(g; P_2) - \underline{S}(g; P_2) < \varepsilon$. Scegliendo la partizione $P = P_1 \cup P_2$, entrambe le disequazioni sono soddisfatte e quindi

$$\overline{S}(c_1f + c_2g; P) - \underline{S}(c_1f + c_2g; P) < (|c_1| + |c_2|) \varepsilon.$$

Dalla Proposizione 1.9, segue che la funzione $c_1f + c_2g$ è integrabile in $[a, b]$.

Resta da dimostrare la formula (64). Data la solita partizione $P = \{x_0, \dots, x_n\}$, scegliamo $\xi_1, \xi_2, \dots, \xi_n$ con $\xi_i \in (x_{i-1}, x_i)$ per $i = 1, \dots, n$. Dato che

$$\underline{S}(f; P) \leq \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) \leq \overline{S}(f; P), \quad \underline{S}(f; P) \leq \int_a^b f(x) dx \leq \overline{S}(f; P),$$

si ha

$$(68) \quad \left| \int_a^b f(x) dx - \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) \right| \leq \overline{S}(f; P) - \underline{S}(f; P)$$

Aggiungendo e sottraendo i termini opportuni, si ricava

$$\begin{aligned} &\left| \int_a^b [c_1f + c_2g] dx - c_1 \int_a^b f dx - c_2 \int_a^b g dx \right| \\ &\leq \left| \int_a^b [c_1f + c_2g] dx - \sum_{i=1}^n [c_1f(\xi_i) + c_2g(\xi_i)] (x_i - x_{i-1}) \right| \\ &\quad + |c_1| \left| \int_a^b f dx - \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) \right| + |c_2| \left| \int_a^b g dx - \sum_{i=1}^n g(\xi_i)(x_i - x_{i-1}) \right|. \end{aligned}$$

Utilizzando (68), si deduce la disuguaglianza

$$\begin{aligned} &\left| \int_a^b [c_1f + c_2g] dx - c_1 \int_a^b f dx - c_2 \int_a^b g dx \right| \\ &\leq \overline{S}(c_1f + c_2g; P) - \underline{S}(c_1f + c_2g; P) \\ &\quad + |c_1| [\overline{S}(f; P) - \underline{S}(f; P)] + |c_2| [\overline{S}(g; P) - \underline{S}(g; P)]. \end{aligned}$$

Scegliendo un partizione P per cui

$$\begin{aligned}\overline{S}(f; P) - \underline{S}(f; P) &< \varepsilon, & \overline{S}(g; P) - \underline{S}(g; P) &< \varepsilon, \\ \overline{S}(c_1 f + c_2 g; P) - \underline{S}(c_1 f + c_2 g; P) &< \varepsilon,\end{aligned}$$

si ottiene

$$\left| \int_a^b [c_1 f + c_2 g] dx - c_1 \int_a^b f dx - c_2 \int_a^b g dx \right| \leq (1 + |c_1| + |c_2|) \varepsilon.$$

Dato che ε è arbitrariamente piccolo, non c'è scampo: (64) è dimostrata.

La linearità garantisce, in particolare, che la somma di funzioni integrabili è integrabile. Lo stesso vale per il prodotto, come enunciato nell'esercizio che segue (ma non è vero che l'integrale del prodotto è il prodotto degli integrali!).

ESERCIZIO 2.2. *Se f e g sono integrabili in $[a, b]$, anche il prodotto fg lo è.*

Soluzione. Le funzioni f e g sono limitate, quindi esistono $M_f, M_g > 0$ tali che $|f(x)| \leq M_f$ e $|g(x)| \leq M_g$ per ogni x . Data la partizione $P = \{x_0, x_1, \dots, x_n\}$ indichiamo con

$$\alpha_i^{fg} = \inf_{[x_{i-1}, x_i]} f(x)g(x) \quad \text{e} \quad \beta_i^{fg} = \sup_{[x_{i-1}, x_i]} f(x)g(x),$$

con significati analoghi per $\alpha_i^f, \alpha_i^g, \beta_i^f$ e β_i^g . Dati $x, y \in (x_{i-1}, x_i)$, si ha

$$f(x)g(x) - f(y)g(y) = [f(x) - f(y)]g(x) + f(y)[g(x) - g(y)] \leq M_g(\beta_i^f - \alpha_i^f) + M_f(\beta_i^g - \alpha_i^g).$$

Prendendo l'estremo superiore in x e l'estremo superiore in y , si ottiene

$$\beta_i^{fg} - \alpha_i^{fg} \leq M_g(\beta_i^f - \alpha_i^f) + M_f(\beta_i^g - \alpha_i^g).$$

Perciò

$$\overline{S}(fg; P) - \underline{S}(fg; P) \leq M_g [\overline{S}(f; P) - \underline{S}(f; P)] + M_f [\overline{S}(g; P) - \underline{S}(g; P)].$$

Dato che il termine a secondo membro può essere reso arbitrariamente piccolo per P opportuno, la conclusione segue dalla solita Proposizione 1.9.

Additività. Se si pensa all'idea geometrica di partenza, la proprietà di additività sembra abbastanza naturale: per calcolare l'area possiamo dividere la regione in due parti e sommare i valori delle aree delle due sottoregioni. Ma qui la cosa è diversa: prima di tutto abbiamo una definizione analitica da rispettare e ogni affermazione deve discendere rigorosamente da quella definizione. In più c'è un particolare non banale: chi garantisce che se una funzione è integrabile in $[a, b]$, allora è anche integrabile in $[a, c]$ e $[c, b]$ per $c \in (a, b)$? Seppur ragionevole, quest'affermazione è tutta da verificare.

Per comodità, introduciamo la funzione $\chi_E : \mathbb{R} \rightarrow \mathbb{R}$

$$\text{funzione caratteristica di } E: \quad \chi_E(x) := \begin{cases} 1 & x \in E, \\ 0 & x \notin E. \end{cases}$$

Se l'insieme E è un intervallo, la funzione χ_E è integrabile. Quindi, se $f : [a, b] \rightarrow \mathbb{R}$ è integrabile in $[a, b]$ e $c \in [a, b]$, allora sono integrabili anche le funzioni prodotto $f\chi_{[a,c]}(x)$ e $f\chi_{[c,b]}(x)$ e vale l'uguaglianza

$$f(x) = f(x)\chi_{[a,c]}(x) + f(x)\chi_{[c,b]}(x),$$

Inoltre, dato che moltiplicare per una funzione caratteristica si traduce nel “troncare a zero” la funzione f fuori dall'insieme corrispondente, si hanno

$$\int_a^b f(x)\chi_{[a,c]}(x) dx = \int_a^c f(x) dx \quad \text{e} \quad \int_a^b f(x)\chi_{[c,b]}(x) dx = \int_c^b f(x) dx.$$

Quindi, grazie alla linearità,

$$\int_a^b f(x) dx = \int_a^b f(x)\chi_{[a,c]}(x) dx + \int_a^b f(x)\chi_{[c,b]}(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx.$$

Monotonia. Dimostrare la monotonia dell'integrale è particolarmente facile. Se $f(x) \leq g(x)$ per ogni $x \in [a, b]$, allora

$$\inf_{x \in [x_{i-1}, x_i]} f(x) \leq \inf_{x \in [x_{i-1}, x_i]} g(x) \quad \forall [x_{i-1}, x_i] \subset [a, b].$$

Quindi, per ogni partizione P di $[a, b]$, vale $\underline{S}(f; P) \leq \underline{S}(g; P)$, e passando all'estremo superiore si ottiene la conclusione.

DIMOSTRAZIONE DELLA PROPOSIZIONE 2.1. La parte più complicata della dimostrazione sta nel verificare che effettivamente la funzione $|f|$ sia integrabile.

Il problema è sempre lo stesso: data $P = \{x_0, \dots, x_n\}$ partizione di $[a, b]$ e posti $A_k = \inf_{[x_{k-1}, x_k]} |f(x)|$, $B_k = \sup_{[x_{k-1}, x_k]} |f(x)|$, quale stima possiamo recuperare per $B_k - A_k$?

Siano $\alpha_k = \inf_{[x_{k-1}, x_k]} f(x)$, $\beta_k = \sup_{[x_{k-1}, x_k]} f(x)$, dati x, y in (x_{k-1}, x_k) , si ha

$$||f(x)| - |f(y)|| \leq |f(x) - f(y)| \leq \beta_k - \alpha_k,$$

e prendendo l'estremo superiore in x e l'inferiore in y si ottiene $B_k - A_k \leq \beta_k - \alpha_k$.

Con questa stima alla mano, si deduce che

$$\overline{S}(|f|; P) - \underline{S}(|f|; P) \leq \sum_{k=1}^n (\beta_k - \alpha_k)(x_k - x_{k-1}) = \overline{S}(f; P) - \underline{S}(f; P).$$

e grazie alla Proposizione 1.9 si arriva alla conclusione.

La stima finale segue da $\pm f \leq |f|$, che, grazie alla monotonia dell'integrale, implica

$$\pm \int_a^b f(x) dx \leq \int_a^b |f(x)| dx,$$

da cui segue la conclusione. \square

ESERCIZIO 2.3. *Date f e g integrabili, dimostrare che anche $f_+(x) := \max\{f(x), 0\}$, $f_-(x) := -\min\{f(x), 0\}$, $\max\{f(x), g(x)\}$ e $\min\{f(x), g(x)\}$ sono integrabili.*

Una volta dato senso al concetto di integrale e determinate le proprietà principali, bisogna dedicarsi a determinare un certo numero di funzioni che siano effettivamente integrabili, altrimenti l'oggetto appena definito risulterebbe sostanzialmente inutile. Come si è detto, per dimostrare l'integrabilità di f , basta mostrare

$$\forall \varepsilon > 0 \quad \exists P_\varepsilon \text{ partizione tale che } \overline{S}(f; P_\varepsilon) - \underline{S}(f; P_\varepsilon) \leq \varepsilon.$$

Dunque seguiremo questa strategia: fissata una partizione P , mostreremo che la differenza $\overline{S}(f; P) - \underline{S}(f; P)$ può essere resa arbitrariamente piccola, a patto di scegliere una partizione P la cui ampiezza $|P|$ sia sufficientemente piccola.

La prima classe che consideriamo è quella delle funzioni monotone in un intervallo chiuso e limitato. Tali funzioni sono sempre limitate (perché?).

TEOREMA 2.4. *Sia $f : [a, b] \rightarrow \mathbb{R}$ monotona, allora è integrabile in $[a, b]$.*

DIMOSTRAZIONE. Supponiamo che f sia una funzione crescente. Data una partizione $P = \{x_0, \dots, x_n\}$, gli estremi inferiori e superiori di f in $[x_{i-1}, x_i]$ sono

$$\alpha_i = \inf_{[x_{i-1}, x_i]} f(x) = f(x_{i-1}) \quad \text{e} \quad \beta_i = \sup_{[x_{i-1}, x_i]} f(x) = f(x_i).$$

Le somme integrali per difetto e per eccesso sono date da

$$\underline{S}(f; P) = \sum_{i=1}^n f(x_{i-1})(x_i - x_{i-1}) \quad \text{e} \quad \overline{S}(f; P) = \sum_{i=1}^n f(x_i)(x_i - x_{i-1}).$$

e la loro differenza è stimata da

$$\overline{S}(f; P) - \underline{S}(f; P) = \sum_{i=1}^n [f(x_i) - f(x_{i-1})] (x_i - x_{i-1})$$

Indicando con $|P|$ l'ampiezza di P , cioè il massimo delle lunghezze $x_{i-1} - x_i$,

$$\overline{S}(f; P) - \underline{S}(f; P) \leq |P| \sum_{i=1}^n [f(x_i) - f(x_{i-1})] = |P| [f(b) - f(a)].$$

Per ogni $\varepsilon > 0$, è possibile scegliere $|P|$ sufficientemente piccola, in modo che la differenza $\overline{S}(f; P) - \underline{S}(f; P)$ sia minore di ε , pertanto, grazie (ancora una volta!) alla Proposizione 1.9, la funzione è integrabile. \square

A partire dalle funzioni monotone e grazie al fatto che combinazioni lineari di funzioni integrabili sono integrabili, è possibile costruire una classe ancora più ampia di funzioni. Ad esempio, sia f una funzione definita in $[a, b]$ tale che, per qualche $c \in [a, b]$, la funzione f è crescente in $[a, c]$ e decrescente in $[c, b]$. Allora la funzione f si può riscrivere come differenza di due funzioni crescenti (vedi Figura 6):

$$f = f_1 - f_2$$

dove

$$f_1(x) = \begin{cases} f(x) & x \in [a, c] \\ f(c) & x \in [c, b] \end{cases} \quad \text{e} \quad f_2(x) = \begin{cases} 0 & x \in [a, c] \\ f(c) - f(x) & x \in [c, b] \end{cases}$$

Quindi, anche funzioni con un cambio di monotonia sono integrabili. Con una costruzione

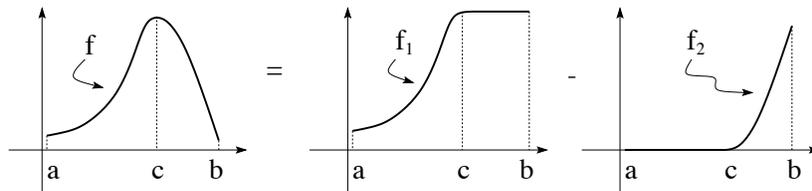


FIGURA 6. Una funzione f con un cambio di monotonia, decomposta come differenza delle funzioni crescenti f_1 e f_2 .

analogha, si mostra che *tutte le funzioni con un numero finito di cambi di monotonia sono integrabili*. Ad esempio, tutti i polinomi e, più in generale, tutte le funzioni razionali sono integrabili.

ESERCIZIO 2.5. *Conoscete una funzione definita in un intervallo limitato che abbia un numero infinito di cambi di monotonia?*

La seconda classe che consideriamo è quella delle *funzioni lipschitziane*, cioè delle funzioni $f : [a, b] \rightarrow \mathbb{R}$ tali che esista $L > 0$ per cui

$$\exists L > 0 \quad \text{t.c.} \quad |f(x) - f(y)| \leq L|x - y| \quad \forall x, y \in [a, b].$$

TEOREMA 2.6. *Sia $f : [a, b] \rightarrow \mathbb{R}$ lipschitziana, allora è integrabile in $[a, b]$.*

DIMOSTRAZIONE. Fissata la beneamata partizione $P = \{x_0, x_1, \dots, x_n\}$,

$$\bar{S}(f; P) - \underline{S}(f; P) = \sum_{i=1}^n (\beta_i - \alpha_i)(x_i - x_{i-1}),$$

con il solito significato per α_i e β_i . Dato che f è lipschitziana, essa è anche continua in $[a, b]$, quindi ammette massimo e minimo in ogni intervallo $[x_{i-1}, x_i]$ e vale

$$\alpha_i = \inf_{(x_{i-1}, x_i)} f = \min_{[x_{i-1}, x_i]} f(x) = f(\eta_i) \quad \text{e} \quad \beta_i = \sup_{(x_{i-1}, x_i)} f = \max_{[x_{i-1}, x_i]} f(x) = f(\xi_i)$$

con η_i e ξ_i , rispettivamente, un punto di massimo ed uno di minimo della funzione f in $[x_{i-1}, x_i]$. Sostituendo nella relazione precedente, si ottiene

$$\overline{S}(f; P) - \underline{S}(f; P) = \sum_{i=1}^n (f(\xi_i) - f(\eta_i))(x_i - x_{i-1}) \leq L \sum_{i=1}^n |\xi_i - \eta_i|(x_i - x_{i-1}).$$

Dato che $\xi_i, \eta_i \in [x_{i-1}, x_i]$, la differenza $|\xi_i - \eta_i|$ è minore o uguale dell'ampiezza $|P|$ della partizione. Quindi

$$\overline{S}(f; P) - \underline{S}(f; P) \leq L|P| \sum_{i=1}^n (x_i - x_{i-1}) \leq L(b-a)|P|.$$

Scegliendo P tale che $|P| < \varepsilon/L(b-a)$, la differenza $\overline{S}(f; P) - \underline{S}(f; P)$ è strettamente minore di ε e, grazie alla Proposizione 1.9, segue la conclusione. \square

Più in generale si può dimostrare il fondamentale

TEOREMA 2.7. *Sia $f : [a, b] \rightarrow \mathbb{R}$ continua, allora f è integrabile in $[a, b]$.*

La dimostrazione passa per il concetto di *uniforme continuità*. Non daremo qui ulteriori dettagli sulla questione.

ESERCIZIO 2.8. *Dimostrare che la composizione di una funzione lipschitziana con una funzione integrabile dà luogo ad una funzione integrabile.*

3. Il Teorema della media integrale

Tutte le funzioni $f : [a, b] \rightarrow \mathbb{R}$ integrabili sono, per definizione, funzioni limitate. In particolare questo significa che data una funzione integrabile, il suo integrale definito può sempre essere stimato in maniera “rude”: per le proprietà di monotonia dell'integrale, se $m \leq f(x) \leq M$ per ogni $x \in [a, b]$,

$$m(b-a) = \int_a^b m \, dx \leq \int_a^b f(x) \, dx \leq \int_a^b M \, dx = M(b-a),$$

Questa formula è intuitivamente ovvia: se pensiamo ad una funzione non negativa e all'integrale come area, le quantità $M(b-a)$ e $m(b-a)$ rappresentano le aree di una rettangolo circoscritto ed inscritto nel sottografico di f (vedi Figura 7).

La formula precedente si può riscrivere come $m \leq \langle f \rangle \leq M$ dove $\langle f \rangle$, detto **media integrale**² di f in $[a, b]$, è definito da

$$(69) \quad \langle f \rangle := \frac{1}{b-a} \int_a^b f(x) \, dx.$$

²La *media aritmetica* dei numeri f_1, f_2, \dots, f_n è, per definizione $(f_1 + f_2 + \dots + f_n)/n$. La media integrale di una funzione in un intervallo è, in un certo senso, limite di medie aritmetiche di valori assunti dalla funzione. Infatti, data la partizione di $[a, b]$ in sottointervalli di lunghezza $\Delta x_i = (b-a)/n$

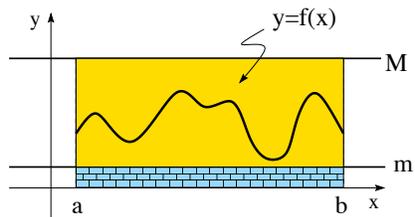


FIGURA 7. Significato geometrico della stima $m(b-a) \leq \int_a^b f(x) dx \leq M(b-a)$.

Quindi, espresso a parole, *la media integrale di una funzione f è sempre compresa tra un qualsiasi minorante ed un qualsiasi maggiorante di f .*

Se f è continua in un intervallo $[a, b]$, si può dire qualcosa di più. Per il teorema di Weierstrass, esistono due punti $\xi, \eta \in [a, b]$ tali che

$$\min_{[a,b]} f(x) = f(\xi) \leq f(x) \leq f(\eta) = \max_{[a,b]} f(x) \quad \forall x \in [a, b].$$

Quindi $\langle f \rangle$, definita in (69) è compresa tra il massimo ed il minimo della funzione f in $[a, b]$. Grazie al Teorema del valore intermedio, si può concludere che *per una funzione f continua in $[a, b]$, la media integrale fa sempre parte dell'insieme immagine $f([a, b])$.*

TEOREMA 3.1. Teorema della Media Integrale. *Sia $f : [a, b] \rightarrow \mathbb{R}$ continua. Allora esiste $\xi \in [a, b]$ tale che $f(\xi) = \langle f \rangle$ con $\langle f \rangle$ definito in (69).*

Nel caso di una funzione non negativa f , il teorema equivale ad affermare che esiste un rettangolo di base $[a, b]$ ed una altezza $f(\xi)$ opportuna con la stessa area del sottografico di f in $[a, b]$.

Non è particolarmente sconvolgente (ma più avanti servirà) osservare che la formula (69) vale anche nel caso $b < a$, infatti

$$\frac{1}{b-a} \int_a^b f(x) dx = \frac{1}{a-b} \int_b^a f(x) dx,$$

e a quest'ultimo termine si può applicare il Teorema della Media Integrale.

per ogni i e scelti $\xi_i \in (x_{i-1}, x_i)$, la media aritmetica di $f(\xi_1), \dots, f(\xi_n)$ è pari a

$$\frac{f(\xi_1) + \dots + f(\xi_n)}{n} = \frac{1}{b-a} \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}).$$

Per $n \rightarrow \infty$, il termine a destra converge proprio alla media integrale di f in $[a, b]$.

CONTROESEMPIO 3.2. Se la funzione f non è continua in tutto $[a, b]$, non è detto che valga la conclusione del Teorema 3.1. Ad esempio, si consideri la funzione

$$f(x) = \operatorname{sgn} x = \begin{cases} -1 & x < 0, \\ 0 & x = 0, \\ +1 & x > 0, \end{cases}$$

nell'intervallo $[-1, 2]$. Allora f è integrabile (per almeno due motivi... quali?)

$$\mu = \frac{1}{2 - (-1)} \int_{-1}^2 \operatorname{sgn} x \, dx = \frac{1}{3} (-1 + 2) = \frac{1}{3},$$

che non fa parte dell'immagine della funzione $\operatorname{sgn} x$.

ESERCIZIO 3.3. Siano $f \in C([a, b])$ e $p \in C([a, b])$ tale che $p(x) > 0$ per ogni x . Dimostrare che esiste $\xi \in [a, b]$ tale che

$$f(\xi) = \frac{\int_a^b f(x)p(x) \, dx}{\int_a^b p(x) \, dx}.$$

Soluzione. Se m e M indicano il minimo ed il massimo di f in $[a, b]$, grazie al fatto che $p(x) > 0$, si ha $mp(x) \leq f(x)p(x) \leq Mp(x)$ per ogni x . Integrate in $[a, b]$, mescolate e concludete...

4. Il Teorema fondamentale del calcolo integrale

Una volta fissata la funzione f , l'integrale definito è una funzione degli estremi di integrazione a e b . Per studiare questa dipendenza, supponiamo l'estremo inferiore fissato al valore a e indichiamo l'estremo superiore (variabile) con x : consideriamo, quindi, la funzione integrale

$$(70) \quad \phi(x) = \int_a^x f(t) \, dt.$$

La formula (70) può essere utilizzata per “generare nuove funzioni” a partire da una funzione integrabile f . Ad esempio, si può definire³

$$\phi(x) := \int_1^x \frac{1}{t} \, dt \quad x > 0,$$

che ha perfettamente senso, dato che la funzione $\frac{1}{t}$ è una funzione continua in ogni intervallo $[a, b]$ con $0 < a \leq b$.

PROPOSIZIONE 4.1. Se f è integrabile in $[a, b]$, la funzione ϕ , definita in (70), è lipschitziana.

³Tale funzione, come vedremo, si dimostra essere uguale al logaritmo naturale e la formula precedente può quindi essere scelta come definizione analitica di $\ln x$.

DIMOSTRAZIONE. Dato che f è integrabile, essa è anche limitata. Sia $M > 0$ tale che $|f(t)| \leq M$ per ogni $t \in [a, b]$, allora,

$$\text{se } x < y, \quad |\phi(x) - \phi(y)| = \left| \int_x^y f(t) dt \right| \leq \int_x^y |f(t)| dt \leq M(y - x);$$

$$\text{se } x > y, \quad |\phi(x) - \phi(y)| = \left| - \int_y^x f(t) dt \right| \leq \int_y^x |f(t)| dt \leq M(x - y).$$

Quindi

$$|f(t)| \leq M \quad \implies \quad |\phi(x) - \phi(y)| \leq M|x - y|.$$

In particolare, se f è integrabile, la funzione ϕ è una funzione continua. \square

ESERCIZIO 4.2. Consideriamo la funzione $f(x) = \operatorname{sgn} x$. Sapete riconoscere la funzione ϕ data in (70) in questo caso?

ESERCIZIO 4.3. Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ integrabile in ogni intervallo $[a, b]$ di \mathbb{R} e sia

$$\phi(x) := \int_0^x f(t) dt$$

- (i) Dimostrare che, se $f(x) \geq 0$ per ogni $x \in \mathbb{R}$, allora ϕ è non decrescente.
- (ii) Disegnare qualitativamente il grafico di ϕ nel caso in cui $xf(x) > 0$ per ogni $x \neq 0$.

Una volta stabilito che la funzione integrale ϕ è sempre lipschitziana, è naturale domandarsi se essa sia anche derivabile. Consideriamone il rapporto incrementale

$$\frac{\phi(x+h) - \phi(x)}{h} = \frac{1}{h} \left[\int_\alpha^{x+h} f(t) dt - \int_\alpha^x f(t) dt \right] = \frac{1}{h} \int_x^{x+h} f(t) dt$$

(l'ultima uguaglianza è conseguenza delle proprietà di additività dell'integrale). Supponendo che la funzione integranda f sia continua, è possibile applicare il Teorema della Media Integrale, Teorema 3.1, e riscrivere il rapporto incrementale come

$$\frac{\phi(x+h) - \phi(x)}{h} = f(\xi)$$

con ξ compreso tra x e $x+h$. Passando al limite $h \rightarrow 0$, dato che $\xi \rightarrow x$, si ha

$$\phi'(x) = \lim_{h \rightarrow 0} \frac{\phi(x+h) - \phi(x)}{h} = \lim_{h \rightarrow 0} f(\xi) = f(x).$$

Abbiamo quindi dimostrato il seguente risultato.

TEOREMA 4.4. Sia $f \in C([a, b])$, $\alpha \in [a, b]$ e $\phi(x) = \int_\alpha^x f(t) dt$ per $x \in [a, b]$. Allora ϕ è derivabile in $[a, b]$ e $\phi'(x) = f(x)$.

Una prima conseguenza (pratica) notevole è che, dato che siamo in grado di calcolarne la derivata, possiamo dedurre molte proprietà qualitative importanti anche per una funzione che non sia espressa direttamente tramite funzioni elementari, ma come integrale di una funzione elementare.

ESEMPIO 4.5. Consideriamo la funzione

$$\text{funzione degli errori: } \operatorname{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt,$$

Dal Teorema 4.4 deduciamo che $D(\operatorname{Erf}(x)) = \frac{2}{\sqrt{\pi}} e^{-x^2} > 0$, quindi $\operatorname{Erf}(x)$ è strettamente crescente. Con una tecnologia più avanzata di quella che ci è disponibile a questo livello, è possibile dimostrare che $\operatorname{Erf}(\pm\infty) = \pm 1$.

Il Teorema 4.4 risolve anche un problema interessante:

data f , trovare una funzione F che risolva l'equazione $F' = f$.

L'equazione $F' = f$ è un'equazione differenziale in cui il dato è la funzione f e l'incognita è la funzione F . Una soluzione F di questa equazione si dice **primitiva** di f . Il Teorema 4.4 afferma che se $f \in C([a, b])$ il problema $F' = f$ ammette almeno una soluzione (data dalla funzione integrale ϕ), cioè esiste sempre almeno una primitiva. Da questo punto di vista *si può interpretare l'operazione di integrazione come l'operazione inversa della derivazione*. Tale operazione di inversione è univocamente definita? In altri termini, data una funzione f , quante primitive esistono?

TEOREMA 4.6. *Sia $f : [a, b] \rightarrow \mathbb{R}$ e siano F e G due sue primitive. Allora esiste $c \in \mathbb{R}$ tale che*

$$F(x) - G(x) = c \quad \text{per ogni } x \in [a, b].$$

DIMOSTRAZIONE. La derivata della funzione differenza $F - G$ è nulla:

$$(F(x) - G(x))' = F'(x) - G'(x) = f(x) - f(x) = 0.$$

Per quanto già visto, la differenza $F - G$ deve essere costante. □

Quindi, se $f \in C([a, b])$, l'equazione $F' = f$ è completamente risolta: tutte le soluzioni sono della forma

$$\int_{\alpha}^x f(t) dt + c \quad c \in \mathbb{R}.$$

La classe delle primitive della funzione f si indica con

$$\int f(x) dx,$$

e si chiama **integrale indefinito** di f . Si noti bene che l'integrale indefinito di una funzione indica una classe di funzioni, e non una singola funzione.

Ad eterna memoria, sintetizziamo i due risultati enunciati in un'unico Teorema.

TEOREMA 4.7. Teorema Fondamentale del Calcolo Integrale. *Data $f \in C([a, b])$, le soluzioni dell'equazione differenziale $F' = f$ sono tutte e sole della forma*

$$F(x) = \int_{\alpha}^x f(t) dt + c, \quad \text{con } \alpha \in [a, b] \quad \text{e } c \in \mathbb{R}.$$

Se si cerca una primitiva F di una funzione f con la richiesta aggiuntiva che la funzione F valga in un punto assegnato x_0 un valore dato y_0 , cioè se si vuole risolvere

$$(71) \quad \text{dati } f \in C([a, b]), x_0 \in [a, b], y_0 \in \mathbb{R}, \text{ trovare } F \text{ tale che } \begin{cases} F'(x) = f(x), \\ F(x_0) = y_0, \end{cases}$$

la soluzione F esiste, è unica ed è data da

$$(72) \quad F(x) = y_0 + \int_{x_0}^x f(t) dt.$$

Il problema (71) rientra nella classe dei *problemi di Cauchy* per equazioni differenziali.

OSSERVAZIONE 4.8. Il problema di Cauchy può essere interpretato in termini di moto di un particella. Se $f(x)$ è la velocità della particella all'istante x e y_0 la sua posizione all'istante iniziale x_0 , la soluzione $F = F(x)$, definita in (72), del problema di Cauchy (71), rappresenta la posizione della particella all'istante x .

Primitive e calcolo degli integrali definiti. Il Teorema fondamentale del calcolo ha una conseguenza interessante che riguarda il calcolo esplicito di integrali definiti. Supponiamo di voler calcolare

$$(73) \quad \int_a^b f(t) dt.$$

e supponiamo di conoscere già (per altre vie) una primitiva della funzione f , cioè una funzione F tale che $F' = f$. Sappiamo che anche la funzione integrale definita in (70) è una primitiva di f e, quindi, per il Teorema Fondamentale del Calcolo, differisce da F per una costante, cioè $\phi(x) = F(x) + c$ per qualche $c \in \mathbb{R}$. La costante c può essere determinata, calcolando in $x = a$:

$$0 = \phi(a) = F(a) + c \quad \implies \quad c = -F(a).$$

Si deduce quindi che $\phi(x) = F(x) - F(a)$ e quindi

$$\int_a^b f(t) dt = \phi(b) = F(b) - F(a).$$

Quindi, se si conosce una primitiva F della funzione f , l'integrale definito di f in $[a, b]$ è uguale alla differenza dei valori della primitiva in b e in a , cioè

$$F' = f \quad \Longrightarrow \quad \int_a^b f(t) dt = F(b) - F(a).$$

La differenza $F(b) - F(a)$ si indica anche con $F(x)\Big|_a^b$, o $[F(x)]_a^b$.

ESEMPIO 4.9. Consideriamo la funzione $f(x) = x^2$. Dato che $D(x^3) = 3x^2$, si ha

$$D\left(\frac{1}{3}x^3\right) = x^2,$$

quindi una primitiva di x^2 è $x^3/3$. Otteniamo perciò

$$\int_a^b x^2 dx = \frac{1}{3}x^3\Big|_a^b = \frac{b^3}{3} - \frac{a^3}{3},$$

che è proprio la formula calcolata ad inizio Capitolo. Più in generale, dato che

$$D\left(\frac{1}{n+1}x^{n+1}\right) = x^n,$$

vale la formula

$$\int_a^b x^n dx = \frac{1}{n+1}x^{n+1}\Big|_a^b = \frac{1}{n+1}(b^{n+1} - a^{n+1}).$$

CAPITOLO 9

Zoologia dell'integrazione

Per iniziare, ricordiamo i fatti principali che abbiamo visto sugli integrali indefiniti.

– Data una funzione f , si dice che F è una **primitiva** di f se $F' = f$; la famiglia di primitive di una funzione f si indica con il simbolo

$$\int f(x) dx.$$

Se la funzione f è considerata in un intervallo, tutte le primitive di f sono uguali a meno di una costante additiva, cioè, data F primitiva di f ,

$$\int f(x) dx = F(x) + C \quad C \in \mathbb{R}.$$

– Se la funzione F è una primitiva della funzione f , allora vale

$$F' = f \quad \implies \quad \int_a^b f(x) dx = F(b) - F(a).$$

In questo Capitolo ci poniamo il problema di determinare *esplicitamente* primitive F di una funzione data f , per opportune classi di funzioni. Cosa si intende qui con “primitive esplicite”? A partire dalle operazioni elementari (addizione, sottrazione, moltiplicazione, divisione) e dalle funzioni trigonometriche ed esponenziali, formando inverse e composte di queste funzioni, è possibile costruire una classe molto ampia di funzioni che possiamo descrivere come “funzioni esplicite”.

Per quanto riguarda l'operazione di derivazione, la derivata di una funzione esplicita è essa stessa una funzione esplicita. Al contrario, per l'integrazione la situazione è differente: *non è vero che tutti gli integrali delle funzioni esplicite si possano scrivere in termini di funzioni esplicite* (ad esempio, non è possibile esprimere in forma “esplicita”, le primitive di e^{-x^2}). Questo risultato può suonare sorprendente, ma è un fatto della vita. Prendere o lasciare.

Una prima classe di funzioni che sono esplicitamente integrabili si determina a partire dal *teorema fondamentale del calcolo*, che, come noto, afferma

$$F'(x) = f(x) \quad \implies \quad \int f(x) dx = F(x) + \text{costante}.$$

In particolare, questa proprietà indica che *ad ogni regola di derivazione corrisponde una regola di integrazione*. Ad esempio

$$D(x^{\alpha+1}) = (\alpha + 1)x^\alpha \quad \Rightarrow \quad \int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha + 1} + \text{costante} \quad \forall \alpha \neq -1.$$

Allo stesso modo si ottengono altre formule elementari:

$$\begin{aligned} \int \sin x dx &= -\cos x + C, & \int \cos x dx &= \sin x + C, \\ \int e^x dx &= e^x + C, & \int \frac{dx}{x} &= \ln |x| + C, \\ \int \frac{dx}{1+x^2} &= \arctan x + C, & \int a^x dx &= \frac{a^x}{\ln a} + C, \\ \int \frac{dx}{\cos^2 x} &= \tan x + C, & \int \frac{dx}{\sin^2 x} &= -\cot x + C, \\ \int \frac{dx}{\sqrt{1-x^2}} &= \arcsin x + C, & \int \frac{dx}{\sqrt{1-x^2}} &= -\arccos x + C, \\ \int \sinh x dx &= \cosh x + C & \int \cosh x dx &= \sinh x + C. \end{aligned}$$

Inoltre, grazie alla linearità dell'integrale, anche combinazioni lineari di funzioni di cui si conosce la primitiva, possono essere integrate esplicitamente. Ad esempio,

$$\int (1 + 2x + 3e^x) dx = \int 1 dx + 2 \int x dx + 3 \int e^x dx = x + x^2 + 3e^x + C.$$

Tramite queste formule è possibile calcolare il valore di certi integrali definiti, senza bisogno di passare per le approssimazioni con somme integrali per eccesso e per difetto.

Passiamo ora a sviluppare i due metodi principali da affiancare alle formule di integrazione elementari: l'integrazione per sostituzione e l'integrazione per parti. Entrambi, in sostanza, discendono da formule di derivazione: il primo discende dalla derivazione di funzione composta, il secondo dalla derivazione della funzione prodotto.

1. Metodo di sostituzione

Il *metodo di sostituzione* consiste nell'introduzione di una nuova variabile, cioè, moralmente, nel "cambiare punto di vista" e osservare lo stesso oggetto da un'altra posizione. La formula di derivazione di funzione composta assicura che

$$(74) \quad (F(\phi(u)))' = F'(\phi(u))\phi'(u).$$

Tale formula, letta in termini di integrazione, diviene

$$(75) \quad \int F'(\phi(x))\phi'(x) dx = F(\phi(x)) + \text{costante}$$

Ad esempio,

$$\int 2x \cos(x^2) dx = \sin(x^2) + \text{costante},$$

dove $F(s) = \sin s$ e $\phi(x) = x^2$. Vediamo un altro esempio. Calcoliamo

$$\int \frac{\ln x}{x} dx.$$

Dato che $D(\ln x) = 1/x$, qui $F(s) = s$ e $\phi(x) = \ln x$. Quindi

$$\int \frac{\ln x}{x} dx = \int \ln x (\ln x)' dx = \frac{1}{2}(\ln x)^2 + \text{costante}.$$

Ma come individuare una decomposizione della funzione integranda come in (75)? Occorre esercizio ed esperienza (anche una certa dose di intuizione non guasta!).

Vediamo la formula di sostituzione per integrali definiti: integriamo la formula (74) nell'intervallo $[\alpha, \beta]$,

$$F(\phi(\beta)) - F(\phi(\alpha)) = \int_{\alpha}^{\beta} (F(\phi(u)))' du = \int_{\alpha}^{\beta} F'(\phi(u))\phi'(u) du.$$

Ponendo $a = \phi(\alpha)$ e $b = \phi(\beta)$,

$$F(\phi(\beta)) - F(\phi(\alpha)) = F(b) - F(a) = \int_a^b F'(x) dx.$$

Quindi otteniamo la formula

$$\int_{\alpha}^{\beta} F'(\phi(u))\phi'(u) du = \int_a^b F'(x) dx,$$

che, chiamando $f = F'$, può essere riscritta come

$$(76) \quad \int_{\alpha}^{\beta} f(\phi(u))\phi'(u) du = \int_a^b f(x) dx \quad \text{con} \quad \begin{cases} a = \phi(\alpha) \\ b = \phi(\beta) \end{cases}$$

Questa formula esprime come si trasforma l'espressione dell'integrale cambiando la variabile di integrazione. Se, per calcolare $\int_a^b f(x) dx$, decidiamo di porre $x = \phi(u)$, dobbiamo sostituire formalmente dx con $\phi'(u) du$ e cambiare gli estremi compatibilmente con la formula che collega x con u , cioè $x = \phi(u)$. Ad esempio, per calcolare

$$\int_0^1 \frac{e^x}{1 + e^{2x}} dx,$$

poniamo $x = \ln u$. Dato che $\phi(u) = \ln u$, dobbiamo sostituire dx con $\frac{1}{u} du$:

$$\int_0^1 \frac{e^x}{1 + e^{2x}} dx = \int_{\alpha}^{\beta} \frac{u}{1 + u^2} \frac{1}{u} du.$$

Rimangono da calcolare α e β che sono soluzione di $0 = \phi(\alpha) = \ln \alpha$ e $1 = \phi(\beta) = \ln \beta$. Invertendo la funzione ϕ otteniamo

$$\alpha = \phi^{-1}(0) = e^0 = 1, \quad \beta = \phi^{-1}(1) = e^1 = e.$$

In definitiva

$$\int_0^1 \frac{e^x}{1+e^{2x}} dx = \int_1^e \frac{u}{1+u^2} \frac{1}{u} du = \int_1^e \frac{du}{1+u^2} = \arctan e - \frac{\pi}{4}.$$

Nel caso degli integrali indefiniti la formula di sostituzione prende la forma (qui F è una primitiva di f , $F' = f$)

$$(77) \quad \int f(\phi(u))\phi'(u) du = \int f(x) dx = F(x) + C = F(\phi(u)) + C$$

Anche in questo caso si può usare, come regola mnemonica, la relazione

$$x = \phi(u) \quad \Rightarrow \quad " \phi'(u) du = dx ".$$

L'uso delle virgolette “ ” sta a ricordare che non è stato dato senso ai simboli du e dx e che la regola suscritta è solo formale¹. Nell'uso di questa formula bisogna ricordarsi di tornare alla fine alla variabile u , sostituendo x con $\phi(u)$.

ESEMPIO 1.1. Sia ϕ una funzione derivabile. Calcoliamo

$$\int \frac{\phi'(u)}{\phi(u)} du.$$

Ponendo $x = \phi(u)$, si ha $dx = \phi'(u) du$, quindi

$$\int \frac{\phi'(u)}{\phi(u)} du = \int \frac{dx}{x} = \ln |x| + C = \ln |\phi(u)| + C.$$

Ad esempio,

$$\int \frac{dx}{x \ln x} = \ln |\ln x| + C, \quad \int \tan x dx = - \int \frac{-\sin x}{\cos x} dx = - \ln |\cos x| + C.$$

ESEMPIO 1.2. Allo stesso modo, ponendo $x = \phi(u)$ (e quindi $dx = \phi'(u) du$),

$$\int [\phi(u)]^\alpha \phi'(u) du = \int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha+1} + C = \frac{1}{\alpha+1} [\phi(u)]^{\alpha+1} + C \quad \alpha \neq -1.$$

Ad esempio,

$$\int \sin^k x \cos x dx = \frac{1}{k+1} \sin^{k+1} x + C.$$

¹Se si pensa all'origine del simbolo dx negli integrali come limite della lunghezza Δx di una data partizione, la sostituzione da dx a du , con il relativo termine moltiplicativo $\phi'(u)$, indica che, nel cambio di variabile, bisogna cambiare opportunamente anche la lunghezza dell'intervallo della partizione, coerentemente con la trasformazione utilizzata.

La formula di sostituzione è sempre conveniente nel caso di funzioni composte di cui l'ultima sia lineare: ponendo $x = au + b$

$$\int f(au + b) du = \frac{1}{a} \int f(x) dx.$$

Anche se l'integrale di destra non fosse risolvibile, l'espressione a secondo membro è comunque più semplice.

Spesso ci si trova a lavorare con espressioni della forma

$$\int h(\phi(u)) du,$$

dove l'integrando è una funzione composta $h(\phi(u))$, senza il fattore moltiplicativo $\phi'(u)$. E' possibile applicare la sostituzione $x = \phi(u)$? Se la funzione ϕ è invertibile, con inversa $u = \psi(x)$, è possibile sostituire a du l'oggetto $\psi'(x) dx$, ottenendo

$$\int h(\phi(u)) du = \int h(x) \psi'(x) dx.$$

Per giustificare in modo rigoroso questa formula, chiamiamo $f(u) := h(\phi(u))$

$$\int h(\phi(u)) du = \int f(u) du = \int f(\psi(x)) \psi'(x) dx = \int h(x) \psi'(x) dx.$$

Nel caso di integrali definiti occorre cambiare gli estremi di integrazione coerentemente con la nuova variabile introdotta:

$$\int_a^b h(\phi(u)) du = \int_{\phi(a)}^{\phi(b)} h(x) \psi'(x) dx.$$

ESEMPIO 1.3. Calcoliamo l'integrale indefinito

$$\int (1 + e^x)^2 dx.$$

Ponendo $t = 1 + e^x$, si ha $x = \ln(t - 1)$ e $dx = \frac{1}{t-1} dt$ e, di conseguenza,

$$\int (1 + e^x)^2 dx = \int \frac{t^2}{t-1} dt.$$

Dato che $\frac{t^2}{t-1} = t + 1 + \frac{1}{t-1}$,

$$\begin{aligned} \int (1 + e^x)^2 dx &= \int \left(t + 1 + \frac{1}{t-1} \right) dt = \frac{1}{2} t^2 + t + \ln |t-1| + C \\ &= \frac{1}{2} (1 + e^x)^2 + 1 + e^x + x + C = \frac{1}{2} e^{2x} + 2e^x + x + C. \end{aligned}$$

Si sarebbe anche potuto procedere utilizzando la decomposizione $(1 + e^x)^2 = 1 + 2e^x + e^{2x}$ e integrando a partire dalle formule elementari.

ESEMPIO 1.4. Calcoliamo l'integrale definito

$$\int_0^{\pi/2} \cos^2 x \, dx.$$

Dato che $\cos(2x) = 2 \cos^2 x - 1$

$$\int_0^{\pi/2} \cos^2 x \, dx = \frac{1}{2} \int_0^{\pi/2} (\cos(2x) + 1) \, dx = \frac{1}{2} \int_0^{\pi/2} \cos(2x) \, dx + \frac{\pi}{4}.$$

Poniamo nell'integrale $t = 2x$:

$$\int_0^{\pi/2} \cos(2x) \, dx = \frac{1}{2} \int_0^{\pi} \cos t \, dt = \frac{1}{2} \sin t \Big|_0^{\pi} = 0.$$

Quindi il valore dell'integrale è $\pi/4$.

ESERCIZIO 1.5. Fissato $a > 0$, calcolare gli integrali (indefinito e definito)

$$\int x \sqrt{a^2 - x^2} \, dx \quad \text{e} \quad \int_0^a x \sqrt{a^2 - x^2} \, dx.$$

ESEMPIO 1.6. Dati $A, B, C \in \mathbb{R}$ tali che $\Delta := B^2 - 4AC < 0$, come risolvere

$$(78) \quad \int \frac{dx}{Ax^2 + Bx + C} \quad ?$$

La condizione $\Delta < 0$, implica che il polinomio è irriducibile (non ha radici reali). In questa classe di integrali rientra un integrale che conosciamo:

$$(79) \quad \int \frac{dx}{x^2 + 1} = \arctan x + C.$$

L'integrale (78) può essere risolto con una sostituzione opportuna che lo riconduce a (79). Invece di dare direttamente la soluzione, proviamo a ricostruire passo passo come possa essere ottenuta.

Il grafico della funzione $\frac{1}{Ax^2+Bx+C}$ è qualitativamente simile a quello della funzione $\frac{1}{x^2+1}$ (Fig.1). È possibile con traslazioni e dilatazioni trasformare il grafico della prima

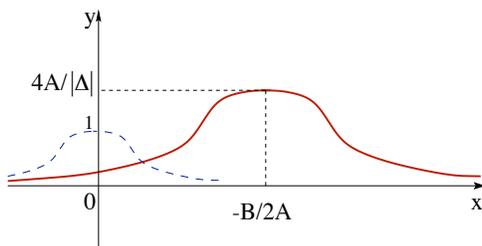


FIGURA 1. Il grafico della funzione $\frac{1}{Ax^2+Bx+C}$ (e, tratteggiato, quello di $\frac{1}{x^2+1}$).

funzione in quello della seconda? Bisogna prima di tutto correggere due “difetti”

evidenti di $f(x) = \frac{1}{Ax^2+Bx+C}$: il valore massimo è $f(-B/2A) = 4A/|\Delta|$ e non 1, l'asse di simmetria è $x = -B/2A$ e non $x = 0$. Per il primo problema, basta utilizzare la linearità dell'integrale: la funzione f può essere riscritta come

$$f(x) = \frac{4A}{|\Delta|} g(x) \quad \text{dove} \quad g(x) := \frac{1}{\frac{4A^2}{|\Delta|} x^2 + \frac{4AB}{|\Delta|} x + \frac{4AC}{|\Delta|}},$$

quindi

$$\int \frac{dx}{Ax^2 + Bx + C} = \frac{4A}{|\Delta|} \int \frac{dx}{\frac{4A^2}{|\Delta|} x^2 + \frac{4AB}{|\Delta|} x + \frac{4AC}{|\Delta|}},$$

dove la funzione all'interno dell'ultimo integrale vale 1 in $x = -B/2A$.

Per fare in modo che l'asse di simmetria sia in $x = 0$, bisogna traslare il grafico. Questo corrisponde ad introdurre una nuova variabile u , legata ad x dalla relazione

$$u = x + \frac{B}{2A}.$$

Con questa scelta, si ha $\frac{4A^2}{|\Delta|} x^2 + \frac{4AB}{|\Delta|} x + \frac{4AC}{|\Delta|} = \frac{4A^2}{|\Delta|} u^2 + 1$, quindi

$$\int \frac{dx}{Ax^2 + Bx + C} = \frac{4A}{|\Delta|} \int \frac{du}{\frac{4A^2}{|\Delta|} u^2 + 1} = \frac{4A}{|\Delta|} \int \frac{du}{\left(\frac{2A}{|\Delta|^{1/2}} u\right)^2 + 1}.$$

Resta da modificare ancora una volta la variabile u , attraverso la posizione

$$v := \frac{2A}{|\Delta|^{1/2}} u,$$

che consiste nel dilatare/comprimere (dipende dalla grandezza di $\frac{2A}{|\Delta|^{1/2}}$) la variabile u per un fattore opportuno. Così facendo si arriva alla conclusione:

$$\int \frac{dx}{Ax^2 + Bx + C} = \frac{4A}{|\Delta|} \int \frac{du}{\left(\frac{2A}{|\Delta|^{1/2}} u\right)^2 + 1} = \frac{2}{|\Delta|^{1/2}} \int \frac{dv}{v^2 + 1} = \frac{2}{|\Delta|^{1/2}} \arctan v + \text{costante}.$$

Per avere l'espressione della primitiva in x (e non in v), basta seguire a ritroso le definizioni di v e di u . In definitiva: (qui $|\Delta| = 4AC - B^2$)

$$\int \frac{dx}{Ax^2 + Bx + C} = \frac{2}{\sqrt{|\Delta|}} \arctan \left(\frac{2Ax + B}{\sqrt{|\Delta|}} \right) + \text{costante}.$$

2. Integrazione per parti

Il metodo di *integrazione per parti* emerge dalla formula di derivazione del prodotto: $(fg)' = f'g + fg'$. Integrando tale formula, si ottiene

$$f(x)g(x) = \int g(x) f'(x) dx + \int g'(x) f(x) dx$$

da cui la formula di integrazione per parti

$$(80) \quad \int g'(x) f(x) dx = f(x)g(x) - \int g(x) f'(x) dx.$$

Questa formula è nota come **integrazione per parti**. Il metodo è vantaggioso se per il termine $g f'$ si conosce un metodo di integrazione.

Per gli integrali definiti, la formula (80) diviene

$$(81) \quad \int_a^b g'(x) f(x) dx = f(b)g(b) - f(a)g(a) - \int_a^b g(x) f'(x) dx.$$

ESEMPIO 2.1. Ecco un primo esempio di applicazione della formula (80)

$$\int x e^x dx = \int x (e^x)' dx = x e^x - \int (x)' e^x dx = x e^x - \int e^x dx = (x - 1)e^x + C.$$

Anche nel caso della funzione $x^2 e^x$ si può procedere in modo analogo, applicando due volte l'integrazione per parti,

$$\begin{aligned} \int x^2 e^x dx &= \int x^2 (e^x)' dx = x^2 e^x - 2 \int x e^x dx = x^2 e^x - 2 \left[x e^x - \int e^x dx \right] \\ &= x^2 e^x - 2 [(x - 1)e^x + C] = (x^2 - 2x + 2)e^x + C. \end{aligned}$$

È possibile risolvere allo stesso modo un qualsiasi integrale del tipo

$$\int p(x) e^x dx \quad p \text{ polinomio di grado } n,$$

infatti basta iterare n volte l'uso della formula di integrazione per parti

$$\begin{aligned} \int p(x) e^x dx &= p(x) e^x - \int p'(x) e^x dx = [p(x) - p'(x)] e^x + \int p''(x) e^x dx \\ &= \dots = [p(x) - p'(x) + \dots + (-1)^n p^{(n)}(x)] e^x + C \end{aligned}$$

Osservando che $e^{\alpha x} = \frac{1}{\alpha} (e^{\alpha x})'$, si possono determinare le primitive anche di funzioni del tipo $p(x) e^{\alpha x}$ con p polinomio. Ad esempio,

$$\begin{aligned} \int x^2 e^{3x} dx &= \frac{1}{3} x^2 e^{3x} - \frac{2}{3} \int x e^{3x} dx \\ &= \frac{1}{3} x^2 e^{3x} - \frac{2}{9} \left(x e^{3x} - \int e^{3x} dx \right) = \left(\frac{1}{3} x^2 - \frac{2}{9} x + \frac{2}{27} \right) e^{3x} + C. \end{aligned}$$

ESEMPIO 2.2. Anche funzioni del tipo prodotto di polinomio e di $\sin x$ o $\cos x$ possono essere risolte integrando per parti. I casi più semplici sono

$$\begin{aligned} \int x \sin x dx &= \int x (-\cos x)' dx = -x \cos x + \int \cos x dx = \sin x - x \cos x + C; \\ \int x \cos x dx &= \int x (\sin x)' dx = x \sin x - \int \sin x dx = \cos x + x \sin x + C. \end{aligned}$$

Iterando il procedimento un certo numero di volte, si calcolano gli integrali

$$\int p(x) \sin(ax) dx, \quad \int p(x) \cos(ax) dx \quad (a \in \mathbb{R}).$$

ESEMPIO 2.3. Proponiamoci di determinare tutte le primitive di $x^2 \sin(2x)$. Applichiamo l'integrazione per parti:

$$\begin{aligned} \int x^2 \sin(2x) dx &= \int x^2 \left(-\frac{1}{2} \cos(2x) \right)' dx = -\frac{1}{2} x^2 \cos(2x) + \int x \cos(2x) dx \\ &= -\frac{1}{2} x^2 \cos(2x) + \frac{1}{2} \int x (\sin(2x))' dx = -\frac{1}{2} x^2 \cos(2x) + \frac{1}{2} x \sin(2x) - \frac{1}{2} \int \sin(2x) dx \\ &= -\frac{1}{2} x^2 \cos(2x) + \frac{1}{2} x \sin(2x) + \frac{1}{4} \cos(2x) + C. \end{aligned}$$

ESEMPIO 2.4. Sempre tramite l'integrazione per parti, si risolvono anche

$$\int p(x) \ln x dx \quad p \text{ polinomio.}$$

Calcoliamo l'integrale di $\ln x$:

$$\begin{aligned} \int \ln x dx &= \int 1 \cdot \ln x dx = \int (x)' \ln x dx = x \ln x - \int x (\ln x)' dx \\ &= x \ln x - \int x \cdot \frac{1}{x} dx = x \ln x - x + C. \end{aligned}$$

Analogamente,

$$\int x \ln x dx = \int \left(\frac{1}{2} x^2 \right)' \ln x dx = \frac{1}{2} x^2 \ln x - \frac{1}{2} \int x dx = \frac{1}{2} x^2 \ln x - \frac{1}{4} x^2 + C.$$

In generale, dato $k \in \mathbb{N}$,

$$\int x^k \ln x dx = \frac{x^{k+1}}{k+1} \ln x - \frac{1}{k+1} \int x^k dx = \frac{x^{k+1}}{k+1} \left(\ln x - \frac{1}{k+1} \right) + C.$$

ESERCIZIO 2.5. Calcolare gli integrali indefiniti

$$\int \arctan x dx, \quad \int x \arctan x dx.$$

ESEMPIO 2.6. Qui usiamo l'integrazione per parti in un modo leggermente diverso: iterando l'applicazione di (80) torniamo all'integrale originale, ottenendo in questo modo un'equazione per la primitiva. In questo modo risolviamo integrali della forma

$$\int e^{ax} \sin(bx) dx, \quad \int e^{ax} \cos(bx) dx.$$

Ad esempio,

$$\begin{aligned}\int e^{2x} \sin(3x) dx &= \frac{1}{3} \int e^{2x} (-\cos(3x))' dx = -\frac{1}{3} \cos(3x)e^{2x} + \frac{2}{3} \int e^{2x} \cos(3x) dx \\ &= -\frac{1}{3} \cos(3x)e^{2x} + \frac{2}{9} \int e^{2x} (\sin(3x))' dx \\ &= \frac{1}{9} (2 \sin(3x) - 3 \cos(3x)) e^{2x} - \frac{4}{9} \int e^{2x} \sin(3x) dx.\end{aligned}$$

Guardando il primo e l'ultimo termine, si ha

$$\int e^{2x} \sin(3x) dx = \frac{1}{9} (2 \sin(3x) - 3 \cos(3x)) e^{2x} - \frac{4}{9} \int e^{2x} \sin(3x) dx,$$

da cui, esplicitando rispetto all'integrale richiesto,

$$\int e^{2x} \sin(3x) dx = \frac{1}{13} (2 \sin(3x) - 3 \cos(3x)) e^{2x} + C.$$

In generale si ottengono le formule (verificare!)

$$\begin{aligned}\int e^{ax} \sin(bx) dx &= \frac{1}{a^2 + b^2} (a \sin(bx) - b \cos(bx)) e^{ax} + C, \\ \int e^{ax} \cos(bx) dx &= \frac{1}{a^2 + b^2} (a \cos(bx) + b \sin(bx)) e^{ax} + C.\end{aligned}$$

Formule ricorsive. Alcune famiglie di integrali (dipendenti da un parametro discreto $n \in \mathbb{N}$), possono essere risolte in modo iterativo, cioè si risolve l'integrale per $n = 1$, e poi si mostra come l'integrale al passo n -esimo si possa ricondurre al calcolo dell'integrale $(n - 1)$ -esimo. Vediamo un paio di esempi. Calcoliamo

$$I_n = \int \sin^{2n} x dx, \quad n \in \mathbb{N}.$$

Allo stesso modo si può calcolare $\int \cos^{2n} x dx$. Calcoliamo I_1 :

$$\begin{aligned}\int \sin^2 x dx &= \int \sin x \cdot \sin x dx = \int \sin x (-\cos x)' dx = -\sin x \cos x + \int \cos^2 x dx \\ &= -\sin x \cos x + \int (1 - \sin^2 x) dx = x - \sin x \cos x - \int \sin^2 x dx.\end{aligned}$$

Abbiamo ottenuto una relazione del tipo $I_1 = x - \sin x \cos x - I_1$, quindi

$$I_1 = \int \sin^2 x dx = \frac{1}{2} (x - \sin x \cos x) + C.$$

Per $n \in \mathbb{N}$,

$$\begin{aligned} I_{n+1} &= \int \sin^{2n+1} x \sin x \, dx = \int \sin^{2n+1} x (-\cos x)' \, dx \\ &= -\sin^{2n+1} x \cos x + (2n+1) \int \sin^{2n} x \cos^2 x \, dx \\ &= -\sin^{2n+1} x \cos x + (2n+1) \int \sin^{2n} x (1 - \sin^2 x) \, dx \\ &= -\sin^{2n+1} x \cos x + (2n+1)I_n - (2n+1)I_{n+1}. \end{aligned}$$

Quindi $I_{n+1} = -\sin^{2n+1} x \cos x + (2n+1)I_n - (2n+1)I_{n+1}$, da cui si deduce

$$I_{n+1} = \frac{1}{2n+2} \{ (2n+1)I_n - \sin^{2n+1} x \cos x \} + C.$$

3. Integrazione di funzioni razionali

Affrontiamo ora il problema di integrare funzioni razionali:

$$\int \frac{P(x)}{Q(x)} \, dx \quad P, Q \text{ polinomi.}$$

Le primitive di una qualsiasi funzione razionale in termini di funzioni esplicite sono note, ma non è in queste Note che troverete i dettagli della questione.

In concreto è possibile completare il calcolo a patto di saper fattorizzare il polinomio a denominatore Q nel prodotto di termini irriducibili, cioè polinomi di primo grado (con molteplicità opportuna) e polinomi di secondo grado irriducibili (con molteplicità opportuna). In questo Paragrafo vedremo come si integrino funzioni razionali nel caso in cui il polinomio Q sia di grado al più due, o sia fattorizzabile in termini di polinomi di grado 1, cioè sia riscrivibile nella forma

$$Q(x) = a(x-x_1)^{k_1} \cdots (x-x_n)^{k_n} \quad a, x_1, \dots, x_n \in \mathbb{R}, \quad a \neq 0, \quad k_1, \dots, k_n \in \mathbb{N}.$$

Denominatore Q di grado 1. Sia $Q(x) = a(x-x_0)$ con $a, x_0 \in \mathbb{R}$ e $a \neq 0$. Se P è un polinomio di grado $p \geq 1$, tramite l'algoritmo di divisione dei polinomi, si determinano un polinomio P_1 di grado $p-1$ e una costante $r \in \mathbb{R}$ tali che

$$\frac{P(x)}{Q(x)} = P_1(x) + \frac{r}{a(x-x_0)}.$$

Quindi l'integrale si può decomporre nella somma di due integrali

$$\int \frac{P(x)}{Q(x)} \, dx = \int P_1(x) \, dx + \frac{r}{a} \int \frac{dx}{x-x_0}.$$

Il polinomio P_1 è integrabile esplicitamente, grazie alla formula

$$\int x^k \, dx = \frac{x^{k+1}}{k+1} + C.$$

Anche l'altro integrale è risolvibile esplicitamente:

$$\frac{r}{a} \int \frac{dx}{x - x_0} = \frac{r}{a} \int \frac{(x - x_0)'}{x - x_0} dx = \frac{r}{a} \ln |x - x_0| + C.$$

Vediamo un esempio. Calcoliamo

$$\int \frac{x^5 + 1}{x - 2} dx.$$

Dato che

$$\frac{x^5 + 1}{x - 2} = x^4 + 2x^3 + 4x^2 + 8x + 16 + \frac{33}{x - 2},$$

si ha

$$\begin{aligned} \int \frac{x^5 + 1}{x - 2} dx &= \int \left(x^4 + 2x^3 + 4x^2 + 8x + 16 + \frac{33}{x - 2} \right) dx \\ &= \frac{1}{5}x^5 + \frac{1}{2}x^4 + \frac{4}{3}x^3 + 4x^2 + 16x + 33 \ln |x - 2| + C. \end{aligned}$$

Denominatore Q di grado 2. Supponiamo che Q sia un polinomio di grado 2. In questo caso Q è scrivibile nella forma

$$Q(x) = a(x^2 + 2bx + c) \quad a, b, c \in \mathbb{R}, \quad a \neq 0.$$

Se il polinomio a numeratore P ha grado $p \geq 2$, allora è possibile applicare l'algoritmo di divisione di polinomi e riscrivere la funzione razionale come somma

$$(82) \quad \frac{P(x)}{Q(x)} = P_1(x) + \frac{R(x)}{Q(x)},$$

dove P_1 è un polinomio di grado $p - 2$ e R è un polinomio di grado minore o uguale a 1. L'integrale della funzione razionale è la somma di due integrali

$$\int \frac{P(x)}{Q(x)} dx = \int P_1(x) dx + \int \frac{R(x)}{Q(x)} dx.$$

Il primo dei due integrali è risolvibile esplicitamente per via elementare. Consideriamo il secondo. Supponiamo che il resto R sia di grado 1 e scriviamolo nella forma $R(x) = \alpha(x + \beta)$ con $\alpha \neq 0$ e $\beta \in \mathbb{R}$. Si tratta di calcolare

$$\int \frac{\alpha(x + \beta)}{a(x^2 + 2bx + c)} dx = \frac{\alpha}{a} \int \frac{x + \beta}{x^2 + 2bx + c} dx.$$

Come primo passo, "costruiamo" a numeratore la derivata del denominatore. Moltiplichiamo e dividiamo per due e, successivamente, sommiamo e sottraiamo $2b$

$$\begin{aligned} \int \frac{\alpha(x + \beta)}{a(x^2 + 2bx + c)} dx &= \frac{\alpha}{2a} \int \frac{2x + 2\beta}{x^2 + 2bx + c} dx \\ &= \frac{\alpha}{2a} \int \frac{(2x + 2b) + 2(\beta - b)}{x^2 + 2bx + c} dx = \dots \end{aligned}$$

L'integrale finale può essere riscritto come somma dei due integrali di cui il primo è della forma $\int \frac{\phi'}{\phi} dx$; quindi

$$\begin{aligned} \dots &= \frac{\alpha}{2a} \int \frac{(x^2 + 2bx + c)'}{x^2 + 2bx + c} dx + \frac{2\alpha(\beta - b)}{2a} \int \frac{dx}{x^2 + 2bx + c} \\ &= \frac{\alpha}{2a} \ln |x^2 + 2bx + c| + \frac{2\alpha(\beta - b)}{2a} \int \frac{dx}{x^2 + 2bx + c}. \end{aligned}$$

Rimane quindi da risolvere l'integrale

$$(83) \quad \int \frac{dx}{x^2 + 2bx + c}.$$

Nel caso in cui R in (82) sia di grado 0 ci si riconduce direttamente a questa situazione. La risoluzione dell'integrale (83) varia a seconda di quante radici reali abbia il denominatore, cioè a seconda che sia $b^2 > c$, $b^2 = c$ o $b^2 < c$. Trattiamo i tre casi separatamente. Ci ricondurremo (sostanzialmente) ai seguenti integrali elementari

$$\begin{aligned} \text{Caso I : } \quad b^2 > c &\quad \longrightarrow \quad \int \frac{dx}{x} = \ln |x| + C, \\ \text{Caso II : } \quad b^2 = c &\quad \longrightarrow \quad \int \frac{dx}{x^2} = -\frac{1}{x} + C, \\ \text{Caso III : } \quad b^2 < c &\quad \longrightarrow \quad \int \frac{dx}{1 + x^2} = \arctan x + C. \end{aligned}$$

Caso I: $b^2 > c$. In questo caso il denominatore ha due radici reali

$$x^2 + 2bx + c = 0 \quad \Longleftrightarrow \quad x = -b \pm \sqrt{b^2 - c}.$$

Indicando le radici con x_1 e x_2 , il polinomio si fattorizza:

$$x^2 + 2bx + c = (x - x_1)(x - x_2).$$

Decomponiamo la funzione integranda nella forma

$$\frac{1}{x^2 + 2bx + c} = \frac{A_1}{x - x_1} + \frac{A_2}{x - x_2},$$

dove $A_1, A_2 \in \mathbb{R}$ sono due costanti da determinare. La somma delle due frazioni a secondo membro è uguale a

$$\frac{(A_1 + A_2)x - (A_1x_2 + A_2x_1)}{x^2 + 2bx + c},$$

e quindi A_1 e A_2 devono essere tali che $(A_1 + A_2)x - (A_1x_2 + A_2x_1) = 1$. Dato che due polinomi coincidono se e solo se coincidono i loro coefficienti, le costanti A_1, A_2 sono le soluzioni del sistema lineare (il cui determinante è $x_1 - x_2$ che, nel caso $b^2 > c$, è diverso da zero)

$$A_1 + A_2 = 0, \quad A_1x_2 + A_2x_1 = -1.$$

Individuati i valori di A_1 e A_2 , l'integrale è risolto:

$$\begin{aligned}\int \frac{dx}{x^2 + 2bx + c} &= A_1 \int \frac{dx}{x - x_1} + A_2 \int \frac{dx}{x - x_2} \\ &= A_1 \ln |x - x_1| + A_2 \ln |x - x_2| + C.\end{aligned}$$

ESERCIZIO 3.1. Calcolare

$$\int \frac{x^3}{x^2 - x - 2} dx.$$

Caso II: $b^2 = c$. In questa situazione, si tratta di risolvere

$$\int \frac{dx}{x^2 + 2bx + b^2}.$$

Questo integrale è immediato, infatti

$$\int \frac{dx}{x^2 + 2bx + b^2} = \int \frac{dx}{(x + b)^2} = -\frac{1}{x + b} + C.$$

ESERCIZIO 3.2. Calcolare

$$\int \frac{x(x + 3)}{(x - 1)^2} dx.$$

Caso III: $b^2 < c$. Questo caso è già stato considerato nell'Esempio 1.6. Ritroviamo qui la stessa soluzione senza fare ricorso al grafico della funzione integranda.

Dato che il polinomio $x^2 + 2bx + c$ è irriducibile, l'obiettivo è di ricondursi, con un opportuno cambiamento di variabili, all'integrale elementare

$$\int \frac{dx}{1 + x^2} = \arctan x + C.$$

Chiamiamo $\nu := \frac{1}{\sqrt{c-b^2}} > 0$ e riscriviamo in maniera opportuna il denominatore

$$x^2 + 2bx + c = x^2 + 2bx + b^2 + (c - b^2) = (x + b)^2 + \frac{1}{\nu^2} = \frac{1}{\nu^2} \{[\nu(x + b)]^2 + 1\}.$$

Ponendo $t = \nu(x + b)$,

$$\int \frac{dx}{x^2 + 2bx + c} = \nu \int \frac{dt}{1 + t^2} = \nu \arctan t + C = \nu \arctan(\nu(x + b)) + C.$$

Dalla definizione di ν si deduce che

$$\int \frac{dx}{x^2 + 2bx + c} = \frac{1}{\sqrt{c - b^2}} \arctan \left(\frac{x + b}{\sqrt{c - b^2}} \right) + C.$$

ESERCIZIO 3.3. Calcolare

$$\int \frac{3x - 2}{x^2 - 2x + 2} dx.$$

Denominatore Q con sole radici reali. Consideriamo prima di tutto un caso semplice: il polinomio Q ha radici reali *distinte*, cioè

$$Q(x) = a(x - x_1) \cdots (x - x_n) \quad x_1, \dots, x_n \in \mathbb{R} \quad \text{con } x_i \neq x_j \quad \text{se } i \neq j.$$

Se $p \geq n$, il primo passaggio è sempre lo stesso: si usa l'algoritmo della divisione di polinomi per riscrivere la funzione razionale nella forma

$$\frac{P(x)}{Q(x)} = P_1(x) + \frac{R(x)}{Q(x)}.$$

dove P_1 è un polinomio di grado $p - n$, e R è un polinomio di grado minore di n . L'integrale si decompone in

$$\int \frac{P(x)}{Q(x)} dx = \int P_1(x) dx + \int \frac{R(x)}{Q(x)} dx.$$

A questo punto sfruttiamo la fattorizzazione di Q per riscrivere la funzione razionale R/Q come somma di funzioni razionali più semplici. Cerchiamo n costanti A_1, \dots, A_n tali che

$$\frac{R(x)}{a(x - x_1) \cdots (x - x_n)} = \frac{1}{a} \left(\frac{A_1}{x - x_1} + \cdots + \frac{A_n}{x - x_n} \right).$$

Per determinare le costanti A_1, \dots, A_n si può imporre l'uguaglianza dei due membri ottenendo un sistema lineare. Equivalentemente si può moltiplicare per $x - x_1$ entrambi i membri

$$\frac{R(x)}{a(x - x_2) \cdots (x - x_n)} = \frac{1}{a} \left\{ A_1 + \frac{A_2(x - x_1)}{x - x_2} + \cdots + \frac{A_n(x - x_1)}{x - x_n} \right\}.$$

e successivamente porre $x = x_1$, ottenendo il valore di A_1

$$A_1 = \frac{R(x_1)}{(x_1 - x_2) \cdots (x_1 - x_n)}$$

Analogamente per A_2, \dots, A_n . Determinate le costanti A_i , si calcola l'integrale:

$$\begin{aligned} \int \frac{R(x)}{Q(x)} dx &= \frac{1}{a} \int \left(\frac{A_1}{x - x_1} + \cdots + \frac{A_n}{x - x_n} \right) dx \\ &= \frac{1}{a} (A_1 \ln |x - x_1| + \cdots + A_n \ln |x - x_n|) + C. \end{aligned}$$

Per digerire la tecnica, calcoliamo

$$\int \frac{dx}{(x + 1)(x + 2)(x + 3)}.$$

Dato che il grado del numeratore è minore del grado del denominatore, non occorre applicare l'algoritmo di divisione di polinomi. Passiamo subito alla decomposizione:

cerchiamo $A_1, A_2, A_3 \in \mathbb{R}$ tali che

$$\frac{1}{(x+1)(x+2)(x+3)} = \frac{A_1}{x+1} + \frac{A_2}{x+2} + \frac{A_3}{x+3}.$$

Moltiplichiamo per $x+1$ e calcoliamo in $x = -1$

$$\frac{1}{(x+2)(x+3)} = A_1 + \frac{A_2(x+1)}{x+2} + \frac{A_3(x+1)}{x+3} \quad \Rightarrow \quad A_1 = \frac{1}{2}.$$

Analogamente

$$\frac{1}{(x+1)(x+3)} = \frac{A_1(x+2)}{x+1} + A_2 + \frac{A_3(x+2)}{x+3} \quad \Rightarrow \quad A_2 = -1.$$

$$\frac{1}{(x+1)(x+2)} = \frac{A_1(x+3)}{x+1} + \frac{A_2(x+3)}{x+2} + A_3 \quad \Rightarrow \quad A_3 = \frac{1}{2}.$$

Quindi

$$\begin{aligned} \int \frac{dx}{(x+1)(x+2)(x+3)} &= \frac{1}{2} \int \left(\frac{1}{x+1} - \frac{2}{x+2} + \frac{1}{x+3} \right) dx \\ &= \frac{1}{2} \ln \left| \frac{(x+1)(x+3)}{(x+2)^2} \right| + C. \end{aligned}$$

Passiamo al caso generale: il denominatore Q si scompone come

$$Q(x) = a(x-x_1)^{k_1} \cdots (x-x_n)^{k_n} \quad a, x_1, \dots, x_n \in \mathbb{R}, \quad k_1, \dots, k_n \in \mathbb{N}.$$

Anche in questo caso, dopo aver applicato (se necessario) l'algoritmo di divisione di polinomi, si deve risolvere un integrale della forma

$$\int \frac{R(x)}{Q(x)} dx,$$

dove R è un polinomio di grado strettamente minore di quello di Q . In questo caso si cerca, analogamente a quanto fatto nel caso di Q di secondo grado con due radici coincidenti, una decomposizione della forma

$$\begin{aligned} \frac{R(x)}{a(x-x_1)^{k_1} \cdots (x-x_n)^{k_n}} &= \frac{1}{a} \left\{ \frac{A_1^1}{x-x_1} + \frac{A_1^2}{(x-x_1)^2} + \cdots + \frac{A_1^{k_1}}{(x-x_1)^{k_1}} + \cdots \right. \\ &\quad \left. + \frac{A_n^1}{x-x_n} + \frac{A_n^2}{(x-x_n)^2} + \cdots + \frac{A_n^{k_n}}{(x-x_n)^{k_n}} \right\}, \end{aligned}$$

dove le costanti $A_i^j \in \mathbb{R}$ con $i = 1, \dots, n$ e $j = 1, \dots, k_i$ sono da determinare. Una volta determinate queste costanti, l'integrale è risolto dato che

$$\int \frac{A dx}{(x-\bar{x})^k} = \begin{cases} A \ln |x-\bar{x}| + C & k = 1, \\ -\frac{A}{k-1} \cdot \frac{1}{(x-\bar{x})^{k-1}} + C & k > 1. \end{cases}$$

ESEMPIO 3.4. Ad esempio, consideriamo l'integrale

$$\int \frac{dx}{x^2(x+1)^2}.$$

In questo caso cerchiamo le costanti $A, B, C, D \in \mathbb{R}$ tali che

$$\frac{1}{x^2(x+1)^2} = \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x+1} + \frac{D}{(x+1)^2}.$$

Imponendo l'uguaglianza, si ottengono

$$A = -2, \quad B = 1, \quad C = 2, \quad D = 1.$$

Perciò

$$\int \frac{dx}{x^2(x+1)^2} = 2 \ln \left| \frac{x+1}{x} \right| - \frac{1}{x} - \frac{1}{x+1} + C.$$

Altre classi di funzioni. Vediamo qualche altra classe di funzioni che si possono ricondurre, tramite un cambio di variabile, all'integrale di funzioni razionali.

ESEMPIO 3.5. Supponiamo di voler calcolare

$$\int R(\sin x, \cos x) dx$$

dove R è una funzione razionale dei suoi argomenti. Dalle relazioni

$$(84) \quad \sin x = \frac{2t}{1+t^2} \quad \text{e} \quad \cos x = \frac{1-t^2}{1+t^2} \quad \text{dove} \quad t = \tan\left(\frac{x}{2}\right),$$

ponendo $t = \tan(x/2)$ o, equivalentemente, $x = 2 \arctan t$, dato che $dx = 2/(1+t^2)dt$, l'integrale si trasforma nell'integrale di una funzione razionale

$$\int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right) \frac{2}{1+t^2} dt,$$

Ad esempio,

$$\int \frac{dx}{\sin x} = \int \frac{1+t^2}{2t} \frac{2}{1+t^2} dt = \int \frac{dt}{t} = \ln |t| + C = \ln |\tan(x/2)| + C.$$

$$\int \frac{dx}{\cos x} = \int \frac{2 dt}{1-t^2} = \int \left(\frac{1}{1+t} + \frac{1}{1-t} \right) dt = \ln \left| \frac{1+\tan(x/2)}{1-\tan(x/2)} \right| + C.$$

Non sempre la sostituzione (84) è conveniente. Ad esempio, consideriamo

$$\int \frac{\sin x}{\cos^{100} x} dx.$$

Tramite (84), l'integrale si trasforma in

$$\int \frac{\sin x}{\cos^{100} x} dx = \int \frac{4t(1+t^2)^{98}}{(1-t^2)^{100}} dt,$$

con un polinomio di grado duecento a denominatore! Che fare? Sarebbe invece stato molto più semplice porre $s = \cos x$, da cui

$$\int \frac{\sin x}{\cos^{100} x} dx = - \int \frac{ds}{s^{100}} = \frac{1}{99s^{99}} + C = \frac{1}{99 \cos^{99} x} + C.$$

ESEMPIO 3.6. Abbiamo un problema: calcolare l'area della regione di piano

$$\Omega = \left\{ (x, y) \in \mathbb{R} : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \right\} \quad a, b > 0.$$

L'area $|\Omega|$ di Ω è pari al valore dell'integrale definito

$$|\Omega| = 4b \int_0^a \sqrt{1 - (x^2/a^2)} dx.$$

Introduciamo la variabile t definita da $x = a \cos t$, da cui $dx = -a \sin t dt$:

$$|\Omega| = 4ab \int_0^{\pi/2} \sqrt{1 - \cos^2 t} \sin t dt = 4ab \int_0^{\pi/2} \sin^2 t dt = 2ab [t - \sin t \cos t]_0^{\pi/2} = \pi ab.$$

Quindi l'area della regione delimitata dall'ellissi di semiassi a e b è πab .

Allo stesso modo è possibile integrare funzioni del tipo

$$R(x, \sqrt{1 - (x^2/a^2)})$$

con R funzione razionale dei suoi argomenti. Infatti

$$\int R(x, \sqrt{1 - (x^2/a^2)}) dx = -a \int R(a \cos t, \sin t) \sin t dt.$$

dove $x = a \cos t$, e il secondo membro è una funzione razionale in $\sin t$ e $\cos t$.

I numeri complessi

Fino ad adesso abbiamo studiato funzioni reali di variabile reale. Vari problemi matematici suggeriscono un'estensione dell'insieme dei numeri reali, particolarmente utile anche per molte applicazioni. Questa estensione dei numeri reali porta ad un nuovo concetto di numero: il *numero complesso*.

1. Nascita dei numeri complessi e loro infanzia

Il problema di partenza è il seguente: *in \mathbb{R} esistono polinomi di grado maggiore di 1 che non hanno radici (reali)*, ad esempio, $x^2 + 1$. Si potrebbe rimanere lì a guardare, risponderci che è un fatto della vita, e passare ad un altro problema. Per chi invece desidera cercare di andare oltre, l'idea plausibile è di costruire un nuovo insieme che estenda l'insieme dei numeri reali e che contenga anche il “numero” i , *unità immaginaria*, soluzione dell'equazione $x^2 + 1 = 0$. Dedichiamoci alla costruzione di questo nuovo insieme, che indichiamo con il simbolo \mathbb{C} .

DEFINIZIONE 1.1. *Si definisce insieme dei numeri complessi \mathbb{C} l'insieme*

$$\mathbb{C} := \{z = a + ib : a, b \in \mathbb{R}\}$$

con la regola di calcolo seguente: i numeri complessi si sommano, sottraggono, moltiplicano, dividono tenendo conto della proprietà $i^2 = -1$.

La regola di calcolo di somma e moltiplicazione per numeri complessi va letta, in concreto, come segue: dati $z_1 = a + ib$ e $z_2 = c + id$, allora

$$z_1 + z_2 = (a + ib) + (c + id) = (a + c) + i(b + d),$$

$$z_1 \cdot z_2 = (a + ib)(c + id) = ac + i(ad + bc) + i^2bd = (ac - bd) + i(ad + bc).$$

Tutte le proprietà di somma e moltiplicazione (commutativa, associativa, distributiva) valgono anche in \mathbb{C} .

Piccolo glossario per i numeri complessi

$z = a + ib$: numero complesso

$\mathbb{R} = \{z \in \mathbb{C} : b = 0\}$: numeri reali

$i\mathbb{R} = \{z \in \mathbb{C} : a = 0\}$: immaginari puri

i : unità immaginaria
 $\operatorname{Re} z = a$: parte reale di z
 $\operatorname{Im} z = b$: parte immaginaria di z

L'insieme \mathbb{C} può essere rappresentato geometricamente come un piano

$$\mathbb{C} = \{(a, b) : a, b \in \mathbb{R}\} = \mathbb{R}^2.$$

La struttura di \mathbb{C} è comunque più ricca di quella del piano dato che in \mathbb{C} è *definita l'operazione di prodotto*.

I numeri complessi si devono poter rappresentare sempre nella forma $a+ib$. Nel caso del rapporto di due numeri, può non essere evidente che questo sia possibile. Vediamo un esempio particolare. Supponiamo di voler scrivere nella forma $a+ib$ con $a, b \in \mathbb{R}$ il numero complesso $z = \frac{1+i}{1-i}$. Moltiplicando e dividendo per $1+i$, e ragionando con le stesse regole algebriche note su \mathbb{R} insieme alla relazione $i^2 = -1$, si ottiene

$$\frac{1+i}{1-i} = \frac{1+i}{1-i} \cdot \frac{1+i}{1+i} = \frac{(1+i)^2}{(1-i)(1+i)} = \frac{1+2i+i^2}{1-i^2} = i.$$

In generale, per scrivere un rapporto di numeri complessi nella forma $a+ib$, procediamo alla stessa maniera. Siano $a, b, c, d \in \mathbb{R}$ con $(c, d) \neq (0, 0)$, allora

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

In particolare l'inverso del numero complesso $c+id$ si scrive come

$$\frac{1}{c+id} = \frac{1}{c+id} \cdot \frac{c-id}{c-id} = \frac{c-id}{c^2+d^2} = \frac{c}{c^2+d^2} - i \frac{d}{c^2+d^2},$$

che mostra che *ogni numero complesso z diverso da 0 è invertibile, cioè esiste un numero z^{-1} tale che $z \cdot z^{-1} = z^{-1} \cdot z = 1$* .

Dato $c+id \in \mathbb{C}$, il numero complesso $c-id \in \mathbb{C}$, che si ottiene cambiando di segno la parte immaginaria ha un ruolo determinante in questa costruzione ed è detto **complesso coniugato**. Il complesso coniugato di z si indica con \bar{z} .

ESERCIZIO 1.2. Verificare la validità di:

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2 \quad \overline{z_1 - z_2} = \bar{z}_1 - \bar{z}_2 \quad \overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2 \quad \overline{\frac{z_1}{z_2}} = \frac{\bar{z}_1}{\bar{z}_2}.$$

Rappresentazione polare. Un numero complesso $z = x+iy$ si può rappresentare come un punto P di coordinate (x, y) nel piano cartesiano e, in quanto tale, può essere individuato anche dalle coordinate polari (r, θ) :

$$x = r \cos \theta, \quad y = r \sin \theta.$$

In questo caso $r = \sqrt{x^2 + y^2}$ è la distanza del punto $P = (x, y)$ dall'origine O e θ

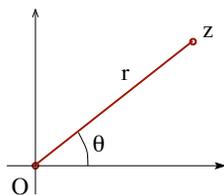


FIGURA 1. Rappresentazione polare di un numero complesso.

misura l'angolo tra il semiasse positivo x e il segmento di estremi 0 e z . Il numero complesso $z = x + iy$ si può, quindi, scrivere come

$$(85) \quad z = r(\cos \theta + i \sin \theta).$$

La quantità r è il **modulo** di z ; l'angolo θ si dice **argomento** di z e si indica con $\arg z$. Per via della periodicità delle funzioni \sin e \cos , l'argomento non è individuato univocamente. Si parla di **argomento principale** se $\theta \in [0, 2\pi)$ e si indica con $\text{Arg } z$ (Attenzione! c'è chi parla di "argomento principale" se $\theta \in (-\pi, \pi]$). Il complesso coniugato \bar{z} di z corrisponde al numero complesso di stesso modulo di z e argomento $-\theta$.

Tramite la rappresentazione polare è facile calcolare il prodotto di numeri complessi:

$$\begin{aligned} zz' &= r(\cos \theta + i \sin \theta) r'(\cos \theta' + i \sin \theta') \\ &= rr'[(\cos \theta \cos \theta' - \sin \theta \sin \theta') + i(\cos \theta \sin \theta' + \sin \theta \cos \theta')] \\ &= rr'[\cos(\theta + \theta') + i \sin(\theta + \theta')]. \end{aligned}$$

Questa formula permette di dare un'interpretazione geometrica del prodotto di due numeri complessi: il risultato è un numero complesso il cui modulo è pari al prodotto dei moduli e l'argomento è la somma degli argomenti. Ad esempio, che succede moltiplicando un generico numero complesso z per l'unità immaginaria i ? Dato che il modulo di i è 1 e il suo argomento è $\frac{\pi}{2}$, il numero iz si trova alla stessa distanza dall'origine di z e ha argomento aumentato di $\frac{\pi}{2}$ rispetto a quello di z . In definitiva, il numero iz si ottiene ruotando il punto z in senso antiorario di un angolo retto.

Dalla formula per il prodotto, segue la formula per la potenza n -esima di z

$$z^n = [r(\cos \theta + i \sin \theta)]^n = r^n [\cos(n\theta) + i \sin(n\theta)].$$

Il modulo nei complessi. Il modulo del numero complesso $a + ib$ è il modulo del punto $(a, b) \in \mathbb{R}^2$, cioè

$$\text{modulo di } z = a + ib : \quad |z| := \sqrt{a^2 + b^2}$$

Il modulo di un numero complesso è sempre un numero reale non negativo e gode delle stesse proprietà di cui gode il modulo reale

PROPOSIZIONE 1.3. *Il modulo complesso gode delle seguenti proprietà:*

- (i) $|z| \geq 0$ per ogni $z \in \mathbb{C}$ e $|z| = 0$ se e solo se $z = 0$;
- (ii) $|z_1 z_2| = |z_1| |z_2|$ per ogni $z_1, z_2 \in \mathbb{C}$;
- (iii) **diseguaglianza triangolare:** $|z_1 + z_2| \leq |z_1| + |z_2|$ per ogni $z_1, z_2 \in \mathbb{C}$.

DIMOSTRAZIONE. La prima proprietà è evidente (basta osservare che una somma di quadrati è nulla se e solo se sono nulli entrambi gli addendi).

La seconda si verifica direttamente, se $z_1 = x_1 + iy_1$ e $z_2 = x_2 + iy_2$, allora

$$\begin{aligned} |z_1 z_2| &= |(x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1)| = \sqrt{(x_1 x_2 - y_1 y_2)^2 + (x_1 y_2 + x_2 y_1)^2} \\ &= \sqrt{x_1^2 x_2^2 + y_1^2 y_2^2 + x_1^2 y_2^2 + x_2^2 y_1^2} \end{aligned}$$

$$|z_1| |z_2| = \sqrt{x_1^2 + y_1^2} \sqrt{x_2^2 + y_2^2} = \sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)} = \sqrt{x_1^2 x_2^2 + x_1^2 y_2^2 + y_1^2 x_2^2 + y_1^2 y_2^2}.$$

Resta da dimostrare la diseguaglianza triangolare. Per la definizione di modulo, la diseguaglianza triangolare si riscrive come

$$(86) \quad \sqrt{(x_1 + x_2)^2 + (y_1 + y_2)^2} \leq \sqrt{x_1^2 + y_1^2} + \sqrt{x_2^2 + y_2^2}$$

Elevando al quadrato

$$(x_1 + x_2)^2 + (y_1 + y_2)^2 \leq x_1^2 + y_1^2 + 2\sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)} + x_2^2 + y_2^2$$

Sviluppando i quadrati dei due binomi a sinistra e semplificando,

$$x_1 x_2 + y_1 y_2 \leq \sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)}$$

Se il termine a sinistra è negativo, la disuguaglianza è verificata; altrimenti, elevando di nuovo al quadrato,

$$x_1^2 x_2^2 + 2x_1 x_2 y_1 y_2 + y_1^2 y_2^2 \leq (x_1^2 + y_1^2)(x_2^2 + y_2^2) = x_1^2 x_2^2 + x_1^2 y_2^2 + y_1^2 x_2^2 + y_1^2 y_2^2.$$

Semplificando si ottiene $2x_1 x_2 y_1 y_2 \leq x_1^2 y_2^2 + y_1^2 x_2^2$, che si può riscrivere come

$$(x_1 y_2 - x_2 y_1)^2 = x_1^2 y_2^2 - 2x_1 x_2 y_1 y_2 + y_1^2 x_2^2 \geq 0,$$

che è sempre vera. □

ESERCIZIO 1.4. *Verificare che, per ogni $z \in \mathbb{C}$, valgono (nella seconda $z \neq 0$)*

$$z \cdot \bar{z} = |z|^2 \quad \text{e} \quad z^{-1} = \frac{\bar{z}}{|z|^2}$$

Il fatto che in \mathbb{C} sia definito il modulo (con le stesse proprietà note in \mathbb{R}) è di fondamentale importanza: con quest'oggetto è possibile estendere ai complessi tutto quello che si è visto nei reali a partire dalla nozione di distanza.

Nel piano complesso \mathbb{C} , la distanza tra due numeri complessi z e z' è data da $|z - z'|$, dove $|\cdot|$ rappresenta il modulo complesso.

Grazie alla presenza della distanza, è possibile introdurre in \mathbb{C} il concetto di intorno.

DEFINIZIONE 1.5. Dato $z_0 \in \mathbb{C}$ e $r > 0$, l'intorno di z_0 di raggio r (indicato con $I(z_0; r)$) è l'insieme dei punti $z \in \mathbb{C}$ che distano da z_0 meno di r :

$$I(z_0; r) := \{z \in \mathbb{C} : |z - z_0| < r\}.$$

Da qui limiti e tutto il resto...

Il Teorema fondamentale dell'algebra. Fin qui abbiamo scoperto che l'insieme \mathbb{C} gode delle stesse proprietà di \mathbb{R} per quanto riguarda le operazioni di somma e prodotto e anche per quanto riguarda la struttura metrica. Ma in \mathbb{C} , c'è qualcosa di più...

TEOREMA 1.6 (Teorema fondamentale dell'algebra). Ogni polinomio $p(z) = a_0 + a_1z + \dots + a_nz^n$ a coefficienti complessi $a_i \in \mathbb{C}$ di grado $n \geq 1$ ammette almeno uno zero $z^* \in \mathbb{C}$, cioè esiste $z^* \in \mathbb{C}$ tale che $p(z^*) = 0$.

Conseguenza di questo Teorema è che un polinomio di grado n in \mathbb{C} ha esattamente n zeri (contati con la loro molteplicità). Ad esempio, il polinomio a coefficienti reali

$$p(z) = z^2 + 2bz + c \quad b, c \in \mathbb{R},$$

ammette in \mathbb{C} sempre due zeri: se $b^2 > c$, gli zeri sono reali e sono $-b \pm \sqrt{b^2 - c}$, se $b^2 = c$ c'è un unico zero, $z = -b$, con molteplicità due, nel caso $b^2 < c$, gli zeri sono due numeri complessi coniugati, $-b \pm i\sqrt{c - b^2}$.

2. Successioni, serie e continuità nei complessi

Una volta definito l'insieme dei numeri complessi \mathbb{C} e introdotta la nozione di modulo, si può ripetere buona parte della teoria sviluppata in \mathbb{R} : successioni, serie, limiti, continuità, derivabilità... Ripercorriamo ora, rapidamente, quello che abbiamo visto nel caso reale, vedendo come sia possibile definire gli stessi oggetti nel caso complesso.

Successioni complesse. Una funzione che associa ad ogni numero naturale $n \in \mathbb{N}$ un valore $z_n \in \mathbb{C}$ è una **successione complessa**. Una successione complessa z_n può essere pensata come una famiglia di valori ordinati in base al loro indice n : z_0, z_1, z_2, \dots

DEFINIZIONE 2.1. Una successione complessa z_n è **convergente a** $\ell \in \mathbb{C}$ se

$$(87) \quad \forall \varepsilon > 0 \quad \exists N \in \mathbb{N} \quad \text{tale che} \quad |z_n - \ell| < \varepsilon \quad \forall n > N.$$

Con una certa dose di fantasia, si scrive $\lim_{n \rightarrow +\infty} z_n = \ell$.

Una successione z_n è **infinitesima** se converge a 0 per $n \rightarrow +\infty$.

Il modulo che compare in (87) è il modulo del numero complesso $z_n - \ell$, e l'essere infinitesima significa semplicemente che comunque si fissi un intorno dell'origine 0 (misurato dal valore di ε), tutti gli elementi della successione tranne al più un numero finito (controllato dall'indice N) si trovano nell'intorno fissato.

Dato che il modulo complesso di un numero reale coincide con l'usuale definizione di modulo di un numero reale, la Definizione 2.1 è, a tutti gli effetti, una estensione di quella già vista per le successioni reali.

OSSERVAZIONE 2.2. A guardar bene, la (87) dice che la successione di numeri reali non negativi $d_n := |z_n - \ell|$ tende a zero per $n \rightarrow +\infty$. Quindi una successione complessa z_n converge ad ℓ se e solo se la distanza di z_n da ℓ è infinitesima per $n \rightarrow +\infty$.

ESEMPIO 2.3. Un esempio interessante di successione complessa è

$$z_n = q^n \quad (q \in \mathbb{C}).$$

Dato che $|z_n| = |q^n| = |q|^n$, il comportamento della successione nel caso $|q| < 1$ è determinato da quello della successione reale $a_n = |q|^n$

$$|q| < 1 \quad \Rightarrow \quad \lim_{n \rightarrow +\infty} q^n = 0.$$

Cosa succede nel caso $|q| \geq 1$? Ad esempio, come si comportano le successioni complesse $z_n = (2i)^n$ e $w_n = i^n$ per $n \rightarrow \infty$?

Ad una successione complessa $z_n = x_n + iy_n$ sono associate in modo naturale le due successioni reali x_n e y_n della sua parte reale e immaginaria, rispettivamente. E' possibile dimostrare che

$$(88) \quad z_n = x_n + iy_n \text{ è convergente in } \mathbb{C} \quad \Longleftrightarrow \quad x_n, y_n \text{ sono convergenti in } \mathbb{R}.$$

Quindi trattare successioni complesse può essere visto come trattare contemporaneamente una coppia di successioni reali.

ESERCIZIO 2.4. (i) *Dimostrare che, per ogni $z \in \mathbb{C}$, vale*

$$\max\{|\operatorname{Re} z|, |\operatorname{Im} z|\} \leq |z| \leq |\operatorname{Re} z| + |\operatorname{Im} z|.$$

(ii) *Utilizzare (i) per dimostrare l'equivalenza (88).*

La differenza sostanziale tra \mathbb{C} e \mathbb{R} sta nel fatto che *in \mathbb{C} non è definito un ordinamento*. Questo vuol dire che, per successioni complesse, non ha senso parlare di monotonia, né di divergenza a $+\infty$ e $-\infty$. Resta ben definito il concetto di divergenza in modulo: *la successione z_n diverge ad ∞ (o diverge in modulo), se la successione di numeri reali positivi $|z_n|$ diverge a $+\infty$.*

Serie complesse. Allo stesso modo, si possono considerare serie a termini complessi. Una serie complessa

$$\sum_{n=0}^{\infty} z_n$$

è **convergente** se, in \mathbb{C} , esiste il limite della successione complessa $s_n := \sum_{k=0}^n z_k$ (successione delle somme parziali). Tale limite è detto **somma della serie**.

ESEMPIO 2.5. Come nel caso reale, è possibile considerare la serie geometrica, definita dalla successione $z_n = q^n$ con $q \in \mathbb{C}$. La successione delle somme parziali è

$$s_n = 1 + q + q^2 + \cdots + q^{n-1} + q^n = \frac{1 - q^{n+1}}{1 - q}.$$

Quindi se $|q| < 1$, la serie converge e la sua somma è

$$(89) \quad \sum_{n=0}^{\infty} q^n = \frac{1}{1 - q} \quad q \in \mathbb{C} \quad \text{tale che} \quad |q| < 1.$$

Otteniamo quindi la stessa formula già vista nel caso reale.

Anche la nozione di convergenza assoluta si estende al caso complesso.

DEFINIZIONE 2.6. La serie complessa $\sum_{n=0}^{\infty} z_n$ è **assolutamente convergente** se la serie reale $\sum_{n=0}^{\infty} |z_n|$ è convergente.

Dato che la serie $\sum_{n=0}^{\infty} |z_n|$ è una serie a termini reali, è possibile, almeno in alcuni casi, ricondurre la convergenza di una serie complessa a quella di una serie reale, tornando quindi in un ambito più familiare.

Inoltre, vale sempre l'implicazione

$$\text{convergenza assoluta} \quad \Rightarrow \quad \text{convergenza (semplice)}.$$

Infatti, se la serie $\sum_{n=0}^{\infty} |z_n|$ è convergente, allora lo sono anche le serie $\sum_{n=0}^{\infty} |\operatorname{Re} z_n|$ e $\sum_{n=0}^{\infty} |\operatorname{Im} z_n|$. Dato che per serie reali la convergenza assoluta implica la convergenza semplice, la precedente affermazione implica che le serie reali di termini generici $\operatorname{Re} z_n$ e $\operatorname{Im} z_n$ sono semplicemente convergenti. Per concludere, basta osservare che

$$\operatorname{Re} s_n = \operatorname{Re} \sum_{k=0}^n z_k = \sum_{k=0}^n \operatorname{Re} z_k \quad \text{e} \quad \operatorname{Im} s_n = \operatorname{Im} \sum_{k=0}^n z_k = \sum_{k=0}^n \operatorname{Im} z_k,$$

e ricordare (88)...

Per serie convergenti, vale la seguente versione della disuguaglianza triangolare

$$\left| \sum_{n=0}^{\infty} z_n \right| \leq \sum_{n=0}^{\infty} |z_n|.$$

ESEMPIO 2.7. Utilizziamo la convergenza assoluta per dimostrare la convergenza di una serie complessa fondamentale:

$$(90) \quad \sum_{n=0}^{\infty} \frac{z^n}{n!} \quad z \in \mathbb{C}.$$

Vi ricorda qualcosa? La convergenza di questa serie discende direttamente dalla convergenza della corrispondente serie nei reali. Infatti, studiamo la convergenza assoluta e consideriamo la serie

$$\sum_{n=0}^{\infty} \left| \frac{z^n}{n!} \right| = \sum_{n=0}^{\infty} \frac{|z|^n}{n!}.$$

Questa serie è convergente (la sua somma è $e^{|z|}$), quindi la serie (90) è convergente assolutamente e, di conseguenza, semplicemente.

Funzioni complesse. Una funzione complessa di variabile complessa è:

$$f : I \subset \mathbb{C} \rightarrow \mathbb{C}$$

dove I è il sottoinsieme di \mathbb{C} in cui è definita la funzione f .

Se, per $\zeta \in \mathbb{C}$, esiste $r > 0$ tale che l'intorno di ζ di raggio r , escluso il punto ζ stesso, è contenuto in I , allora si può dare senso alla nozione di limite di f per $z \rightarrow \zeta$ utilizzando la stessa definizione già data nel caso reale.

DEFINIZIONE 2.8. *La funzione f tende ad $\ell \in \mathbb{C}$ per $z \rightarrow \zeta \in I$, e si scrive $\lim_{z \rightarrow \zeta} f(z) = \ell$, se*

$$\forall \varepsilon > 0, \quad \exists \delta > 0 \quad \text{tale che} \quad |f(z) - \ell| < \varepsilon \quad \forall z \in I, \quad 0 < |z - \zeta| < \delta.$$

Se (come faremo sempre nel seguito) si suppone che per ogni $\zeta \in I$ esista $r > 0$ tale che l'intorno di ζ di raggio r sia interamente contenuto in I , in altre parole, se si suppone che ogni punto $\zeta \in I$ sia interno ad I , allora è possibile estendere la definizione di *continuità* al caso di funzioni complesse di variabile complessa.

DEFINIZIONE 2.9. *La funzione f è continua in $\zeta \in I$ se $\lim_{z \rightarrow \zeta} f(z) = f(\zeta)$, cioè se*

$$\forall \varepsilon > 0, \quad \exists \delta > 0 \quad \text{tale che} \quad |f(z) - f(\zeta)| < \varepsilon \quad \forall z \in I, \quad |z - \zeta| < \delta.$$

Né più né meno della definizione di continuità già vista nel caso reale, con l'unica accortezza che i moduli che compaiono qui sono moduli di numeri complessi.

Ad esempio, la funzione $f(z) = 2iz - 1$ è continua in ζ per ogni $\zeta \in \mathbb{C}$. Infatti

$$|f(z) - f(\zeta)| = |(2iz - 1) - (2i\zeta - 1)| = |2i||z - \zeta| = 2|z - \zeta|,$$

quindi $|f(z) - f(\zeta)| < \varepsilon$ per tutti valori z per cui $|z - \zeta| < \delta = \varepsilon/2$.

ESEMPIO 2.10. Consideriamo un altro esempio: $f(z) = \bar{z}$, cioè la funzione che associa ad un numero complesso $z = x + iy$, il suo complesso coniugato $x - iy$. Fissato $\zeta = \xi + i\eta$, se $|z - \zeta| \equiv \sqrt{(x - \xi)^2 + (y - \eta)^2} < \delta$,

$$\begin{aligned} |f(z) - f(\zeta)| &= |(x - iy) - (\xi - i\eta)| = |(x - \xi) - i(y - \eta)| \\ &= \sqrt{(x - \xi)^2 + (y - \eta)^2} < \delta. \end{aligned}$$

La condizione $|f(z) - f(\zeta)| < \varepsilon$ è soddisfatta per $\delta = \varepsilon$.

Quali funzioni complesse sono continue? Ragioniamo come già fatto in \mathbb{R} . Somma, prodotto di funzioni continue sono funzioni continue. Quindi, dal fatto che le costanti e la funzione z sono funzioni continue, possiamo dedurre che tutti i polinomi in \mathbb{C} sono funzioni continue. Analogamente per le funzioni razionali.

Infine è possibile definire la derivata di funzioni complesse (qui è fondamentale la possibilità di dividere per un numero complesso).

DEFINIZIONE 2.11. Una funzione f è derivabile in $\zeta \in I$ (o olomorfa in $\zeta \in I$) se esiste finito il limite

$$(91) \quad f'(\zeta) = \lim_{z \rightarrow \zeta} \frac{f(z) - f(\zeta)}{z - \zeta}.$$

Ad esempio, per calcolare la derivata della funzione $f(z) = z^n$ ($n \in \mathbb{N}$), basta riscrivere il rapporto incrementale nella forma

$$\frac{f(z) - f(\zeta)}{z - \zeta} = \frac{z^n - \zeta^n}{z - \zeta} = z^{n-1} + z^{n-2}\zeta + \dots + \zeta^{n-1},$$

e passare al limite per $z \rightarrow \zeta$, ottenendo

$$(z^n)' = nz^{n-1} \quad \forall n \in \mathbb{N}.$$

Anche in quest'ambito è vera l'implicazione:

$$\text{derivabilità} \quad \Rightarrow \quad \text{continuità}.$$

Perché? Basta ragionare come in \mathbb{R} :

$$\lim_{z \rightarrow \zeta} [f(z) - f(\zeta)] = \lim_{z \rightarrow \zeta} \frac{f(z) - f(\zeta)}{z - \zeta} \cdot (z - \zeta) = f'(\zeta) \cdot 0 = 0$$

Nel caso reale, a partire dall'operazione di derivazione abbiamo dedotto informazioni relative alla monotonia delle funzioni. E' possibile fare lo stesso nel caso delle funzioni

complesse? Assolutamente no! Dato che in \mathbb{C} non è definito l'ordinamento, non ha senso parlare di funzioni crescenti/decrescenti. Analogamente concetti come massimo, minimo, convessità e concavità sono privi di senso per funzioni $f : I \subset \mathbb{C} \rightarrow \mathbb{C}$.

3. L'esponenziale complesso

Una funzione particolarmente importante è la funzione esponenziale. Dato che la serie considerata nell'Esempio 2.7 è convergente per ogni $z \in \mathbb{C}$, è più che ragionevole utilizzarla per definire la funzione esponenziale anche nei complessi:

$$(92) \quad e^z := \sum_{n=0}^{\infty} \frac{z^n}{n!} \quad z \in \mathbb{C}.$$

Come si può immaginare/sperare, anche per (92), vale la proprietà

$$(93) \quad e^{w+z} = e^w e^z \quad \forall w, z \in \mathbb{C}.$$

La dimostrazione si basa sull'uso dei prodotti di serie. Niente dettagli in questa sede.

La funzione esponenziale è continua? È derivabile? L'uno e l'altro. Dimostriamo prima di tutto la continuità in 0. Dato che $e^0 = 1$, bisogna stimare $|e^z - 1|$:

$$|e^z - 1| = \left| \sum_{n=0}^{\infty} \frac{z^{n+1}}{(n+1)!} \right| = \left| z \sum_{n=0}^{\infty} \frac{z^n}{(n+1)!} \right| \leq |z| \sum_{n=0}^{\infty} \frac{|z|^n}{(n+1)!} \leq |z| \sum_{n=0}^{\infty} \frac{|z|^n}{n!} = |z|e^{|z|},$$

quindi, dato che $e^{|z|} \leq e$ per $|z| \leq 1$,

$$|e^z - 1| \leq e|z| \quad \forall |z| \leq 1,$$

da cui si deduce la continuità in $\zeta = 0$. Se $\zeta \in \mathbb{C}$, allora

$$\lim_{z \rightarrow \zeta} e^z = \lim_{z \rightarrow \zeta} e^{\zeta + (z - \zeta)} = e^\zeta \lim_{z \rightarrow \zeta} e^{z - \zeta} = e^\zeta \lim_{h \rightarrow 0} e^h = e^\zeta,$$

quindi la funzione è continua in ζ .

Anche per la derivabilità, consideriamo prima $\zeta = 0$: si ha

$$\frac{e^z - 1}{z} - 1 = \sum_{n=1}^{\infty} \frac{z^n}{(n+1)!} = z \sum_{n=0}^{\infty} \frac{z^n}{(n+2)!},$$

quindi, per $|z| \leq 1$,

$$\left| \frac{e^z - 1}{z} - 1 \right| \leq |z| \sum_{n=0}^{\infty} \frac{|z|^n}{(n+2)!} \leq |z| \sum_{n=0}^{\infty} \frac{|z|^n}{n!} = |z|e^{|z|} \leq e|z|,$$

che mostra che la derivata di e^z in 0 è 1.

Per ζ qualsiasi,

$$\lim_{z \rightarrow \zeta} \frac{e^z - e^\zeta}{z - \zeta} = \lim_{z \rightarrow \zeta} \frac{e^{\zeta + (z - \zeta)} - e^\zeta}{z - \zeta} = e^\zeta \lim_{z \rightarrow \zeta} \frac{e^{z - \zeta} - 1}{z - \zeta} = e^\zeta \lim_{h \rightarrow 0} \frac{e^h - 1}{h} = e^\zeta.$$

Quindi

$$(e^z)' = e^z \quad \forall z \in \mathbb{C}.$$

Fissato $\lambda \in \mathbb{C}$, per la funzione $f(z) = e^{\lambda z}$, vale

$$\frac{d(e^{\lambda z})}{dz} = e^{\lambda z} \frac{d(\lambda z)}{dz} = \lambda e^{\lambda z}.$$

Per derivare la funzione f , basta moltiplicare per λ la funzione stessa: in altri termini, $f(z) = e^{\lambda z}$ verifica l'equazione (differenziale) $f' = \lambda f$. Lo stesso è vero per le funzioni $A e^{\lambda z}$ con $A \in \mathbb{C}$.

Quale equazione è soddisfatta dalla derivata seconda di $f(z) = A e^{\lambda z}$? Facile:

$$f''(z) = \frac{d(A \lambda e^{\lambda z})}{dz} = A \lambda \frac{d(e^{\lambda z})}{dz} = A \lambda^2 e^{\lambda z} = \lambda^2 f,$$

quindi $f'' - \lambda^2 f = 0$.

Il fatto che le funzioni del tipo $A e^{\lambda z}$ soddisfino le semplicissime equazioni $f' - \lambda f = 0$ e $f'' - \lambda^2 f = 0$ fa di questa classe il mattone fondamentale per la costruzione delle soluzioni di equazioni differenziali lineari a coefficienti costanti, come vedremo tra un paio di fogli.

Formula di Eulero. Prima di concludere, deduciamo una formula estremamente interessante. Se calcoliamo l'esponenziale e^z in $z = ix \in i\mathbb{R}$, otteniamo

$$e^{ix} = \sum_{n=0}^{\infty} \frac{i^n x^n}{n!}.$$

Dato che

$$i^0 = i^4 = \dots = 1, \quad i^1 = i^5 = \dots = i, \quad i^2 = i^6 = \dots = -1, \quad i^3 = i^7 = \dots = -i,$$

possiamo scrivere

$$e^{ix} = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}.$$

Dalle formule di Taylor per le funzioni $\cos x$ e $\sin x$, segue la formula di Eulero:

$$(94) \quad e^{ix} = \cos x + i \sin x \quad \forall x \in \mathbb{R}.$$

Chi è dunque l'esponenziale del numero complesso $z = x + iy$? Una sintesi opportuna di esponenziale reale e funzioni trigonometriche: applicando (94),

$$e^z = e^{x+iy} = e^x e^{iy} = e^x (\cos y + i \sin y) = e^x \cos y + i e^x \sin y.$$

In particolare $\operatorname{Re} e^z = e^x \cos y$ e $\operatorname{Im} e^z = e^x \sin y$.

La formula di Eulero, interessante di per sé, è estremamente comoda per rappresentare i numeri complessi quando si usino le coordinate polari (ρ, θ) :

$$z = \rho (\cos \theta + i \sin \theta) = \rho e^{i\theta}$$

Le formule per il prodotto e le potenze diventano, facendo uso della rappresentazione esponenziale, conseguenze immediate di (93). Ad esempio,

$$z_1 \cdot z_2 = (\rho_1 e^{i\theta_1}) \cdot (\rho_2 e^{i\theta_2}) = \rho_1 \rho_2 e^{i(\theta_1 + \theta_2)}.$$

Da (94) si deduce, in particolare, che l'esponenziale complesso non è una funzione iniettiva e quindi nemmeno invertibile. La definizione della funzione logaritmo nei numeri complessi richiede quindi una particolare attenzione. Qui non approfondiremo la questione. Anzi, non la toccheremo proprio.

Equazioni differenziali

Il problema più semplice che abbiamo affrontato e che rientra nella categoria delle equazioni differenziali è la determinazione di primitive

data $f(x)$ trovare tutte le funzioni $F(x)$ tali che $F'(x) = f(x)$.

In generale, un'equazione differenziale è un oggetto del tipo:

$$\mathcal{F}(x, y, y', y'', \dots) = 0,$$

cioè è una relazione che collega una funzione incognita y con la variabile indipendente x e con le derivate y', y'', \dots . Il caso della ricerca delle primitive corrisponde alla scelta $\mathcal{F} = y - F(x)$. Una funzione y , sufficientemente regolare, che verifica

$$\mathcal{F}(x, y(x), y'(x), y''(x), \dots) = 0,$$

per ogni scelta della variabile x , è detta **soluzione** (o **integrale**) dell'equazione.

In quel che segue, ci limitiamo a considerare alcuni tipi specifici di equazioni differenziali: le equazioni lineari. Il significato, in quest'ambito, del vocabolo “lineare” sarà più chiaro tra qualche pagina.

1. Equazioni lineari del I ordine a coefficiente costante

Equazioni omogenee. Per iniziare, consideriamo un'equazione differenziale particolarmente semplice

$$(95) \quad y' + ay = 0 \quad a \in \mathbb{R}.$$

L'equazione (95) è *del primo ordine, lineare, a coefficiente costanti, omogenea*:

- del *primo ordine* perché nell'equazione compare solo la derivata prima;
- *lineare* perché la dipendenza da y, y' è lineare;
- a *coefficiente costante* perché il coefficienti di a è indipendente da t ;
- *omogenea*, perché la funzione $y = 0$ è una soluzione dell'equazione.

OSSERVAZIONE 1.1. *Il termine “lineare”.* Supponiamo che y_1 e y_2 siano due soluzioni dell'equazione (95). Dati $\alpha, \beta \in \mathbb{R}$, consideriamo $w(t) := \alpha y_1(t) + \beta y_2(t)$, combinazione lineare di y_1 e y_2 . Allora:

$$w' + aw = \alpha y_1' + \beta y_2' + \alpha a y_1 + \beta a y_2 = \alpha(y_1' + a y_1) + \beta(y_2' + a y_2) = 0.$$

Quindi, dal fatto che l'equazione è lineare e omogenea discende la proprietà seguente: *ogni combinazione lineare di soluzioni di (95) è essa stessa una soluzione di (95).*

Passiamo al problema di trovare tutte le soluzioni dell'equazione (95). Dato che tale equazione è risolta da tutte le funzioni della forma Ae^{-at} con $A \in \mathbb{R}$, poniamo $y(t) = A(t)e^{-at}$ e cerchiamo tutte le funzioni A per cui y è soluzioni. Dato che $y' = (A' - aA)e^{-at}$, l'equazione (95) equivale a

$$(A' - aA)e^{-at} + aAe^{-at} = 0 \quad \iff \quad A' = 0.$$

Dato che la derivata di A è identicamente nulla, A è costante. Ricapitolando:

$$y'(t) + ay(t) = 0 \quad \forall t \in \mathbb{R} \quad \iff \quad y = Ae^{-at} \quad \text{per qualche } A \in \mathbb{R}.$$

Se la soluzione cercata deve soddisfare, oltre all'equazione differenziale (95), la condizione iniziale $y(t_0) = y_0$ per t_0, y_0 assegnati, si parla di **problema di Cauchy** per (95):

$$(96) \quad \text{trovare } y = y(t) \text{ tale che } \begin{cases} y' + ay = 0, \\ y(t_0) = y_0 \end{cases}$$

Determinare la soluzione è più che facile: dato che conosciamo tutte le soluzioni dell'equazione differenziale, basta determinare quali di queste verificano il vincolo imposto dalla *condizione iniziale* $y(t_0) = y_0$. Se $y(t) = Ae^{-at}$, sostituendo

$$y_0 = y(t_0) = Ae^{-at_0} \quad \Rightarrow \quad A = y(t_0)e^{at_0} = y_0e^{at_0},$$

quindi c'è un'unica soluzione per il problema di Cauchy (96) ed è $y(t) = y_0e^{-a(t-t_0)}$.

Equazioni non omogenee. Data una funzione f integrabile, consideriamo

$$(97) \quad y' + ay = f(t) \quad a \in \mathbb{R}, \quad f : \mathbb{R} \rightarrow \mathbb{R}.$$

L'equazione (97) è *del primo ordine, lineare, a coefficienti costanti, non omogenea*. La presenza del termine noto $f = f(t)$ fa in modo che $y = 0$ non sia soluzione (tranne nel caso $f \equiv 0$). Per cercare le soluzioni di questa equazione utilizziamo il metodo di *variazione delle costanti*: cerchiamo soluzioni della forma

$$y(t) = A(t)e^{-at}$$

dove $A = A(t)$ è una funzione da determinare. Dato che $y' = (A' - aA)e^{-at}$, sostituendo nell'equazione differenziale:

$$f = y' + ay = (A' - aA)e^{-at} + aAe^{-at} = A'e^{-at} \quad \iff \quad A'(t) = f(t)e^{at}.$$

Se la funzione f è integrabile in ogni intervallo di \mathbb{R} , la funzione A è:

$$A(t) = \int_0^t f(s)e^{as} ds + C \quad C \in \mathbb{R}$$

e quindi anche la funzione y è determinata:

$$y(t) = A(t)e^{-at} = Ce^{-at} + \int_0^t f(s)e^{a(s-t)} ds \quad C \in \mathbb{R}.$$

La soluzione dell'equazione non omogenea è quindi della forma

$$y(t) = y_0(t) + \bar{y}(t) \quad \text{dove} \quad \begin{cases} y_0(t) = Ce^{-at} & \text{soluzione generale di (95)} \\ \bar{y}(t) = \int_0^t f(s)e^{a(s-t)} ds & \text{soluzione particolare di (97)}. \end{cases}$$

L'aver scelto in \bar{y} come primo estremo di integrazione $s = 0$ è del tutto arbitrario, un qualsiasi altro numero fissato sarebbe andato altrettanto bene. Ad esempio, per trovare la soluzione del problema di Cauchy

$$(98) \quad \text{trovare } y = y(t) \text{ tale che} \quad \begin{cases} y' + ay = f(t), \\ y(t_0) = y_0 \end{cases}$$

è più comodo scegliere come primo estremo di integrazione t_0

$$y(t) = Ce^{-at} + \int_{t_0}^t f(s)e^{a(s-t)} ds \quad C \in \mathbb{R},$$

di modo che

$$y_0 = y(t_0) = Ce^{-at_0} + \int_{t_0}^{t_0} f(s)e^{a(s-t_0)} ds = Ce^{-at_0} \quad \Rightarrow \quad C = y_0 e^{at_0},$$

e la soluzione richiesta è

$$y(t) = y_0 e^{-a(t-t_0)} + \int_{t_0}^t f(s)e^{a(s-t)} ds.$$

Sistemi disaccoppiati di equazioni lineari. Il caso di un'unica funzione incognita è evidentemente il caso più semplice. In generale, è utile considerare situazioni in cui sono presenti più funzioni incognite: y_1, y_2, \dots . Risolvere sistemi nel caso generale è complicato e non è questa la sede in cui affrontare la questione. Esiste però una situazione facile in cui possibile ricondursi al caso studiato in precedenza: assegnate le costanti a, b, c , cerchiamo le soluzioni y_1, y_2 del sistema

$$\begin{cases} y_1' + ay_1 = 0, \\ y_2' + by_1 + cy_2 = 0. \end{cases}$$

La peculiarità di questo sistema è che la prima equazione $y_1' + ay_1 = 0$ è disaccoppiata dalla seconda, cioè è una equazione per la sola incognita y_1 (y_2 non compare). Qui si può risolvere la prima equazione e, successivamente, risolvere la seconda dopo aver inserito l'espressione esplicita per y_1 . Procediamo

$$y_1' + ay_1 = 0 \quad \Rightarrow \quad y_1 = Ae^{-at} \quad A \in \mathbb{R},$$

quindi l'equazione per y_2 diviene $y_2' + cy_2 = -Abe^{-at}$, cioè un'equazione non omogenea. Perciò, se $c \neq a$,

$$y_2(t) = Be^{-ct} - Ab \int_0^t e^{-as+c(s-t)} ds = Be^{-at} + \frac{Ab}{c-a} (e^{-at} - e^{-ct}) \quad A, B \in \mathbb{R}.$$

Invece, se $c = a$,

$$y_2(t) = Be^{-at} - Ab \int_0^t e^{-at} ds = (B - Abt) e^{-at} \quad A, B \in \mathbb{R}.$$

Si faccia attenzione alle costanti in gioco: a, b, c sono dati del problema, mentre A, B sono costanti arbitrarie (ogni scelta di A, B corrisponde ad una soluzione).

Equazioni lineari del I ordine a coefficiente variabile. Torniamo, ora, al caso di una singola equazione differenziale, e consideriamo il caso

$$(99) \quad y' + a(t)y = b(t) \quad t \in \mathbb{R},$$

dove a e b sono funzioni reali continue e la funzione incognita $y = y(x)$ è anch'essa una funzione reale. L'equazione differenziale (99) si dice *equazione differenziale lineare del primo ordine a coefficiente variabile*.

Come sempre, il problema di Cauchy consiste nel trovare la soluzione di

$$(100) \quad y' + a(t)y = b(t) \quad y(t_0) = y_0,$$

dove t_0 e y_0 sono assegnati.

Ecco come procedere per determinare l'integrale generale di (99). Indicata con A una primitiva della funzione a , moltiplichiamo l'equazione per la funzione $e^{A(t)}$, ottenendo

$$e^{A(t)} y'(t) + e^{A(t)} a(t)y(t) = e^{A(t)} b(t).$$

Il termine a primo membro è la derivata della funzione $e^{A(t)}y(t)$, infatti

$$(e^{A(t)}y(t))' = e^{A(t)}y'(t) + (e^{A(t)})'y(t) = e^{A(t)}y'(t) + e^{A(t)}a(t)y(t).$$

Quindi l'equazione prende la forma

$$(e^{A(t)}y(t))' = e^{A(t)}b(t).$$

Calcoliamo l'integrale indefinito del secondo membro

$$\int e^{A(t)}b(t) dt = F(t) + C.$$

Dato che la derivata di $e^{A(t)}y(t)$ è la funzione $e^{A(t)}b(t)$, deve valere

$$e^{A(t)}y(t) = F(t) + C.$$

Esplicitando rispetto a y , otteniamo la soluzione generale di (99):

$$(101) \quad y(t) = (F(t) + C)e^{-A(t)} \quad \text{dove} \quad \begin{cases} A(t) = \int a(t) dt, \\ F(t) = \int e^{A(t)} b(t) dt, \\ C \in \mathbb{R}. \end{cases}$$

ESERCIZIO 1.2. Risolvere l'equazione differenziale $y' + y \cos t = \frac{1}{2} \sin(2t)$.

Nel caso in cui si debba risolvere un problema di Cauchy, quindi se è assegnata anche la condizione iniziale $y(t_0) = y_0$, basta imporre che la generica soluzione data da (101) soddisfi la condizione data.

Una strada alternativa, che utilizza l'uso di integrali definiti (anziché indefiniti), segue lo stesso schema visto in precedenza:

– si calcola A , primitiva di a , e si moltiplica l'equazione per $e^{A(t)}$ in modo da ottenere

$$(y e^{A(t)})' = b(t) e^{A(t)}.$$

– si integra in (t_0, t) e, per il teorema fondamentale del calcolo integrale,

$$y(t) e^{A(t)} - y(t_0) e^{A(t_0)} = \int_{t_0}^t b(t) e^{A(t)} dt,$$

– si usa la condizione $y(t_0) = y_0$ e si esplicita $y(t)$,

$$y(t) = y_0 e^{A(t_0)-A(t)} + e^{-A(t)} \int_{t_0}^t b(t) e^{A(t)} dt,$$

Se $A(t)$ è scelta come $\int_{t_0}^t a(s) ds$, la formula può essere riscritta come

$$y(t) = y_0 \exp\left(\int_{t_0}^t a(s) ds\right) + \int_{t_0}^t b(t) \exp\left(-\int_t^t a(s) ds\right) dt.$$

ESERCIZIO 1.3. Risolvere il problema di Cauchy

$$ty' - y + t \ln t = 0 \quad y(1) = 1.$$

2. Equazioni lineari del II ordine a coefficienti costanti omogenee

Consideriamo una particella, cioè un punto in cui supponiamo concentrata la massa m , e supponiamo che si muova lungo una retta. La posizione della particella al tempo t è individuata da una funzione $y = y(t)$, che ne rappresenta la coordinata rispetto ad un sistema di riferimento fissato. La funzione $y'(t)$ è la *velocità* della particella, mentre $y''(t)$ ne è l'*accelerazione*.

In meccanica, il moto della particella è causato dalla presenza di forze, descritte qui da una funzione F , che agiscono sulla particella stessa. La *legge di Newton* afferma

che: la massa m moltiplicata per l'accelerazione è uguale alla forza risultante che agisce sulla particella:

$$(102) \quad my'' = F.$$

Modellizzare l'azione di una certa forza F sulla particella corrisponde ad esprimere F in termini della posizione y , della velocità y' , dell'istante t . Di conseguenza, la legge di Newton prende la forma di una equazione differenziale

$$(103) \quad my'' = F(t, y, y').$$

Ad esempio, consideriamo una particella che si sposti lungo l'asse y , su cui agisca una forza elastica F_e , diretta verso l'origine e proporzionale alla distanza dall'origine. In altre parole, supponiamo $F_e = -ky$ con $k > 0$ (k è il coefficiente di proporzionalità e misura la grandezza delle forza elastica). La legge di Newton dice che il moto della particella è descritto dall'equazione

$$(104) \quad my'' = -ky \quad (\text{equazione dell'oscillatore armonico})$$

Se sulla particella agisce anche la forza d'attrito F_a , che supponiamo proporzionale alla velocità y' , ma diretta in senso contrario, cioè $F_a = -ry'$ con $r \geq 0$, otteniamo l'equazione

$$(105) \quad my'' = -ky - ry' \quad (\text{equazione dell'oscillatore con attrito}).$$

Questo è l'esempio tipo di equazione di cui vogliamo determinare le soluzioni.

Studiamo, perciò, l'equazione differenziale *del secondo ordine, lineare, a coefficienti costanti, omogenea*:

$$(106) \quad y'' + 2by' + cy = 0 \quad b, c \in \mathbb{R}.$$

L'equazione è:

- del *secondo ordine* perché compaiono derivate fino al secondo ordine;
- *lineare* perché la dipendenza da y , y' e y'' è lineare;
- a *coefficienti costanti* perché i coefficienti di b e c non dipendono dalla variabile t ;
- *omogenea*, perché la funzione $y = 0$ è una soluzione dell'equazione.

Un esempio di equazione di questo genere è $y'' + y = 0$. Per ogni $A, B \in \mathbb{R}$, la funzione $y = A \cos t + B \sin t$ ne è soluzione (verificatelo!). Questo fatto è generale: date y_1 e y_2 soluzioni di (106), *tutte le combinazioni lineari di y_1 e y_2 sono soluzioni della stessa equazione*.

DEFINIZIONE 2.1. *Due funzioni y_1 e y_2 sono linearmente dipendenti se una è proporzionale all'altra, cioè se esistono due costanti α, β , non entrambe nulle, tali che*

$\alpha y_1(t) + \beta y_2(t) = 0$ per ogni t . Si dicono **linearmente indipendenti** se non sono **linearmente dipendenti**.

ESEMPIO 2.2. Ad esempio, $y_1 = \sin t$ e $y_2 = \cos t$ sono soluzioni linearmente indipendenti di $y'' + y = 0$. Infatti, se, per qualche $\alpha, \beta \in \mathbb{R}$, si ha $\alpha \sin t + \beta \cos t = 0$ per ogni t , allora

$$\begin{aligned} t = 0 & \Rightarrow \alpha \sin 0 + \beta \cos 0 = \beta = 0, \\ t = \frac{\pi}{2} & \Rightarrow \alpha \sin \frac{\pi}{2} + \beta \cos \frac{\pi}{2} = \alpha = 0. \end{aligned}$$

Quindi $\alpha = \beta = 0$.

Il risultato che segue (di cui omettiamo la dimostrazione) descrive la struttura dell'insieme delle soluzioni di (106).

TEOREMA 2.3. Se y_1 e y_2 sono soluzioni linearmente indipendenti di (106), tutte le soluzioni di (106) sono della forma

$$y(t) = c_1 y_1(t) + c_2 y_2(t) \quad c_1, c_2 \in \mathbb{R}.$$

Quindi, per conoscere tutte le soluzioni dell'equazione differenziale lineare (106) basta individuare una coppia di soluzioni linearmente indipendenti. Tutte le altre soluzioni si ottengono per combinazione lineare. Ad esempio, dato che $\sin t$ e $\cos t$ sono linearmente indipendenti, tutte le soluzioni di $y'' + y = 0$ sono del tipo $A \sin t + B \cos t$ con $A, B \in \mathbb{R}$.

Costruzione delle soluzioni linearmente indipendenti. L'obiettivo ora è determinare una coppia di soluzioni linearmente indipendenti dell'equazione (106). A questo scopo procediamo in una maniera che può sembrare strana, ma molto efficace:

- consideriamo la stessa equazione differenziale per funzioni a valori complessi, cioè cerchiamo $z = z(t) \in \mathbb{C}$ tali che $z'' + 2bz' + cz = 0$;
- determiniamo due soluzioni complesse z_1 e z_2 opportune;
- scegliamo $y_1 = \operatorname{Re} z_1$ e $y_2 = \operatorname{Re} z_2$.

Consideriamo quindi l'equazione

$$(107) \quad z'' + 2bz' + cz = 0 \quad b, c \in \mathbb{R},$$

dove $z : \mathbb{R} \rightarrow \mathbb{C}$ e cerchiamo una soluzione della forma $z = e^{\lambda t}$ dove $\lambda \in \mathbb{C}$ è da determinare. Dato che $z' = \lambda e^{\lambda t}$ e $z'' = \lambda^2 e^{\lambda t}$, λ deve essere tale che

$$\lambda^2 e^{\lambda t} + 2b\lambda e^{\lambda t} + ce^{\lambda t} = 0 \quad \forall t \in \mathbb{R}.$$

Dato che $e^{\lambda t} \neq 0$, λ deve essere soluzione dell'equazione di secondo grado complessa

$$(108) \quad \text{polinomio caratteristico:} \quad \lambda^2 + 2b\lambda + c = 0.$$

Studiamo tre casi a seconda del tipo di radici del polinomio caratteristico.

Caso I. $b^2 > c$. Il polinomio caratteristico (108) ha due radici reali

$$\lambda_1 = -b + \sqrt{b^2 - c}, \quad \lambda_2 = -b - \sqrt{b^2 - c}.$$

Otteniamo in corrispondenza due soluzioni

$$z_1(t) = e^{\lambda_1 t} \quad z_2(t) = e^{\lambda_2 t}.$$

Prendendo la parte reale otteniamo le soluzioni y_1 e y_2 cercate, ma... z_1 e z_2 sono funzioni a valori reali! Quindi in realtà y_1 e y_2 coincidono con z_1 e z_2 , rispettivamente:

$$y_1(t) = \operatorname{Re} z_1(t) = e^{\lambda_1 t} \quad y_2(t) = \operatorname{Re} z_2(t) = e^{\lambda_2 t}.$$

Per verificare che y_1, y_2 sono indipendenti, supponiamo che esistano $\alpha, \beta \in \mathbb{R}$ tali che

$$\alpha e^{\lambda_1 t} + \beta e^{\lambda_2 t} = 0.$$

Derivando e calcolando entrambe le relazioni per $t = 0$, si ottiene un sistema lineare omogeneo per le incognite α, β

$$\alpha + \beta = 0, \quad \alpha \lambda_1 + \beta \lambda_2 = 0.$$

Dato che il determinante di questo sistema è $\lambda_2 - \lambda_1 = 2\sqrt{b^2 - c} > 0$, l'unica soluzione è $\alpha = \beta = 0$.

In conclusione, nel caso $b^2 > c$, tutte le soluzioni di (106) sono della forma

$$y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}.$$

ESERCIZIO 2.4. *Determinare le soluzioni dell'equazione $y'' - 3y' + 2y = 0$.*

Caso II. $b^2 = c$. In questo caso il polinomio caratteristico è il quadrato di un binomio

$$\lambda^2 + 2b\lambda + b^2 = (\lambda + b)^2,$$

ed ha un'unica soluzione $\lambda = -b$. In corrispondenza troviamo un'unica soluzione $z_1(t) = e^{-bt}$ dell'equazione differenziale. Per determinare una seconda soluzione (indipendente da z_1), consideriamo $z_2(t) = te^{-bt}$. Dato che

$$z_2'(t) = (1 - bt)e^{-bt} \quad z_2''(t) = (-2b + b^2 t)e^{-bt},$$

si ha

$$z_2''(t) + 2bz_2'(t) + b^2 z_2(t) = [(-2b + b^2 t) + 2b(1 - bt) + b^2 t]e^{-bt} = 0.$$

Quindi anche z_2 è soluzione dell'equazione.

OSSERVAZIONE 2.5. La scelta di z_2 è suggerita da un procedimento di limite, immaginando il caso di due radici coincidenti come limite di un caso con due radici distinte. Se $\lambda_1 \neq \lambda_2$, l'espressione $\frac{e^{\lambda_2 t} - e^{\lambda_1 t}}{\lambda_2 - \lambda_1}$, è soluzione di (107). Per $\lambda_2 \rightarrow \lambda_1$ otteniamo

$$\lim_{\lambda_2 \rightarrow \lambda_1} \frac{e^{\lambda_2 t} - e^{\lambda_1 t}}{\lambda_2 - \lambda_1} = \lim_{\lambda_2 \rightarrow \lambda_1} \frac{e^{\lambda_2 t} - e^{\lambda_1 t}}{\lambda_2 t - \lambda_1 t} \cdot t = te^{\lambda_1 t},$$

che è la soluzione determinata in precedenza.

Come nel caso precedente, le due soluzioni complesse $z_1(t) = e^{-bt}$ e $z_2(t) = te^{-bt}$ sono a valori reali. Inoltre si tratta di soluzioni indipendenti. Infatti, siano $\alpha, \beta \in \mathbb{R}$ tali che $\alpha e^{-bt} + \beta te^{-bt} = 0$, per ogni t . Dividendo per e^{-bt} , si deduce $\alpha + \beta t = 0$, cioè $\alpha = \beta = 0$.

In conclusione, tutte le soluzioni di (106) nel caso $b^2 = c$, sono del tipo

$$y(t) = c_1 e^{-bt} + c_2 t e^{-bt} = (c_1 + c_2 t) e^{-bt} \quad c_1, c_2 \in \mathbb{R}.$$

ESERCIZIO 2.6. *Determinare tutte le soluzioni dell'equazione $y'' + 4y' + 4y = 0$.*

Caso III. $b^2 < c$. Qui, finalmente!, è necessario l'uso dei complessi. Se $b^2 < c$, il polinomio caratteristico $\lambda^2 + 2b\lambda + c$ ha due radici complesse coniugate

$$\lambda^2 + 2b\lambda + c = 0 \quad \iff \quad \lambda \in \{\lambda_1, \lambda_2\} \quad \text{dove} \quad \begin{cases} \lambda_1 = -b + i\sqrt{c - b^2} \\ \lambda_2 = -b - i\sqrt{c - b^2} \end{cases}$$

Nel seguito, indichiamo con $\nu = \sqrt{c - b^2} > 0$. L'equazione differenziale complessa (107) ha due soluzioni indipendenti date da

$$\begin{aligned} z_1(t) &= e^{\lambda_1 t} = e^{-bt} \left\{ \cos(\nu t) + i \sin(\nu t) \right\}, \\ z_2(t) &= e^{\lambda_2 t} = e^{-bt} \left\{ \cos(\nu t) - i \sin(\nu t) \right\} \end{aligned}$$

La parte reale di queste due (distinte) soluzioni è però la stessa: $e^{-bt} \cos(\nu t)$. Occorre quindi sostituire la soluzione z_2 con un'altra soluzione \tilde{z}_2 diversa. Dato che $e^{\lambda_1 t}$ è soluzione di (107), anche tutte le funzioni della forma $Ae^{\lambda_1 t}$ sono soluzioni della stessa equazione per ogni scelta di $A \in \mathbb{C}$. Scegliamo $A = -i$, ossia poniamo $\tilde{z}_2(t) = -i e^{\lambda_1 t} = -i z_1(t)$, e scegliamo come soluzioni dell'equazione (107) le funzioni

$$z_1(t) = e^{-bt} \left\{ \cos(\nu t) + i \sin(\nu t) \right\}, \quad \tilde{z}_2(t) = e^{-bt} \left\{ \sin(\nu t) - i \cos(\nu t) \right\}.$$

Le soluzioni linearmente indipendenti di (106) sono date dalle parti reali di z_1 e \tilde{z}_2 :

$$\begin{aligned} y_1(t) &= \operatorname{Re} z_1(t) = e^{-bt} \cos(\nu t), \\ y_2(t) &= \operatorname{Re} \tilde{z}_2(t) = e^{-bt} \sin(\nu t). \end{aligned}$$

Quindi, nel caso $b^2 < c$, tutte le soluzioni di (106) sono della forma

$$y(t) = e^{-bt} \left(c_1 \cos(\nu t) + c_2 \sin(\nu t) \right) \quad \nu := \sqrt{c - b^2}.$$

Dato che $-b = \operatorname{Re} \lambda_1$ e $\nu = \operatorname{Im} \lambda_1$, se preferite la le soluzioni si possono scrivere così:

$$y(t) = e^{(\operatorname{Re} \lambda) t} \left(c_1 \cos[(\operatorname{Im} \lambda) t] + c_2 \sin[(\operatorname{Im} \lambda) t] \right) \quad \text{con } \lambda \text{ tale che } \lambda^2 + 2b\lambda + c = 0$$

ESERCIZIO 2.7. *Dimostrare che, se $\nu \neq 0$, le funzioni $e^{-bt} \cos(\nu t)$ e $e^{-bt} \sin(\nu t)$ sono linearmente indipendenti.*

ESERCIZIO 2.8. *Determinare tutte le soluzioni di $y'' + y' + y = 0$.*

Analizziamo le soluzioni nel caso $b, c \geq 0$. Nel modello di oscillatore con attrito, b e c sono positivi dato che rappresentano

$$b = \frac{r}{2m} \quad c = \frac{k}{m},$$

dove m è la massa, r misura l'attrito e k misura la forza elastica.

Nei Casi I e II, in cui $b^2 \geq c$, ossia $r^2 \geq 4km$ la soluzione è data o dalla sovrapposizione di due funzioni esponenziali

$$y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} \quad \text{con} \quad \begin{cases} \lambda_1 = -b + \sqrt{b^2 - c} < 0, \\ \lambda_2 = -b - \sqrt{b^2 - c} < 0. \end{cases}$$

oppure dal prodotto di una funzione lineare con un'esponenziale

$$y(t) = (c_1 + c_2 t) e^{-bt}.$$

In entrambi i casi, la soluzione tende asintoticamente a zero per $t \rightarrow +\infty$:

$$\lim_{t \rightarrow +\infty} y(t) = 0.$$

Fisicamente questo corrisponde al fatto che l'effetto di attrito è talmente forte da impedire alla forza elastica di generare un moto oscillatorio.

Al contrario, nel Caso III, $b^2 < c$, cioè per $r^2 < 4km$, le soluzioni sono

$$y(t) = e^{-bt} (c_1 \cos(\nu t) + c_2 \sin(\nu t)) \quad \nu = \sqrt{c - b^2}.$$

Se $b > 0$ (presenza di attrito), le soluzioni sono il prodotto di un termine oscillatorio e di un termine esponenziale che regola l'ampiezza delle oscillazioni e che tende a 0 per $t \rightarrow +\infty$. In questo caso si parla di *oscillazioni armoniche smorzate*. Il valore $\nu = \sqrt{c - b^2}$ è la *frequenza* delle oscillazioni.

Nel caso in cui $b = 0$ (assenza di attrito), le soluzioni sono date da una combinazione lineare di $\sin(\sqrt{c}t)$ e $\cos(\sqrt{c}t)$, e quindi danno luogo a oscillazioni periodiche *non smorzate*.

Problema di Cauchy. Per individuare un'unica soluzione occorre fissare le *costanti di integrazione* c_1 e c_2 . Questo fatto è sensato rispetto al problema fisico: l'evoluzione del moto dipende dalle *condizioni iniziali*, cioè da posizione e velocità ad un istante t_0 . Il problema di Cauchy per l'equazione (106) è il problema di determinare la soluzione di

$$(109) \quad \begin{cases} y'' + 2by' + cy = 0, \\ y(t_0) = y_0, \quad y'(t_0) = y_1. \end{cases}$$

dove $t_0, y_0, y_1 \in \mathbb{R}$ sono costanti assegnate. Il problema è quindi di trovare, tra tutte le soluzioni dell'equazione $y'' + 2by' + cy = 0$, l'unica che soddisfi le condizioni iniziali $y(t_0) = y_0$ e $y'(t_0) = y_1$.

ESEMPIO 2.9. Cerchiamo la soluzione del problema di Cauchy

$$y'' + y = 0 \quad y(0) = 1, \quad y'(0) = 1.$$

Dato che le soluzioni di $y'' + y = 0$ sono tutte e sole combinazioni lineari di $\sin t$ e $\cos t$, consideriamo una generica soluzione e imponiamo che soddisfi le condizioni iniziali. In questo modo otteniamo un sistema lineare per le costanti c_1, c_2 :

$$y(x) = c_1 \cos t + c_2 \sin t, \quad y(0) = 1, \quad y'(0) = 1,$$

implica

$$\begin{cases} 1 \cdot c_1 + 0 \cdot c_2 = 1 \\ 0 \cdot c_1 + 1 \cdot c_2 = 1. \end{cases}$$

Quindi $c_1 = c_2 = 1$ e l'unica soluzione del problema è $y(t) = \cos t + \sin t$.

In generale, se si sono determinate due soluzioni linearmente indipendenti y_1 e y_2 di (106), è sempre possibile risolvere il problema di Cauchy imponendo le condizioni iniziali date in (109): sia

$$y(x) = c_1 y_1(t) + c_2 y_2(t), \quad y(t_0) = y_0, \quad y'(t_0) = y_1,$$

allora

$$\begin{cases} y_1(0) c_1 + y_2(0) c_2 = y_0 \\ y_1'(0) c_1 + y_2'(0) c_2 = y_1. \end{cases}$$

Il fatto che le soluzioni siano linearmente indipendenti garantisce che il determinante del sistema sia sempre non nullo e che il sistema ammetta un'unica soluzione.

ESECIZIO 2.10. *Determinare le soluzioni dei seguenti problemi di Cauchy*

$$\begin{cases} y'' - y' - 2y = 0 \\ y(0) = 1, \quad y'(0) = -2, \end{cases} \quad \begin{cases} y'' + y' - 6y = 0 \\ y(0) = -1, \quad y'(0) = 3, \end{cases}$$

3. Equazioni lineari del II ordine a coefficienti costanti non omogenee

Torniamo all'esempio dell'oscillatore con attrito. Se supponiamo che sulla particella agisca anche una forza esterna f che sia funzione solo del tempo, cioè $f = f(t)$, l'equazione del moto diviene $my'' + ry' + ky = f(t)$. Nel caso in cui la funzione f sia nulla, si parla di *moto libero*, se invece $f \neq 0$, si parla di *moto forzato*.

Prendiamo il toro per le corna e consideriamo l'equazione

$$(110) \quad y'' + 2by' + cy = f(t) \quad b, c \in \mathbb{R},$$

dove f è una funzione assegnata. L'equazione differenziale (110) è del secondo ordine, lineare, a coefficienti costanti, *non omogenea*.

Se w e v sono due soluzioni, allora la funzione differenza $u := w - v$ soddisfa l'equazione omogenea (106), come si può verificare con una semplice sostituzione. Viceversa, se u è una soluzione dell'equazione omogenea (106) e v dell'equazione non omogenea (110), la somma $w := u + v$ è soluzione dell'equazione (110). Quindi la *soluzione generale dell'equazione non omogenea (110)* è

$$y(t) = y_0(t) + \bar{y}(t) \quad \text{dove} \quad \begin{cases} y_0 & \text{soluzione generale di (106)} \\ \bar{y} & \text{soluzione particolare di (110)}. \end{cases}$$

Dato che abbiamo già una strategia generale per trovare tutte le soluzioni dell'equazione omogenea (106), per risolvere il problema nel caso non omogeneo, basta trovare una *singola* soluzione di (110).

Principio di sovrapposizione. Supponiamo di decomporre la funzione f come somma di due funzioni f_1 e f_2 , cioè supponiamo che sia

$$f(t) = f_1(t) + f_2(t).$$

Se y_1 è una soluzione di $y'' + 2by' + cy = f_1(t)$ e y_2 è una soluzione di $y'' + 2by' + cy = f_2(t)$, allora la funzione somma $y(t) := y_1(t) + y_2(t)$ è soluzione di (110). Questo fatto è noto come **principio di sovrapposizione**. Sostanzialmente significa che se abbiamo una forza “complicata” che, però, può essere decomposta come somma di singole parti “elementari”, possiamo ottenere l'evoluzione del fenomeno completo sommando le evoluzioni relative ad ogni singola parte. In termini matematici possiamo dire che la soluzione generale di

$$y'' + 2by' + cy = f_1(t) + \cdots + f_n(t)$$

è data da

$$y(t) = y_0(t) + \bar{y}_1(t) + \cdots + \bar{y}_n(t)$$

dove

$$y_0 \quad \text{soluzione generale di} \quad y'' + 2by' + cy = 0$$

$$\bar{y}_i \quad \text{soluzione particolare di} \quad y'' + 2by' + cy = f_i(t).$$

L'obiettivo è quindi quello di determinare una soluzione particolare di (110) nel caso di f particolarmente semplici e poi determinare la soluzione particolare per funzioni più complicate che si decompongano come somma di funzioni "semplici".

Forzanti periodiche. $f(t) = Ce^{i\omega t}$ Il caso più importante è quello di forzante periodica, cioè del tipo

$$C \cos(\omega t) \quad \text{o} \quad C \sin(\omega t) \quad (C, \omega \in \mathbb{R}).$$

Invece di lavorare con queste funzioni trigonometriche, è estremamente più semplice ed elegante lavorare nell'ambito complesso. Poniamo $f(t) = Ce^{i\omega t}$ e studiamo

$$(111) \quad z'' + 2bz' + cz = Ce^{i\omega t} \quad C, \omega \in \mathbb{R},$$

in cui l'incognita è la funzione $z = z(t)$ a valori in \mathbb{C} . Prima di risolvere questa equazione, poniamoci il problema del "ritorno": *una volta trovata una soluzione z a valori complessi, come determiniamo la soluzione reale y che cerchiamo?* Indichiamo la funzione z , soluzione di (111), nella forma

$$z(t) = y_1(t) + iy_2(t),$$

dove $y_1 = \operatorname{Re} z$ e $y_2 = \operatorname{Im} z$. Inserendo in (111), abbiamo

$$y_1'' + 2by_1' + cy_1 + i(y_2'' + 2by_2' + cy_2) = C \cos(\omega t) + iC \sin(\omega t).$$

Quindi, uguagliando la parte reale e la parte immaginaria,

$$y_1'' + 2by_1' + cy_1 = C \cos(\omega t)$$

$$y_2'' + 2by_2' + cy_2 = C \sin(\omega t),$$

cioè y_1 , parte reale di z , verifica l'equazione con termine forzante $C \cos(\omega t)$ e y_2 , parte immaginaria di z , con termine forzante $C \sin(\omega t)$. A seconda della funzione f che stiamo considerando, sceglieremo la parte reale o la parte immaginaria della soluzione complessa.

Con la tranquillità di poter tornare nell'ambito reale, riprendiamo lo studio dell'equazione (111) con l'obiettivo di trovarne una soluzione particolare. È sensato immaginare che esista una soluzione con la stessa frequenza di oscillazione di f :

$$z(t) = \sigma e^{i\omega t} \quad \sigma \in \mathbb{C}.$$

Il numero complesso σ è da determinare imponendo la condizione che z sia soluzione dell'equazione cercata.

ESEMPIO 3.1. Partiamo da un problema specifico: trovare una soluzione di

$$y'' - y = \cos t.$$

Dato che $\cos t = \operatorname{Re} e^{it}$, cerchiamo $\sigma \in \mathbb{C}$ tale che $z(t) = \sigma e^{it}$ sia soluzione di $z'' - z = e^{it}$. Dato che $z'' = -\sigma e^{it}$, sostituendo, otteniamo $(-\sigma - \sigma)e^{it} = e^{it}$ da cui $\sigma = -\frac{1}{2}$. Quindi la soluzione complessa z è $-\frac{1}{2}e^{it}$. La soluzione particolare \bar{y} dell'equazione richiesta è data dalla parte reale di z , cioè

$$\bar{y}(t) = \operatorname{Re} z(t) = -\frac{1}{2} \cos t.$$

Nel caso generale, dato che $z'(t) = i\omega\sigma e^{i\omega t}$ e $z''(t) = -\omega^2\sigma e^{i\omega t}$, sostituendo ed eliminando il termine $e^{i\omega t}$, otteniamo

$$-\omega^2\sigma + 2bi\omega\sigma + c\sigma = C \quad \Rightarrow \quad (-\omega^2 + 2bi\omega + c)\sigma = C.$$

Quindi, se $c - \omega^2 + 2bi\omega \neq 0$,

$$\sigma = \frac{C}{c - \omega^2 + 2bi\omega} = C \frac{c - \omega^2 - 2bi\omega}{(c - \omega^2)^2 + 4b^2\omega^2}.$$

Questo valore di σ può essere riscritto nella forma esponenziale

$$\sigma = C\alpha e^{-i\omega\delta} \quad \text{dove} \quad \begin{cases} \alpha = |\sigma| = \frac{1}{\sqrt{(c - \omega^2)^2 + 4b^2\omega^2}} \\ \cos(\omega\delta) = (c - \omega^2)\alpha, \quad \sin(\omega\delta) = -2b\omega\alpha \end{cases}$$

La soluzione complessa z è quindi

$$z(t) = C\alpha e^{i\omega(t-\delta)}.$$

di cui scegliere la parte reale o la parte immaginaria a seconda della funzione f data.

ESERCIZIO 3.2. Determinare una soluzione particolare di $y'' + 2y' + 2y = 2 \sin(2t)$.

Soluzione. Sostituendo $z(t) = \sigma e^{2it}$ in $z'' + 2z' + 2z = 2e^{2it}$, segue $-4\sigma + 4i\sigma + 2\sigma = 2$ da cui $\sigma = \frac{1}{2i-1} = -\frac{1+2i}{5}$. Quindi

$$z(t) = -\frac{1+2i}{5} [\cos(2t) + i \sin(2t)]$$

La soluzione particolare è data dalla parte immaginaria di z

$$y(t) = \operatorname{Im} z(t) = -\frac{1}{5} [2 \cos(2t) + \sin(2t)].$$

Attenzione! Il metodo appena esposto funziona se $c - \omega^2 + 2bi\omega \neq 0$. Rimane quindi da capire il caso in cui

$$c = \omega^2, \quad b\omega = 0.$$

Per cominciare, supponiamo $\omega \neq 0$. La seconda condizione implica $b = 0$ e quindi

$$z'' + cz = Ce^{i\omega t} \quad C, \omega \in \mathbb{R}.$$

Cercando soluzioni $z(t) = \sigma e^{i\omega t}$ con $\sigma \in \mathbb{C}$, otteniamo l'equazione $(c - \omega^2)\sigma = C$. Se $\omega^2 = c$ la relazione precedente dice che non esiste nessuna soluzione della forma $z(t) = \sigma e^{i\omega t}$! Non ci perdiamo d'animo e, in analogia con quanto visto per le equazioni omogenee, cerchiamo una soluzione della forma

$$z(t) = \sigma t e^{i\omega t} \quad \sigma \in \mathbb{C}.$$

Dato che $z'(t) = \sigma(1 + i\omega t)e^{i\omega t}$ e $z''(t) = \sigma(2i\omega - \omega^2 t)e^{i\omega t}$, sostituendo e semplificando il termine $e^{i\omega t}$,

$$2\sigma i\omega = C \quad \implies \quad \sigma = \frac{C}{2i\omega} = -C \frac{i}{2\omega}$$

avendo tenuto conto di $\omega^2 = c$.

ESERCIZIO 3.3. Determinare una soluzione particolare di $y'' + y = \sin t$.

Soluzione. Studiamo l'equazione complessa $z'' + z = e^{it}$ e cerchiamone soluzioni della forma $z(t) = \sigma t e^{it}$. Sostituendo nell'equazione $z'(t) = \sigma(1 + it)e^{it}$ e $z''(t) = \sigma(2i - t)e^{it}$,

$$\sigma(2i - t)e^{it} + \sigma t e^{it} = e^{it} \quad \Rightarrow \quad 2\sigma i = 1.$$

Quindi la soluzione complessa è

$$z(t) = -\frac{it}{2} e^{it} = -\frac{it}{2}(\cos t + i \sin t) = \frac{t}{2} \sin t - i \frac{t}{2} \cos t.$$

La soluzione desiderata è la parte immaginaria di z : $y(t) = \text{Im } z(t) = -\frac{1}{2} t \cos t$.

Resta escluso un solo caso: quello in cui $c - \omega^2 + 2bi\omega = 0$ e $\omega = 0$. Allora $c = 0$ e l'equazione diventa $y'' + 2by' = C$. Ecco le soluzioni cercate in questi casi

$$b \neq 0 : \quad \bar{y} = \frac{Ct}{2b} \quad \text{e} \quad b = 0 : \quad \bar{y} = \frac{Ct^2}{2}.$$

Verificate!

Risonanza. Se $c - \omega^2 + 2bi\omega \neq 0$, una soluzione particolare dell'equazione complessa è

$$z(t) = C\alpha e^{i\omega(t-\delta)} \quad \text{dove} \quad \begin{cases} \alpha = |\sigma| = \frac{1}{\sqrt{(c - \omega^2)^2 + 4b^2\omega^2}} \\ \cos(\omega\delta) = (c - \omega^2)\alpha, \quad \sin(\omega\delta) = -2b\omega\alpha \end{cases}$$

Il fattore α corrisponde ad un fattore di “distorsione” e il fattore $\omega\delta$ ad un fattore di “spostamento di fase” rispetto al segnale esterno originale $Ce^{i\omega t}$.

Il fattore α descrive quanto il “segnale esterno” $Ce^{i\omega t}$ viene amplificato dal sistema. Immaginiamo b, c fissati, e supponiamo di far variare la frequenza ω del segnale. Il grafico di

$$(112) \quad \phi(\omega) = \frac{1}{\sqrt{(c - \omega^2)^2 + 4b^2\omega^2}}$$

è detto *curva di risonanza*.

Nel caso $b, c \neq 0$, ϕ è definita dappertutto, pari e tende a zero per $\omega \rightarrow \pm\infty$ con ordine 2:

$$\lim_{\omega \rightarrow \pm\infty} \frac{1/\sqrt{(c-\omega^2)^2 + 4b^2\omega^2}}{1/\omega^2} = \lim_{\omega \rightarrow \pm\infty} \left[\left(1 - \frac{c}{\omega^2}\right)^2 + \frac{4b^2}{\omega^2} \right]^{-1/2} = 1.$$

La derivata prima di ϕ è

$$\phi'(\omega) = -\frac{4\omega[\omega^2 - (c - 2b^2)]}{[(c - \omega^2)^2 + 4b^2\omega^2]^{3/2}}.$$

– Se $2b^2 \geq c$, la funzione ha un andamento particolarmente semplice: decresce per $\omega > 0$ e cresce per $\omega < 0$, il suo valore massimo è assunto in $\omega = 0$

$$\max_{\omega \in \mathbb{R}} \phi(\omega) = \phi(0) = \frac{1}{|c|}.$$

– Se $2b^2 < c$, la funzione ha due punti di massimo simmetrici nei valori

$$\omega = \pm\sqrt{c - 2b^2},$$

il massimo della funzione è dato da

$$\max_{\omega \in \mathbb{R}} \phi(\omega) = \phi(\pm\sqrt{c - 2b^2}) = \frac{1}{2|b|\sqrt{c - b^2}}.$$

Il fatto significativo è che questo valore massimo tende a $+\infty$ per $b \rightarrow 0$, cioè se l'effetto dell'attrito è piccolo, per una scelta di ω particolare (vicino a $\sqrt{c - b^2}$), il sistema determinato dall'equazione differenziale risponde al segnale esterno $e^{i\omega t}$ con una distorsione estremamente grande. Questo fenomeno è detto *fenomeno della risonanza* e appare in natura in moltissime situazioni differenti.

Lo spazio reale multidimensionale

Lo spazio reale d -dimensionale, indicato con \mathbb{R}^d (dove $d \in \mathbb{N}$), è l'insieme costituito da d -ple ordinate, dette **punti** (o **vettori**), e indicate, nel seguito, con la notazione

$$P = (x_1, \dots, x_d)$$

dove x_1, \dots, x_d indicano le coordinate del punto P . Per indicare un secondo punto verrà spesso utilizzata la lettera Q (con corrispondenti coordinate date da (y_1, \dots, y_d)). Nei casi, particolarmente significativi, planare, i.e. $d = 2$, e spaziale, i.e. $d = 3$, le corrispondenti coordinate saranno indicate con

$$P = (x, y), \quad P = (x, y, z),$$

a seconda della situazione considerata.

L'insieme \mathbb{R}^d è uno spazio vettoriale su \mathbb{R} grazie alla presenza delle operazioni di somma e prodotto per uno scalare: dati $P = (x_1, \dots, x_d)$, $Q = (y_1, \dots, y_d) \in \mathbb{R}^d$ e $\lambda \in \mathbb{R}$,

$$P + Q = (x_1 + y_1, \dots, x_d + y_d), \quad \lambda P = (\lambda x_1, \dots, \lambda x_d)$$

Il vettore nullo $(0, \dots, 0)$ viene indicato con la lettera O .

La differenza fondamentale tra il caso uni-dimensionale, $d = 1$, e il caso multidimensionale, $d > 1$, sta nel fatto che, mentre a differenza di \mathbb{R} , nello spazio \mathbb{R}^d , $d > 1$, non è definita una nozione di ordine e simboli del tipo $<$, \leq non hanno senso in relazione a vettori di dimensione superiore ad uno¹.

1. Serve una struttura metrica

Il modulo è la distanza di un punto dall'origine del sistema di riferimento. Una volta definita tale distanza, grazie alla struttura di spazio vettoriale, è possibile definire la distanza tra due punti qualsiasi dello spazio. In altri termini, dato il modulo, lo spazio considerato viene dotato di una **struttura metrica**. Da tale struttura discende, a catena, il concetto di limite con annessi e connessi.

Come definire il modulo? Supponiamo di avere scelto come misura delle distanze nel caso unidimensionale, quella definita attraverso il valore assoluto: i punti di \mathbb{R} di

¹In alcuni contesti specifici vengono introdotte relazioni di ordine (parziale) nel caso di \mathbb{R}^d , $d > 1$, ma nessuna di queste sarà considerata in queste Note.

coordinate x e y distano, per definizione, $|x - y|$. Passiamo al caso planare: $d = 2$. Il Teorema di Pitagora afferma che, dato un triangolo rettangolo in \mathbb{R}^2 di vertici $O = (0, 0)$, $P = (x, y)$ e $Q = (x, 0)$ con $x, y > 0$ fissati, vale la relazione

$$\ell(OP)^2 = \ell(OQ)^2 + \ell(QP)^2$$

dove $\ell(AB)$ indica la lunghezza del segmento AB . La lunghezza di segmenti unidimensionali è nota ed è data dal modulo unidimensionale della differenza delle coordinate: nel caso in considerazione $\ell(OQ) = |x - 0|$ e $\ell(QP) = |y - 0|$. Quindi $\ell(OP)^2 = x^2 + y^2$, cioè

$$\ell(OP) = \sqrt{x^2 + y^2}.$$

Cosa cambia nel caso tridimensionale? Fissiamo il punto $P = (x, y, z)$ e consideriamo il triangolo rettangolo di vertici $O = (0, 0, 0)$, $P = (x, y, z)$ e $Q = (x, y, 0)$. Il calcolo del caso bidimensionale, mostra che $\ell(OQ)^2 = x^2 + y^2$, quindi, sempre dando fiducia al Teorema di Pitagora, l'espressione della distanza di P da O è data da

$$\ell(OP) = \sqrt{\ell(OQ)^2 + \ell(QP)^2} = \sqrt{x^2 + y^2 + z^2}.$$

A questo punto, non sorprende la **definizione** seguente.

DEFINIZIONE 1.1. Dato $P = (x_1, \dots, x_d) \in \mathbb{R}^d$ si chiama **norma** (o modulo, o valore assoluto) di P il numero reale

$$(113) \quad |P|_d := \sqrt{x_1^2 + \dots + x_d^2} = \left(\sum_{k=1}^d x_k^2 \right)^{1/2}.$$

Nel caso unidimensionale $d = 1$, la relazione (113) diviene $|x|_1 = \sqrt{x^2}$ che coincide con la usuale definizione di modulo di un numero reale.

OSSERVAZIONE 1.2. Nello spazio \mathbb{R}^d è possibile definire anche il **prodotto scalare**: dati $P = (x_1, \dots, x_d)$ e $Q = (y_1, \dots, y_d)$, si pone $P \cdot Q := x_1 y_1 + \dots + x_d y_d$. La norma di $P \in \mathbb{R}^d$ può essere, allora, riscritta nella forma

$$(114) \quad |P|_d = (P \cdot P)^{1/2}.$$

Il prodotto scalare permette di dare senso al concetto di **ortogonalità** e, per questo motivo, si dice che lo spazio \mathbb{R}^d , considerato con il prodotto scalare \cdot , è dotato di una **struttura euclidea**. Dato che la norma $|\cdot|_d$ definita in (113) discende, attraverso (114), dall'introduzione del prodotto scalare, essa viene anche detta **norma euclidea**.

Per fini pratici, occorre prima di tutto convincersi del fatto che la norma $|\cdot|_d$ gode di proprietà analoghe a quella utilizzate in lungo e in largo nel caso unidimensionale.

PROPOSIZIONE 1.3. (*Proprietà della norma*) Per la norma $|\cdot|_d$ valgono:

- i. (*positività*) $|P|_d \geq 0$ per ogni $P \in \mathbb{R}^d$ e $|P|_d = 0$ se e solo se $P = O$;
- ii. (*omogeneità*) $|\lambda P|_d = |\lambda| |P|_d$ per ogni $\lambda \in \mathbb{R}$ e $P \in \mathbb{R}^d$;
- iii. (*disuguaglianza triangolare*) per ogni $P, Q \in \mathbb{R}^d$,

$$(115) \quad |P + Q|_d \leq |P|_d + |Q|_d.$$

Premettiamo alla dimostrazione di queste proprietà il risultato seguente.

LEMMA 1.4 (Disuguaglianza di Cauchy–Schwarz). Per ogni $P, Q \in \mathbb{R}^d$, vale

$$(116) \quad |P \cdot Q| \leq |P|_d |Q|_d \quad \forall P, Q \in \mathbb{R}^d.$$

DIMOSTRAZIONE DEL LEMMA 1.4. Dato $t \in \mathbb{R}$, si ha $|P + tQ|_d^2 \geq 0$. Inoltre

$$|P + tQ|_d^2 = (P + tQ) \cdot (P + tQ) = |P|_d^2 + 2P \cdot Qt + |Q|_d^2 t^2.$$

Dato che il polinomio di secondo grado $|P|_d^2 + 2P \cdot Qt + |Q|_d^2 t^2$ è sempre non negativo, il suo discriminante $(P \cdot Q)^2 - |P|_d^2 |Q|_d^2$ deve essere minore od uguale a zero; ne segue la disuguaglianza (116). \square

DIMOSTRAZIONE DELLA PROPOSIZIONE 1.3. Le dimostrazioni di **i.** e **ii.** non riservano particolari sorprese e sono lasciate per esercizio al lettore. La proprietà **iii.** segue dalla disuguaglianza di Cauchy–Schwarz. Infatti, utilizzando (116), si ottiene

$$|P + Q|_d^2 = |P|_d^2 + 2P \cdot Q + |Q|_d^2 \leq |P|_d^2 + 2|P \cdot Q| + |Q|_d^2 \leq (|P|_d + |Q|_d)^2.$$

Applicando la radice quadrata al primo e all'ultimo membro, si ottiene (115). \square

La norma euclidea $|\cdot|_d$ gode di un'altra proprietà analoga a quella già vista nel caso unidimensionale.

COROLLARIO 1.5. Per ogni $P, Q \in \mathbb{R}^d$, vale la disuguaglianza

$$(117) \quad ||P|_d - |Q|_d| \leq |P - Q|_d.$$

DIMOSTRAZIONE. La disequazione (117) equivale a

$$|P|_d - |P - Q|_d \leq |Q|_d \leq |P|_d + |P - Q|_d.$$

Entrambe le disuguaglianze sono conseguenza delle proprietà della norma. Infatti:

$$|P|_d - |P - Q|_d = |Q + (P - Q)|_d - |P - Q|_d \leq |Q|_d + |P - Q|_d - |P - Q|_d = |Q|_d$$

$$|Q|_d = |P + (Q - P)|_d \leq |P|_d + |Q - P|_d = |P|_d + |P - Q|_d.$$

La dimostrazione è completa. \square

In prima battuta, lo spazio vettoriale \mathbb{R}^d può essere pensato come uno spazio *isotropo*, cioè uno spazio in cui tutti i punti siano uguali tra di loro. In particolare, non c'è nessun motivo per preferire il punto origine O ad altri punti del piano. Si potrebbe decidere, ad esempio, di cambiare sistema di riferimento applicando una traslazione allo spazio \mathbb{R}^d , cioè si può pensare di scegliere un nuovo punto origine P^0 e di individuare tutti i punti di \mathbb{R}^d attraverso il vettore $P - P^0$ anziché tramite il vettore $P = P - O$. Questa idea suggerisce di utilizzare la traslazione e la norma per introdurre una nozione di distanza tra punti qualsiasi di \mathbb{R}^d .

DEFINIZIONE 1.6. *Dati $P, Q \in \mathbb{R}^d$ si chiama distanza di P da Q il numero reale*

$$(118) \quad d(P, Q) := |P - Q|_a.$$

Dalla definizione (118) e dalla Proposizione 1.3, discendono alcune proprietà per la distanza (la cui dimostrazione è lasciata per indispensabile esercizio).

PROPOSIZIONE 1.7. *(Proprietà della distanza) Per la distanza definita in (118) valgono le seguenti proprietà:*

- i. *(positività) $d(P, Q) \geq 0$ per ogni P, Q e $d(P, Q) = 0$ se e solo se $P = Q$;*
- ii. *(simmetria) $d(P, Q) = d(Q, P)$ per ogni $P, Q \in \mathbb{R}^d$;*
- iii. *(disuguaglianza triangolare) per ogni $P, Q, R \in \mathbb{R}^d$,*

$$(119) \quad d(P, Q) \leq d(P, R) + d(R, Q).$$

OSSERVAZIONE 1.8. Conseguenza immediata della definizione (118), è che la distanza d è *invariante per traslazioni*, cioè per ogni $P, Q \in \mathbb{R}^d$ e per ogni $h \in \mathbb{R}^d$ vale

$$d(P + h, Q + h) = d(P, Q).$$

La distanza in \mathbb{R}^d gode anche di altre proprietà di invarianza. Sia A una trasformazione lineare di \mathbb{R}^n in sè che sia *unitaria*², cioè tale che $A^{-1} = A^t$. Da (114) segue

$$d(AP, AQ) = AP \cdot AQ = P \cdot A^tAQ = P \cdot Q = d(P, Q).$$

Si noti che mentre l'invarianza per traslazioni segue solamente dalla relazione che intercorre tra distanza e norma (si veda la formula (118)), l'invarianza per rotazioni discende dalla specifica scelta operata nella definizione di norma. Con norme diverse, tale invarianza potrebbe non valere.

²Un esempio particolarmente interessante di trasformazione unitaria, nel caso $d = 2$, è dato da

$$A_\theta := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \theta \in \mathbb{R}$$

La matrice A_θ descrive la trasformazione del piano \mathbb{R}^2 in sè data da una rotazione in senso antiorario di angolo θ attorno al punto O .

Una volta introdotte le nozioni di norma e distanza, si possono definire, senza difficoltà, i concetti di insiemi limitati³ e di intorni di un punto.

DEFINIZIONE 1.9. *Un insieme $A \subset \mathbb{R}^d$ è limitato se esiste $M > 0$ tale che*

$$|P|_d \leq M \quad \forall P \in A.$$

Dato $P_0 \in \mathbb{R}^d$ e $r > 0$, si chiama intorno aperto di centro P_0 e raggio r l'insieme

$$I_r(P_0) := \{P \in \mathbb{R}^d : |P - P_0|_d < r\};$$

si chiama intorno chiuso di centro P_0 e raggio r l'insieme

$$\bar{I}_r(P_0) := \{P \in \mathbb{R}^d : |P - P_0|_d \leq r\}.$$

La condizione di limitatezza può essere formulata anche come segue: un insieme $A \subset \mathbb{R}^d$ è limitato se e solo se è contenuto in un intorno di centro O (e raggio M sufficientemente grande).

Altre norme. Oltre alla norma euclidea, definita in (113), si possono considerare altre espressioni che continuano a soddisfare le stesse proprietà e che pertanto meritano anchesse il nome di norme. Consideriamo, ad esempio, il caso di \mathbb{R}^2 (analoghe definizioni si possono dare nel caso generale). Dato $p \geq 1$, consideriamo

$$|(x, y)|_{2,p} := (|x|^p + |y|^p)^{1/p}.$$

oppure

$$|(x, y)|_{2,\infty} := \max\{|x|, |y|\}.$$

E' possibile dimostrare che sia ogni norma del tipo $|\cdot|_{2,p}$, $p \in [1, \infty]$, verifica le stesse proprietà espone nella Proposizione 1.3. Le definizioni di intorno di un punto descrivono insiemi di tipo diverso.

Sussiste, al riguardo di tali varie norme diverse, il seguente risultato.

PROPOSIZIONE 1.10. *Dato $p \in [1, +\infty]$, esistono due costanti m_p e M_p tali che*

$$(120) \quad m_p |(x, y)|_{2,2} \leq |(x, y)|_{2,p} \leq M_p |(x, y)|_{2,2} \quad \forall (x, y) \in \mathbb{R}^2.$$

DIMOSTRAZIONE. La disuguaglianza di destra è valida con $M_p = 2^{1/p}$, infatti

$$|(x, y)|_{2,p} = (|x|^p + |y|^p)^{1/p} \leq 2^{1/p} \max\{|x|, |y|\} \leq 2^{1/p} (|x|^2 + |y|^2)^{1/2}.$$

Per l'altra, si può procedere in maniera analoga:

$$|(x, y)|_{2,2} = (|x|^2 + |y|^2)^{1/2} \leq 2^{1/2} \max\{|x|, |y|\} \leq 2^{1/2} (|x|^p + |y|^p)^{1/p},$$

ottenendo la disuguaglianza di sinistra con $m_p := 2^{-1/2}$. □

³Dato che in \mathbb{R}^d , $d > 1$, non sono definite nozioni di ordinamento, non ha senso parlare di insiemi superiormente/inferiormente limitati.

Dalle disuguaglianze (120) segue che:

- gli insiemi limitati rispetto alla norma euclidea e quelli rispetto alla norma $|\cdot|_{2,p}$ coincidono;
- anche se gli intorni relativi a norme diverse sono diversi, all'interno di ogni intorno relativo ad una norma, è sempre contenuto un intorno (se necessario di raggio più piccolo) relativo all'altra norma. Questa proprietà è particolarmente rilevante nell'ambito delle proprietà di convergenza (che vedremo tra poche righe).

Versioni analoghe della Proposizione 1.10 valgono per qualsiasi norma in \mathbb{R}^d che verifichi le proprietà elencate nella Proposizione 1.3.

2. Quattro salti in \mathbb{R}^d : successioni di punti

Una successione di punti in \mathbb{R}^d è un'applicazione da \mathbb{N} a \mathbb{R}^d che associa ad ogni numero naturale n un punto P^n di \mathbb{R}^d . Tramite la definizione di distanza è possibile definire il concetto di limite di una successione $P^n \in \mathbb{R}^d$ utilizzando la stessa definizione del caso unidimensionale.

DEFINIZIONE 2.1. *Siano $P^n, \bar{P} \in \mathbb{R}^d$ per ogni $n \in \mathbb{N}$. La successione P^n converge a \bar{P} per $n \rightarrow +\infty$ e si scrive $\lim_{n \rightarrow +\infty} P^n = \bar{P}$ se*

$$(121) \quad \forall \varepsilon > 0, \quad \exists n_\varepsilon \in \mathbb{N}, \quad \forall n \geq n_\varepsilon, \quad \text{vale } |P^n - \bar{P}|_d < \varepsilon.$$

Che cosa vuol dire, in concreto, che la successione P^n converge a \bar{P} per $n \rightarrow +\infty$? La definizione (121) va letta a partire dalla fine: alcuni punti della successione P^n distano meno di ε dal punto limite \bar{P} . Quindi, la convergenza di P^n a \bar{P} indica una proprietà di "vicinanza" di (alcuni) punti della successione al corrispondente punto limite. Occorre precisare, però, **quanto** siano vicini tali punti al punto limite e **quali** siano i punti a godere di tale proprietà. È a questo fine che vengono introdotti i (misteriosi) valori ε e n_ε . La "vicinanza" è quantificata dal parametro ε , che è **qualsiasi**. L'arbitrarietà di ε fa in modo che la definizione sia interessante qualsiasi sia la scala significativa considerata, indipendentemente dal gusto e dalla preferenza dell'utente di turno. Più il valore di ε è piccolo e più la condizione $|P^n - \bar{P}|_d < \varepsilon$ diviene restrittiva e, di conseguenza, soddisfatta da un minor numero di punti della successione. Il valore n_ε permette di individuare punti della successione per cui la relazione è soddisfatta. La richiesta, nella definizione di limite, è che la distanza di P_n da \bar{P} sia minore di ε tutte le volte che si scelgano indici n sufficientemente grandi. In definitiva, fissata una qualsiasi soglia di errore $\varepsilon > 0$, i termini della successione P_n sono ben approssimati da \bar{P} per tutte le scelte di n maggiori di n_ε .

OSSERVAZIONE 2.2. La nozione di convergenza si basa sulla definizione di distanza. Disponendo di metriche diverse potrebbe quindi accadere che una stessa successione $\{P^n\}$ sia convergente rispetto a certe metriche e non convergente rispetto ad altre. Il risultato della Proposizione 1.10 consente di riconoscere che se una successione $\{P^n\}$ è convergente rispetto alla metrica euclidea lo è anche rispetto alle metriche $|\cdot|_{2,p}$. Per questo motivo, metriche che verifichino stime del tipo (120) si dicono essere **equivalenti** alla metrica euclidea.

Come nel caso del limite di successioni reali, anche il limite di successioni di punti di \mathbb{R}^d è lineare: se P^n e Q^n sono due successioni convergenti, rispettivamente, a P e Q ,

$$\lim_{n \rightarrow +\infty} \alpha P^n + \beta Q^n = \alpha P + \beta Q \quad \text{per ogni } \alpha, \beta \in \mathbb{R}.$$

La dimostrazione è conseguenza immediata della disuguaglianza

$$|(\alpha P^n + \beta Q^n) - (\alpha P + \beta Q)|_d \leq |\alpha| |P^n - P|_d + |\beta| |Q^n - Q|_d.$$

Inoltre, come nel caso reale, il limite di una successione, quando esiste, è unico.

Una successione di punti P^n è individuata univocamente dalle successioni reali delle sue coordinate. Che relazione intercorre tra la convergenza della successione P^n e quella delle successioni delle coordinate?

PROPOSIZIONE 2.3. *Sia $P^n = (x_1^n, \dots, x_d^n)$ e $\bar{P} = (\bar{x}_1, \dots, \bar{x}_d)$. Allora la successione P^n converge a \bar{P} se e solo se*

$$\lim_{n \rightarrow +\infty} x_j^n = \bar{x}_j \quad \text{per ogni } j = 1, \dots, d.$$

DIMOSTRAZIONE. Dalla definizione di distanza in \mathbb{R}^d , segue

$$|x_j^n - \bar{x}_j| \leq |P^n - \bar{P}|_d = \left(\sum_{k=1}^d (x_k^n - \bar{x}_k)^2 \right)^{1/2}$$

per ogni $j \in \{1, \dots, d\}$. Di conseguenza, se $|x_k^n - \bar{x}_k|$ tende a 0 per ogni k , allora anche $|P^n - \bar{P}|_d$ tende a zero. Viceversa, se $|P^n - \bar{P}|_d$ tende a zero, anche la successione reale $|x_j^n - \bar{x}_j|$ tende a zero per $n \rightarrow +\infty$ per ogni j . \square

La Proposizione 2.3 indica che determinare il limite di una successione di punti in \mathbb{R}^d equivale a determinare i limiti di d successioni reali. Ad esempio, per le successioni $P_n := (1/n^2, e^{1/n})$ e $Q_n := (n \sin(1/n), \sin n/n)$ valgono i limiti $\lim_{n \rightarrow +\infty} P_n = (0, 1)$ e $\lim_{n \rightarrow \infty} Q_n = (1, 0)$.

Come nel caso uni-dimensionale, ogni successione convergente è anche limitata e, di conseguenza, una successione non limitata non può essere convergente.

PROPOSIZIONE 2.4. Sia $\{P^n\}$ una successione convergente. Allora la successione $\{P^n\}$ è limitata.

DIMOSTRAZIONE. Sia \bar{P} il limite della successione P^n . Per definizione, esiste n_1 tale che $|P^n - \bar{P}|_d < 1$ per ogni $n \geq n_1$. Vale quindi la stima

$$|P^n|_d \leq |\bar{P}|_d + |P^n - \bar{P}|_d < |\bar{P}|_d + 1 \quad \forall n \geq n_1.$$

Per dimostrare la limitatezza della successione, basta controllare la norma dei termini P^0, \dots, P^{n_1-1} . Quindi, scegliendo

$$M := \max\{|P^0|_d, \dots, |P^{n_1-1}|_d, |\bar{P}|_d + 1\},$$

si ha che $|P^n| \leq M$ per ogni $n \in \mathbb{N}$. □

ESERCIZIO 2.5. Siano P^n e Q^n due successioni in \mathbb{R}^d che convergono, rispettivamente, a \bar{P} e \bar{Q} . Dimostrare che la successione $a_n := P^n \cdot Q^n$ converge a $\bar{P} \cdot \bar{Q}$.

La limitatezza non è una condizione sufficiente per la convergenza: si pensi all'esempio facile della successione $a_n = (-1)^n$. Occorre, quindi, qualche ipotesi aggiuntiva oltre alla limitatezza per essere certi della convergenza della successione in questione. Nel caso uni-dimensionale, ad esempio, vale il *criterio di monotonia*: una successione di numeri reali limitata e monotona è convergente. Nel caso di successioni vettoriali, dato che in \mathbb{R}^n non è possibile introdurre un ordinamento totale se $n \geq 2$, non esiste un analogo di questo criterio.

L'ipotesi di limitatezza, dunque, non implica la convergenza della successione assegnata. E' vero però una proprietà più debole che necessita la definizione seguente.

DEFINIZIONE 2.6. Data la successione di punti $\{P^n\}$ e una successione $\{n_k\}$ di numeri naturali, strettamente crescente, la successione $\{P^{n_k}\}$ si dice essere una **sottosuccessione** della successione $\{P^n\}$.

Il termine sottosuccessione può essere chiarito tramite alcuni esempi. Data la successione $\{n\} = \{0, 1, 2, 3, \dots\}$ dei numeri naturali:

– le successioni

$$\{1, 3, 5, 7, \dots, 2k+1, \dots\}, \quad \{1, 10, 100, 1000, \dots, 10^k, \dots\}$$

sono sottosuccessioni della successione assegnata;

– le successioni

$$\{1, 2, 3, \pi, 4, \dots\}, \quad \{1, 3, 2, 4, \dots\}, \quad \{1, 2, 2, 3, 4, \dots\}$$

non sono sottosuccessioni della successione dei naturali $\{n\}$.

Dalla definizione di convergenza, segue immediatamente che se una successione di punti $\{P^n\}$ converge a \bar{P} allora tutte le sottosuccessioni di $\{P^n\}$ convergono allo stesso punto limite \bar{P} .

TEOREMA 2.7 (*Teorema di Bolzano–Weierstrass*). *Da ogni successione limitata $\{P^n\}$ è possibile estrarre almeno una sottosuccessione $\{P^{n_k}\}$ convergente.*

DIMOSTRAZIONE. La dimostrazione è divisa in due passi: prima si considera il caso di \mathbb{R} e poi il caso generale.

1. **Caso uni-dimensionale.** Dimostriamo prima di tutto il Teorema nel caso uni-dimensionale. Sia $\{x^n\}$ una successione reale limitata, cioè interamente contenuta nell'intervallo $I_0 := [a, b]$ per qualche $a, b \in \mathbb{R}$, $a < b$. Consideriamo i sotto-intervalli I_0^\pm definiti da

$$I_0^- := \left[a, \frac{a+b}{2} \right], \quad I_0^+ := \left[\frac{a+b}{2}, b \right].$$

Dato che x^n è composta da un numero infinito di elementi, almeno uno dei due sotto-intervalli I_0^\pm contiene infiniti elementi della successione x^n . Indichiamo con I_1 , quello dei due sotto-intervalli per cui è verificata tale proprietà e con a_1, b_1 i suoi estremi, $I = [a_1, b_1]$ e ripetiamo il ragionamento precedente. Consideriamo i sotto-intervalli I_1^\pm definiti da

$$I_1^- := \left[a_1, \frac{a_1 + b_1}{2} \right], \quad I_1^+ := \left[\frac{a_1 + b_1}{2}, b_1 \right].$$

Dato che un numero infinito di elementi $\{x^n\}$ si trovano in I_1 , almeno uno dei due sotto-intervalli I_1^\pm contiene infiniti elementi della successione x^n . Indichiamo con I_2 tale sotto-intervallo.

Iterando il procedimento, si costruisce una successione di intervalli chiusi e limitati I_n , ciascuno contenente un numero infinito di elementi della successione $\{x^n\}$. Per l'assioma degli intervalli incapsulati, esiste $\ell \in \mathbb{R}$ contenuto in I_n per ogni n . Costruiamo ora una sottosuccessione $\{x^{n_k}\}$ di $\{x^n\}$ come segue:

- scegliamo n_1 in modo che $x^{n_1} \in I_1$;
- scegliamo $n_2 > n_1$ in modo che $x^{n_2} \in I_2$;
- in generale, scegliamo $n_{k+1} > n_k$ in modo che $x^{n_{k+1}} \in I_{k+1}$.

Per costruzione, $|x^{n_k} - \ell| < (b - a)/2^k$, e, di conseguenza, la sottosuccessione x^{n_k} converge ad ℓ .

2. **Caso multi-dimensionale.** Consideriamo il caso di successioni in \mathbb{R}^2 (il caso generale è analogo). Sia $\{P^n\} = \{(x^n, y^n)\}$ una successione limitata in \mathbb{R}^2 . Ne segue che anche le due successioni di numeri reali delle coordinate $\{x^n\}$ e $\{y^n\}$ sono limitate. Applicando alla successione $\{x^n\}$ il teorema di Bolzano–Weierstrass uni-dimensionale, si deduce che esiste una sotto-successione $\{x^{n_k}\}$ convergente.

La sottosuccessione $\{y^{n_k}\}$, estratta dalla successione $\{y^n\}$ con gli stessi indici della sotto-successione $\{x^{n_k}\}$, è limitata. Applicando nuovamente il teorema di Bolzano–Weierstrass unidimensionale, si deduce l'esistenza di una sotto-sotto-successione $\{y_{n_{k_j}}\}$

convergente. La successione $P^{n_{k_j}} := (x_{n_{k_j}}, y_{n_{k_j}})$, avendo entrambe le successioni delle coordinate convergenti, è una successione convergente. \square

Il criterio di Cauchy. Tramite il Teorema di Bolzano–Weierstrass si dimostra una condizione necessaria e sufficiente per la convergenza di successioni di punti: il **criterio di Cauchy**.

TEOREMA 2.8 (Criterio di Cauchy). *Una successione P^n in \mathbb{R}^d è convergente se e solo se vale la condizione seguente*

$$(122) \quad \forall \varepsilon > 0, \quad \exists n_\varepsilon \in \mathbb{N}, \quad \forall m, n \geq n_\varepsilon, \quad \text{vale} \quad |P^m - P^n|_d < \varepsilon.$$

DIMOSTRAZIONE. Se la successione P^n converge a \bar{P} , vale la stima

$$|P^m - P^n|_d \leq |P^m - \bar{P}|_d + |P^n - \bar{P}|_d,$$

da cui segue che la condizione (122).

Viceversa, se la successione $\{P^n\}$ è di Cauchy, allora è anche limitata. Infatti, scegliendo $\varepsilon = 1$, si deduce, da (122), che, per un opportuno $n_1 \in \mathbb{N}$ vale

$$|P^n|_d \leq |P^{n_1}|_d + |P^n - P^{n_1}|_d < |P^{n_1}|_d + 1$$

per ogni $n \geq n_1$. Posto $M := \max\{|P^0|_d, |P^1|_d, \dots, |P^{n_1-1}|_d, |P^{n_1}|_d + 1\}$, si ha

$$|P^n|_d \leq M \quad \forall n \in \mathbb{N}.$$

Grazie al Teorema di Bolzano–Weierstrass, si deduce che esiste una sottosuccessione $\{P^{n_k}\}$ convergente ad un opportuno punto limite \bar{P} . Per dimostrare che tutta la successione P^n converge a \bar{P} , basta utilizzare la stima

$$|P^n - \bar{P}|_d \leq |P^n - P^{n_k}|_d + |P^{n_k} - \bar{P}|_d$$

e lavorare con accortezza con ε e n_ε . I dettagli sono lasciati al volenteroso lettore. \square

ESERCIZIO 2.9. Sia $\{P^n\}$ tale che, per qualche $\theta \in [0, 1)$ valga la condizione

$$|P^{n+1} - P^n|_d \leq \theta |P^n - P^{n-1}|_d \quad \forall n \in \mathbb{N} \setminus \{0\}.$$

Dimostrare che la successione P^n converge ad un $\bar{P} \in \mathbb{R}^d$.

La condizione (122) non necessita la conoscenza esplicita del limite \bar{P} . Tale fatto permette di utilizzare il criterio di Cauchy come mattone fondamentale per una costruzione rigorosa dell'insieme dei numeri reali a partire dai numeri razionali (o, più in generale, per una costruzione dell'insieme \mathbb{R}^d a partire da \mathbb{Q}^d). L'idea di base sta nel fatto che un numero reale è individuato da una successione (razionale) che lo approssimi con un errore piccolo quanto si vuole. Ad esempio, per approssimare/definire il numero (irrazionale) π si può considerare la successione di numeri (razionali)

$$a_0 = 3 \quad a_1 = 3,1 \quad a_2 = 3,14 \quad a_3 = 3,141 \quad a_4 = 3,1415 \quad a_5 = 3,14159 \quad \dots,$$

cioè l'elemento a_n è quello che si ottiene considerando i primi n decimali del numero π .

Da questo punto di vista, un numero è individuato da una **successione di Cauchy** in \mathbb{Q} , cioè da una successione a_n tale che

$$\forall \varepsilon \in \mathbb{Q}, \quad \varepsilon > 0, \quad \exists n_\varepsilon \in \mathbb{N}, \quad \text{tale che} \quad |a_m - a_n| < \varepsilon \quad \forall m, n \geq n_\varepsilon.$$

L'insieme delle successioni di Cauchy

$$S = \{ \{x_n\} : \text{successione di Cauchy in } \mathbb{Q} \}.$$

consta, però, di troppi elementi: esistono successioni diverse che rappresentano lo stesso numero (basta pensare a due diverse successioni di razionali che tendano a zero). Occorre quindi raggruppare gli elementi di S in classi di equivalenza, in modo che ogni singola classe rappresenti un singolo numero reale. La relazione d'equivalenza si definisce così: date $\{x_n\}$ e $\{y_n\}$ in S

$$\{x_n\} \sim \{y_n\} \quad \text{se} \quad \lim_{n \rightarrow +\infty} (x_n - y_n) = 0.$$

E' facile verificare che si tratta di una relazione di equivalenza (riflessività e simmetria sono banali, la transitività discende dalla disuguaglianza triangolare). A questo punto il gioco è fatto: l'insieme dei numeri reali è il quoziente $\mathbb{R} := S / \sim$.

3. Informazioni di base sulla topologia di \mathbb{R}^d

Partiamo da un problema prototipo: data una funzione $f : A \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$, determinare una soluzione dell'equazione

$$(123) \quad f(P) = 0.$$

Supponiamo di essere in grado di dimostrare l'esistenza di P^n tali che $f(P^n)$ non sia nullo, ma verifichi la stima

$$|f(P^n)| \leq \frac{1}{n}.$$

Il limite \bar{P} della successione P^n è un buon candidato ad essere soluzione di (123), ma...

- ...la successione P^n è convergente ad un limite P ?
- ...il limite P appartiene all'insieme A ?
- ...è vero che $f(P) = 0$?

Ciascuna di queste domande porta all'introduzione di un concetto importante: la compattezza, la chiusura, la continuità. A seguire, vediamo i primi due, mentre il terzo sarà esplorato in lungo e in largo più avanti.

DEFINIZIONE 3.1. *Un insieme $C \subset \mathbb{R}^d$ è chiuso (rispetto all'operazione di limite) se per ogni successione $P^n \in A$ convergente il punto limite \bar{P} appartiene all'insieme A .*

ESEMPIO 3.2. Il prodotto cartesiano $[a, b] \times [c, d]$ è un insieme chiuso di \mathbb{R}^2 , per ogni scelta di $a, b, c, d \in \mathbb{R}$. Infatti, sia $P^n = (x^n, y^n) \in [a, b] \times [c, d]$ una successione convergente al limite $\bar{P} = (\bar{x}, \bar{y})$. Dato che

$$a \leq x^n \leq b, \quad c \leq y^n \leq d, \quad \forall n \in \mathbb{N},$$

passando al limite per $n \rightarrow +\infty$, si deduce che $(\bar{x}, \bar{y}) \in [a, b] \times [c, d]$.

ESEMPIO 3.3. L'insieme dei numeri naturali \mathbb{N} in \mathbb{R} è un insieme chiuso. Infatti, se x_n è una successione di numeri naturali convergente ad \bar{x} , si ha che, per qualche \bar{n} , vale la stima $|x_n - \bar{x}| < 1/2$ per ogni $n \geq \bar{n}$. Se $m, n \geq \bar{n}$, si ha

$$|x_n - x_m| \leq |x_n - \bar{x}| + |x_m - \bar{x}| < \frac{1}{2} + \frac{1}{2} = 1,$$

e, di conseguenza, $x_n = x_m$. Quindi la successione è costante per $n \geq \bar{n}$ e, pertanto, convergente.

ESERCIZIO 3.4. Sia C un insieme finito di punti. Dimostrare che C è chiuso.

In generale, una successione contenuta in un insieme chiuso potrebbe non essere convergente. Occorre, quindi, una condizione che garantisca, quanto meno, la convergenza di una qualche sottosuccessione della successione data. Eccola.

DEFINIZIONE 3.5. Un insieme $K \subset \mathbb{R}^d$ è compatto (per successioni)⁴ se da ogni successione $P^n \in K$ è sempre possibile estrarre una sottosuccessione P^{n_k} convergente a qualche $\bar{P} \in K$.

ESEMPIO 3.6. Il prodotto cartesiano $[a, b] \times [c, d]$ è un insieme compatto di \mathbb{R}^2 , per ogni scelta di $a, b, c, d \in \mathbb{R}$. Infatti, sia $P^n = (x^n, y^n) \in [a, b] \times [c, d]$. Dato che la successione P^n è limitata, essa ammette, per il Teorema di Bolzano–Weierstrass una sottosuccessione P^{n_k} convergente ad un qualche punto \bar{P} e, dato che $[a, b] \times [c, d]$ è chiuso, il punto \bar{P} appartiene ad $[a, b] \times [c, d]$.

Tutti gli insiemi compatti sono anche chiusi. Infatti, se $P^n \in K$ è una successione convergente a \bar{P} , ogni sottosuccessione P^{n_k} converge allo stesso limite \bar{P} . Per definizione, quindi, \bar{P} appartiene ad K . Il viceversa non è vero: esistono insiemi chiusi che non sono compatti.

ESEMPIO 3.7. L'insieme dei numeri naturali \mathbb{N} in \mathbb{R} non è compatto. Infatti, tutte le sottosuccessioni di $a_n = n$ sono illimitate e quindi non convergenti.

La limitatezza impedisce alle successioni di "scappare" all'infinito e, insieme alla chiusura, garantisce l'esistenza di sottosuccessioni convergenti.

⁴Esistono anche altre definizioni di compattezza, utilizzate in ambiti più generali, che comunque, nel caso di \mathbb{R}^d coincidono con la compattezza per successioni.

TEOREMA 3.8 (*Insiemi compatti di \mathbb{R}^d*). *Un sottoinsieme K di \mathbb{R}^d è compatto (per successioni) se e solo se è chiuso e limitato.*

DIMOSTRAZIONE. Sia K compatto e, di conseguenza, chiuso. Supponiamo, per assurdo, che K non sia limitato. Allora, per ogni $k \in \mathbb{N}$ esiste P^{n_k} tale che $|P^{n_k}|_d \geq n$ e, di conseguenza,

$$(124) \quad \lim_{k \rightarrow +\infty} |P^{n_k}|_d = +\infty.$$

Dato che K è compatto, esiste una sottosuccessione di P^{n_k} , che indichiamo per semplicità (e con abuso di notazione...) sempre con P^{n_k} convergente a qualche $\bar{P} \in K$. In particolare, dato che

$$||P^{n_k}|_d - |\bar{P}|_d| \leq |P^{n_k} - \bar{P}|_d,$$

la successione reale $|P^{n_k}|_d$ converge a $|\bar{P}|_d$, in contraddizione con (124).

Viceversa, sia A chiuso e limitato e sia P^n una successione in K . Per il Teorema di Bolzano–Weierstrass, Teorema 2.7, è possibile estrarre una sottosuccessione P^{n_k} convergente ad un punto \bar{P} . Dato che l'insieme K è chiuso, il punto limite appartiene ad K ed è quindi verificata la proprietà di compattezza. \square

Parallelamente alla nozione di insieme chiuso, si introduce anche la nozione di insieme aperto.

DEFINIZIONE 3.9. *Sia $S \subseteq \mathbb{R}^d$. Un punto $P \in \mathbb{R}^d$ è interno ad S se esiste $r > 0$ tale che $I_r(P) = \{P \in \mathbb{R}^d : |P - P_0|_d < r\} \subseteq S$. Un punto $P \in \mathbb{R}^d$ è esterno ad S se è interno all'insieme complementare $\mathbb{R}^d \setminus S$, cioè se esiste $r > 0$ tale che $S \cap I_r(P) = \emptyset$.*

Un insieme $A \subseteq \mathbb{R}^d$ si dice aperto se i suoi punti sono tutti interni.

ESEMPIO 3.10. Dati $a, b, c \in \mathbb{R}$ l'insieme $A := \{ax + by + c > 0\} \subset \mathbb{R}^2$ è aperto. Infatti, dato $P_0 = (x_0, y_0) \in A$, si ha $I_r(P_0) \subseteq A$ per ogni scelta di r strettamente minore della distanza P_0 dalla retta di equazione $ax + by + c = 0$.

ESEMPIO 3.11. Dati $a, b, c \in \mathbb{R}$ l'insieme $C := \{(x, y) \in \mathbb{R}^2 : ax + by + c \geq 0\}$ non è aperto. Infatti, tutti i punti della forma $P_0 = (x_0, y_0)$ con $ax_0 + by_0 + c = 0$ appartengono a C , ma ogni intorno di tali punti contiene sempre punti non appartenenti a C . Analogo discorso vale per l'insieme $C_2 := \{(x, y) \in \mathbb{R}^2 : ax + by + c = 0\}$.

Che relazione intercorre tra insiemi aperti e insiemi chiusi?

PROPOSIZIONE 3.12. *Un insieme A è aperto se e solo se il suo complementare $C := \mathbb{R}^d \setminus A$ è chiuso.*

DIMOSTRAZIONE. Supponiamo A aperto e dimostriamo che $C := \mathbb{R}^d \setminus A$ è chiuso. Sia $\{P^n\}$ una successione di punti in C convergente a \bar{P} . Se, per assurdo, \bar{P} appartenesse ad A , esisterebbe $r > 0$ tale che $I_r(\bar{P}) \subseteq A$. In particolare, per ogni n si avrebbe $|P^n - \bar{P}| \geq r > 0$ in contraddizione con il fatto che la successione P^n converge a \bar{P} .

Supponiamo $C = \mathbb{R}^d \setminus A$ chiuso e dimostriamo che A è aperto. Se, per assurdo, A non fosse aperto, esisterebbe $\bar{P} \in A$ non interno ad A . Quindi, per ogni $n \in \mathbb{N}$, esisterebbe $P^n \in C$ tale che $P^n \in I_{1/n}(\bar{P})$. La successione P^n sarebbe allora convergente al punto \bar{P} , in contraddizione con l'ipotesi. \square

Gli insiemi aperti e gli insiemi chiusi sono chiusi rispetto all'operazione di unione e di intersezione: cioè valgono le affermazioni seguenti

- se A_1, A_2 sono aperti, allora anche $A_1 \cup A_2, A_1 \cap A_2$ sono aperti;
- se C_1, C_2 sono chiusi, allora anche $C_1 \cup C_2, C_1 \cap C_2$ sono chiusi.

OSSERVAZIONE 3.13. Nel caso di unione o intersezione di un numero infinito di aperti/chiusi, la situazione è un po' più delicata. Valgono le implicazioni seguenti:

- se $A_n, n \in \mathbb{N}$, è una successione di aperti, allora l'unione $\bigcup_{n \in \mathbb{N}} A_n$ è un aperto;
- se $C_n, n \in \mathbb{N}$, è una successione di chiusi, allora l'intersezione $\bigcap_{n \in \mathbb{N}} C_n$ è un chiuso.

Non è difficile costruire esempi di successioni di aperti la cui intersezione sia un chiuso e successioni di chiusi la cui unione sia un aperto. Provateci.

Per ogni insieme $S \subset \mathbb{R}^d$ è possibile distinguere tre categorie di punti: i punti interni, i punti esterni e... gli altri, cioè i punti che non sono né interni né esterni ad S .

DEFINIZIONE 3.14. Un punto $P \in \mathbb{R}^d$ è un **punto di frontiera** dell'insieme S se ogni intorno di P contiene punti di S e punti non appartenenti ad S : per ogni $r > 0$,

$$S \cap I_r(P_0) \neq \emptyset, \quad (\mathbb{R}^d \setminus S) \cap I_r(P_0) \neq \emptyset.$$

L'insieme dei punti di frontiera di S si chiama **frontiera** di S e si denota con ∂S .

Si chiama **chiusura** dell'insieme S l'insieme $\bar{S} := S \cup \partial S$.

La chiusura \bar{S} di un insieme S è il più piccolo insieme chiuso contenente S .

ESEMPIO 3.15. Per gli insiemi

$$A := \{(x, y) : ax + by + c < 0\}, \quad C := \{(x, y) : ax + by + c \leq 0\}.$$

con $a, b, c \in \mathbb{R}$, si ha $\partial A = \partial C = \{(x, y) : ax + by + c = 0\}$. A voi la verifica.

ESEMPIO 3.16. Per gli insiemi

$$A := \{(x, y) : x^2 + y^2 < r^2\}, \quad C := \{(x, y) : x^2 + y^2 \leq r^2\},$$

con $r > 0$, si ha $\partial A = \partial C = \{(x, y) : x^2 + y^2 = r^2\}$.

ESERCIZIO 3.17. Determinare la frontiera e la chiusura dei seguenti insiemi

$$\begin{aligned} S_1 &= \{(x, y) : 0 < x^2 + y^2 \leq 1\}, & S_2 &= \{(x, y) : x + y + 1 < 0, x + 1 < 0\}, \\ S_3 &= \{(0, 0), (1, 0), (0, 1)\}, & S_4 &= \{(x, y) : x, y \geq 0, x + y \leq 0\} \end{aligned}$$

ESERCIZIO 3.18. Dimostrare che la frontiera di un insieme S e del suo complementare $T := \mathbb{R}^d \setminus S$ coincidono, cioè $\partial S = \partial T$.

Come espresso dal seguente risultato (la cui dimostrazione è lasciata come esercizio), gli insiemi aperti e gli insiemi chiusi hanno una relazione particolare con la loro frontiera.

PROPOSIZIONE 3.19. *Un insieme A è aperto se e solo se $A \cap \partial A = \emptyset$; un insieme C è chiuso se e solo se $\partial C \subseteq C$.*

4. Natura non facit saltus: le curve

Una successione di punti $P^n \in \mathbb{R}^d$ può essere interpretata come una sequenza di posizioni nello spazio d -dimensionale di una particella in movimento che viene osservata ad istanti successivi $t_0 < t_1 < \dots < t_n < \dots$: il punto P^n è la posizione all'istante t_n . Una maniera altrettanto (o più?) interessante per descrivere il moto di un punto è quella di considerare che la variabile temporale non vari in un insieme discreto, ma in un insieme continuo⁵. Questo corrisponde, in concreto a considerare funzioni del tipo

$$\phi : I \subseteq \mathbb{R} \longrightarrow \mathbb{R}^d$$

dove I è un intervallo di \mathbb{R} , che associano ad uno scalare $t \in I$ il vettore $\phi(t) = (\phi_1(t), \dots, \phi_d(t))$. La funzione ϕ_i che associa a t la i -esima componente di $\phi(t)$ si chiama i -esima componente di ϕ .

Tenendo conto della definizione di modulo in \mathbb{R}^d , è possibile, senza fatica, introdurre il concetto di continuità anche in questo contesto.

DEFINIZIONE 4.1. *Sia I un intervallo di \mathbb{R} . Una funzione $\phi : I \subseteq \mathbb{R} \longrightarrow \mathbb{R}^d$ è continua in $t_0 \in I$ se*

$$(125) \quad \forall \varepsilon > 0, \quad \exists \delta > 0 \quad \text{tale che} \quad |\phi(t) - \phi(t_0)|_d < \varepsilon \quad \forall t \in I, |t - t_0| < \delta.$$

Una funzione $\phi : I \subseteq \mathbb{R} \longrightarrow \mathbb{R}^d$ è continua in I se è continua in tutti i punti di I . L'insieme delle curve $\phi : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^d$ continue in I si indica con il simbolo $C(I; \mathbb{R}^d)$.

⁵La locuzione latina *Natura non facit saltus*, formulata per la prima volta da Carl von Linné, nella *Philosophia Botanica*, è stata utilizzata da Leibniz come assioma per negare l'esistenza di quantità discrete indivisibili ed esprimere che i processi naturali variano in maniera continua.

Dalla definizione di modulo e seguendo una strategia analoga a quella utilizzata nella Proposizione 2.3, si dimostra che la continuità della funzione ϕ equivale alla continuità di tutte le sue componenti ϕ_i .

Una sottoclasse delle funzioni continue è data dalle cosiddette *funzioni lipschitziane*: una funzione $\phi \in C(I; \mathbb{R}^d)$ si dice **lipschitziana** se esiste $L > 0$ tale che

$$(126) \quad |\phi(t) - \phi(\tau)|_d \leq L|t - \tau| \quad \forall t, \tau \in I.$$

In questo caso, la condizione di continuità (125) è verificata scegliendo $\delta = \varepsilon/L$. Come si vede, il valore di δ è indipendente dal punto t_0 di continuità. Sempre ragionando come nella Proposizione 2.3, si può mostrare che una funzione ϕ è lipschitziana se e solo se tutte le sue componenti ϕ_i sono lipschitziane.

Nella logica della descrizione di un moto nello spazio, è ragionevole restringere l'attenzione a funzioni ϕ le cui componenti ϕ_1, \dots, ϕ_d sono derivabili. In questo modo, per ogni valore $t \in I$ è ben definito il **vettore velocità** $\phi'(t) = (\phi'_1(t), \dots, \phi'_d(t))$. L'insieme delle curve $\phi : I \subset \mathbb{R} \rightarrow \mathbb{R}^d$ con derivata ϕ' continua, si indica con il simbolo $C^1(I; \mathbb{R}^d)$.

DEFINIZIONE 4.2. *Sia $I \subseteq \mathbb{R}$ un intervallo di estremi a e b . Una curva parametrizzata regolare è un'applicazione $\phi \in C^1(I; \mathbb{R}^d)$ tale che*

$$(127) \quad \phi'(t) \neq 0 \quad \forall t \in I.$$

Una curva parametrizzata regolare a tratti è un'applicazione $\phi \in C(I; \mathbb{R}^d)$ tale che esiste un insieme $S = \{a < t_0 < t_1 < \dots < t_N < b\} \subset I$ per cui vale

$$\phi \text{ è regolare a tratti in } [a, t_0] \cap I, [t_1, t_2], \dots, [t_{N-1}, t_N], [t_N, b] \cap I$$

*In entrambi i casi, l'insieme immagine $\phi(I)$ è detto **supporto della curva**.*

Da un punto di vista cinematico, la funzione ϕ che definisce la curva è la **legge oraria** del moto in questione, mentre l'insieme immagine $\phi(I)$ ne è la **traiettoria**. Si noti che la stessa traiettoria può essere percorsa con leggi orarie diverse e che, parallelamente, curve parametrizzate diverse possono avere lo stesso supporto.

La condizione $\phi'(t_0) \neq 0$ equivale a richiedere che il moto del punto descritto dalla legge oraria ϕ non abbia mai velocità nulla; dal punto di vista geometrico, essa garantisce l'esistenza in ogni punto della curva di una retta tangente, la cui equazione parametrica è $P = \phi(t_0) + \phi'(t_0)h$ con $h \in \mathbb{R}$.

Segmenti e poligonali. Consideriamo due punti distinti $P_0, P_1 \in \mathbb{R}^d$. Come descrivere un moto da P_0 a P_1 lungo una linea retta (cioè per la via più rapida)? In

altre parole, come parametrizzare il segmento $\overline{P_0P_1}$? E' immediato riconoscere che la definizione seguente

$$\phi(t) = P_0 + t(P_1 - P_0) \quad t \in [0, 1],$$

risolve la questione. Come si nota facilmente, si ha $\phi'(t) = P_1 - P_0 \neq 0$ per ogni t e, di conseguenza, la curva ϕ è effettivamente regolare.

Ovviamente, esistono (infinite) altre maniere di parametrizzare un segmento.

ESEMPIO 4.3. Il segmento nel piano di estremi $(0, 1)$ e $(1, 0)$ è parametrizzato da

$$\begin{cases} x = x(t) = t, \\ y = y(t) = 1 - t \end{cases} \quad t \in [0, 1].$$

Lo stesso segmento può essere parametrizzato in tante maniere diverse, ad esempio da

$$\begin{cases} x = x(\tau) = 1 - \cos(\tau), \\ y = y(\tau) = \cos(\tau) \end{cases} \quad \tau \in [0, \pi/2].$$

Le due rappresentazioni parametriche offerte differiscono dal punto di vista cinematico. Calcolando i corrispondenti vettori velocità si ha

$$\begin{aligned} (x', y')(t) = (1, -1) &\Rightarrow \sqrt{x'(t)^2 + y'(t)^2} = \sqrt{2}, \\ (x', y')(\tau) = \sin(\tau)(1, -1) &\Rightarrow \sqrt{x'(\tau)^2 + y'(\tau)^2} = \sqrt{2} \sin(\tau). \end{aligned}$$

Quindi, nel primo caso, il moto è rettilineo uniforme (i.e. la velocità è costante), nel secondo, il moto è ancora rettilineo, ma la velocità cresce al crescere di $\tau \in [0, \pi/2]$.

Consideriamo tre punti distinti $P_0, P_1, P_2 \in \mathbb{R}^d$. Come parametrizzare il percorso da P_0 a P_2 passando per P_1 , ottenuto percorrendo i segmenti $\overline{P_0P_1}$ e $\overline{P_1P_2}$? Basta operare piccole modifiche alla parametrizzazione vista per il caso di un singolo segmento:

$$\phi(t) = \begin{cases} P_0 + t(P_1 - P_0) & t \in [0, 1], \\ P_1 + (t - 1)(P_2 - P_1) & t \in (1, 2] \end{cases}$$

In questo caso,

$$\phi'(t) = \begin{cases} P_1 - P_0 & t \in [0, 1), \\ P_2 - P_1 & t \in (1, 2] \end{cases}$$

quindi, in generale, la parametrizzazione è solo regolare a tratti. In questo caso l'insieme S della Definizione 4.2 è composto da un singolo elemento: $S = \{1\}$.

Passiamo al caso generale di $n + 1$ punti.

DEFINIZIONE 4.4. *Assegnati $n + 1$ punti $P_0, P_1, P_2, \dots, P_n$ si dice poligonale S da essi determinata l'insieme unione degli n segmenti $\overline{P_0P_1}, \dots, \overline{P_{n-1}P_n}$. I punti $P_0, P_1, P_2, \dots, P_n$ sono detti vertici della poligonale e i due punti P_0 e P_n si dicono estremi della poligonale S .*

Parametrizzare una poligonale Π non è così proibitivo: basta porre

$$\phi(t) = P_k + (t - k)(P_{k+1} - P_k) \quad \forall t \in [k, k + 1], \quad k \in [0, n - 1].$$

Anche, in questo caso, in generale, la parametrizzazione è solo regolare a tratti con $S = \{1, 2, \dots, n\}$.

Altri esempi di curve. Il caso più noto è quello dei grafici di funzioni. Data $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ di classe C^1 , il grafico della funzione f si parametrizza banalmente ponendo

$$\phi(t) = (t, f(t)) \quad t \in I.$$

Dato che $\phi'(t) = (1, f'(t))$, la curva è regolare.

Non serve un grande sforzo per trovare qualche esempio di curva che non sia un grafico di funzione. Dati $P_0 = (x_0, y_0) \in \mathbb{R}^2$ ed $r > 0$, la circonferenza di centro P_0 e raggio r si rappresenta con

$$\begin{cases} x = x(t) = x_0 + r \cos(t), \\ y = y(t) = x_0 + r \sin(t) \end{cases} \quad t \in [0, 2\pi).$$

In maniera simile, dati $P_0 = (x_0, y_0) \in \mathbb{R}^2$ ed $a, b > 0$, l'ellissi di semiassi a e b , centrata in P_0 e simmetrica rispetto agli assi coordinati, si rappresenta parametricamente con

$$\begin{cases} x = x(t) = x_0 + a \cos(t), \\ y = y(t) = x_0 + b \sin(t) \end{cases} \quad t \in [0, 2\pi).$$

Una curva più originale è quella data da $\phi(t) = (t^2, t^3)$ con $t \in [-1, 1]$. Il grafico presenta una *cuspid*e nel punto $(0, 0)$. Si noti che, dato che $\phi'(t) = (2t, 3t^2)$, la condizione (127) non è soddisfatta in $t = 0$. La curva in esame, quindi, non è regolare. Si tratta, comunque, di una curva regolare a tratti, con $S = \{0\}$.

Tanto per gradire, vediamo anche un esempio di curva nello spazio tridimensionale. Dati $a > 0$ e $r > 0$, la curva parametrizzata data da

$$\begin{cases} x = x(t) = r \cos(t), \\ y = y(t) = r \sin(t), \\ z = z(t) = at \end{cases} \quad t \in \mathbb{R}.$$

ha come supporto un'elica circolare che si avvolge attorno all'asse z . A seconda della scelta di a , l'elica è più o meno ripida. Dato che $\phi'(t) = (-r \sin t, r \cos t, a) \neq 0$ per ogni t , la curva è regolare.

Lunghezza di curve. Il concetto di lunghezza è intuitivo: immaginando una curva come un oggetto deformabile, basta “raddrizzare” tale oggetto e confrontarlo con l’unità di misura prescelta. Ma, come nel caso degli integrali definiti, occorre dare una definizione rigorosa di lunghezza di una curva che, auspicabilmente, ne permetta anche il calcolo esplicito, almeno in un certo numero di situazioni semplici. Come procedere?

Nella costruzione dell’integrale, abbiamo ricondotto il calcolo di un’area con un bordo curvilineo a quello di insiemi dati da unioni di rettangoli, passando poi al limite in maniera opportuna. Nel caso di curve, l’idea naturale è di utilizzare come approssimanti le poligonalari. In effetti, scegliendo un numero finito di punti della curva e considerando la poligonale determinata da tali punti, si ottiene una nuova curva la cui lunghezza è minore o uguale della lunghezza della curva originale in esame. La lunghezza della curva è data dall’estremo superiore delle lunghezze di tali poligonali approssimanti. Vediamo di formalizzare precisamente quanto descritto.

Consideriamo, una curva parametrizzata regolare ϕ definita in $I = [a, b] \subset \mathbb{R}$. Data la partizione $\pi = \{a = t_0 < t_1 < \dots < t_{n-1} < t_n = b\}$, in corrispondenza dei quali saranno assegnati sulla curva i punti $P_i = \phi(t_i)$. I punti P_0, P_1, \dots, P_n definiscono la poligonale $S(\pi)$ la cui lunghezza $\ell(\phi, \pi)$ è data da

$$\ell(\phi, \pi) = \sum_{k=0}^{n-1} |P_{k+1} - P_k|_d = \sum_{k=0}^{n-1} |\phi(t_{k+1}) - \phi(t_k)|_d.$$

DEFINIZIONE 4.5. La lunghezza ℓ di una curva parametrizzata regolare a tratti $\phi : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^d$ è data da

$$\ell(\phi) := \sup\{\ell(\phi, \pi) : \pi \text{ partizione di } [a, b]\}.$$

Le curve regolari hanno tutte lunghezza finita. Infatti, dato che la funzione ϕ appartiene a $C^1([a, b]; \mathbb{R}^d)$, le componenti ϕ_i sono di classe C^1 ed hanno quindi derivata limitata. Sia $M > 0$ tale che $|\phi'_i(t)| \leq M$ per ogni $t \in [a, b]$ e per ogni $i \in \{1, \dots, d\}$. Dati $t, \tau \in [a, b]$, applicando il Teorema di Lagrange, si deduce

$$|\phi(t) - \phi(\tau)|_d = \left(\sum_{k=1}^d (\phi_k(t) - \phi_k(\tau))^2 \right)^{1/2} \leq M\sqrt{d}|t - \tau|.$$

Data la partizione $\pi = \{a = t_0 < t_1 < \dots < t_{n-1} < t_n = b\}$, si ha

$$\ell(\phi, \pi) = \sum_{k=0}^{n-1} |\phi(t_{k+1}) - \phi(t_k)|_d \leq M\sqrt{d} \sum_{k=0}^{n-1} (t_{k+1} - t_k) = M\sqrt{d}(b - a).$$

L’insieme $\{\ell(\pi) : \pi \text{ partizione di } [a, b]\}$ è quindi superiormente limitato e, di conseguenza, ammette estremo superiore finito. Una curva regolare a tratti può essere

pensata come unione finita di curve regolari e, pertanto, ammette anch'essa lunghezza finita.

La definizione di lunghezza come estremo superiore è pienamente soddisfacente dal punto di vista teorico, ma lascia qualche perplessità dal punto di vista operativo: come calcolare in concreto la lunghezza di una curva?

TEOREMA 4.6. *Sia $\phi : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^d$ una curva parametrizzata regolare a tratti. Vale l'uguaglianza*

$$(128) \quad \ell(\phi) = \int_a^b |\phi'(t)|_a dt.$$

OSSERVAZIONE 4.7. Dato che la funzione ϕ è regolare a tratti e si ha

$$||\phi'(t)|_a - |\phi'(t_0)|_a| \leq |\phi'(t) - \phi'(t_0)|_a,$$

è possibile decomporre l'intervallo di integrazione $[a, b]$ in unione finita di sotto-intervalli chiusi in cui la funzione $|\phi'(\cdot)|_a$ è continua. Quindi, la funzione $|\phi'(\cdot)|_a$ è integrabile (secondo Riemann) in $[a, b]$ e l'integrale in (128) è ben definito.

La dimostrazione della relazione (128) si basa sull'approssimazione

$$\phi(\tau) - \phi(t) \approx \phi'(t) (\tau - t) \quad \tau \rightarrow t,$$

che consiste nel sostituire il segmento di estremi $\phi(t)$ e $\phi(t+h)$ con un segmento tangente al supporto della curva in $\phi(t)$ e di lunghezza opportuna. Il seguente Lemma si propone di stimare con precisione l'errore commesso in tale approssimazione, in un caso leggermente più semplice.

LEMMA 4.8. *Sia $I \subseteq \mathbb{R}$ un intervallo e sia $\phi \in C^1(I; \mathbb{R}^d)$ tale che ϕ' è lipschitziana in I . Allora, esiste $C > 0$ tale che*

$$(129) \quad |\phi(\tau) - \phi(t) - \phi'(t) (\tau - t)|_a \leq C |\tau - t|^2$$

per ogni $t, \tau \in I$.

DIMOSTRAZIONE. Per ogni $i \in \{1, \dots, d\}$, dal Teorema di Lagrange segue

$$|\phi_i(\tau) - \phi_i(t) - \phi'_i(t) (\tau - t)| = |(\phi_i(\sigma_i) - \phi'_i(t)) (\tau - t)|$$

per qualche σ_i compreso tra τ e t . Utilizzando l'ipotesi di lipschitzianità di ϕ' , si deduce che, per qualche $L > 0$, vale

$$(130) \quad |\phi_i(\tau) - \phi_i(t) - \phi'_i(t) (\tau - t)| \leq L |\sigma_i - t| |\tau - t| \leq L |\tau - t|^2.$$

Utilizzando (130), si ottiene

$$|\phi(\tau) - \phi(t) - \phi'(t) (\tau - t)|_a \leq L \sqrt{d} |\tau - t|^2.$$

La dimostrazione è completa. □

Grazie al Lemma 4.8, è possibile dimostrare il Teorema 4.6 supponendo, in aggiunta, ϕ' lipschitziana.

DIMOSTRAZIONE (SEMPLIFICATA) DEL TEOREMA 4.6. Data una partizione π dell'intervallo I , si ha, utilizzando (117) e (129),

$$(131) \quad \left| \ell(\phi, \pi) - \sum_{k=0}^{n-1} |\phi'(t_k)|_a (t_{k+1} - t_k) \right| \leq \sum_{k=0}^{n-1} \left| |\phi(t_{k+1}) - \phi(t_k)|_a - |\phi'(t_k)|_a (t_{k+1} - t_k) \right|$$

$$\leq \sum_{k=0}^{n-1} \left| \phi(t_{k+1}) - \phi(t_k) - \phi'(t_k)(t_{k+1} - t_k) \right|_a$$

$$\leq C \sum_{k=0}^{n-1} (t_{k+1} - t_k)^2 \leq C (b - a) |\pi|,$$

dove $|\pi| := \max_{k \in \{0, n-1\}} (t_{k+1} - t_k)$. Quando l'ampiezza $|\pi|$ della partizione tende a zero, la

lunghezza $\ell(\phi, \pi)$ converge alla lunghezza $\ell(\pi)$, mentre la sommatoria $\sum_{k=0}^{n-1} |\phi'(t_k)|_a (t_{k+1} - t_k)$ tende all'integrale della funzione $|\phi'(t)|_a$ rispetto a t nell'intervallo $[a, b]$. Grazie alla stima (131), ne segue la formula (128). \square

ESEMPIO 4.9. Come prima verifica/applicazione della formula (128), consideriamo la circonferenza parametrizzata da $\phi(t) = (\cos t, \sin t)$ con $t \in [0, 2\pi]$. In questo caso, si ha $\phi'(t) = (-\sin t, \cos t)$ e, di conseguenza, $|\phi'(t)|_a = 1$ per ogni t . Perciò, la lunghezza richiesta è

$$\ell = \int_0^{2\pi} 1 \, dt = 2\pi,$$

coerentemente con quanto noto dalla geometria elementare.

ESEMPIO 4.10. Calcoliamo la lunghezza della curva $\phi(t) = (t^2, t^3)$ con $t \in [-1, 1]$. Dato che $\phi'(t) = (2t, 3t^2)$, la lunghezza della curva è

$$L = \int_{-1}^1 \sqrt{4t^2 + 9t^4} \, dt = \frac{1}{9} \int_4^{13} \tau^{1/2} \, d\tau = \frac{2}{27} (13\sqrt{13} - 8).$$

avendo posto $\tau = 4 + 9t^2$.

ESEMPIO 4.11. Dati $a > 0$ e $r > 0$, consideriamo l'arco di elica descritto dalla parametrizzazione $\phi(t) = (r \cos t, r \sin t, at)$ con $t \in [t_0, t_1]$, $t_0 < t_1$. Dato che $\phi'(t) = (-r \sin t, r \cos t, a) \neq 0$ per ogni t , la lunghezza della curva è data da

$$\ell = \int_{t_0}^{t_1} |\phi'(t)|_a \, dt = \int_{t_0}^{t_1} \sqrt{r^2 + a^2} \, dt = \sqrt{r^2 + a^2} (t_1 - t_0).$$

dato che $|\phi'(t)|_a = \sqrt{r^2 + a^2}$.

Nel caso di parametrizzazione della forma $\phi(t) = (t, f(t))$ con $t \in [a, b]$, dato che $\phi(t) = (1, f'(t))$, l'espressione per la lunghezza della curva prende la forma

$$\ell = \int_a^b \sqrt{1 + (f'(t))^2} dt,$$

che, quindi, esprime la formula per la lunghezza di un grafico di funzione.

ESEMPIO 4.12 (Lunghezza di un arco di parabola). Sia $f(t) = t^2$ con $t \in [0, a]$ con $a > 0$. La lunghezza del grafico della funzione f è pari a

$$\begin{aligned} \ell &= \int_0^a \sqrt{1 + (f'(t))^2} dt = \int_0^a \sqrt{1 + 4t^2} dt = \frac{1}{2} \int_0^{\text{settsinh}(2a)} \cosh^2 \tau dt \\ &= \frac{1}{2} \left(\text{settsinh}(2a) + 2a \sqrt{1 + 4a^2} \right), \end{aligned}$$

avendo posto $\tau = \text{settsinh}(2t)$.

ESEMPIO 4.13 (Lunghezza di un arco di curva esponenziale). Sia $f(t) = e^t$ con $t \in [0, a]$ con $a > 0$. In questo caso, la lunghezza del grafico della funzione f è data da

$$\begin{aligned} \ell &= \int_0^a \sqrt{1 + (f'(t))^2} dt = \int_0^a \sqrt{1 + e^{2t}} dt = \int_{\sqrt{2}}^{\sqrt{1+e^{2a}}} \left(1 + \frac{1}{\tau^2 - 1} \right) dt \\ &= \tau + \frac{1}{2} \ln \left(\frac{\tau - 1}{\tau + 1} \right) \Big|_{\sqrt{2}}^{\sqrt{1+e^{2a}}} = \sqrt{1 + e^{2a}} + a - \sqrt{2} + \ln \left(\frac{\sqrt{2} + 1}{\sqrt{1 + e^{2a}} + 1} \right) \end{aligned}$$

avendo posto $\tau = \sqrt{1 + e^{2t}}$ ed avendo applicato qualche trattamento cosmetico...

Cambiamenti di parametro. La lunghezza della curva è un concetto *intrinseco*, cioè non dipende dalla particolare scelta della parametrizzazione: uno stesso tragitto, anche se percorso con velocità diverse, indica sul contachilometri la stessa lunghezza di strada. Formalizziamo precisamente la questione.

DEFINIZIONE 4.14. Due curve regolari $\phi \in C^1([a, b]; \mathbb{R}^d)$ e $\psi \in C^1([\alpha, \beta]; \mathbb{R}^d)$ si dicono **equivalenti** se esiste una funzione reale di variabile reale $g \in C^1([\alpha, \beta]; [a, b])$ biunivoca tale che $g'(\tau) \neq 0$ per ogni τ e

$$(132) \quad \psi(\tau) = \phi(g(\tau)) \quad \forall \tau \in [\alpha, \beta].$$

In concreto, la funzione g consiste in una *riparametrizzazione* della curva: lo stesso supporto viene descritto sia dal parametro t che dal parametro τ e la funzione g definisce la corrispondenza tra i due parametri. In termini cinematici, le due funzioni ϕ e ψ corrispondono a due leggi orarie distinte e la funzione g mette in corrispondenza istanti relativi alla stessa posizione.

Derivando rispetto a τ la relazione (132), si deduce

$$\frac{d\psi}{d\tau} = \frac{d\phi}{dt} \frac{dg}{d\tau},$$

che descrive la relazione che intercorre tra i vettori velocità delle due curve. Utilizzando la formula 128 per la curva ψ ed introducendo la variabile $t = g(\tau)$, si ottiene

$$\ell(\psi) = \int_{\alpha}^{\beta} \left| \frac{d\psi}{d\tau} \right|_a d\tau = \int_{g(\alpha)}^{g(\beta)} \left| \frac{d\phi}{dt} \right|_a \frac{|g'|}{g'} dt$$

Dato che $g' \neq 0$ per ogni τ , la funzione g' o è sempre strettamente positiva o è sempre strettamente negativa. Nel primo caso, si ha $g(\alpha) = a$ e $g(\beta) = b$, quindi

$$\ell(\psi) = \int_a^b \left| \frac{d\phi}{dt} \right|_a dt = \ell(\phi).$$

Se $g' < 0$, si ha $g(\alpha) = b$ e $g(\beta) = a$, quindi

$$\ell(\psi) = \int_b^a \left| \frac{d\phi}{dt} \right|_a \frac{-g'}{g'} dt = - \int_b^a \left| \frac{d\phi}{dt} \right|_a dt = \int_a^b \left| \frac{d\phi}{dt} \right|_a dt = \ell(\phi).$$

Quindi, le formule precedenti mostrano che *curve equivalenti hanno la stessa lunghezza*. Oltre a dare ulteriore conforto e coerenza alla definizione di lunghezza espressa in (128), la precedente affermazione permette di scegliere la parametrizzazione che si preferisce nel calcolo della lunghezza di una curva.

Gli insiemi connessi. Una proprietà significativa degli insiemi di \mathbb{R}^d è quella di essere *connessi*, cioè composti da un singolo pezzo. Sinonimi del vocabolo “connesso” sono i termini “congiunto”, “allacciato”, perciò una maniera per formalizzare tale concetto è di richiedere che ogni coppia di punti dell’insieme possa essere collegata da un percorso interamente contenuto nell’insieme stesso. A seconda del tipo di percorsi ammissibili per realizzare tale collegamento tra punti, si utilizzano terminologie diverse.

La versione più semplice di tutte è quella che considera come unici percorsi ammissibili i segmenti.

DEFINIZIONE 4.15. *Un insieme $S \subset \mathbb{R}^d$ è convesso se, per ogni $P, Q \in S$, si ha*

$$tP + (1-t)Q \in S \quad \forall t \in [0, 1],$$

cioè se il segmento di estremi P e Q è interamente contenuto in S .

ESEMPIO 4.16. Ecco un piccolo elenco di insiemi convessi “facili”:

- gli intervalli in \mathbb{R} ;
- i sopragrafici di funzioni convesse, cioè gli insiemi della forma $S = \{(x, y) : y \geq f(x)\} \subset \mathbb{R}^2$, per $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ convessa;
- i rettangoli nel piano, cioè insiemi della forma $[a, b] \times [c, d] \subset \mathbb{R}^2$;

- i parallelepipedi nello spazio, cioè insiemi della forma $[a, b] \times [c, d] \times [e, f] \subset \mathbb{R}^3$;
- gli intorni di un punto, cioè insiemi della forma $\{P \in \mathbb{R}^d : |P - P_0|_d < r\}$ con $P_0 \in \mathbb{R}^d$ e $r > 0$.

OSSERVAZIONE 4.17. Se $S_1 \subset \mathbb{R}^{d_1}$ e $S_2 \subset \mathbb{R}^{d_2}$ sono due insiemi convessi, anche l'insieme $S := S_1 \times S_2 \subset \mathbb{R}^{d_1+d_2}$ è convesso. Infatti, siano $P = (x_P, y_P), Q = (x_Q, y_Q) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ due punti dell'insieme $S := S_1 \times S_2$. Dato che i punti x_P, x_Q appartengono ad S_1 , il punto $tx_P + (1-t)x_Q$ appartiene ad S_1 per ogni $t \in [0, 1]$. Analogamente, dato che i punti y_P, y_Q appartengono ad S_2 , anche il punto $ty_P + (1-t)y_Q$ è in S_2 . Di conseguenza, il punto $tP + (1-t)Q = (tx_P + (1-t)x_Q, ty_P + (1-t)y_Q)$ appartiene a $S_1 \times S_2$ per ogni $t \in [0, 1]$.

ESERCIZIO 4.18. Siano A e B due insiemi convessi di \mathbb{R}^n . L'insieme $A \cup B$ è convesso? L'insieme $A \cap B$ è convesso?

La convessità è una richiesta molto forte. Tanto forte che basta apportare piccole modifiche ad un insieme convesso per ottenere un insieme non convesso. Ad esempio, se $S \subseteq \mathbb{R}^d$ è un insieme convesso e P_0 è interno ad S , l'insieme $S \setminus \{P_0\}$ non è convesso. Infatti, se $r > 0$ è tale che $I_r(P_0) \subset S$ e $v \in \mathbb{R}^d$ è un qualsiasi vettore tale che $|v|_d < r$, i due punti $P_{\pm} := P_0 \pm v$ appartengono ad $S \setminus \{P_0\}$, ma, dato che $tP_- + (1-t)P_+ = P_0 + (1-2t)v$ coincide con P_0 per $t = 1/2$, il segmento che li congiunge non è interamente contenuto nello stesso insieme. In generale, tutto le volte che si crea un buco in un insieme convesso eliminando un sottoinsieme interno si ottiene un insieme non convesso.

Una condizione più flessibile della convessità è espressa dalla prossima definizione.

DEFINIZIONE 4.19. Un insieme S si dice **connesso per poligonali** se per ogni coppia $P, Q \in S$ esiste una poligonale di estremi P e Q interamente contenuta in S .

Tutti gli insiemi convessi sono anche connessi per poligonali. In questo caso, infatti, basta scegliere come poligonale il segmento che congiunge i due punti P e Q dell'insieme. Esistono anche tantissimi insiemi che sono connessi per poligonali, ma che non sono convessi. Ad esempio, le corone circolari nel piano, i ferri di cavallo, e tutti gli insiemi ottenibili a partire da un insieme connesso con modificazioni elasto-plastiche, ma senza fratture...

ESEMPIO 4.20. Sia I un intervallo di \mathbb{R} e sia $f : I \rightarrow \mathbb{R}$ una funzione continua. Il sopragrafico $S := \{(x, y) \in \mathbb{R}^2 : y \geq f(x)\}$ è un insieme connesso per poligonali. Infatti, dati $(x_1, y_1), (x_2, y_2) \in S$ con $x_1 \leq x_2$, sia $M := \max_{x \in [x_1, x_2]} f(x)$, che esiste finito per il Teorema di Weierstrass. Allora, la poligonale di vertici $P_0 = (x_1, y_1), P_1 = (x_1, M), P_2 = (x_2, M)$ e $P_3 = (x_2, y_2)$ è interamente contenuta in S .

Allo stesso modo, si mostra che il sopragrafico di una funzione superiormente limitata è connesso per poligonali.

ESEMPIO 4.21. Un insieme convesso a cui siano stati sottratti un punto interno non è più convesso, ma è connesso per poligonalità. Infatti, sia S l'insieme convesso di partenza, sia P_0 interno ad S . Se P, Q sono elementi di S , ci sono due eventualità: o il segmento di estremi P e Q non contiene il punto P_0 , o P_0 appartiene a tale segmento. Nel primo caso non c'è nulla da aggiungere. Nel secondo, occorre costruire una poligonale che connetta P e Q , aggirando il punto P_0 . Volete provvedere voi?

In generale, un insieme convesso a cui siano stati sottratti un numero finito di punti interni è un insieme connesso per poligonalità.

ESERCIZIO 4.22. Trovare un insieme $S \subseteq \mathbb{R}^d$ convesso tale che esista $P_0 \in S$ (non interno ad S) per cui $S \setminus \{P_0\}$ non sia connesso per poligonalità.

ESERCIZIO 4.23. Siano $A, B \subseteq \mathbb{R}^d$ due insiemi connessi per poligonalità. L'insieme $A \cup B$ è convesso? L'insieme $A \cap B$ è convesso?

Esistono numerosi insiemi fatti di un solo pezzo che non sono connessi per poligonalità: basta pensare ad un arco S di circonferenza. E' evidente che comunque si prendano due punti $P, Q \in S$ non esiste alcuna poligonale di estremi P e Q contenuta in S , per il semplice motivo che non esistono poligonalità contenute in S . Per inserire anche questo tipo di situazione, si introduce un'ulteriore nuova definizione di connesso.

DEFINIZIONE 4.24. Un insieme S si dice **connesso per archi** se per ogni coppia $P, Q \in S$ esiste una curva $\phi : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^d$ regolare a tratti tale che

$$\phi(a) = P, \quad \phi(b) = Q, \quad \phi(t) \in S \quad \forall t \in [a, b].$$

Dato che le poligonalità sono particolari scelte di curve regolari a tratti, ogni insieme connesso per poligonalità è anche connesso per archi. Tutti gli archi di curva (che non siano poligonalità) sono insiemi connessi per archi (ovviamente), ma non connessi per poligonalità.

Funzioni di più variabili

1. Questioni elementari

Le funzioni sono algoritmi che fanno corrispondere ad ogni punto di un insieme (l'insieme di definizione) un altro punto di un altro insieme (il codominio). In quel che segue, ci interesseremo al caso in cui l'insieme di definizione I sia un sottoinsieme del dominio \mathbb{R}^d e il codominio sia l'insieme \mathbb{R}^p dove d, p sono numeri interi positivi opportuni. In notazione, consideriamo funzioni della forma seguente

$$f : I \subseteq \mathbb{R}^d \rightarrow \mathbb{R}^p.$$

Si tratta di una regola che associa alla variabile indipendente (o *input*) $P \in I$ la variabile dipendente (o *output*) $Q \in \mathbb{R}^p$. Passiamo in rassegna i tipi fondamentali di funzioni di questo genere.

Funzioni reali di una variabile reale: $d = p = 1$. Sono funzioni $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$, già stato considerate in lungo e in largo nei Capitoli precedenti.

Le curve: $d = 1, p > 1$. Si tratta di funzioni del tipo $\phi : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^p$, di cui si è discusso nel Capitolo precedente.

Funzioni reali di più variabili reali: $d > 1, p = 1$. Si tratta di funzioni

$$f : I \subseteq \mathbb{R}^d \rightarrow \mathbb{R}.$$

Nel caso $d = 2$, si tratta di funzioni che fanno corrispondere ai punti $(x, y) \in \mathbb{R}^2$ un numero reale $z = f(x, y) \in \mathbb{R}$; nel caso $d = 3$, sono funzioni che associano ai punti $(x, y, z) \in \mathbb{R}^3$ un numero reale $w \in \mathbb{R}$.

In generale, una funzione f da \mathbb{R}^d in \mathbb{R}^p equivale all'assegnazione di p funzioni, indicate con f_1, \dots, f_p , da \mathbb{R}^d in \mathbb{R} che associano al punto $P \in \mathbb{R}^d$ la i -esima coordinata del punto immagine $f(P) = (f(P)_1, \dots, f(P)_p)$ con $i \in \{1, \dots, p\}$:

$$f_i(P) := f(P)_i \quad P \in I \subseteq \mathbb{R}^d.$$

Per questo motivo, le funzioni reali di più variabili reale sono particolarmente interessanti da analizzare e, in quel che segue, ci dedicheremo con particolare attenzione proprio a questa classe.

Le trasformazioni di \mathbb{R}^d in sé: $d = p \geq 1$. Nel caso $d = 2$, si tratta di trasformazioni del piano, cioè di funzioni che fanno corrispondere alla coppia $(x, y) \in \mathbb{R}^2$ una nuova coppia $(u, v) \in \mathbb{R}^2$. Una trasformazione del piano in sé può essere vista come una coppia di funzioni da $I \subseteq \mathbb{R}^2$ in \mathbb{R} :

$$(x, y) \in I \mapsto u = u(x, y) \in \mathbb{R}, \quad (x, y) \in I \mapsto v = v(x, y) \in \mathbb{R}.$$

In generale, si tratta di trasformazione che associano ad un punto $P = (x_1, \dots, x_d)$ un nuovo punto $Q = (y_1, \dots, y_d)$. Un'ampia classe di trasformazioni di \mathbb{R}^d è quella delle **trasformazioni lineari**, cioè funzioni della forma $f(P) := AP$, dove A è una matrice di dimensione $d \times d$.

Assegnata una funzione specifica, come procedere per riconoscerne le proprietà più significative? Nel caso di funzioni reali di una variabile reale, una strategia abbastanza generale consiste nel rappresentare l'andamento qualitativo della funzione disegnandone il grafico. Cosa cambia quando la dimensione aumenta? Per ora, accontentiamoci del caso di funzioni reali di più variabile reale.

Il primo passo è quello di determinare l'insieme di definizione della funzione, qualora non sia specificato esplicitamente. Una funzione è generalmente assegnata indicando un procedimento di calcolo che può essere applicabile ad alcuni punti di \mathbb{R}^d e non ad altri. Qualora non sia specificato diversamente, l'insieme dei punti P per cui il procedimento relativo alla funzione è ben definito si considera essere l'insieme di definizione della funzione. Ad esempio, consideriamo le funzioni

$$f(x, y) = \frac{1}{x + y}, \quad g(x, y) = \ln(1 - x^2 - y^2), \quad h(x, y) = \sqrt{3x^2 + 2y^2 - 5}.$$

La funzione f è definita in \mathbb{R}^2 privato della retta di equazione $x + y = 0$; la funzione g ha per insieme di definizione I il disco aperto di centro l'origine e raggio 1: $I := \{(x, y) : x^2 + y^2 < 1\}$; la funzione h è definita in tutto \mathbb{R}^2 privato dell'interno dell'ellisse di equazione $3x^2 + 2y^2 = 5$.

Una volta noti i punti per cui la funzione ha senso, si potrebbe essere interessati ad analizzare, in maniera qualitativa, l'andamento della funzione. Generalizzando la definizione già vista nel caso di funzioni di una variabile reale, si chiama **grafico della funzione f** l'insieme

$$\Gamma_f := \{(P, Q) \in \mathbb{R}^d \times \mathbb{R} : P \in I, Q = f(P)\}.$$

Nel caso $d = 2$, il grafico di una funzione $z = f(x, y)$ ha, quasi sempre l'aspetto di una superficie dello spazio. Se $d \geq 3$, il grafico è un sottoinsieme di uno spazio con dimensione maggiore o uguale a 4 e, di conseguenza, la sua visualizzazione grafica risulta sensibilmente più complicata... In quel che segue, ci concentriamo principalmente sul caso di funzioni da \mathbb{R}^2 in \mathbb{R} .

Profili altimetrici. Il grafico Γ_f di una funzione f da \mathbb{R}^2 in \mathbb{R} può essere immaginato come una superficie (di montagna o di pianura, a seconda della funzione) su cui si può liberamente passeggiare, trovandosi ad altezza $z = f(x, y)$ non appena le coordinate (cartesiane) della posizione siano date dalla coppia (x, y) .

Uno specifico cammino sul grafico Γ_f è quindi individuato da una legge oraria che associa alla variabile tempo t una corrispondente coppia di coordinate $(x(t), y(t))$. In corrispondenza, all'istante t , ci si trovare ad altezza $h(t) := f(x(t), y(t))$. Qualcosa suona familiare... la coppia $(x(t), y(t))$ è una curva a valori nell'insieme di definizione della funzione f e l'altezza h non è altro che la composizione tra la funzione $t \mapsto \phi(t) := (x(t), y(t))$ e la funzione $(x, y) \mapsto z = f(x, y)$. Per coerenza con il punto di vista escursionistico, chiamiamo nel seguito la funzione composta $h = f \circ \phi$ il **profilo altimetrico** di f relativo alla curva ϕ . Un profilo altimetrico fornisce sempre un'informazione parziale relativa alla struttura di un grafico di funzione, dato che si riferisce ad uno specifico cammino tra gli infiniti cammini possibili. Comunque, utilizzando l'informazione relativa ad un certo numero di profili altimetrici, si può sperare di ottenere un'idea qualitativa della struttura globale del grafico di una funzione.

ESEMPIO 1.1. Consideriamo la funzione $f(x, y) = x - y$, definita in tutto \mathbb{R}^2 . Proviamo ad esplorare il grafico di tale funzione considerando una curva della forma $\phi(t) = (t, 0)$ con $t \in \mathbb{R}$. Si tratta quindi di un'esplorazione che consiste nel tenere costante la variabile y e spostarsi nella direzione determinata dall'asse x . Il corrispondente profilo altimetrico è

$$h(t) = f(x(t), y(t)) = x(t) - y(t) = t.$$

Al crescere del tempo (e quindi al crescere della variabile x), l'altezza corrispondente cresce in maniera lineare (la funzione h è un polinomio di grado 1). In altre parole, l'intersezione tra il grafico della funzione f e il piano (x, z) nello spazio \mathbb{R}^3 è dato da una retta.

Se consideriamo la curva $\psi(\tau) = (0, \tau)$ con $\tau \in \mathbb{R}$, il profilo altimetrico è

$$h(\tau) = f(x(\tau), y(\tau)) = x(\tau) - y(\tau) = -\tau.$$

Anche in questo caso il profilo altimetrico è dato da un polinomio di grado 1. A differenza del caso precedente, al crescere di τ l'altezza diminuisce.

In generale, considerando una curva della forma della forma $\phi(t) = tv$ con $t \in \mathbb{R}$, dove $v = (v_1, v_2) \in \mathbb{R}^2$ è un vettore fissato, si ottiene un profilo altimetrico dato da un polinomio di primo grado

$$h(t) = f(x(t), y(t)) = v_1 t - v_2 t + 1 = (v_1 - v_2) t.$$

Tutte le sezioni del grafico della funzione con piani verticali, danno luogo a rette e, quindi, il grafico della funzione è un piano nello spazio.

ESEMPIO 1.2. Consideriamo la funzione $f(x, y) = x^2 + y^2$, definita in tutto \mathbb{R}^2 . Dato un vettore $v = (v_1, v_2) \in \mathbb{R}^2$ di norma unitaria, consideriamo la curva $\phi(t) = tv$ con $t \in \mathbb{R}$. Il corrispondente profilo altimetrico è

$$h(t) = f(x(t), y(t)) = v_1^2 t^2 + v_2^2 t^2 = (v_1^2 + v_2^2) t^2 = t^2$$

Si tratta cioè di una parabola con concavità rivolta verso l'alto e minimo nel punto $t = 0$. Come si vede, il profilo altimetrico non varia al variare della scelta del vettore unitario v , indice del fatto che il grafico della funzione considerata è *invariante per rotazioni* attorno all'asse z . Il grafico completo si ottiene quindi facendo ruotare il grafico della parabola $z = t^2$ attorno all'asse z ed ha quindi la forma di una scodella (a sezione parabolica) rivolta verso l'alto.

ESEMPIO 1.3. Piccola (ma non tanto...) variante dell'esempio precedente: studiamo la funzione $f(x, y) = x^2 - y^2$, definita in tutto \mathbb{R}^2 . Di nuovo, scegliamo come legge oraria dell'esplorazione la curva $\phi(t) = tv$ con $t \in \mathbb{R}$, dove $v = (v_1, v_2) \in \mathbb{R}^2$, $|v|_d = 1$. Si ha

$$h(t) = f(x(t), y(t)) = v_1^2 t^2 - v_2^2 t^2 = (v_1^2 - v_2^2) t^2.$$

In questo caso, a seconda della scelta di v il comportamento del profilo altimetrico varia. Precisamente:

- se $|v_1| > |v_2|$, il profilo è quello di una parabola con concavità verso l'alto e minimo in $t = 0$;
- se $|v_1| = |v_2|$, il profilo è identicamente nullo;
- se $|v_1| < |v_2|$, il profilo è quello di una parabola con concavità verso il basso e massimo in $t = 0$.

Inoltre, la convessità è massima per $v = (\pm 1, 0)$ e minima per $v = (0, \pm 1)$. Il grafico della funzione f ha la forma di una sella.

ESEMPIO 1.4. Studiamo la funzione

$$f(x, y) = \frac{xy}{x^2 + y^2} \quad (x, y) \neq (0, 0).$$

Consideriamo come curve di percorrenza della superficie, le rette per l'origine, espresse dalle funzioni $\phi(t) = tv$ con $t \in \mathbb{R}$, dove $v = (v_1, v_2) \in \mathbb{R}^2$, $|v|_d = 1$. Il profilo altimetrico è

$$h(t) = f(v_1 t, v_2 t) = \frac{v_1 v_2}{v_1^2 + v_2^2} = v_1 v_2.$$

L'altezza di ogni singolo profilo è costante, con altezza dipendente dalla retta scelta. Il grafico è rappresentato in Figura 1.

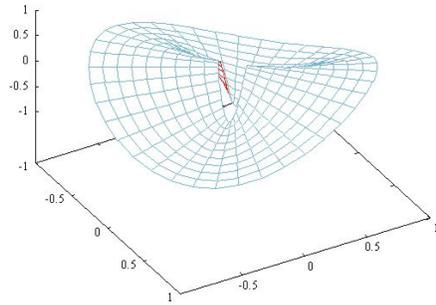


FIGURA 1. Il grafico della funzione $f(x, y) = \frac{xy}{x^2 + y^2}$.

I profili altimetrici possono essere considerati anche per funzioni di d variabili con $d \geq 2$. Date una funzione $f : I \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$ e una curva $\phi : J \subset \mathbb{R} \rightarrow \mathbb{R}^d$ tale che $\phi(J) \subset I$, la funzione composta $f \circ \phi : J \subseteq \mathbb{R} \rightarrow \mathbb{R}$ si chiama **profilo altimetrico** della funzione f lungo la curva ϕ .

Curve di livello. Tra i profili altimetrici di una funzione data, ne esistono alcuni particolarmente interessanti: sono quelli costanti. Una maniera alternativa per rappresentare una funzione $f = f(x, y)$ è quello di disegnare nel piano (x, y) gli insiemi di livello $\gamma_1 := \{(x, y) : f(x, y) = 1\}$, $\gamma_2 := \{(x, y) : f(x, y) = 2\}$, $\gamma_3 := \{(x, y) : f(x, y) = 3\}$,... e, in generale, $\gamma_c := \{(x, y) : f(x, y) = c\}$ con $c \in \mathbb{R}$. Tale metodo, detto **metodo delle linee di livello**, è quello utilizzato nelle carte geografiche¹.

Determinare insiemi di livello di una funzione è, in qualche modo, un procedimento inverso a quello dei profili altimetrici: data la costante c , si cercano tutti i percorsi possibili che diano come profilo altimetrico il valore costante c .

ESEMPIO 1.5. Consideriamo, di nuovo, la funzione $f(x, y) = x - y$. Dato il livello $c \in \mathbb{R}$, l'insieme $\gamma_c := \{(x, y) : f(x, y) = c\}$ è determinato da

$$x - y = c \quad \iff \quad y = x - c.$$

Si tratta quindi di rette parallele alla bisettrice del primo e terzo quadrante. Il fatto che variazioni di livello c uguali corrispondano a traslazioni uguali, si traduce nel fatto che la pendenza della superficie è sempre la stessa.

ESEMPIO 1.6. Gli insiemi di livello della funzione $f(x, y) = x^2 + y^2$ sono particolarmente semplici: data $c \geq 0$, si ha $\gamma_c := \{(x, y) : x^2 + y^2 = c\}$, si tratta cioè di

¹Per arricchire la rappresentazione delle linee di livello, in topografia, si utilizzano anche scale cromatiche che rappresentano i rilievi: toni di marrone via via più deciso per le catene montuose, toni di blu sempre più intenso per le profondità degli oceani e così via...

circonferenze centrate nell'origine e di raggio \sqrt{c} . Per $c < 0$, l'insieme γ_c è vuoto. Dalla struttura delle curve di livello si riconosce l'invarianza del grafico rispetto a rotazioni attorno all'asse z .

A differenza del caso precedente, a variazioni uguali del livello c non corrispondono variazioni uguali degli insiemi di livello: il raggio della circonferenza γ_c è \sqrt{c} , quindi il raggio cresce più rapidamente in prossimità del punto $(0, 0)$ e sempre più lentamente per valori di c grandi. Nella rappresentazione per curve di livello (che in concreto viene effettuata disegnando un numero finito di curve relative a livelli a distanza costante), si vedrà una minore concentrazione di curve vicino al punto $(0, 0)$. In corrispondenza, il grafico risulta essere meno pendente vicino all'origine e più pendente man mano che ci si allontana dal punto $(0, 0)$.

ESEMPIO 1.7. Consideriamo la funzione $f(x, y) = x^2 - y^2$. L'insieme di livello relativa al livello $c = 0$ è particolarmente semplice: dato che

$$f(x, y) = x^2 - y^2 = (x - y)(x + y),$$

si ha $\gamma_0 = \{y = x\} \cup \{y = -x\}$, cioè l'insieme di livello è dato dalle due bisettrici del piano (x, y) . Scegliendo $c \neq 0$, si ha $\gamma_c = \{x^2 - y^2 = c\}$, il cui grafico nel piano rappresenta una iperbole equilatera con asintoti dati dalle bisettrici $y = \pm x$. Se $c > 0$, l'insieme γ_c è contenuto nella regione $\{|y| < |x|\}$, mentre se $c < 0$, l'insieme γ_c è contenuto nella regione $\{|x| < |y|\}$.

Nel caso di funzioni di più variabili, la definizione di insieme di livello è del tutto analoga. In genere, dato un livello c , l'insieme di livello $\gamma_c = \{P \in \mathbb{R}^d : f(P) = c\}$ è un'unione di superfici di dimensione $d - 1$.

Funzioni con simmetrie. Disegnare il grafico di una funzione che possiede qualche forma di simmetria è, come al solito, più facile. Partiamo da un caso semplice semplice. Data la funzione reale di una variabile reale $F : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$, consideriamo la funzione di due variabili reali:

$$f(x, y) = F(x).$$

La funzione f è definita in $I \times \mathbb{R}$, infatti, dato che non c'è dipendenza esplicita dalla variabile y , il valore della seconda coordinata può essere qualsiasi. Inoltre, sempre per l'indipendenza della funzione f dalla variabile y , il grafico Γ_f risulta essere invariante rispetto a traslazioni nella direzione dell'asse y :

$$(133) \quad \Gamma_f + \lambda \mathbf{j} = \Gamma_f \quad \forall \lambda \in \mathbb{R}.$$

dove $\mathbf{j} = (0, 1)$. La proprietà (133) indica che per disegnare il grafico della funzione f basta disegnare il grafico della funzione F nel piano (x, z) e, poi, traslare parallelamente tale grafico nella direzione dell'asse y (vedi Figura 2). Il caso di funzioni della

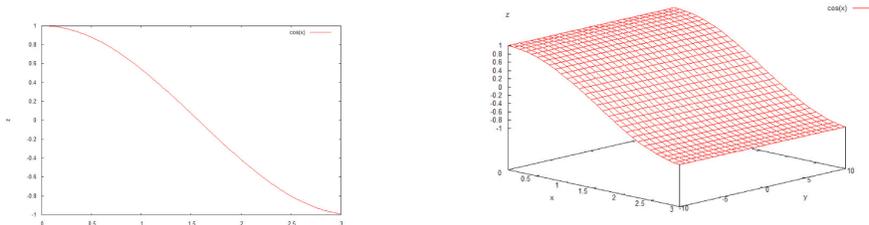


FIGURA 2. Il grafico della funzione $F(x) = \cos x$ e quella della funzione $f(x, y) = \cos x$.

forma $f(x, y) = F(y)$ è del tutto analogo, con l'unica differenza che l'invarianza è nella direzione dell'asse x .

Leggermente più in generale, data una funzione $F : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ e date due costanti $a, b \in \mathbb{R}$, non entrambe nulle, si può considerare la funzione di due variabili reali:

$$f(x, y) = F(ax + by).$$

Le curve di livello di una funzione di questo genere sono (unioni di) rette della forma $ax + by = \text{costante}$. Di conseguenza, il grafico della funzione gode dell'invarianza seguente

$$(134) \quad \Gamma_f + \lambda v = \Gamma_f \quad \forall \lambda \in \mathbb{R}.$$

dove $v = (-b, a)$. Per disegnare il grafico della funzione f , basta disegnare il grafico della funzione F in uno qualsiasi dei piani verticali ortogonali alla direzione di invarianza e poi traslare parallelamente nella direzione di v . Sperimentate a vostro gradimento.

Consideriamo un secondo tipo di simmetria. Data $F : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$, studiamo la funzione di due variabili reali definita nel modo seguente

$$f(x, y) = F(\sqrt{x^2 + y^2}).$$

Gli insiemi di livello di una funzione di questo genere sono (unioni di) circonferenze centrate in $(0, 0)$. Pertanto il grafico della funzione f risulta essere invariante rispetto a rotazioni attorno all'asse z . Formalmente, definendo

$$A_\theta := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

la matrice che descrive la rotazione di angolo θ in senso antiorario del piano (x, y) , il grafico della funzione f gode della proprietà

$$(135) \quad A_\theta \Gamma_f = \Gamma_f \quad \forall \theta \in [0, 2\pi).$$

Concretamente, la proprietà (135) indica che per disegnare il grafico della funzione f , basta disegnare il grafico della funzione F nel semipiano (x, z) con $x \geq 0$ e, successivamente, far ruotare tale grafico attorno all'asse z (vedi Figura 3).

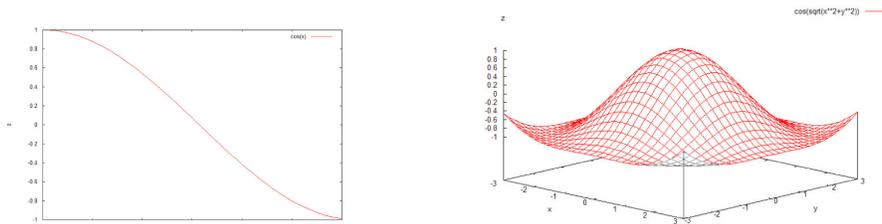


FIGURA 3. Il grafico delle funzioni $F(x) = \cos x$ e $f(x, y) = \cos(\sqrt{x^2 + y^2})$.

Operazioni elementari su grafici. A partire da grafici noti, è possibile dedurre la forma di altri grafici che si ottengono dai primi attraverso un certo numero di operazioni elementari. Qui, consideriamo il caso di dilatazioni/compressioni (o omotetie) e traslazioni.

Supponiamo di conoscere il grafico della funzione $f = f(x, y)$. Dato $\lambda > 0$, il grafico della funzione

$$g(x, y) = f(\lambda x, y)$$

si ottiene a partire dal grafico della funzione f applicando una compressione, nel caso $\lambda > 1$, o una dilatazione, nel caso $0 < \lambda < 1$, nella direzione dell'asse x .

ESEMPIO 1.8. A mo' d'esempio, si considerino i grafici delle funzioni

$$g(x, y) = 4x^2 + y^2, \quad h(x, y) = \frac{x^2}{4} + y^2,$$

che corrispondono, rispettivamente, ad una compressione e una dilatazione del grafico della funzione $f(x, y) = x^2 + y^2$, con fattori $\lambda = 2$ e $\lambda = 1/2$. Gli insiemi di livello delle funzioni g ed h relativi al livello $c = 1$, ad esempio, sono dati da ellissi di equazioni

$$\frac{x^2}{(1/2)^2} + y^2 = 1, \quad \frac{x^2}{2^2} + y^2 = 1,$$

quindi ellissi con assi di simmetria gli assi coordinati x e y , e semiassi di lunghezza $1/2$ e 1 , nel caso della funzione g , e di lunghezza 2 e 1 , nel caso della funzione f .

Considerazioni analoghe possono essere fatte per funzioni della forma

$$g(x, y) = f(\lambda_1 x, \lambda_2 y)$$

con $\lambda_1, \lambda_2 > 0$ dati. I fattori λ_1 e λ_2 danno luogo a dilatazioni e/o compressioni nelle direzioni degli assi principali, e gli insiemi di livello della funzione g si ottengono a partire da quelli della funzione f applicando le corrispondenti deformazioni.

Continuando a supporre il grafico della funzione f noto, consideriamo una funzione del tipo

$$g(x, y) = f(x + x_0, y + y_0)$$

con $(x_0, y_0) \in \mathbb{R}^2$ dato. Gli insiemi di livello della funzione g sono dati da

$$\begin{aligned} \gamma_c^g &= \{(x, y) : g(x, y) = c\} = \{(x, y) : f(x + x_0, y + y_0) = c\} \\ &= \{(\xi, \eta) : f(\xi, \eta) = c\} - (x_0, y_0) = \gamma_c^f - (x_0, y_0). \end{aligned}$$

avendo posto $(\xi, \eta) = (x + x_0, y + y_0)$. Le curve di livello della funzione g si ottengono per traslazione delle curve di livello della funzione f e, di conseguenza, il grafico della funzione g si ricava da quella della funzione f applicando un'analogia traslazione.

I casi considerati fin qui corrispondono a trasformazioni applicate alla variabile indipendente (x, y) . Analogamente, si può considerare il caso di trasformazioni applicate alla variabile dipendente. Al volenteroso lettore il compito di meditare sui casi

$$g(x, y) = \lambda f(x, y), \quad g(x, y) = f(x, y) + c$$

supponendo il grafico della funzione f noto e $\lambda > 0$, $c \in \mathbb{R}$ assegnati.

2. Le derivate parziali

Un possibile approccio allo studio di funzioni reale di più variabili reali è quello di analizzare i profili altimetrici relativi a percorsi diversi e cercare di “incollare” le informazioni parziali ottenute in questo modo. Il vantaggio di questo punto di vista è che studiare uno specifico profilo altimetrico, vuol dire, in concreto, studiare una funzione reale di variabile reale ed è quindi applicabile tutta la teoria e la tecnica vista nei Capitoli precedenti.

Ad esempio, data una funzione $f : I \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ ed un punto $P_0 = (x_0, y_0)$ interno ad I , può essere interessante studiare il comportamento della funzione f lungo le rette passanti per tale punto, ovvero analizzare i profili altimetrici

$$h(t) = f(x_0 + t v_1, y_0 + t v_2)$$

con $v = (v_1, v_2) \in \mathbb{R}^2$, $|v|_a = 1$, assegnato. Ad esempio, la monotonia lungo la retta per P_0 e di direzione v è determinata dal segno della derivata della funzione h . A tale derivata, ottenuta considerando la funzione lungo una retta fissata, si dà un nome ben preciso.

DEFINIZIONE 2.1. Sia $f : I \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$ e sia P_0 un punto interno ad I . Dato $v \in \mathbb{R}^d$, $|v|_d = 1$, si chiama derivata direzionale della funzione f nel punto P_0 rispetto alla direzione v il limite seguente

$$\frac{df}{dv}(P_0) = \lim_{t \rightarrow 0} \frac{f(P_0 + tv) - f(P_0)}{t}.$$

Scegliendo $v = \mathbf{e}_i$ per qualche i , indicando con $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$ uno dei vettori della base canonica, la derivata direzionale si chiama derivata parziale della funzione f nel punto P_0 rispetto a x_i . Le derivate parziali si indicano, a seconda dei gusti, con i simboli

$$\frac{\partial f}{\partial x_i}, \quad f_{x_i}, \quad D_{x_i}f, \quad \partial_{x_i}f.$$

Si chiama vettore gradiente (o semplicemente gradiente) della funzione f , il vettore composto dalle derivate parziali prime

$$\nabla f := (f_{x_1}, \dots, f_{x_n}).$$

Il simbolo ∇ si legge *nabla*.

Il calcolo delle derivate parziali non comporta nessuna difficoltà aggiuntiva rispetto al calcolo delle derivate di funzioni di una variabile. Consideriamo infatti il caso di una funzione di due variabili $f = f(x, y)$. La derivata parziale rispetto alla variabile x è la derivata del profilo altimetrico ottenuto fissando la variabile y e facendo variare la sola variabile x . Pertanto, derivare rispetto ad x corrisponde a derivare la funzione $F(x) = f(x, y)$ dove y è considerato costante. Esempi:

$$\frac{\partial}{\partial x}(x + y^2) = 1, \quad \frac{\partial}{\partial x}(x y^2) = y^2, \quad \frac{\partial}{\partial x}(e^{x y^2}) = y^2 e^{x y^2}.$$

Analogamente:

$$\frac{\partial}{\partial y}(x + y^2) = 2y, \quad \frac{\partial}{\partial x}(x y^2) = 2xy, \quad \frac{\partial}{\partial x}(e^{x y^2}) = 2xy e^{x y^2}.$$

Il calcolo delle derivate direzionali a partire dalla definizione è meno immediato (vedremo più avanti una regola di calcolo, valida per una ampia classe di funzioni, che rende tale operazione molto più semplice). Ad esempio, calcoliamo le derivate direzionali della funzione $f(x, y) = x + y^2$. Dato $v \in \mathbb{R}^2$ con $|v|_2 = 1$, si ha

$$\frac{f(x + t v_1, y + t v_2) - f(x, y)}{t} = \frac{x + t v_1 + (y + t v_2)^2 - x - y^2}{t} = v_1 + 2 v_2 y + t v_2^2,$$

e, passando al limite per $t \rightarrow 0$, si ottiene

$$\frac{df}{dv} = v_1 + 2 v_2 y.$$

Nei casi particolari $v = (1, 0)$ e $v = (0, 1)$, si ottengono nuovamente le derivate parziali rispetto ad x e rispetto ad y .

Analogamente, per $g(x, y) = x y^2$, si ha

$$\begin{aligned} \frac{g(x + t v_1, y + t v_2) - g(x, y)}{t} &= \frac{(x + t v_1)(y + t v_2)^2 - x y^2}{t} \\ &= v_1 y^2 + 2 v_2 x y + (v_2^2 x + 2 v_1 v_2 y)t + v_1 v_2^2 t^2, \end{aligned}$$

da cui, passando al limite per $t \rightarrow 0$, segue

$$\frac{df}{dv} = v_1 y^2 + 2 v_2 x y.$$

Una conseguenza immediata della definizione di derivata direzionale e derivata direzionale è la seguente condizione necessaria per punti estremo relativo.

PROPOSIZIONE 2.2. *Sia $f : I \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$ e sia $P_0 \in I$, interno ad I . Se la funzione f ha un massimo (o un minimo) locale in P_0 , cioè se per qualche $r > 0$ si ha*

$$f(P) \leq (\geq) f(P_0) \quad \forall P \in I, |P - P_0|_d < r,$$

allora tutte le derivate direzionali di f in P_0 , qualora esistano, sono nulle.

DIMOSTRAZIONE. La dimostrazione è particolarmente semplice. Fissata una direzione $v \in \mathbb{R}^2$, $|v|_d = 1$, il profilo altimetrico $h(t) = f(P_0 + t v)$ descrive una funzione reale di variabile reale che ha, per ipotesi, un massimo (minimo) locale in $t = 0$. Di conseguenza, la derivata prima $h'(0)$, se esiste, è nulla. Dato che, per definizione, $h'(0)$ coincide con $\frac{df}{dv}(P_0)$, la proposizione è dimostrata. \square

Operativamente, la precedente Proposizione indica che, nella ricerca di massimi e minimi di una funzione data, occorre considerare come candidati possibili tutti i punti in cui le derivate direzionali, o semplicemente le derivate parziali (più semplici da calcolare!), siano nulle. Torneremo sulla questione tra qualche Capitolo.

Fissata una direzione $v \in \mathbb{R}^d$, la derivata direzionale rispetto alla direzione v è funzione del punto di derivazione e, di conseguenza, è essa stessa, qualora esista una funzione reale di più variabili reali. E' quindi possibile definire le derivate direzionali seconde, cioè le derivate direzionali di una derivata direzionale. Nel seguito, considereremo solamente il caso delle derivate parziali seconde, cioè delle derivate direzionali rispetto agli assi principali ed utilizzeremo notazioni coerenti con quelle utilizzate per le derivate parziali prime. Ad esempio, per le derivate seconde si usa comunemente scrivere

$$\frac{\partial^2 f}{\partial x_i \partial x_j}, \quad f_{x_i x_j}, \quad D_{x_i x_j}^2 f, \quad \partial_{x_i x_j} f.$$

Ad esempio, per le funzioni $x + y^2$ e $x y^2$, valgono, rispettivamente, le formule seguenti

$$\begin{aligned} \frac{\partial^2}{\partial x^2}(x + y^2) &= 0, & \frac{\partial^2}{\partial y \partial x}(x + y^2) &= 0, & \frac{\partial^2}{\partial x \partial y}(x + y^2) &= 0, & \frac{\partial^2}{\partial y^2}(x + y^2) &= 2, \\ \frac{\partial^2}{\partial x^2}(x y^2) &= 0, & \frac{\partial^2}{\partial y \partial x}(x y^2) &= 2 y, & \frac{\partial^2}{\partial x \partial y}(x y^2) &= 2 y, & \frac{\partial^2}{\partial y^2}(x y^2) &= 2 x. \end{aligned}$$

A voi, il gusto di calcolare tutte le derivate terze delle stesse due funzioni. Una funzione f da \mathbb{R}^d in \mathbb{R} possiede d derivate parziali prime, d^2 derivate seconde e, in generale, d^k derivate parziali k -esime.

DEFINIZIONE 2.3. *La matrice con elementi dati dalle derivate seconde della funzione $f : I \subseteq \mathbb{R}^d \rightarrow \mathbb{R}$*

$$d^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \frac{\partial^2 f}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{pmatrix}$$

si chiama *matrice hessiana* (o semplicemente *hessiano*) della funzione f .

Ad esempio, si hanno

$$d^2(x + y^2) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{e} \quad d^2(x y^2) = \begin{pmatrix} 0 & 2 y \\ 2 y & 2 x \end{pmatrix}.$$

Come vedremo più avanti, la matrice hessiana dà informazioni relative alla curvatura del grafico della funzione f .

Guardare ad una funzione di più variabili studiandone profili altimetrici e/o sezioni con piani verticali fornisce informazioni rilevanti, ma resta, purtroppo, un approccio parziale. Ecco due esempi indicativi a riguardo.

ESEMPIO 2.4 (Una retta non basta!). Consideriamo la funzione

$$f(x, y) = \frac{xy}{x^2 + y^2} \quad (x, y) \neq (0, 0).$$

Dato che la funzione non è definita nell'origine, è interessante studiare cosa avvenga in prossimità di tale punto. Consideriamo come profilo altimetrico, quello determinato dall'asse x : data $\phi(t) = (t, 0)$, studiamo

$$h(t) = f(t, 0) = 0 \quad \forall t \neq 0.$$

La funzione è identicamente nulla lungo l'asse delle x e, quindi, tende a 0 lungo tale direzione. Cosa succede se si sceglie una direzione diversa? Dato $v = (v_1, v_2) \in \mathbb{R}^2$, $|v|_2 = 1$, lungo $\phi(t) = tv$, si ha

$$h(t) = f(tv_1, tv_2) = v_1 v_2 \quad \forall t \neq 0,$$

come già visto nell'Esempio 1.4. Lungo la direzione determinata dal vettore unitario v , la funzione f tende al valore $v_1 v_2$. Direzioni diverse danno luogo a limiti diversi.

ESEMPIO 2.5 (Tutte le rette non bastano!). Come secondo esempio, consideriamo la funzione

$$f(x, y) = \frac{x^2 y}{x^4 + y^2} \quad (x, y) \neq (0, 0).$$

Dato $v = (v_1, v_2) \in \mathbb{R}^2$, $|v|_2 = 1$, lungo $\phi(t) = tv$, il profilo altimetrico è dato da

$$h(t) = f(tv_1, tv_2) = \frac{t v_1^2 v_2}{t^2 v_1^4 + v_2^2}. \quad \forall t \neq 0,$$

Per $t \rightarrow 0$, si ha, per ogni scelta del vettore unitario v ,

$$\lim_{t \rightarrow 0} \frac{t v_1^2 v_2}{t^2 v_1^4 + v_2^2} = 0.$$

Lungo tutte le rette per l'origine, la funzione si avvicina al valore 0. Cosa succede lungo altri cammini che passano per l'origine? Consideriamo, ad esempio, la parabola $\phi(t) = (t, t^2)$: si ha

$$h(t) = f(t, t^2) = \frac{t^4}{t^4 + t^4} = \frac{1}{2}. \quad \forall t \neq 0;$$

e lungo le parabole la parabola $\phi(t) = (t, at^2)$, $a \in \mathbb{R}$

$$h(t) = f(t, at^2) = \frac{a^2 t^4}{t^4 + a^2 t^4} = \frac{a^2}{1 + a^2}. \quad \forall t \neq 0,$$

Quindi, percorsi diversi dai percorsi rettilinei portano a valori limite diversi.

La morale è semplice: una funzione di più variabili non può essere semplicemente considerata come l'incollamento delle sue sezioni planari o dei suoi profili altimetrici considerati separatamente. Una visione globale è indispensabile per avere una comprensione globale della funzione considerata. Tale punto di vista globale verrà analizzato a partire dal prossimo Capitolo.

3. Integrali curvilinei

Data una funzione $f : I \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ a valori positivi, cioè tale che $f(x, y) \geq 0$ per ogni (x, y) e data una curva $\phi : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^2$ tale che $\phi([a, b]) \subset I$, è ben definito il profilo altimetrico $f \circ \phi$. Consideriamo il sottoinsieme di $\mathbb{R}^2 \times \mathbb{R}$, delimitato dall'immagine del profilo $f \circ \phi$ e dal piano $z = 0$, cioè

$$\Sigma_{f, \phi} := \{(x, y, z) \in \mathbb{R}^3 : (x, y) = \phi(t), 0 \leq z \leq f(\phi(t)), t \in [a, b]\}.$$

L'insieme $\Sigma_{f, \phi}$ può essere immaginato come un nastro in \mathbb{R}^3 posato verticalmente sul piano $z = 0$: il supporto della curva ϕ determina la base d'appoggio del nastro ed il profilo altimetrico $f \circ \phi$ ne descrive lo spessore. Qual è l'area di tale insieme?

L'allenamento fatto nella costruzione dell'integrale di Riemann e ripreso successivamente nel calcolo delle lunghezze di curve suggerisce la possibile strategia... Data una partizione $\pi = \{a = t_0 < t_1 < \dots < t_n = b\}$ dell'intervallo $[a, b]$, consideriamo come approssimante dell'area cercata il valore:

$$\sum_{k=0}^{n-1} f(\phi(t_k)) |\phi'(t_k)|_2 (t_{k+1} - t_k).$$

Scegliendo partizioni π con ampiezza infinitesima, si ottiene formalmente l'integrale della funzione $f(\phi(t)) |\phi'(t)|_2$ nell'intervallo $[a, b]$. Dimentichiamo la richiesta di segno e passando a dimensione qualsiasi, si ottiene, in maniera naturale, la seguente definizione.

DEFINIZIONE 3.1. *Data una funzione $f : I \subset \mathbb{R}^d \rightarrow \mathbb{R}$ e una curva parametrizzata regolare a tratti $\phi : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^d$ tale che $\phi([a, b]) \subset I$, si chiama integrale curvilineo di f lungo ϕ il valore dell'integrale definito (qualora esista)*

$$(136) \quad \int_{\phi} f ds := \int_a^b f(\phi(t)) |\phi'(t)|_d dt.$$

La richiesta che la funzione ϕ sia regolare a tratti, indica che la funzione $|\phi'(t)|_d$ è integrabile secondo Riemann, quindi, affinché l'integrale curvilineo (136) sia ben definito, è sufficiente che il profilo altimetrico $f \circ \phi$ determini una funzione integrabile secondo Riemann.

ESEMPIO 3.2. Consideriamo le funzioni $f(x, y) = x^2 + 4y^2$ e $\phi : [0, \pi/2] \rightarrow \mathbb{R}^2$ definita da $\phi(t) = (\cos t, \sin t)$, e calcoliamo l'integrale curvilineo di f lungo ϕ . Utilizzando la formula (136), si ottiene

$$\begin{aligned} \int_{\phi} f ds &= \int_0^{\pi/2} (\cos^2 t + 4 \sin^2 t) \sqrt{\sin^2 t + \cos^2 t} dt \\ &= \int_0^{\pi/2} (1 + 3 \sin^2 t) dt = \frac{1}{2} \pi + \frac{3}{4} \pi = \frac{5}{4} \pi. \end{aligned}$$

ESEMPIO 3.3. Consideriamo un caso in dimensione più alta. Sia $f(x, y, z) = y \sin z$ e consideriamo la curva ϕ definita da

$$\phi(t) = (\cos t, \sin t, t) \quad t \in [0, \pi].$$

Si ha

$$\int_{\phi} f ds = \int_0^{\pi} \sin t \sin t \sqrt{\sin^2 t + \cos^2 t + 1} dt = \sqrt{2} \int_0^{\pi} \sin^2 t dt = \frac{\sqrt{2}}{2} \pi.$$

Niente di speciale...

Non fatevi ingannare dai due esempi precedenti: in genere, il calcolo esplicito di un integrale curvilineo è difficile se non infattibile...

La costruzione che abbiamo visto dell'integrale curvilineo lascia sospettare che, cambiando parametrizzazione della curva, il valore dell'integrale curvilineo non cambi. La dimostrazione di tale proprietà è una semplice modifica di quella già vista nel caso della lunghezza di curve. A voi ripercorrerla implementando le modifiche del caso.

Come conseguenza del fatto che l'integrale curvilineo è, in realtà, un integrale di una funzione reale di variabile reale, esso gode delle stesse proprietà viste per l'integrale usuale. Eccole in rapida rassegna.

Linearità. Data una curva ϕ , due funzioni f e g e due costanti $\alpha, \beta \in \mathbb{R}$, se le funzioni f e g sono integrabili lungo ϕ , anche la funzione $\alpha f + \beta g$ è integrabile lungo ϕ e si ha

$$\int_{\phi} (\alpha f + \beta g) ds = \alpha \int_{\phi} f ds + \beta \int_{\phi} g ds$$

Monotonia. Data una curva ϕ e due funzioni f e g , integrabili lungo ϕ e tali che $f \leq g$,

$$\int_{\phi} f ds \leq \int_{\phi} g ds.$$

Additività. Data una curva $\phi : [a, b] \rightarrow \mathbb{R}^d$ ed una funzione f , integrabile lungo ϕ , per ogni $c \in (a, b)$, la funzione f risulta integrabile anche lungo ϕ_- e ϕ_+ , dove ϕ_{\pm} indicano le restrizioni di ϕ a $[a, c]$ e $[c, b]$ e si ha

$$\int_{\phi} f ds = \int_{\phi_-} f ds + \int_{\phi_+} f ds.$$