

Raggruppamento delle
NOTE DI ANALISI I
fornite al corso di laurea
triennale in Matematica.

L. Orsina - P. D'Ancona
Canali A-K, L-Z 2024/2025
La Sapienza

CAPITOLO 1

La retta reale

1. I numeri naturali. Gli interi relativi.

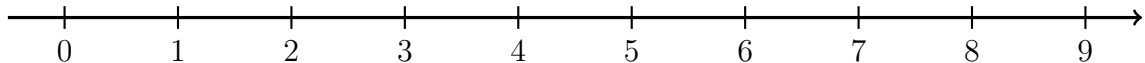
L'operazione di contare è una delle più naturali che esistano. Ognuno di noi, prima ancora di sapere che cosa vogliono dire “uno” e “due”, ha già la capacità di distinguere tra (almeno) due persone differenti. Solo dopo un po' di tempo si rende conto che le persone sono in realtà “tre”, ed inizia a contare sé stesso⁽¹⁾. Per complicare un po' le cose, invece di contare da uno, decideremo di contare da zero⁽²⁾.

Per meglio identificare i numeri che possiamo ottenere contando a partire da zero, definiamo l'insieme dei numeri **naturali**:

$$(1.1) \quad \mathbb{N} = \{0, 1, 2, 3, \dots, n, \dots\}.$$

In questo caso i secondi puntini di sospensione “...” indicano che siamo in grado di contare “fino a quando vogliamo”; in altre parole che l'insieme dei numeri naturali è “infinito”, ovvero che non ammette “massimo”. Non esiste, cioè, il più grande numero naturale: non appena pensiamo di averlo trovato, è sufficiente “aggiungere 1” per trovarne uno ancora più grande.

Se rappresentiamo i numeri naturali su una retta, scegliendo in maniera arbitraria la distanza tra 0 ed 1 (ovvero, l'unità di misura), avremo una immagine “discreta”: per passare da un numero al successivo si deve compiere un passo lungo 1.



L'operazione di “aggiungere 1”, restando ancora all'interno dello stesso insieme è una proprietà di \mathbb{N} , che può essere così generalizzata: dati due numeri naturali n ed m qualsiasi, la loro **somma** $n + m$ è ancora un numero naturale: quello ottenuto, appunto, partendo da n ed effettuando m passi verso destra. In linguaggio matematico, “l'insieme dei numeri naturali è chiuso rispetto alla somma”. Siccome partire da n e fare m passi è la stessa cosa che partire da m e fare n passi, l'addizione è **commutativa**:

$$\forall n, m \in \mathbb{N}, \quad n + m = m + n.$$

Inoltre, esiste un numero naturale “speciale”, lo zero, che gode della proprietà di lasciare “immutato” qualsiasi numero rispetto alla somma (come è evidente: sommare zero vuol dire fare 0 passi, cioè non muoversi):

$$\forall n \in \mathbb{N}, \quad n + 0 = n.$$

⁽¹⁾In alcune popolazioni aborigene dell'Amazzonia, il concetto di tre (ovvero, della paternità) non esiste.

⁽²⁾Storicamente, il concetto di zero compare molto dopo quello di uno, due e tre.

Lo zero viene pertanto detto **elemento neutro** dell'operazione di somma. La chiusura rispetto alla somma permette di ordinare i numeri naturali: diremo che un numero naturale n è **maggiore** di m se esiste un numero naturale p , diverso da zero, tale che $n = m + p$. In simboli, scriveremo $n > m$. Data questa definizione, si può affermare che dati due numeri naturali **diversi** n ed m , si ha $n > m$, oppure $m > n$ (in definitiva, uno dei due sarà a destra dell'altro sulla retta). Ammettendo che il numero naturale p possa essere 0, viene definito il concetto di "maggiore od uguale"; in simboli $n \geq m$. La relazione di " \geq " è antisimmetrica: se $n \geq m$ e $m \geq n$, allora $n = m$ ⁽³⁾.

Infine, dato un numero naturale n , e fissato un secondo numero naturale m che avrà la funzione di "contatore", è possibile sommare n a sé stesso ripetendo l'operazione m volte. Il risultato, ovvero " m volte n ", verrà indicato con

$$m \cdot n = \overbrace{n + n + \dots + n}^{m \text{ volte}}:$$

il **prodotto** di m con n . Il numero 1 è l'elemento neutro del prodotto, essendo $1 \cdot n = n$ per ogni n in \mathbb{N} . Le relazioni tra somma, prodotto e ordinamento sono quelle note:

$$\begin{aligned} \forall n, m \in \mathbb{N}, \quad m \cdot n &= n \cdot m, \\ \forall n, m, p \in \mathbb{N}, \quad p \cdot (n + m) &= p \cdot n + p \cdot m, \\ \forall n, m, p \in \mathbb{N}, \quad n \geq m &\Rightarrow p \cdot n \geq p \cdot m, \\ \forall n, m, p \in \mathbb{N}, \quad n > m, \quad p \neq 0 &\Rightarrow p \cdot n > p \cdot m. \end{aligned}$$

Il prodotto di un numero naturale con sé stesso, ovvero $n \cdot n$, viene convenzionalmente indicato con n^2 , il quadrato di n . Analogamente, il prodotto di n per sé stesso m volte viene indicato con n^m , la potenza m -sima di n . Ovviamente $n^1 = n$, mentre, per convenzione, $n^0 = 1$ per ogni n in \mathbb{N} . Per le potenze valgono le regole fondamentali:

$$\begin{aligned} \forall n, m, p \in \mathbb{N}, \quad n^{m+p} &= n^m \cdot n^p, \\ \forall n, m, p \in \mathbb{N}, \quad (n \cdot m)^p &= n^p \cdot m^p, \\ \forall n, m, p \in \mathbb{N}, \quad (n^m)^p &= n^{m \cdot p}. \end{aligned}$$

Supponiamo ora di avere un sottoinsieme E di \mathbb{N} ; ad esempio

$$\begin{aligned} E_1 &= \{\text{numeri pari}\} = \{0, 2, 4, 6, 8, \dots\}, \\ E_2 &= \{\text{numeri dispari}\} = \{1, 3, 5, 7, 9, \dots\}, \\ E_3 &= \{\text{multipli non nulli di cinque}\} = \{5, 10, 15, 20, 25, \dots\}, \\ E_4 &= \{\text{potenze di due}\} = \{1, 2, 4, 8, 16, \dots\}, \\ E_5 &= \{\text{numeri primi}\} = \{2, 3, 5, 7, 11, \dots\}, \\ E_6 &= \{\text{numeri di Fibonacci}\} = \{1, 1, 2, 3, 5, 8, 13, 21, \dots\}, \end{aligned}$$

Ognuno di questi insiemi non ammette un numero più grande: dato un numero (pari o dispari) qualsiasi, è sufficiente aggiungere due per ottenere un numero più grande della stessa parità; sommando 5 ad un multiplo di 5 se ne ottiene uno più grande; raddoppiando una potenza di due se ne trova una più grande; dimostrare che esistono infiniti numeri primi (numeri, cioè, che non sono ottenibili come prodotto di naturali diversi da 1) è più complicato, ma è comunque vero; anche i numeri di Fibonacci (ognuno dei quali è somma dei due precedenti, a partire da 1, 1), perché è sufficiente sommare ad uno di essi quello che lo precede nella lista per ottenere un numero di Fibonacci maggiore. Invece, per tutti gli insiemi considerati in precedenza esiste un

⁽³⁾Esercizio!

numero — appartenente all'insieme — più piccolo di tutti gli altri (rispettivamente, 0, 1, 5, 1, 2 e 1).

DEFINIZIONE 1.1. Sia E un sottoinsieme non vuoto di \mathbb{N} . Diremo che E ammette **minimo** se esiste m in E tale che $n \geq m$ per ogni n in E . In altre parole, se esiste un elemento di E che è il più piccolo (il più a sinistra) di tutti. Se E ammette minimo, tale minimo è unico; se infatti ne esistessero due, si avrebbe $m_1 \geq m_2$ (perché m_2 è minimo) e $m_2 \geq m_1$ (perché m_1 è minimo); per antisimmetria, $m_1 = m_2$.

Si osservi che la richiesta di appartenenza ad E è fondamentale, dato che esiste sempre un numero naturale che è più piccolo di tutti gli elementi di E : lo zero. Inoltre, se il minimo non appartenesse ad E , l'unicità del minimo verrebbe meno: ad esempio l'insieme E_3 definito prima ammette 0, 1, 2, 3, 4 e 5 come numeri naturali “più piccoli” di tutti i numeri in E_3 (di questi, però, solo un appartiene ad E_3 , ed è il maggiore di tutti).

ESERCIZIO 1.2. Costruire un sottoinsieme non vuoto di \mathbb{N} che non ammette minimo.

Svolto l'esercizio precedente? Se non lo avete fatto, fatelo o, almeno, tentate di farlo. Non ci riuscite? Non c'è da meravigliarsi: non si riesce a trovare un sottoinsieme non vuoto di \mathbb{N} che non ammetta minimo, per il semplice fatto che un tale sottoinsieme non esiste. Questo è il contenuto del seguente principio.

PRINCIPIO DEL BUON ORDINAMENTO: *Qualunque sottoinsieme non vuoto di \mathbb{N} ammette minimo.*

È possibile *dimostrare* la precedente affermazione? Proviamo con il seguente ragionamento.

Ragionamento. Sia E un sottoinsieme non vuoto di \mathbb{N} . Allora, per definizione, E contiene almeno un elemento; sia esso n_0 . A questo punto, esistono due possibilità: o n_0 è più piccolo di tutti gli altri elementi di E , oppure no. Se è più piccolo, n_0 è il minimo (ed allora abbiamo finito); se non è più piccolo, esiste n_1 in E , con $n_0 > n_1$ (strettamente maggiore, perché deve essere diverso). Se n_1 è più piccolo di tutti gli elementi di E , allora è il minimo; se non lo è, allora esiste n_2 in E con $n_1 > n_2$. Possiamo continuare il ragionamento, ed è chiaro che “ci troviamo nei pasticci” se la scoperta di un numero n_k in E tale che $n_{k-1} > n_k$ prosegue indefinitamente: ovvero se continuiamo a trovare elementi di E più piccoli del precedente, ma ancora maggiori di qualche altro elemento di E . La nostra fortuna è che una tale “discesa infinita” non è possibile perché, fissato n_0 (ed n_0 è fissato una volta per tutte dall'essere E non vuoto), esistono al più n_0 numeri di cui n_0 è maggiore e che possono appartenere ad E : 0, 1, 2, ..., $n_0 - 1$. Pertanto, dopo al più n_0 “scelte”, non avremo più a disposizione numeri naturali; il che vuol dire che l'ultima scelta che abbiamo fatto non può essere migliorata: E ammette minimo. ■

Il ragionamento precedente è molto convincente, e avrebbe soddisfatto un matematico dei primi del novecento. Ma ad un matematico moderno non basta più: cos'è una discesa infinita? E perché non è possibile? In realtà non si può rispondere a queste domande in modo soddisfacente: ossia, non è possibile *dimostrare* il principio del buon ordinamento in modo rigoroso! Quello che possiamo fare è assumere la validità del

principio *a priori*, come punto di partenza dei ragionamenti seguenti. In altri termini, *il principio del buon ordinamento è un assioma della nostra teoria*. Notiamo che si potrebbe enunciare questo principio in varie altre forme equivalenti; ne vedremo una più avanti.

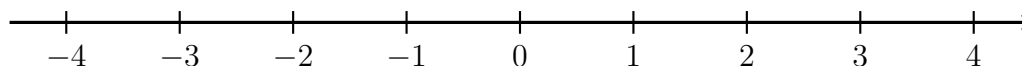
Fino ad ora ci siamo mossi sulla retta “andando verso destra”, nella direzione cioè della somma. Che succede se, partendo da un numero n qualsiasi, ad esempio 5, ci muoviamo verso sinistra? Ovviamente, dopo un salto siamo atterrati su 4 (che è caratterizzato — rispetto alla nostra operazione di salto — dall’essere l’unico numero naturale il cui successore è 5), e se continuiamo a spostarci verso sinistra troviamo (nell’ordine), 3, 2, 1, e 0. A questo punto l’operazione di salto verso sinistra ci viene impedita dal fatto che “prima di zero” non c’è nulla. Non esiste alcun numero naturale il cui successore sia zero. In altre parole, se interpretiamo l’operazione di saltare verso sinistra come “sottrazione” (così come avevamo interpretato l’azione di saltare verso destra come addizione), l’insieme dei numeri naturali non è chiuso rispetto a tale operazione: non sempre sottraendo un numero naturale da un altro naturale si ottiene un elemento di \mathbb{N} . A questo punto, si aprono davanti a noi due strade: lasciare tutto così com’è, emanare un editto che vieti la sottrazione⁽⁴⁾, e proseguire a lavorare con i naturali⁽⁵⁾, oppure decidere che una tale restrizione non va bene, ed “aggiungere” ad \mathbb{N} tutti i numeri mancanti.

Già, ma quanti sono? E come indicarli? Per convenzione, si decide di usare le stesse cifre usate per i naturali, ovvero 1, 2, 3, e di farle precedere da un segno “-” (a ricordare che sono numeri che si possono ottenere solo tramite sottrazioni, appunto). Inoltre, si sceglie di indicare con $-n$ il numero ottenuto facendo n passi a sinistra a partire dallo zero (così come n è il numero che si ottiene facendo n passi verso destra a partire da zero). Abbiamo allora l’insieme degli **interi relativi**:

$$(1.2) \quad \mathbb{Z} = \{\dots, -n, \dots, -3, -2, -1, 0, 1, 2, 3, \dots, n, \dots\}.$$

La definizione stessa di $-n$ fa poi sì che $n - n = n + (-n)$ sia uguale a zero, e che quindi $-n$ sia l’**inverso** di n rispetto all’operazione di somma (e, simmetricamente, n sia l’inverso di $-n$).

Come i naturali, anche gli interi sono rappresentabili su una retta. Siccome \mathbb{N} è contenuto in \mathbb{Z} , conviene “prolungare” a sinistra in maniera simmetrica la rappresentazione grafica di \mathbb{N} , per ottenere quella di \mathbb{Z} :



Per costruzione, l’insieme dei numeri interi relativi è chiuso sia rispetto alla somma che alla sottrazione; tramite la sottrazione, è possibile definire la relazione di **minore**: dati n ed m in \mathbb{Z} , diciamo che n è minore di m , e scriviamo $n < m$, se esiste p in $\mathbb{N} \setminus \{0\}$ tale che $n = m - p$. Se p può essere anche 0, abbiamo la relazione di **minore od uguale**, che indichiamo \leq . Dal momento che se $n = m - p$, allora $n + p = m - p + p = m + 0 = m$, ne segue che $n < m$ ($n \leq m$) se e solo se

⁽⁴⁾Per alcuni tipi di sottrazione tale “editto” esiste già, e viene studiato in un’altra facoltà. . .

⁽⁵⁾In definitiva, il teorema di matematica più famoso di tutti i tempi riguarda i numeri naturali!

$m > n$ ($m \geq n$). Mentre per i sottoinsiemi dei naturali non avevamo necessariamente il numero più grande, ma avevamo il minimo (per il principio di buon ordinamento), per \mathbb{Z} la situazione è differente: esistono sottoinsiemi di \mathbb{Z} che non posseggono minimo; ad esempio, \mathbb{Z} stesso: se supponiamo che m sia il minimo, $m - 1$ è un numero intero più piccolo di m . È anche possibile definire il prodotto tra due numeri interi, e si verifica che le proprietà di somma e prodotto (come la commutatività e la distributività) si estendono anche a \mathbb{Z} . Notiamo, in particolare che $(-m) \cdot (-n) = m \cdot n$; infatti, essendo $0 = (m + (-m)) \cdot (-n)$, si ha che $-(m \cdot (-n)) = (-m) \cdot (-n)$. Analogamente si prova che $m \cdot (-n) = -(m \cdot n)$, da cui $m \cdot n = -(-m \cdot n) = (-m) \cdot (-n)$.

2. Calcolo combinatorio. Principio di induzione.

Quanti sono i numeri che si possono scrivere usando una ed una sola volta le cifre da 1 a 9? In altre parole, vogliamo contare quante “stringhe” come “135792468” si possono formare. È chiaro che se procediamo in maniera empirica, scrivendo numeri a caso su un foglio di carta e verificando che il numero che abbiamo appena scritto non compaia nella lista, corriamo il rischio da un lato di perdere molto tempo, dall’altro di non sapere quando li abbiamo scritti tutti. Per rispondere alla domanda, affrontiamo per il momento un problema più semplice: quanti sono i numeri che si possono scrivere usando una ed una sola volta le cifre 1 e 2? In questo caso la risposta è semplice: sono esattamente 2, e precisamente “12” e “21”. Iniziamo con l’osservare che uno dei due numeri ha come prima cifra “1”, mentre l’altro inizia con “2”. Una volta scelta la prima cifra (e lo possiamo fare in due modi diversi — tanti quante sono le cifre a nostra disposizione), la seconda è obbligata (avendo sotto mano solo due cifre). E se le cifre sono 1, 2, e 3? In questo caso abbiamo tre scelte possibili per la prima cifra: “1xx”, “2xx” e “3xx”. Una volta scelta la prima cifra, quante scelte abbiamo per la seconda? Non più tre, perché una cifra l’abbiamo già utilizzata, ma solo due: ad esempio, “12x” e “13x”. È ovvio che, scelte le prime due cifre, la terza è obbligata. In definitiva, abbiamo tre scelte per la prima cifra, due per la seconda, ed una per la terza: in tutto $6 (= 3 \cdot 2 \cdot 1)$ possibilità.

Possiamo allora rispondere alla domanda iniziale: usando una ed una sola volta le cifre da 1 a 9 possiamo scrivere $9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 326880$ numeri diversi; è chiaro che se avessimo affrontato il problema usando la forza bruta, non avremmo mai ottenuto un risultato preciso.

Siccome scrivere $9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ è faticoso, è stata “inventata” una notazione apposita: dato un numero naturale n , indichiamo con $n!$, che si legge **n fattoriale**⁽⁶⁾, il numero

$$(2.1) \quad n! = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 3 \cdot 2 \cdot 1.$$

Una delle proprietà più importanti di $n!$, che discende direttamente dalla definizione, è il fatto che $n! = n \cdot (n - 1)!$; questa proprietà ci dice che, se partiamo a fare le moltiplicazioni da 1, e non da n , otteniamo via via $1!$, $2!$ e così via fino ad $n!$. I primi dieci fattoriali sono 1, 2, 6, 24, 120, 720, 5040, 40320, 362880, 3628800: come si vede, i fattoriali crescono abbastanza rapidamente all’aumentare di n .

⁽⁶⁾E non “enne!” con enfasi, come narrano le leggende metropolitane...

Cosa conta esattamente $n!$? Dati n numeri (o oggetti), $n!$ conta il numero delle possibili **permutazioni** degli n numeri, dove per permutazione si intende uno dei possibili ordinamenti degli n numeri dati. Ad esempio, “12345” è una permutazione di “54321”, così come lo sono “13524” e “24135”, mentre (cane, gatto, topo) è una permutazione di (topo, cane, gatto).

ESERCIZIO 2.1. A partire da $5! = 120$, $n!$ termina con almeno uno zero. Al crescere di n , aumenta il numero degli zeri di $n!$. Con quanti zeri termina $30!$? Evidentemente, di calcolare $30!$ non se ne parla neppure. . .

Risposta 2.1: 7. Infatti, tra i fattori di $30!$ ci sono 2 e 5 (uno zero), 10 (uno zero), 4 e 15 (uno zero), 20 (uno zero), 16 e 25 (due zeri) e 30 (uno zero). Per i curiosi, $30! = 265252859812191058636308480000000$.

ESERCIZIO 2.2. Si ha $20! = 243290200x176640000$. Quanto vale x ?

Risposta 2.2: 8. Infatti $20!$, avendo come fattore 9, deve essere divisibile per 9, e quindi la somma delle sue cifre deve essere, come è noto, divisibile per 9. La somma delle cifre di $20!$ è $46 + x$ da cui $x = 8$.

Supponiamo adesso di avere n oggetti, e di sceglierne m tra questi: ad esempio, date le cifre da 1 a 9, ne scegliamo 5. Quanti numeri possiamo formare in questo modo? Ancora una volta, conviene ragionare per “numeri piccoli”, per capire cosa succede. Supponiamo di avere a disposizione 1, 2 e 3, e di dover scegliere due cifre. Ci sono, evidentemente, sei possibilità: “12”, “21”, “13”, “31”, “23” e “32”. Se le cifre a disposizione sono 1, 2, 3 e 4, e dobbiamo sempre scegliere due cifre, abbiamo dodici possibilità: “12”, “21”, “13”, “31”, “14”, “41”, “23”, “32”, “24”, “42”, “34” e “43”. Come abbiamo ottenuto questi numeri? Se, nel primo caso, consideriamo le $3!$ permutazioni di 1, 2 e 3, possiamo “scegliere” in tutti i modi possibili due cifre semplicemente “cancellando” l’ultima; in questa maniera da “123” si ottiene “12”, da “213” si ottiene “21”, e così via. Siccome abbiamo 6 permutazioni dalle quali cancellare l’ultima cifra, abbiamo 6 numeri che si possono ottenere scegliendo due cifre tra 1, 2 e 3. Che succede nel secondo caso? Le possibili permutazioni di 1, 2, 3 e 4 sono $4!$, cioè 24, ma scegliendone una qualsiasi e cancellando le ultime due cifre, non sempre si ottiene (come resto) lo stesso numero: ad esempio, sia “1234” che “1243” generano “12” per cancellazione delle ultime due cifre; e questo perché le ultime due cifre possono essere permutate tra loro in due modi possibili (o meglio, in $2!$ modi possibili).

A questo punto, possiamo rispondere alla domanda: se abbiamo 9 cifre, e ne dobbiamo scegliere 5, è sufficiente considerare una qualsiasi delle $9!$ permutazioni di 1, 2, . . . , 9; cancellare le ultime 4 cifre, ed osservare che queste 4 cifre possono permutarsi (nelle ultime quattro posizioni) in $4!$ modi diversi; pertanto, il numero richiesto è $\frac{9!}{4!} = \frac{362880}{24} = 15120$.

Definiamo allora **disposizioni** di n oggetti presi a gruppi di m la quantità

$$(2.2) \quad {}_nD_m = \frac{n!}{(n-m)!} = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-m+2) \cdot (n-m+1).$$

Supponiamo ora di avere n cifre, e di chiederci (un’altra volta!) quanti numeri si possono creare scegliendo esattamente n cifre. Da un lato, sappiamo già la risposta:

$n!$; dall'altro, stiamo considerando le disposizioni di n oggetti presi a gruppi di n , e quindi la risposta è ${}_n D_n$; la formula ci dà ${}_n D_n = \frac{n!}{0!}$; sappiamo che deve essere uguale ad $n!$, ma non abbiamo la più pallida idea di cosa sia $0!$. Per cavarci d'impaccio, *definiamo* $0! = 1$. In questa maniera, pur forzando l'interpretazione del fattoriale come numero delle "permutazioni", salviamo sia la correttezza della formula delle disposizioni, sia il suo significato combinatorio.

Ed ora, la prossima domanda: abbiamo sempre le cifre da 1 a 9, e vogliamo contare in quanti modi ne possiamo scegliere 5, senza tenere conto dell'ordine nel quale le scegliamo. In altre parole, non distinguiamo "12345" da "12354" e "53421". Invece di ridurci a casi semplici, questa volta ragioniamo direttamente: i numeri ottenuti scegliendo 5 cifre tra le 9 a nostra disposizione sono ${}_9 D_5$. Scelto uno qualsiasi dei ${}_9 D_5$ numeri, dobbiamo scartare tutti quelli che hanno le stesse cifre del nostro numero; siccome le cifre del numero che abbiamo scelto sono permutabili in $5!$ modi diversi, dovremo "buttare via" esattamente $5! - 1$ numeri. È allora evidente che il risultato voluto si ottiene dividendo ${}_9 D_5$ per $5!$: $\frac{{}_9 D_5}{5!} = \frac{15120}{120} = 126$.

Definiamo **combinazioni** di n oggetti presi a gruppi di m la quantità

$$(2.3) \quad {}_n C_m = \frac{n!}{m!(n-m)!} = \binom{n}{m}^{(7)}.$$

Dalla definizione stessa di combinazione, si deduce che m deve essere compreso tra 0 e n . Tenendo conto del fatto che $0! = 1$, la formula per ${}_n C_m$ è sempre ben definita (anche per $n = 0$) e — miracolosamente — restituisce sempre un numero intero. I ${}_n C_m$, o $\binom{n}{m}$, sono anche detti **coefficienti binomiali**⁽⁸⁾.

Una delle proprietà più notevoli di $\binom{n}{m}$ è la seguente: dati n ed m (con $0 \leq m \leq n$), si ha

$$(2.4) \quad \binom{n+1}{m+1} = \binom{n}{m} + \binom{n}{m+1}.$$

La dimostrazione di questa formula usa in maniera cruciale il fatto che $n! = n \cdot (n-1)!$. Infatti

$$\begin{aligned} \binom{n}{m} + \binom{n}{m+1} &= \frac{n!}{m!(n-m)!} + \frac{n!}{(m+1)!(n-m-1)!} \\ &= \frac{n!}{(m+1)!(n-m)!} [(m+1) + (n-m)] \\ &= \frac{n!}{(m+1)!(n-m)!} (n+1) \\ &= \frac{(n+1)!}{(m+1)!(n-m)!} = \binom{n+1}{m+1}. \end{aligned}$$

⁽⁷⁾ $\binom{n}{m}$ si legge "n sopra m".

⁽⁸⁾Per motivi che saranno chiariti tra breve.

Questa proprietà permette di scrivere i coefficienti binomiali in maniera “triangolare”, formando il cosiddetto **triangolo di Tartaglia**:

$$\begin{array}{cccccccc}
 n & & & & & & & & \\
 0 & & & & & & & & \binom{0}{0} \\
 1 & & & & & & & & \binom{1}{0} & \binom{1}{1} \\
 2 & & & & & & & & \binom{2}{0} & \binom{2}{1} & \binom{2}{2} \\
 3 & & & & & & & & \binom{3}{0} & \binom{3}{1} & \binom{3}{2} & \binom{3}{3} \\
 4 & & & & & & & & \binom{4}{0} & \binom{4}{1} & \binom{4}{2} & \binom{4}{3} & \binom{4}{4}
 \end{array}$$

Ogni numero nel triangolo è somma dei due numeri che si trovano a sinistra e a destra nella riga precedente, come si vede eseguendo i calcoli:

$$\begin{array}{cccccccc}
 n & & & & & & & & \\
 0 & & & & & & & & 1 \\
 1 & & & & & & & & 1 & 1 \\
 2 & & & & & & & & 1 & 2 & 1 \\
 3 & & & & & & & & 1 & 3 & 3 & 1 \\
 4 & & & & & & & & 1 & 4 & 6 & 4 & 1
 \end{array}$$

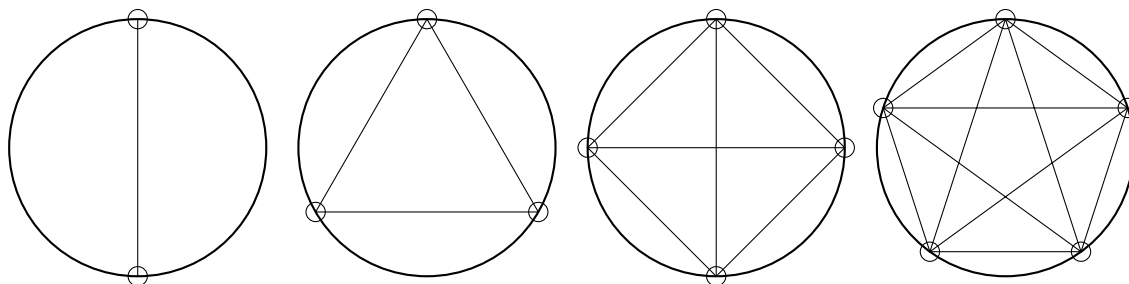
Affrontiamo ora un altro problema: dati due numeri, a e b , ed un naturale n , vogliamo calcolare la potenza n -sima del binomio $a + b$, e trovare una formula che, dato n , ci permetta di scrivere $(a + b)^n$ senza dover eseguire i calcoli tutte le volte. Alcuni casi particolari sono già noti⁽⁹⁾:

$$\begin{aligned}
 (a + b)^0 &= 1 \\
 (a + b)^1 &= a + b \\
 (a + b)^2 &= a^2 + 2ab + b^2 \\
 (a + b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3 \\
 (a + b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4
 \end{aligned}$$

A questo punto, non v'è chi non veda la relazione strettissima che intercorre tra le potenze di $a + b$ e il triangolo di Tartaglia: i coefficienti delle potenze di a e b sono gli stessi che compaiono nel triangolo di Tartaglia.

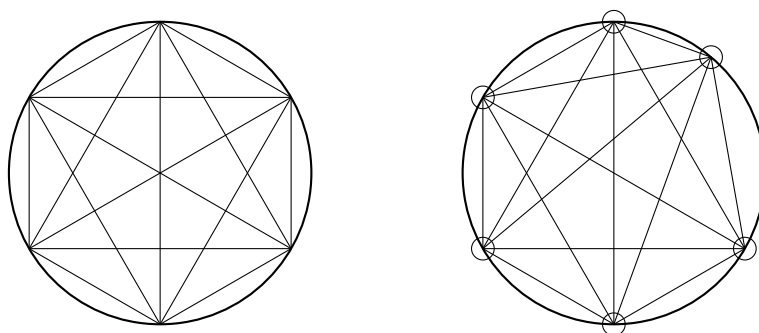
ESERCIZIO 2.3. Prima di andare avanti, supponiamo di avere una circonferenza, e di voler disegnare n punti su di essa in modo tale che, congiungendo i punti in tutti i modi possibili, il cerchio risulti diviso nel maggior numero di parti. Quante sono queste parti? Ovvero, in che modo dipendono da n ? Come al solito, adottiamo la strategia dei “numeri piccoli”; se $n = 1$, siccome non c'è nulla da congiungere, abbiamo una sola parte (tutto il cerchio); il caso $n = 2$ è molto semplice: presi due punti sulla circonferenza, unendoli si divide in due il cerchio (e chiaramente questo è il massimo numero possibile di parti). Se $n = 3$ si ottengono 4 parti, mentre se ne hanno 8 per 4 punti. Se i punti sono 5, ne otteniamo 16.

⁽⁹⁾o, almeno, dovrebbero...



Siamo pronti per scrivere la formula?

Risposta 2.3: No. Congetturando che il numero delle parti sia 2^{n-1} , se $n = 6$ non riusciamo in nessun modo ad ottenerne 32; se adottiamo la “strategia simmetrica” e scegliamo i sei punti come vertici di un esagono regolare, otteniamo 30 parti. Se spostiamo un po’ uno dei punti arriviamo a 31, ma non di più.



La formula corretta è

$$\text{Parti}(n) = \frac{n^4 - 6n^3 + 23n^2 - 18n + 24}{24} = \binom{n}{4} + \binom{n}{2} + 1,$$

ed è evidente che il numero delle parti è molto minore di 2^{n-1} .

Risolto l’esercizio precedente⁽¹⁰⁾, appare chiaro che la frase “i coefficienti delle potenze di a e b sono gli stessi che compaiono nel triangolo di Tartaglia” è forse un po’ affrettata: in definitiva, abbiamo verificato che le prime cinque righe del triangolo di Tartaglia danno i coefficienti dello sviluppo della potenza n -sima di $a + b$, ma non abbiamo in nessun modo giustificato l’affermazione che tali coefficienti siano dati dalle righe del triangolo di Tartaglia **per ogni n** . Certamente, possiamo provare a scrivere lo sviluppo di $(a + b)^5$, e controllare se le cose vanno bene, e proseguire con $(a + b)^6$, $(a + b)^7$, ma per quante “verifiche” si facciamo, non saremo mai in grado di concludere che i coefficienti dello sviluppo di $(a + b)^n$ sono dati, **per ogni n** , dalla $(n + 1)$ -sima riga del triangolo di Tartaglia.

Prima di proseguire, però, cerchiamo di formalizzare matematicamente la “congettura” che stiamo facendo: cerchiamo cioè di scrivere in termini matematici la frase “i coefficienti dello sviluppo di $(a + b)^n$ sono dati dalla $(n + 1)$ -sima riga del triangolo di Tartaglia”. Iniziamo con scrivere $(a + b)^3$:

$$\begin{aligned} (a + b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3 \\ &= \binom{3}{0} a^3 b^0 + \binom{3}{1} a^2 b^1 + \binom{3}{2} a^1 b^2 + \binom{3}{3} a^0 b^3. \end{aligned}$$

⁽¹⁰⁾Non lo avete risolto? Provateci!

Quali sono gli elementi comuni all'ultima formula? Innanzitutto, l'indice n di $\binom{n}{m}$ è sempre 3 (che, poi, è la potenza cui viene elevato il binomio $a + b$). Poi, la somma delle potenze di a e b è sempre 3, ed inoltre l'indice m di $\binom{n}{m}$ cresce da 0 a 3, ed è sempre uguale all'esponente di b . Se ripetiamo lo stesso ragionamento per $(a + b)^2$, il risultato è lo stesso, così come è lo stesso per $(a + b)^4$. A questo punto, possiamo iniziare a formalizzare la nostra "congettura": per la potenza n -sima avremo bisogno di un indice m che "conta" da 0 ad n , e che di volta in volta viene utilizzato sia come esponente di b , che in $\binom{n}{m}$; inoltre, il coefficiente sarà sempre della forma $\binom{n}{m}$, e l'esponente da dare ad a sarà tale che la somma degli esponenti di a e b sia n , e quindi darà $n - m$. In definitiva,

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

Questa formula — sulla cui correttezza non possiamo ancora giurare — è però molto "scomoda" da manipolare; basti pensare che ha $n + 1$ termini, e se n è molto grande è necessaria parecchia carta (ed altrettanto inchiostro) per poterla scrivere⁽¹¹⁾. Fortunatamente, è stato inventato un modo più "compatto" per scrivere formule di questo tipo, basandosi sul fatto che siamo in presenza di un'operazione matematica ripetuta (in questo caso l'addizione) al variare di un "contatore". Avremo così

$$(2.5) \quad (a + b)^n \stackrel{?}{=} \sum_{m=0}^n \binom{n}{m} a^{n-m} b^m.$$

Il simbolo

$$\sum_{m=0}^n$$

si legge "sommatoria per m che va da 0 ad n di..."., o "somme per m da 0 ad n di...". Il suo effetto è quello di generare un contatore (m) che "corre" sui naturali muovendosi da 0 ad n (lo si può pensare come un "cronometro", ad esempio); ad ogni valore di m associamo una quantità (in questo caso il monomio $\binom{n}{m} a^{n-m} b^m$) che va sommata al totale ottenuto in precedenza (il totale iniziale essendo zero).

ESERCIZIO 2.4. Calcolare

$$\sum_{m=1}^5 m, \quad \sum_{\text{Qfwfq}=1}^5 (\text{Qfwfq})^2, \quad \sum_{m=1}^5 n^2, \quad \sum_{m=0}^5 \binom{5}{m}.$$

Risposta 2.4: 15, 55, $5n^2$, 32.

ESERCIZIO 2.5. Calcolare

$$\sum_{m=1}^n m, \quad \sum_{m=0}^n n^2, \quad \sum_{m=0}^n \binom{n}{m};$$

per l'ultima somma, si supponga valida la (2.5) (e si scelgano due valori opportuni per a e b).

Risposta 2.5: $\frac{n(n+1)}{2}$, $n^2(n+1)$ e 2^n , scegliendo $a = 1 = b$.

⁽¹¹⁾Se la formula serve per altri calcoli, l'uso dei puntini di sospensione è impossibile...

ESERCIZIO 2.6. Scrivere sotto forma di sommatoria le seguenti quantità

$$\begin{aligned} & 1 + 2 + 3 + 4 + \dots + n, \\ & 1 + 2 + 4 + 8 + \dots + 2^n, \\ & 1 \cdot 1 + 2 \cdot 4 + 3 \cdot 6 + 4 \cdot 4 + 5 \cdot 1, \\ & 10 + 10^{10} + 10^{10^{10}} + 10^{10^{10^{10}}} + 10^{10^{10^{10^{10}}}} + \dots \end{aligned}$$

Risposta 2.6:

$$\sum_{m=1}^n m, \quad \sum_{m=0}^n 2^m, \quad \sum_{m=0}^4 (m+1) \cdot \binom{4}{m},$$

mentre l'ultima somma non può essere scritta in maniera compatta, a meno di non osservare che l' n -simo addendo è 10 elevato all' $(n-1)$ -simo.

ESERCIZIO 2.7. Sapendo che \sum sta a \prod (che si legge “produttoria”) come la somma sta al prodotto, in quale modo possiamo anche scrivere i numeri

$$\prod_{m=0}^n m, \quad \text{e} \quad \prod_{m=1}^n m?$$

Risposta 2.7: 0 ed $n!$.

Dopo aver lavorato un po' con le sommatorie, siamo finalmente in grado di affrontare il problema di dimostrare la (2.5) per ogni n in \mathbb{N} . Come abbiamo detto (ammaestrati anche dall'Esercizio 2.3), il fatto di verificare che la (2.5) è vera per qualche valore di n non implica in nessun modo che lo sia per tutti. Possiamo però fare un ragionamento differente: supponiamo di aver — con lunghi e laboriosi calcoli — dimostrato che la formula è vera per un certo valore di n , e proviamo *ad usare questo fatto per dimostrare che la formula è valida per $n+1$* . In altre parole, invece di sviluppare $(a+b)^{n+1}$, e di verificare successivamente che lo sviluppo che otteniamo è proprio quello dato dal secondo membro della (2.5) con n sostituito da $n+1$, proviamo ad usare l'informazione — già nota — che $(a+b)^n$ è esattamente uguale al secondo membro della (2.5). Per utilizzare quello che già sappiamo, basta osservare che se

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

allora

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

Distribuendo la somma,

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

Lavoriamo sulla prima sommatoria, isolando il primo termine:

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

Nella seconda sommatoria, isoliamo l'ultimo termine,

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

e successivamente osserviamo che, se chiamiamo $m+1 = p$, quando m si muove da 0 a $n-1$, p si muove da 1 a n e quindi, sostituendo m con $p-1$,

$$\sum_{m=0}^n \binom{n}{m} a^{n-m} b^{m+1} = \sum_{p=1}^n \binom{n}{p-1} a^{n-p+1} b^p + b^{n+1}.$$

Ricordando che l'indice di sommatoria è "muto" (si veda l'Esercizio 2.4), possiamo riscrivere la formula precedente scrivendo m al posto di $p^{(12)}$, ottenendo

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

In definitiva, abbiamo

$$\begin{aligned} (a+b)^{n+1} &= \sum_{m=0}^n \binom{n}{m} a^{n-m+1} b^m + \sum_{m=0}^n \binom{n}{m} a^{n-m} b^{m+1} \\ &= a^{n+1} + \sum_{m=1}^n \binom{n}{m} a^{n-m+1} b^m \\ &\quad + \sum_{m=1}^n \binom{n}{m-1} a^{n-m+1} b^m + b^{n+1} \\ &= a^{n+1} + \sum_{m=1}^n \left[\binom{n}{m} + \binom{n}{m-1} \right] a^{n-m+1} b^m + b^{n+1}. \end{aligned}$$

Grazie alla formula (2.4), il contenuto delle parentesi quadre non è altro che $\binom{n+1}{m}$, e quindi

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

Dal momento che i coefficienti di a^{n+1} e b^{n+1} si possono scrivere rispettivamente come $\binom{n+1}{0}$ e come $\binom{n+1}{n+1}$, possiamo raccogliere tutto insieme ed affermare — non senza un pizzico di orgoglio — che

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

e questa è esattamente la formula (2.5) con n sostituito da $n+1$.

Ricapitolando: se la (2.5) vale per un certo valore di n , allora vale anche per il successivo. Il che vuol dire che abbiamo finito: siccome la formula è vera per $n=0$ (non si tratta di niente altro che dell'identità $(a+b)^0 = 1$), allora lo sarà per $n=1$, e siccome è vera per $n=1$, lo sarà per $n=2$, e siccome è vera per $n=2$, lo sarà anche per $n=3$, e così via... La (2.5) è vera per ogni n in \mathbb{N} !

⁽¹²⁾Questa operazione di "cambio di indici" è abbastanza frequente, per cui cercate di capire bene cosa sta succedendo!

Ne siamo proprio sicuri? In altre parole, la frase “e così via. . .” è matematicamente accettabile? Dire che una determinata proprietà, una formula, vale “per ogni n in \mathbb{N} ” vuol dire fare un’infinità di affermazioni: chi può controllarle tutte? Per ovviare al problema di verificare la (2.5) per ogni n , abbiamo “inventato” una via alternativa: quella di supporre la formula vera per un certo valore di n , e di dedurre da questo fatto la veridicità per $n + 1$; questo fatto, unitamente al verificarsi della formula per $n = 0$, ci ha portato a concludere che la formula era vera per ogni n usando la frase “e così via. . .”.

Ahimé, nel mondo della matematica gli “e così via. . .” non sono accettati⁽¹³⁾, perché le proprietà (soprattutto quelle che coinvolgono infiniti elementi) devono essere dimostrate. Ovvero, ci serve un meccanismo matematicamente corretto che permetta di concludere che se la (2.5) è vera per $n = 0$, e se dal supporla vera per n segue che è vera per $n + 1$, allora la (2.5) è vera per ogni n .

Tale strumento esiste, è chiamato **induzione matematica**, ed è il soggetto del prossimo teorema, chiamato il *Principio di induzione*.

Prima di enunciare il risultato, facciamo un’osservazione importante. Dalla dimostrazione che segue, si vede come il principio di induzione sia una conseguenza diretta del principio di buon ordinamento. Ma è vero anche il contrario: partendo dal principio di induzione è molto facile dimostrare la validità del principio di buon ordinamento. In altri termini, i due principi sono *equivalenti*. Questo vuol dire che, nella costruzione della nostra teoria, possiamo prendere come assioma indifferentemente uno dei due principi. Naturalmente, a questo punto l’altro principio si può dimostrare e diventa un teorema vero e proprio.

TEOREMA 2.8. *Sia $\mathcal{P}(n)$ una proprietà, dipendente da un indice naturale n , tale che*

- i) $\mathcal{P}(0)$ è vera;
- ii) se $\mathcal{P}(n)$ è vera, allora $\mathcal{P}(n + 1)$ è vera.

Allora $\mathcal{P}(n)$ è vera per ogni n in \mathbb{N} .

Dimostrazione. Definiamo

$$F = \{n \in \mathbb{N} : \mathcal{P}(n) \text{ è falsa}\},$$

e dimostriamo che F è vuoto. Supponiamo per assurdo che F non sia vuoto, e vediamo come da questo fatto si ottenga una contraddizione. Dal momento che F è un sottoinsieme non vuoto di \mathbb{N} , per il principio del buon ordinamento (Teorema 1) esso ammette minimo. Esiste cioè \bar{m} in F tale che \bar{m} è più piccolo di tutti gli elementi di F . Dal momento che \bar{m} appartiene ad F ,

$$\mathcal{P}(\bar{m}) \text{ è falsa.}$$

Inoltre, \bar{m} non è zero (perché, per la i), 0 non appartiene ad F). Ne segue che $\bar{m} - 1$ è un numero naturale (perché \bar{m} è almeno 1), e che $\mathcal{P}(\bar{m} - 1)$ è **vera** (dal momento che $\bar{m} - 1$ non appartiene ad F). Ma allora, per la ii),

$$\mathcal{P}((\bar{m} - 1) + 1) = \mathcal{P}(\bar{m}) \text{ è vera.}$$

⁽¹³⁾Anche nella vita reale: il tacchino induttivista di Bertrand Russell sostiene: “Tutti i giorni mi hanno dato da mangiare; quindi anche domani mi daranno da mangiare”; ma il giorno di Natale. . .

Siamo così arrivati ad un assurdo, generato dall'ipotesi che F non sia vuoto. ■

Si osservi che la proprietà chiave, quella che fa “scattare” il meccanismo dell’“e così via...”, è la ii), ed è dunque necessario comprendere che cosa richieda. Affinché la ii) sia vera, **non** si deve dimostrare che $\mathcal{P}(n)$ e $\mathcal{P}(n+1)$ sono vere, ma che dal supporre $\mathcal{P}(n)$ vera, segue che $\mathcal{P}(n+1)$ è vera. In altre parole, si deve trovare un modo di far “comparire” la proprietà $\mathcal{P}(n)$ all'interno della dimostrazione della proprietà $\mathcal{P}(n+1)$. Per meglio comprendere il meccanismo dell'induzione matematica, dimostriamo il seguente teorema.

TEOREMA 2.9 (Disuguaglianza di Bernoulli). *Sia $h \geq -1$. Dimostrare che per ogni n in \mathbb{N} si ha*

$$(2.6) \quad (1+h)^n \geq 1+nh.$$

Dimostrazione. La proprietà $\mathcal{P}(n)$ è dunque la (2.6). Se $n=0$ la (2.6) diventa $(1+h)^0 = 1 \geq 1 = 1+0h$, ed è dunque vera; pertanto, la i) è soddisfatta. Supponiamo ora che la (2.6) sia vera per un certo valore di n , ovvero che

$$(2.7) \quad (1+h)^n \geq 1+nh.$$

Prima di procedere, cerchiamo di capire che cosa stiamo facendo: è chiaro che si potrebbe dire che la (2.7) è “esattamente” la (2.6), e che quindi stiamo usando la tesi per dimostrare la tesi⁽¹⁴⁾. In realtà, non c'è alcuna contraddizione: non stiamo usando la (2.7) per dimostrare la (2.6), ma la stiamo supponendo vera, e la useremo per dimostrare la (2.6) per $n+1$. In altre parole, non stiamo dimostrando che “(2.6) è vera”, ma che “se (2.7) è vera per un certo valore di n , allora è vera anche per il valore successivo”. Proseguendo nella dimostrazione, si ha

$$\begin{aligned} (1+h)^{n+1} &= (1+h)(1+h)^n \\ \left[\begin{array}{l} 1+h \geq 0 \\ \text{per la (2.7)} \end{array} \right] &\geq (1+h)(1+nh) \\ \text{[sviluppando il prodotto]} &= 1+(n+1)h+nh^2 \\ \left[\begin{array}{l} nh^2 \geq 0 \end{array} \right] &\geq 1+(n+1)h, \end{aligned}$$

che è esattamente la (2.6) con n sostituito da $n+1$. Abbiamo così dimostrato che

$$\mathcal{P}(n) \text{ vera} \implies \mathcal{P}(n+1) \text{ vera}.$$

Per il principio di induzione, la (2.6) è vera per ogni n in \mathbb{N} . ■

TEOREMA 2.10 (Sviluppo del binomio).

$$(2.5) \quad (a+b)^n = \sum_{m=0}^n \binom{n}{m} a^{n-m} b^m, \quad \forall n \in \mathbb{N}.$$

OSSERVAZIONE 2.11. Nello sviluppo di $(a+b)^4$, da dove viene il coefficiente di $a^2 b^2$? Se scriviamo

$$(a+b)^4 = (a+b)(a+b)(a+b)(a+b),$$

è evidente che, per ottenere $a^2 b^2$ dobbiamo “scegliere” 2 volte a e 2 volte b nei quattro fattori $a+b$. In quanti modi diversi possiamo farlo? Abbiamo 4 “caselle”, numerate 1, 2, 3 e 4 (i fattori del prodotto), e dobbiamo sceglierne due (dalle quali prenderemo il

⁽¹⁴⁾Se volete, che ci stiamo mordendo la coda!

fattore b), indipendentemente dall'ordine, dato che la moltiplicazione è commutativa. Questo problema è lo stesso del determinare quante coppie di due numeri si possano estrarre dalle cifre 1, 2, 3 e 4, e già conosciamo la risposta: $\binom{4}{2}$. A questo punto è chiaro perché il coefficiente di $a^{n-m} b^m$ è $\binom{n}{m}$: per ottenere $a^{n-m} b^m$, dobbiamo scegliere m volte b da n fattori, indipendentemente dall'ordine. . .

ESERCIZIO 2.12. Dimostrare la (2.6) nel caso $h \geq 0$ usando la (2.5).

Risposta 2.12: Usando la (2.5) si ha

$$(1+h)^n = \sum_{m=0}^n \binom{n}{m} 1^{n-m} h^m = 1 + n h + \sum_{m=2}^n \binom{n}{m} h^m.$$

Dal momento che il valore della somma è non negativo (perché h è maggiore o uguale a zero), si ha la (2.6).

Dell'induzione matematica si può dare una forma più debole (di dimostrazione analoga), che risponde alla necessità di verificare che una determinata proprietà valga per ogni n maggiore di un certo n_0 fissato.

TEOREMA 2.13. *Sia $\mathcal{P}(n)$ una proprietà, dipendente da un indice naturale n , tale che esista n_0 in \mathbb{N} per cui*

- i) $\mathcal{P}(n_0)$ è vera;
- ii) se $n \geq n_0$, e se $\mathcal{P}(n)$ è vera, allora $\mathcal{P}(n+1)$ è vera.

Allora $\mathcal{P}(n)$ è vera per ogni n in \mathbb{N} , con $n \geq n_0$.

ESERCIZIO 2.14. Si dimostri, per induzione, che

$$\begin{aligned} \sum_{m=0}^n m &= \frac{n^2}{2} + \frac{n}{2}, & \forall n \in \mathbb{N}, \\ \sum_{m=0}^n m^2 &= \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}, & \forall n \in \mathbb{N}, \\ \sum_{m=0}^n m^3 &= \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4}, & \forall n \in \mathbb{N}, \\ \sum_{m=0}^n m^3 &= \left(\sum_{m=0}^n m \right)^2, & \forall n \in \mathbb{N}, \\ \text{se } q \neq 1, \sum_{m=0}^n q^k &= \frac{1 - q^{n+1}}{1 - q}, & \forall n \in \mathbb{N}. \end{aligned}$$

Risposta 2.14: Se $n = 0$ abbiamo $1 = 1$, che è vera. Se supponiamo l'uguaglianza vera per n , l'uguaglianza per $n+1$ diventa

$$\sum_{m=0}^{n+1} q^k = \sum_{m=0}^n q^k + q^{n+1} = \frac{1 - q^{n+1}}{1 - q} + q^{n+1} = \frac{1 - q^{n+2}}{1 - q},$$

come volevasi dimostrare.

ESERCIZIO 2.15. Si dimostri, per induzione, che

$$\begin{aligned} \binom{2n}{n} &\geq 2^n, & \forall n \in \mathbb{N}, \\ n! &\geq 2^{n-1}, & \forall n \in \mathbb{N}, \\ n^n &\geq 2^{n-1} n!, & \forall n \in \mathbb{N}, n \geq 1, \end{aligned}$$

determinare n_0 tale che $n^n \geq 2^n n!$, $\forall n \in \mathbb{N}, n \geq n_0$.

Risposta 2.15: 3) se $n = 1$, la disuguaglianza si riduce all'identità $1 = 1$, ed è pertanto vera. Supponiamo ora che sia vera per un certo n ; la disuguaglianza per $n + 1$ è $(n + 1)^{n+1} \geq 2^n (n + 1)!$. Dal momento che $2^n (n + 1)! = 2(n + 1) [2^{n-1} n!]$, ed essendo per ipotesi induttiva $2^{n-1} n! \leq n^n$, avremo dimostrato la nostra disuguaglianza se dimostreremo che

$$(n + 1)^{n+1} \geq 2(n + 1) n^n \iff (n + 1)^n \geq 2n^n.$$

Ricordando la (2.5), abbiamo

$$(n + 1)^n = \sum_{m=0}^n \binom{n}{m} n^{n-m} 1^m = \binom{n}{0} n^n + \binom{n}{1} n^{n-1} + \sum_{m=2}^n \binom{n}{m} n^{n-m}.$$

Essendo $\binom{n}{0} = 1$, e $\binom{n}{1} = n$, ed essendo tutti i termini della sommatoria non negativi, si ha

$$(n + 1)^n \geq n^n + n n^{n-1} = 2n^n,$$

che è quanto si voleva dimostrare. 4) $n_0 = 6$ ($6^6 = 46656$ e $2^6 6! = 46080$, mentre $5^5 = 3125$ e $2^5 5! = 3840$).

3. I razionali.

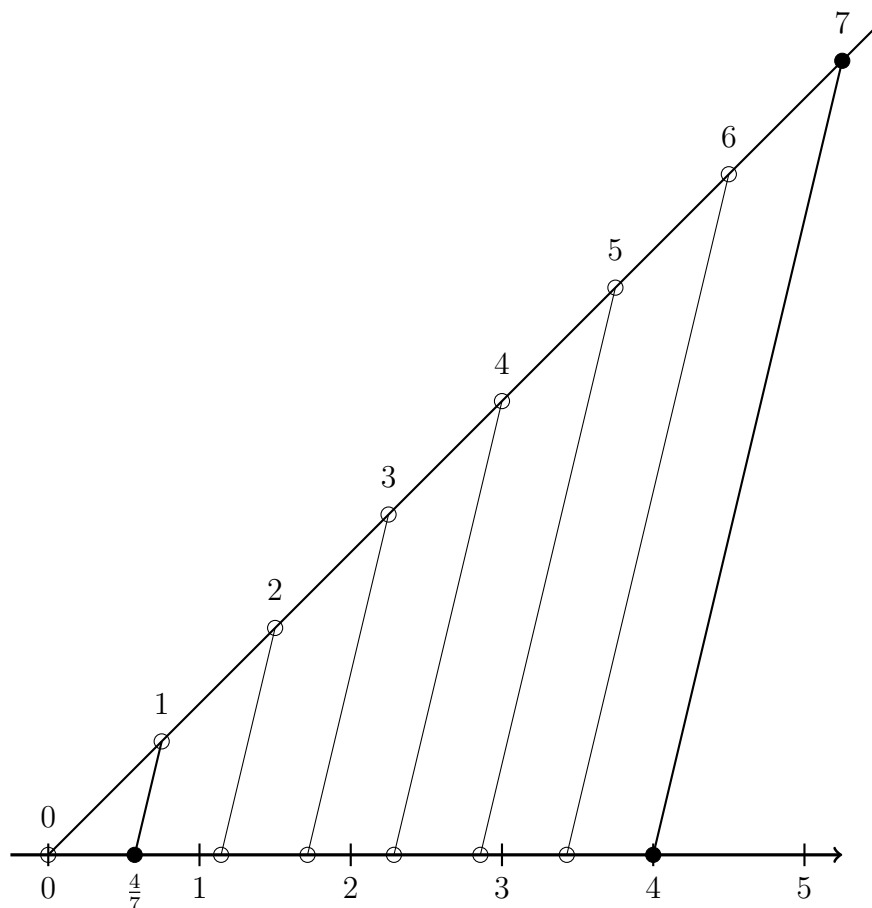
Come abbiamo visto, i numeri interi vengono introdotti per poter ben definire la sottrazione tra due numeri naturali. Una volta introdotti, ad ogni numero intero n viene associato un inverso $(-n)$, con la proprietà che $n + (-n) = 0$, l'elemento neutro dell'addizione. Ora, tra numeri interi abbiamo definito un'altra operazione, il prodotto, che ha come elemento neutro il numero 1. Dato un numero intero n , esiste un numero intero m che ne sia l'inverso rispetto al prodotto, cioè tale che $n \cdot m = 1$? A meno che il numero n non sia ± 1 , la risposta è **no**. Se, ad esempio, $n = 2$, non esiste alcun numero intero m tale che $2 \cdot m = 1$. Ed infatti, se m fosse positivo, sarebbe $2 \cdot m \geq 2 > 1$, se m fosse negativo, allora $2 \cdot m \leq -2 < 1$, mentre se $m = 0$ si ha $2 \cdot m = 0 < 1$. Lo stesso ragionamento si può ripetere per n qualsiasi (diverso da ± 1). In altre parole, l'operazione di inversione rispetto al prodotto non è ben definita all'interno di \mathbb{Z} .

A questo punto, scartata l'idea di lasciare tutto così com'è, anche perché alle volte è necessario "dividere" (ovvero "fare le parti"), ragioniamo come abbiamo fatto nel passaggio da \mathbb{N} a \mathbb{Z} : aggiungiamo tutte le quantità mancanti. Quello che otteniamo è l'insieme dei numeri **razionali**:

$$(3.1) \quad \mathbb{Q} = \left\{ \frac{p}{q}, p \in \mathbb{Z}, q \in \mathbb{N} \setminus \{0\} \right\}.$$

D'accordo, adesso abbiamo definito \mathbb{Q} , ma cosa vuol dire $\frac{p}{q}$? Che "numero" è? Per capirlo, torniamo alla rappresentazione di \mathbb{N} sulla retta. Se, per il momento, supponiamo $p \geq 1$, possiamo rappresentare p come un punto sulla retta, a distanza p dall'origine (nel senso che per trovare p dobbiamo effettuare p salti partendo da 0). Analogamente, q è un altro punto sulla retta. Per comodità, lo rappresentiamo su una seconda retta, sghemba rispetto alla prima, con l'origine in comune. Congiungiamo ora il punto p con

il punto q , e tracciamo, dal punto 1 sulla retta che contiene il punto q , la parallela alla retta che congiunge p con q . Questa retta taglierà la retta che contiene p in un punto, che per noi è $\frac{p}{q}$. D'ora in poi il numero razionale $\frac{p}{q}$ sarà il punto così costruito sulla retta.



$p = 4$ e $q = 7$.

Lo stesso ragionamento si può ripetere nel caso in cui p sia negativo.

Ricordando il Teorema di Talete⁽¹⁵⁾ è facile verificare (ad esempio dalla figura) che sommando q volte $\frac{p}{q}$ (ovvero, facendo q salti sulla retta di lunghezza $\frac{p}{q}$) si ottiene p . Come conseguenza, se $n \geq 1$, moltiplicando n per $\frac{1}{n}$ si ottiene 1, e quindi $\frac{1}{n}$ è l'inverso moltiplicativo di n , mentre l'inverso moltiplicativo di $-n$ è $\frac{-1}{n}$ (per mantenere la convenzione che vuole denominatori non negativi). Un modo alternativo di indicare $\frac{1}{n}$ è scrivere n^{-1} . In questo modo si conserva la regola dei prodotti di potenze con la stessa base: $n^1 \cdot n^{-1} = n^{1-1} = n^0 = 1$. Analogamente, n^{-m} è l'inverso moltiplicativo di n^m , ovvero $\frac{1}{n^m}$ (ovviamente, deve essere $n \neq 0$).

Dati due numeri razionali $\frac{p}{q}$ e $\frac{r}{s}$, si può stabilire sempre se uno dei due è maggiore dell'altro:

$$\frac{p}{q} \geq \frac{r}{s} \iff p \cdot s \geq r \cdot q.$$

⁽¹⁵⁾Un fascio di rette parallele tagliato da due trasversali stacca su di esse segmenti a due a due proporzionali.

Dati due numeri razionali, sono ben definiti sia la somma, che il prodotto:

$$\frac{p}{q} + \frac{r}{s} = \frac{p \cdot s + r \cdot q}{qs}, \quad \frac{p}{q} \cdot \frac{r}{s} = \frac{p \cdot r}{q \cdot s}.$$

L'inverso additivo di $\frac{p}{q}$ è $\frac{-p}{q}$, mentre l'inverso moltiplicativo di $\frac{p}{q}$ (con $p \neq 0$) è $\frac{q}{p}$ (se $p > 0$), oppure $\frac{-q}{-p}$ (se $p < 0$).

L'insieme degli interi \mathbb{Z} è un sottoinsieme di \mathbb{Q} ; infatti, se m è un intero, allora è facile vedere che la costruzione geometrica precedente restituisce m se si calcola il numero razionale $\frac{m}{1}$.

Abbiamo dunque definito i razionali rifacendoci alla loro rappresentazione sulla retta; una tale definizione, anche se ci permette di “misurare” i razionali, è però molto poco pratica per lavorare: non è immaginabile dover prendere una riga ed un compasso tutte le volte che è necessario fare un'addizione. In altre parole, vogliamo un modo più comodo, possibilmente svincolato dalla rappresentazione geometrica dei razionali, per “scrivere” $\frac{p}{q}$. Tale modo è — ovviamente — la rappresentazione decimale.

Come possiamo scrivere la rappresentazione decimale di $\frac{p}{q}$, partendo da p e q ? Iniziamo a trattare il caso $p \geq 0$. Innanzitutto, è sempre possibile scrivere p come $p = m \cdot q + r_1$ con m in \mathbb{N} , r_1 in \mathbb{N} e $0 \leq r_1 < q$. Ad esempio,

$$10 = 3 \cdot 3 + 1, \quad 23 = 0 \cdot 43 + 23, \quad 31 = 5 \cdot 6 + 1.$$

Il numero intero m viene detto **parte intera** di $\frac{p}{q}$, e viene di solito indicato nel seguente modo:

$$m = \left[\frac{p}{q} \right].$$

Dividendo la relazione $p = m \cdot q + r_1$ per q si ottiene $\frac{p}{q} = m + \frac{r_1}{q}$. Essendo $0 \leq \frac{r_1}{q} < 1$, ne segue che m è il più grande intero minore di $\frac{p}{q}$. Esaminiamo ora la frazione $\frac{r_1}{q}$; se $r_1 = 0$, abbiamo finito: $\frac{p}{q} = m$ è già un numero intero, e con gli interi sappiamo lavorare. Se $r_1 > 0$, consideriamo il numero intero $10 \cdot r_1$ ⁽¹⁶⁾, e scriviamo anch'esso nella forma $10 \cdot r_1 = d_1 \cdot q + r_2$, con d_1 e r_2 in \mathbb{N} , e $0 \leq r_2 < q$. Essendo $0 < 10 \cdot r_1 \leq 10 \cdot (q - 1) < 10 \cdot q$, d_1 sarà un numero compreso tra 0 e 9: ovvero, una cifra decimale. Se $r_2 = 0$, ci fermiamo, mentre se $r_2 > 0$ andiamo avanti, scriviamo $10 \cdot r_2 = d_2 \cdot q + r_3$, ed otteniamo una seconda cifra compresa tra 0 e 9, d_2 . Proseguendo, a meno di non trovare uno degli $r_i = 0$, otteniamo d_3, d_4, d_5 e così via.

E così via... Se così fosse, avremmo peggiorato la situazione: se nessuno degli r_i è zero, staremmo associando ad un numero razionale $\frac{p}{q}$ un numero intero m , e poi “infinite” cifre d_1, d_2, d_3, \dots , tutte comprese tra 0 e 9. Il che vuol dire che, invece di portarci appresso una riga ed un compasso, dovremmo portarci appresso molta, molta carta per scrivere tutte le cifre d_i ⁽¹⁷⁾. Fortunatamente, non c'è bisogno di calcolare infinite cifre d_i per “conoscere” il numero razionale $\frac{p}{q}$, perché le cifre d_i si ripetono da un certo punto in poi. Vale a dire che — se nessuno degli r_i è zero — esistono due indici j e k (dipendenti da p e q) tali che $d_i = d_{i+j}$ per ogni $i \geq k$. Perché? Perché le cifre d_i sono determinate dagli r_i : ricordiamo che d_i non è altro che la parte intera di $\frac{10 \cdot r_i}{q}$; pertanto, se esistono j e k tali che $r_{j+k} = r_k$, si avrà sia $d_{j+k} = d_k$ che $r_{j+k+1} = r_{k+1}$, e

⁽¹⁶⁾La scelta di $10 \cdot r_1$ ci porta ai numeri decimali. Se avessimo scelto $2 \cdot r_1$ avremmo avuto la rappresentazione binaria, eccetera.

⁽¹⁷⁾Per non parlare del tempo necessario a calcolare la somma di due razionali...

quindi le cifre decimali inizieranno a ripetersi nello stesso ordine. Siccome, qualsiasi sia i , si ha $0 < r_i \leq q - 1$, ne segue che i “resti possibili” sono al massimo $q - 1$; pertanto, dopo aver calcolato al più $q - 1$ cifre decimali, uno dei resti “dovrà” ripetersi, e con loro le cifre decimali di $\frac{p}{q}$.

Possiamo allora affermare che, dato un qualsiasi numero razionale $\frac{p}{q}$ positivo, esistono un intero m ed un numero finito d_1, d_2, \dots, d_j di cifre comprese tra 0 e 9 tali che

$$\frac{p}{q} = m, d_1 d_2 d_3 \dots d_{j-1} d_j 0$$

oppure

$$\frac{p}{q} = m, d_1 d_2 d_3 \dots d_{k-1} d_k d_{k+1} \dots d_{j-1} d_j d_k d_{k+1} \dots \text{ per qualche } k \geq 1.$$

Nel caso in cui le cifre d_k, \dots, d_j si ripetono, scriviamo

$$\frac{p}{q} = m, d_1 d_2 d_3 \dots d_{k-1} \overline{d_k d_{k+1} \dots d_{j-1} d_j}.$$

Le cifre $d_1 \dots d_{k-1}$ si dicono **antiperiodo**, mentre le cifre $d_k \dots d_j$ si dicono **periodo** del numero razionale $\frac{p}{q}$. Se lo sviluppo è della forma $m, d_1 \dots d_j 0$, possiamo considerarlo periodico con periodo $\overline{0}$:

$$m, d_1 \dots d_j 0 = m, d_1 \dots d_j \overline{0},$$

anche se in genere il periodo $\overline{0}$ si omette.

Ad esempio,

$$\frac{1}{4} = 0, 25\overline{0} = 0, 25, \quad \frac{1}{3} = 0, \overline{3}, \quad \frac{1}{7} = 0, \overline{142857},$$

$$\frac{230}{170} = 1, \overline{35882352941176470}, \quad \frac{1}{19} = 0, \overline{526315789473684210}.$$

Se il numero razionale considerato è negativo, ovvero se $p < 0$, la rappresentazione decimale di $\frac{p}{q}$ si ottiene scrivendo la rappresentazione decimale di $-\frac{p}{q}$ (o, meglio, di $\frac{-p}{q}$), e “cambiando segno”. Ad esempio, $-\frac{1}{3} = -0, \overline{3}$.

A questo punto le operazioni con i razionali (somma, prodotto, confronto) si possono eseguire “manipolando” opportunamente le cifre decimali; ad esempio, si può stabilire quale tra due numeri razionali sia il più grande.

DEFINIZIONE 3.1. Dati due numeri razionali non negativi $\frac{p}{q}$ e $\frac{r}{s}$ con rappresentazione decimale

$$\frac{p}{q} = m, d_1 d_2 d_3 \dots d_n \dots, \quad \frac{r}{s} = m', d'_1 d'_2 d'_3 \dots d'_n \dots,$$

diremo che $\frac{p}{q} > \frac{r}{s}$ se $m > m'$, oppure se $m = m'$ ed esiste $j \geq 1$ tale che $d_i = d'_i$ per ogni $i < j$ ma $d_j > d'_j$. Se $\frac{p}{q} > 0$ e $\frac{r}{s} < 0$, allora sarà $\frac{p}{q} > \frac{r}{s}$ (indipendentemente dalla rappresentazione decimale: è il “segno” a decidere l’ordine). Se sia $\frac{p}{q}$ che $\frac{r}{s}$ sono negativi, avremo

$$\frac{p}{q} = -m, d_1 d_2 d_3 \dots d_n \dots, \quad \frac{r}{s} = -m', d'_1 d'_2 d'_3 \dots d'_n \dots,$$

e diremo che $\frac{p}{q} > \frac{r}{s}$ se $m < m'$, oppure se $m = m'$ ed esiste $j \geq 1$ tale che $d_i = d'_i$ per ogni $i < j$ ma $d_j < d'_j$.

Ad esempio, $1,2345\bar{6} > 1,234\bar{5}$ e $-1,234\bar{5} > -1,2345\bar{6}$. L'ordinamento così definito è compatibile sia con l'ordinamento definito precedentemente, sia con l'ordinamento stabilito dalla rappresentazione dei razionali come punti sulla retta.

Supponiamo ora di aver preso un razionale $\frac{p}{q}$, di aver eseguito il procedimento precedente, e di aver ottenuto la rappresentazione decimale $\frac{p}{q} = 0,\bar{9}$. È possibile arrivare ad una rappresentazione siffatta, o abbiamo commesso un errore? Se abbiamo ottenuto 0 come parte intera, questo vuol dire che $p = 0 \cdot q + r_1$, e quindi $r_1 = p$. Siccome la prima cifra decimale è 9, allora $10 \cdot p = 9 \cdot q + r_2$; essendo anche la seconda cifra decimale 9, abbiamo $10 \cdot r_2 = 9 \cdot q + r_3$, e quindi

$$100 \cdot p = 90 \cdot q + 10 \cdot r_2 = 90 \cdot q + 9 \cdot q + r_3 = 99 \cdot q + r_3.$$

Continuando, troviamo

$$1000 \cdot p = 999 \cdot q + r_4, \quad 10000 \cdot p = 9999 \cdot q + r_5,$$

e, in generale,

$$(3.2) \quad 10^n \cdot p = (10^n - 1) \cdot q + r_{n+1}, \quad \forall n \geq 1.$$

ESERCIZIO 3.2. Sapendo che, per ogni $n \geq 1$, $10 \cdot r_n = 9 \cdot q + r_{n+1}$, dimostrare per induzione la (3.2).

Risposta 3.2: Se $n = 1$, la (3.2) è vera. Supponiamo che sia vera per un certo valore di n . Essendo $10 \cdot r_{n+1} = 9 \cdot q + r_{n+2}$ per ipotesi, si ha

$$\begin{aligned} 10^{n+1} \cdot p &= 10 \cdot (10^n - 1) \cdot q + 10 \cdot r_{n+1} \\ &= 10^{n+1} \cdot q - 10 \cdot q + 9 \cdot q + r_{n+2} \\ &= (10^{n+1} - 1) \cdot q + r_{n+2}, \end{aligned}$$

come volevasi dimostrare.

Dividiamo ora la (3.2) per 10^n ; otteniamo

$$p = q - \frac{q}{10^n} + \frac{r_{n+1}}{10^n} \quad \Rightarrow \quad p - q = \frac{r_{n+1}}{10^n} - \frac{q}{10^n}.$$

Prendiamo ora n tale che $10^n > q$ (ricordiamo che q è fissato una volta per tutte); una tale scelta di n è semplice: basta infatti contare le cifre di q — siano esse m — e scegliere $n = m$. Siccome $10^n > q$, allora $\frac{q}{10^n} < 1$; d'altra parte, essendo $0 < r_{n+1} \leq q - 1$, si ha $\frac{r_{n+1}}{10^n} < 1$. Di conseguenza

$$-1 < \frac{r_{n+1}}{10^n} - \frac{q}{10^n} < 1.$$

Essendo però $p - q = \frac{r_{n+1}}{10^n} - \frac{q}{10^n}$ un numero intero, l'unica possibilità è $p - q = 0$, ovvero $p = q$. Ma se $p = q$, il procedimento per il calcolo dell'espansione decimale di $\frac{p}{q}$ dà 1, non $0,\bar{9}$! In altre parole, scrivendo lo sviluppo decimale di un qualsiasi razionale $\frac{p}{q}$, non troveremo mai $0,\bar{9}$; in realtà, non troveremo mai uno sviluppo della forma $m, d_1 \dots d_j \bar{9}$. Alternativamente, non tutte le "stringhe" della forma $m, d_1 d_2 d_3 \dots d_{k-1} d_k d_{k+1} \dots d_{j-1} d_j d_k d_{k+1} \dots$ sono ottenibili tramite lo sviluppo di un numero razionale: mancano quelle di periodo "9".

Siano ora $\frac{p}{q} = \frac{1}{3} = 0,\bar{3}$ e $\frac{r}{s} = \frac{2}{3} = 0,\bar{6}$. Se calcoliamo $\frac{p}{q} + \frac{r}{s} = \frac{1}{3} + \frac{2}{3}$, troviamo ovviamente 1, ma se sommiamo gli sviluppi decimali troviamo $0,\bar{9}$, ed abbiamo appena visto che questo non è uno sviluppo "ammissibile". Come fare? Ci serve un meccanismo che ci permetta, dato lo sviluppo decimale (dato **anche** uno sviluppo decimale

“scorretto” come $0, \overline{9}$), di risalire al razionale che lo ha generato (o ad un razionale “compatibile” con le operazioni che hanno generato lo sviluppo scorretto). Il metodo è noto, ed è il seguente: sia $m, d_1 d_2 d_3 \dots d_{k-1} \overline{d_k d_{k+1} \dots d_{j-1} d_j}$ uno sviluppo decimale periodico. Il numeratore della frazione $\frac{p}{q}$ che “corrisponde” allo sviluppo dato si ottiene considerando il numero intero

$$p = md_1 d_2 d_3 \dots d_{k-1} d_k d_{k+1} \dots d_{j-1} d_j - md_1 d_2 d_3 \dots d_{k-1},$$

mentre il denominatore q è il numero intero

$$q = \underbrace{99 \dots 99}_{j-k \text{ volte}} \quad \underbrace{00 \dots 00}_{k-1 \text{ volte}} \quad .$$

Ad esempio,

$$0, \overline{3} = \frac{3-0}{9} = \frac{3}{9} = \frac{1}{3}, \quad 0, \overline{142857} = \frac{142857-0}{999999} = \frac{142857}{999999} = \frac{1}{7},$$

e

$$1, \overline{16} = \frac{116-11}{90} = \frac{105}{90} = \frac{7}{6}, \quad 0, \overline{9} = \frac{9-0}{9} = \frac{9}{9} = 1.$$

In generale, se lo sviluppo decimale è $m, d_1 \dots d_k \overline{9}$, con $d_k < 9$, allora la procedura appena descritta costruisce il numero razionale

$$\frac{p}{q} = \frac{md_1 \dots (d_k + 1)}{10 \dots 0} = m, d_1 \dots (d_k + 1),$$

dove al denominatore sono presenti esattamente k zeri. Infatti,

$$\begin{aligned} m, d_1 \dots d_k \overline{9} &= \frac{md_1 \dots d_k 9 - md_1 \dots d_k}{90 \dots 0} \\ &= \frac{10 \cdot md_1 \dots d_k - md_1 \dots d_k + 9}{90 \dots 0} \\ &= \frac{9 \cdot (md_1 \dots d_k + 1)}{90 \dots 0} = \frac{md_1 \dots (d_k + 1)}{10 \dots 0}. \end{aligned}$$

Ad esempio, $1, 234\overline{9} = 1, 235$.

Si noti che la procedura di “ricostruzione” di $\frac{p}{q}$ a partire da uno sviluppo decimale funziona anche se lo sviluppo è errato: ad esempio,

$$0, 123\overline{23} = \frac{12323-123}{9900} = \frac{12200}{9900} = \frac{122}{99}, \quad 0, \overline{123} = \frac{123-1}{99} = \frac{122}{99}.$$

In definitiva, ad ogni numero razionale della forma $\frac{p}{q}$ può essere associato uno sviluppo decimale periodico “ben formato” (ovvero, con un periodo diverso da $\overline{9}$), mentre ad un qualsiasi sviluppo decimale periodico può essere associato un numero razionale, che però non è unico: a $1 = 1, \overline{0}$ e a $0, \overline{9}$ corrisponde lo stesso numero, che è però compatibile con le operazioni algebriche eseguibili con i razionali: ad esempio $1 = 3 \cdot \frac{1}{3} = 3 \cdot 0, \overline{3} = 0, \overline{9}$. Che cosa è la rappresentazione decimale di un razionale della forma $\frac{p}{q}$? È un oggetto che racchiude in sé le proprietà di $\frac{p}{q}$ e che “misura” la distanza del punto $\frac{p}{q}$, costruito come spiegato all’inizio, dall’origine (così come un qualsiasi numero naturale misura la distanza dall’origine). E la misura in modo tale che se $\frac{r}{s}$ è un altro punto sulla retta, a sinistra di $\frac{p}{q}$, allora il numero ottenuto rappresentando $\frac{p}{q}$ in forma decimale è minore del numero ottenuto rappresentando $\frac{r}{s}$.

Osserviamo infine che mentre esiste il più piccolo naturale maggiore di zero (ed è 1), non esiste il primo numero razionale maggiore di zero. Se, infatti, supponiamo che

esista, sia esso $\frac{p}{q}$; ebbene, considerando $\frac{p}{q+1}$ si ottiene un numero razionale positivo (perché p è non negativo), ma minore strettamente di $\frac{p}{q}$ dal momento che $p \cdot (q+1) = p \cdot q + p > p \cdot q$. Se pensassimo che questo nuovo numero razionale fosse il più piccolo maggiore di zero, potremmo sempre ottenerne uno più piccolo prendendo $\frac{p}{q+2}$, e così via...⁽¹⁸⁾. Non solo: dati comunque due razionali diversi $\frac{p}{q} < \frac{r}{s}$, esiste sempre un razionale compreso tra i due:

$$\frac{p}{q} < \frac{\frac{p}{q} + \frac{r}{s}}{2} = \frac{p \cdot s + r \cdot q}{2qs} < \frac{r}{s},$$

come si verifica facilmente. Questo vuol dire che se tentassimo una rappresentazione grafica dei razionali come abbiamo fatto con gli interi, dovremmo avere a disposizione un computer (o un pittore) in grado di disegnare linee estremamente sottili: tali che, tra due di esse, se ne deve poter mettere sempre un'altra. In altre parole, continuando ad ingrandire la nostra rappresentazione, non arriveremo mai ad una situazione simile a quella dei naturali, con dei "salti" discreti.

A questo punto, ci poniamo una domanda: prendiamo, su una retta sulla quale siano stati fissati 0 ed 1, un punto qualsiasi: esistono p e q tali che la procedura geometrica spiegata in precedenza "costruisca" questo punto? Altrimenti detto: presa una lunghezza qualsiasi sulla retta, questa lunghezza è rappresentata da un razionale?

ESERCIZIO 3.3. Prima di andare avanti, provate a rispondere alla domanda precedente, ragionando non sui razionali come "punti sulla retta", ma come sviluppi decimali: quale numero razionale (se ne esiste uno) corrisponde allo sviluppo

$$0,10100100010000 \dots 1 \overbrace{0 \dots 0}^{n \text{ zeri}} 1 \overbrace{0 \dots 0}^{n+1 \text{ zeri}} 1 \dots,$$

o allo sviluppo

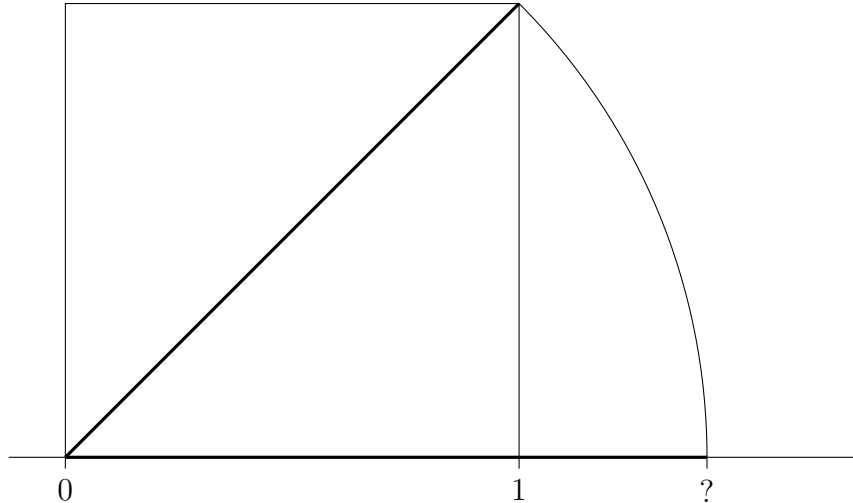
$$0,12345678910111213141516 \dots n(n+1)(n+2) \dots ?$$

Risposta 3.3: Dato che ogni razionale dà luogo ad uno sviluppo decimale periodico, e non essendo i due sviluppi precedenti periodici, se ne deduce che nessun razionale li genera. Cosa sono, allora?

4. Non tutto ciò che è reale è razionale.

Costruiamo un quadrato di lato 1, con uno dei lati coincidente con il segmento unitario dell'asse dei numeri razionali, e disegniamone la diagonale. Riportiamo la diagonale sull'asse dei numeri, e chiediamoci se l'estremo destro del segmento cada o meno su di un razionale.

⁽¹⁸⁾Si, questa volta è lecito dire "e così via ..."!



Quanto è lunga la diagonale? Usando il teorema di Pitagora, sappiamo che il quadrato costruito sulla diagonale ha come area la somma delle aree dei quadrati costruiti sui due lati del quadrato. Essendo il quadrato unitario, il quadrato costruito sulla diagonale ha area 2. Pertanto, la lunghezza della diagonale è tale che, elevata al quadrato, vale 2. Supponiamo ora che tale lunghezza sia un numero razionale $\frac{p}{q}$. Pertanto, $\frac{p^2}{q^2} = 2$, da cui segue che $p^2 = 2q^2$, ovvero che p^2 è pari. Dal momento che il prodotto tra due numeri dispari è sempre un numero dispari, dall'essere p^2 pari segue che anche p è pari, e quindi $p = 2p_1$, con p_1 un numero intero strettamente minore di p . Sostituendo nella relazione $p^2 = 2q^2$, si ottiene $4p_1^2 = 2q^2$ e, dividendo per 2, $q^2 = 2p_1^2$. Ripetendo il ragionamento precedente, se ne deduce che q è pari, e quindi $q = 2q_1$, con q_1 numero intero strettamente minore di q . Ri-sostituendo, otteniamo (dopo aver diviso per due) che $p_1^2 = 2q_1^2$, ovvero che $\frac{p_1^2}{q_1^2} = 2$. Ricapitolando, dati p e q interi e tali che $\frac{p^2}{q^2} = 2$, abbiamo trovato $1 \leq p_1 < p$ e $q_1 < q$ tali che $\frac{p_1^2}{q_1^2} = 2$. È evidente che possiamo allora ripetere il ragionamento precedente, e determinare altri due numeri interi $1 \leq p_2 < p_1$ e $q_2 < q_1$ tali che $\frac{p_2^2}{q_2^2} = 2$, e poi $1 \leq p_3 < p_2$ e $q_3 < q_2$ tali che $\frac{p_3^2}{q_3^2} = 2$ e così via. Il difetto del ragionamento consiste nel fatto che di numeri naturali p_i che verificano la disuguaglianza $1 \leq p_i < p$ ne esistono esattamente $p - 1$, e che quindi non è possibile continuare a far decrescere indefinitamente i numeratori p_i : ad un certo punto uno dei p_i non può esistere. Il fatto che non esista discende dall'aver supposto che esistano p e q tali che $\frac{p^2}{q^2} = 2$, ovvero che la lunghezza della diagonale del quadrato di lato 1 sia un numero razionale.

ESERCIZIO 4.1. Sia m un numero naturale che non sia un quadrato perfetto (ovvero tale che non esista n in \mathbb{N} per il quale $n^2 = m$). Dimostrare che non esistono p e q in \mathbb{N} tali che $\frac{p^2}{q^2} = m$ (se $m = n^2$ è un quadrato perfetto, è sufficiente scegliere $p = n$ e $q = 1$). Suggerimento: fattorizzare m , ed osservare che esiste almeno uno dei fattori primi di m che compare con un esponente dispari; dimostrare successivamente che p è multiplo di questo fattore, da cui...

Risposta 4.1: Scriviamo $m = m_0 \cdot m_0^{r_0} \cdot m_1^{r_1} \cdot \dots \cdot m_k^{r_k}$, con m_i primo per ogni i , $m_i \neq m_j$ per $i \neq j$, r_0 pari e gli r_i interi qualsiasi. Dimostriamo che p è divisibile per m_0 . Scriviamo $p = km_0 + r$, con $0 \leq r < m_0$. Allora $p^2 = k^2m_0^2 + 2km_0r + r^2 = m_0 \cdot m_0^{r_0} \cdot m_1^{r_1} \cdot \dots \cdot m_k^{r_k} q^2$,

da cui segue che

$$m_0(k^2 m_0 + 2kr - m_0^{r_0} \cdot m_1^{r_1} \cdot \dots \cdot m_k^{r_k} q^2) + r^2 = 0.$$

Dividendo per m_0 , e cambiando segno si ottiene

$$\frac{r^2}{m_0} = m_0^{r_0} \cdot m_1^{r_1} \cdot \dots \cdot m_k^{r_k} q^2 - k^2 m_0 - 2kr,$$

e quindi $\frac{r^2}{m_0}$ è un intero. Ne segue che m_0 divide r^2 , ed essendo m_0 primo, m_0 divide r , il che è possibile, essendo $m_0 > r$, se e solo se $r = 0$. Ma allora $p = km_0$. Sostituendo, si trova

$$k^2 m_0^2 = m_0 \cdot m_0^{r_0} \cdot m_1^{r_1} \cdot \dots \cdot m_k^{r_k} q^2.$$

Osservando che k^2 ha solo potenze pari di m_0 (o non ne ha nessuna), e a destra essendo presente una potenza dispari di m_0 , si ottiene

$$m_1^{r_1} \cdot \dots \cdot m_k^{r_k} q^2 = m_0 m_0^{s_0} k_1^2,$$

da cui segue (analogamente a prima, ed usando il fatto che gli m_i sono a due a due distinti), che anche q deve essere un multiplo di m_0 .

Pertanto, non tutto quello che possiamo “costruire” può essere misurato con i numeri razionali: alcune quantità, come la lunghezza della diagonale di un quadrato di lato 1, non possono essere rappresentate mediante un numero razionale.

Dal momento, poi, che ogni razionale della forma $\frac{p}{q}$ genera una rappresentazione decimale periodica, è chiaro che se vogliamo rappresentare come numero la lunghezza della diagonale del quadrato di lato 1, dovremo ricorrere ad una lista di decimali che “non si ripetono”; ad esempio, a qualcosa del tipo

$$1,41421356237309504880168872420969807856967187537694\dots$$

Vedremo nel prossimo paragrafo come risolvere il problema.

ESERCIZIO 4.2. Sapendo che la diagonale del cubo è un numero il cui quadrato è 3, quale diagonale ha una lunghezza il cui quadrato è n ?

Risposta 4.2: Quella del cubo n dimensionale, qualsiasi cosa esso sia...

5. I numeri reali.

Riassumiamo quello che abbiamo scoperto sul concetto di numero. Siamo partiti dai numeri naturali, che non abbiamo veramente definito ma piuttosto richiamato alla mente basandoci sulla nostra intuizione degli insiemi con un numero finito di oggetti. Su \mathbb{N} abbiamo considerato le operazioni di somma e prodotto, che non ci costringono a uscire da \mathbb{N} ; se invece vogliamo calcolare la differenza di due naturali, siamo portati naturalmente a concepire un insieme più grande di numeri interi relativi \mathbb{Z} . Somma, prodotto e differenza hanno senso su \mathbb{Z} e non ci costringono a uscire da \mathbb{Z} . Se però vogliamo dare un significato al rapporto di due numeri interi, dobbiamo estendere ulteriormente il nostro insieme di numeri e arriviamo ad immaginare l'insieme \mathbb{Q} dei numeri razionali; possiamo estendere a \mathbb{Q} somma, prodotto e differenza, e in più possiamo definire il rapporto di due numeri razionali qualunque (purché il denominatore non sia nullo) e quindi in particolare anche di due numeri relativi.

Infine abbiamo descritto due metodi per visualizzare in modo unitario e studiare le proprietà degli insiemi di numeri che stiamo costruendo:

- metodo *geometrico*: abbiamo identificato \mathbb{N} , \mathbb{Z} , \mathbb{Q} con sottoinsiemi di una retta della geometria euclidea, cioè con le *lunghezze dei segmenti* dati dalle costruzioni della geometria euclidea. Oltre a fornirci una potente intuizione su come sono fatti i nostri insiemi, questo punto di vista ci ha rivelato che l'insieme \mathbb{Q} , per quanto sia “molto fitto” nella retta, non è ancora sufficiente per descrivere tutte le lunghezze dei segmenti che compaiono nelle costruzioni geometriche, persino quelle semplicissime come la diagonale di un quadrato. Questo probabilmente è il punto di vista più antico sui numeri (quello dei geometri greci).
- metodo *decimale*: il grosso vantaggio di questo metodo è che ci consente di maneggiare i numeri concretamente, operare su di essi e confrontarli, tutte cose piuttosto complicate da fare usando il metodo geometrico. Anche in questo caso, si vede chiaramente che manca qualcosa: i numeri razionali producono sviluppi decimali *periodici*, e non abbiamo difficoltà a immaginare la possibilità di sviluppi decimali *aperiodici*. Il punto di vista decimale sui numeri è molto più recente: fu introdotto da Simon Stevin nel 1585.

Siamo così portati in modo naturale a concepire un insieme ancora più grande di numeri che chiameremo *l'insieme dei numeri reali* \mathbb{R} . Potremmo definire ad esempio \mathbb{R} geometricamente come

$$(5.1) \quad \mathbb{R} = \{\text{punti sulla retta}\},$$

oppure potremmo dire che \mathbb{R} è l'insieme di tutti i numeri decimali

$$x = a, b_1 b_2 b_3 \dots b_n \dots,$$

dove la sequenza di cifre $b_1 b_2 b_3 \dots$ può essere periodica da un certo punto in poi, nel qual caso otteniamo un numero razionale, ma può anche non essere mai periodica, nel qual caso otteniamo quello che si chiama un *numero irrazionale*.

La prima definizione è molto suggestiva (\mathbb{R} è semplicemente una retta della geometria euclidea!) ma come sappiamo bene il metodo geometrico non ci consente di operare agevolmente sui numeri.

La seconda definizione è abbastanza soddisfacente, e si può tranquillamente utilizzare per fare analisi senza preoccuparsi troppo del significato esatto della nostra costruzione. Tuttavia, un vero matematico non può essere completamente soddisfatto neanche in questo caso. Cos'è *esattamente* uno sviluppo decimale non periodico? Cosa vuol dire *confrontare* o *sommare* due sviluppi decimali? Se gli sviluppi sono periodici, dobbiamo eseguire solo un numero finito di operazioni quindi il significato è abbastanza evidente, ma per sviluppi non periodici dovremmo eseguire un numero infinito di operazioni e c'è qualcosa di poco chiaro.

Dopo profonde riflessioni sulla natura dei numeri (anzi, di tutta la matematica), tra la fine dell'ottocento e la prima metà del novecento i matematici si sono resi conto che per trovare una via d'uscita soddisfacente bisogna seguire un terzo metodo: quello *assiomatico* (per essere proprio sinceri, non si trattava di una completa novità dato che anche Euclide lo utilizzava...). In questo punto di vista non esistono più verità assolute da scoprire; le dimostrazioni matematiche sono semplicemente delle deduzioni che portano da un'affermazione P a un'altra affermazione Q . A sua volta, l'affermazione

P è stata dedotta da un'affermazione precedente N , N da un'affermazione precedente M , eccetera. Naturalmente, non possiamo andare indietro all'infinito: serve un punto di partenza, abbiamo bisogno di un certo numero di enunciati di base sui quali edificare tutta la teoria. Questi enunciati di base sono detti gli assiomi della teoria.

Il matematico deve concentrarsi solo sulla correttezza delle deduzioni, mentre chiedersi se gli assiomi siano “veri o falsi” diventa un'operazione esterna alla matematica, una riflessione sui fondamenti (e sulla psicologia umana, o sulla natura). Tutt'al più, dagli assiomi si pretende che siano ragionevoli, intuitivi, facili da usare, e condivisi dalla comunità matematica. Però nulla vieta di scegliere assiomi diversi e costruire una teoria diversa. La teoria comunemente utilizzata oggi per fare matematica è quella di Zermelo-Fraenkel (ZF), a cui solitamente si aggiunge l'assioma della scelta (ZFC).

Attenzione: anche se l'approccio assiomatico risolve completamente alcuni problemi, ne apre altri. Per esempio, come facciamo ad essere sicuri che gli assiomi che abbiamo scelto non siano in contraddizione fra di loro? Questo si chiama il problema della *consistenza* della teoria. Un altro problema è se gli assiomi che abbiamo scelto siano sufficienti per dimostrare tutte le affermazioni “interessanti”; questo (all'incirca) è il problema della *completezza* della teoria. A entrambe queste domande furono date risposte pressoché definitive da Kurt Gödel durante la prima metà del novecento.

Per i nostri scopi, partire proprio da zero (cioè dagli assiomi ZFC) sarebbe un pò esagerato, oltre che estremamente faticoso. Ci accontentiamo di partire da uno stadio più avanzato, ossia definiamo in modo assiomatico che cosa è l'insieme dei numeri reali. Vediamo in dettaglio quali sono gli assiomi.

\mathbb{R} è un'insieme su cui sono definiti i seguenti oggetti.

- (1) Una *addizione* $+$ che ad ogni coppia di elementi a, b in \mathbb{R} fa corrispondere un elemento $a + b$ in \mathbb{R} tale che:
 - (a) $a + b = b + a$ per tutti gli a e b in \mathbb{R} (commutativa);
 - (b) $a + (b + c) = (a + b) + c$ per tutti gli a, b e c in \mathbb{R} (associativa);
 - (c) esiste un unico elemento 0 in \mathbb{R} tale che $a + 0 = a$ per ogni a in \mathbb{R} (zero, l'elemento neutro per $+$);
 - (d) per ogni a in \mathbb{R} esiste un unico $-a$ in \mathbb{R} tale che $a + (-a) = 0$. Si scrive anche $a + (-b) = a - b$.
- (2) Una *moltiplicazione* \cdot che ad ogni coppia di elementi a e b in \mathbb{R} fa corrispondere un elemento $a \cdot b$ in \mathbb{R} (spesso scriveremo soltanto ab) tale che:
 - (a) $a \cdot b = b \cdot a$ per tutti gli a e b in \mathbb{R} (commutativa);
 - (b) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ per tutti gli a, b e c in \mathbb{R} (associativa);
 - (c) esiste un unico elemento 1 in \mathbb{R} tale che $a \cdot 1 = a$ per ogni a in \mathbb{R} (uno, l'elemento neutro per \cdot);
 - (d) per ogni a in \mathbb{R} diverso da zero esiste un unico a^{-1} in \mathbb{R} tale che $a \cdot a^{-1} = 1$. Si scrive anche $a^{-1} = 1/a$ e $a \cdot b^{-1} = a/b$;
 - (e) $a \cdot (b + c) = a \cdot b + a \cdot c$ per tutti gli a, b e c in \mathbb{R} (distributiva).
- (3) Un *ordinamento totale*, ossia una relazione \leq tra coppie di elementi di \mathbb{R} tale che per ogni a, b e c in \mathbb{R} si ha:
 - (a) se $a \leq b$ e $a \leq c$ allora $a \leq c$ (transitiva);
 - (b) se $a \leq b$ e $b \leq a$ allora $a = b$ (antisimmetrica);

- (c) si ha sempre $a \leq a$ (riflessiva);
- (d) per ogni coppia a e b in \mathbb{R} si ha $a \leq b$ oppure $b \leq a$ (totale);
- (e) se $a \leq b$ allora $a + c \leq b + c$ (cancellazione);
- (f) se $0 \leq a$ e $0 \leq b$, allora $0 \leq a \cdot b$.

Attenzione! manca un ultimo assioma che si chiama **assioma di completezza**. Però notiamo che se ci fermassimo qua, avremmo già un insieme che soddisfa gli assiomi (1),(2) e (3): l'insieme dei numeri razionali \mathbb{Q} . Un insieme di questo tipo si chiama un *corpo ordinato*. Ma come abbiamo notato in entrambi i modelli esaminati in precedenza, fra un razionale e l'altro esistono come delle "lacune", dei numeri mancanti. Possiamo immaginare di colmarle prendendo tutta la retta per il modello geometrico, o tutti i numeri decimali anche non periodici per il modello decimale. Ma se vogliamo definire rigorosamente l'insieme dei numeri reali, è necessario aggiungere alla lista un assioma che ci garantisca che queste lacune non esistono.

6. Max, min, sup e inf: la definizione di \mathbb{R}

Prima di tutto dobbiamo introdurre alcuni concetti di importanza fondamentale nel seguito. Notiamo che, partendo dalla relazione d'ordine \leq , gli altri simboli di disuguaglianza si possono definire così: $x \geq y$ significa $y \leq x$, $x < y$ significa $x \leq y$ e $x \neq y$, $x > y$ significa $y < x$.

DEFINIZIONE 6.1. Sia E un sottoinsieme di \mathbb{R} .

Diciamo che E ha **massimo** se esiste \bar{x} in \mathbb{R} tale che

- 1) per ogni y in E , si ha $y \leq \bar{x}$.
- 2) \bar{x} appartiene ad E ;

Il numero reale \bar{x} si indica anche $\max E$.

Diciamo che E ha **minimo** se esiste \underline{x} in \mathbb{R} tale che

- 3) per ogni y in E , si ha $\underline{x} \leq y$.
- 4) \underline{x} appartiene ad E ;

Il numero reale \underline{x} si indica anche $\min E$.

Di sottoinsiemi di \mathbb{R} che ammettono minimo ne abbiamo già incontrati: ad esempio qualunque sottoinsieme non vuoto di \mathbb{N} . Osserviamo — e ripetiamo — che se E ammette massimo (o minimo), allora tale massimo (o minimo) è unico. La proprietà 1) afferma che \bar{x} è più grande di tutti gli elementi di E (il che non implica che appartenga all'insieme); unita alla 2), si ottiene che \bar{x} è "il" più grande elemento di E . Analogamente, 3) e 4) affermano che \underline{x} è "il" più piccolo elemento di E .

ESERCIZIO 6.2. Trovare massimo e minimo dei seguenti insiemi:

$$\begin{aligned} E &= \{x \in \mathbb{R} : 0 \leq x \leq 1\}, \\ E &= \{x \in \mathbb{R} : x^2 \leq 1\}, \\ E &= \{x \in \mathbb{R} : (x^2 - 1)(x^2 - 4) \leq 0\}, \\ E &= \{x \in \mathbb{R} : x \in \mathbb{Q}, -3 \leq x \leq 4\}, \\ E &= \{x \in \mathbb{R} : 0 < x \leq 1\}. \end{aligned}$$

Una volta svolto l'esercizio, è evidente che c'è qualcosa che non va nell'ultimo insieme: se è facile vedere che 1 è il massimo dell'insieme dei numeri reali x tali che

$0 < x \leq 1$, è altrettanto evidente che tale insieme non ha minimo. Se lo avesse, sarebbe per definizione il più piccolo numero reale maggiore di zero, ma un tale numero non esiste: se pensassimo che \underline{x} fosse il più piccolo reale maggiore di zero, potremmo considerare $\frac{\underline{x}}{2}$, che è ancora maggiore di zero, ma minore di \underline{x} , e così via. Pertanto, a differenza dei naturali, e come i razionali, non tutti i sottoinsiemi di \mathbb{R} ammettono minimo: aggiungendo a \mathbb{Q} altri “oggetti”, non abbiamo migliorato la situazione.

ESERCIZIO 6.3. Trovare un sottoinsieme E di \mathbb{R} , contenuto in $\{x \in \mathbb{R} : 0 \leq x \leq 1\}$, che non abbia massimo, ed uno che non abbia né massimo né minimo.

Abbiamo dunque visto che non tutti i sottoinsiemi di \mathbb{R} ammettono massimo e (o) minimo. Torniamo per un momento a considerare l'insieme senza minimo $E = \{x \in \mathbb{R} : 0 < x \leq 1\}$. Perché non ha minimo? In altre parole, quale tra le proprietà 3) e 4) della Definizione 6.1 non è verificata? Se prendiamo $\underline{x} = -1$, è evidente che la 3) è vera, ma la 4) è falsa; se prendiamo $\underline{x} = -\frac{1}{2}$, è sempre la 4) a saltare, e continua a non essere verificata se scegliamo $\underline{x} = 0$. Se, invece, la 4) fosse vera per un certo \underline{x} , che sarebbe quindi maggiore di zero, la 3) non potrebbe essere vera, perché (come già detto prima) $\frac{\underline{x}}{2}$ è ancora in E ma non è maggiore di \underline{x} .

Dal momento che non riusciamo a far valere contemporaneamente la 3) e la 4), accontentiamoci di una sola di esse, e vediamo se è possibile salvare qualche proprietà interessante. Dal momento che la 4) da sola non ci dice niente di più sull'insieme (è solo la condizione di appartenenza ad E), manteniamo solo la 3), e andiamo a considerare l'insieme dei numeri reali che verificano tale proprietà:

$$m(E) = \{x \in \mathbb{R} : x \leq y \text{ per ogni } y \text{ in } E\}.$$

Tale insieme è non vuoto (abbiamo già detto che contiene -1 , $-\frac{1}{2}$ e 0), e può essere così caratterizzato:

$$m(E) = \{x \in \mathbb{R} : x \leq 0\}.$$

Infatti, se $x \leq 0$, e se y è in E , allora $x \leq 0 < y$ e quindi x è in $m(E)$; viceversa, se x è in $m(E)$, allora x deve essere minore od uguale a zero: se fosse maggiore di zero, $\frac{x}{2}$ sarebbe contemporaneamente minore di x e appartenente ad E . Tutti gli elementi di $m(E)$ sono “candidati minimo” per E , dato che ognuno di essi soddisfa la 3); tra tutti, ne esiste uno che è “migliore” di tutti gli altri? Uno che è più “vicino” ad E degli altri? Uno per il quale non esiste un altro numero in $m(E)$ più grande di lui? Evidentemente, sì: il numero reale 0 — il massimo di $m(E)$ — è, tra tutti i numeri in $m(E)$, il migliore; ovvero, non appena ci si sposta un po' verso destra, anche di pochissimo, si trovano elementi di E (quelli, cioè, che soddisfano la 4), e per i quali la 3) non vale).

Ricapitolando: l'insieme $E = \{x \in \mathbb{R} : 0 < x \leq 1\}$ non ha minimo; abbiamo però identificato un numero reale, 0 , il massimo di $m(E)$, che è quanto di più “vicino” al minimo sia possibile ottenere. Ovviamente, siamo riusciti ad identificarlo in questo caso: dal momento che abbiamo trovato dei sottoinsiemi di \mathbb{R} che non ammettono massimo, potrebbe essere senza speranza fare affermazioni del tipo: “D'accordo, l'insieme non ha minimo, ma si può considerare al suo posto il massimo di $m(E)$ ” perché non è detto che tale massimo esista. Prima di proseguire, definiamo in maniera precisa gli insiemi come $m(E)$.

DEFINIZIONE 6.4. Sia E un sottoinsieme di \mathbb{R} . Definiamo $M(E)$, l'insieme dei **maggioranti** di E , come

$$M(E) = \{x \in \mathbb{R} : x \geq y \text{ per ogni } y \text{ in } E\}.$$

Definiamo $m(E)$, l'insieme dei **minoranti** di E , come

$$m(E) = \{x \in \mathbb{R} : x \leq y \text{ per ogni } y \text{ in } E\}.$$

Siamo allora interessati a cercare, se esiste, il massimo di $m(E)$ e, specularmente, il minimo di $M(E)$. Un primo caso di “esistenza” si ha quando l'insieme E ammette minimo (o massimo).

TEOREMA 6.5. *Sia E un sottoinsieme di \mathbb{R} . Se E ammette massimo \bar{x} , allora $M(E)$ è non vuoto, ammette minimo, e tale minimo è \bar{x} . Se E ammette minimo \underline{x} , allora $m(E)$ è non vuoto, ammette massimo, e tale massimo è \underline{x} .*

Dimostrazione. Essendo \bar{x} il minimo di E , \bar{x} è un numero reale tale che $y \leq \bar{x}$ per ogni y in E ; pertanto \bar{x} è un maggiorante di E , e quindi $M(E)$ è non vuoto. Se dimostriamo che, preso comunque z in $M(E)$, si ha $\bar{x} \leq z$, avremo allora dimostrato che \bar{x} è il minimo di $M(E)$. Sia allora z in $M(E)$, e supponiamo per assurdo che $z < \bar{x}$; se così fosse, z non potrebbe essere un maggiorante di E perché sarebbe minore di un elemento di E , che è proprio \bar{x} . Quindi, $\bar{x} \leq z$. L'altra metà della dimostrazione⁽¹⁹⁾ è lasciata per esercizio. ■

ESERCIZIO 6.6. Determinare $m(E)$ e $M(E)$ per i seguenti sottoinsiemi di \mathbb{R} :

$$\begin{aligned} E &= \{x \in \mathbb{R} : -2 < x \leq 1\} \cup \{2\}, \\ E &= \{x \in \mathbb{R} : x^2 - 1 < 0\}, \\ E &= \{x \in \mathbb{R} : x^2 - 1 \geq 0\} \cap \{x \in \mathbb{R} : x^2 < 4\}, \\ E &= \{x \in \mathbb{Q} : 3 < x < 4\}. \end{aligned}$$

Per ognuno degli insiemi precedenti determinare (se esistono) il massimo di $m(E)$ ed il minimo di $M(E)$.

Una volta svolto l'esercizio precedente⁽²⁰⁾, siamo fortemente tentati⁽²¹⁾ di affermare che, per ogni sottoinsieme E di \mathbb{R} , ed indipendentemente dall'esistenza o meno del massimo o del minimo di E , esistono sia il massimo di $m(E)$ che il minimo di $M(E)$. Tanto per smontare subito le nostre certezze, consideriamo il seguente insieme:

$$E = \{x \in \mathbb{R} : x^2 > 1\}.$$

Risolvendo la disequazione, vediamo subito che

$$E = \{x \in \mathbb{R} : x < -1\} \cup \{x \in \mathbb{R} : x > 1\},$$

e che quindi sia $m(E)$ che $M(E)$ sono vuoti, come si verifica facilmente. Essendo vuoti, $m(E)$ non ha massimo e $M(E)$ non ha minimo, e quindi non possiamo più affermare che, qualsiasi sia il sottoinsieme E di \mathbb{R} , esistono sia il massimo dei minoranti che il minimo dei maggioranti. Quello che possiamo fare è modificare l'affermazione “ $m(E)$

⁽¹⁹⁾Che potrebbe essere scritta da un *word processor* capace di eseguire sostituzioni nel testo...

⁽²⁰⁾Se l'abbiamo svolto, ovviamente!

⁽²¹⁾Ma già temiamo che la cosa non sarà comunque facile da dimostrare...

ammette massimo per ogni sottoinsieme E di \mathbb{R} ” nell’affermazione (comunque tutta da provare) “ $m(E)$ ammette massimo per ogni sottoinsieme E di \mathbb{R} tale che $m(E)$ non sia vuoto”.

Cosa vuol dire che $m(E)$ non è vuoto? Vuol dire che esiste almeno un minorante di E , ovvero un numero reale x tale che $x \leq y$ per ogni y di E . Se rappresentiamo E sulla retta, questo vuol dire che x è una sorta di “barriera”, di “limitazione” per E : l’insieme E si trova tutto a destra di x ; analogamente, se $M(E)$ è non vuoto, esiste z in $M(E)$ e quindi (essendo z più grande di tutti gli elementi di E), E si trova tutto a sinistra di z . Viceversa, se E ammette una “barriera” a destra (o a sinistra), $M(E)$ (o $m(E)$) non sarà vuoto.

DEFINIZIONE 6.7. Sia E un sottoinsieme di \mathbb{R} . E si dice **limitato superiormente** se esiste M in \mathbb{R} tale che $x \leq M$ per ogni x in E ; equivalentemente, se $M(E)$ non è vuoto. E si dice **limitato inferiormente** se esiste m in \mathbb{R} tale che $m \leq x$ per ogni x in E ; equivalentemente, se $m(E)$ non è vuoto. Un sottoinsieme E di \mathbb{R} che sia limitato sia inferiormente che superiormente si dice **limitato**.

Alla luce della definizione precedente, le nostre affermazioni diventano: “se E è un sottoinsieme di \mathbb{R} limitato inferiormente, esiste il massimo dei minoranti” e “se E è un sottoinsieme di \mathbb{R} limitato superiormente, esiste il minimo dei maggioranti”. E adesso viene il difficile: ci “sembra”, dopo aver fatto molti esempi, che le due affermazioni precedenti siano vere; adesso è necessario **dimostrare** che sono vere. Ed invece. . .

ASSIOMA DI COMPLETEZZA (4) *Se $E \subseteq \mathbb{R}$ è limitato superiormente, allora esiste il minimo dei maggioranti. Se $E \subseteq \mathbb{R}$ è limitato inferiormente, allora esiste il massimo dei minoranti.*

ESERCIZIO 6.8. Dimostrare che le due proprietà date nell’assioma sono equivalenti (e quindi basterebbe dare solo la prima, o solo la seconda).

Il nostro sistema di assiomi finalmente è completo e possiamo dare la definizione che aspettavamo:

DEFINIZIONE 6.9. \mathbb{R} è un insieme che verifica gli assiomi (1), (2), (3) e (4) (assiomi per addizione, moltiplicazione, ordinamento e completezza).

Ricordiamo che un insieme che verifica solo (1), (2), (3) si chiama un corpo ordinato, e se verifica anche (4) si chiama un *corpo ordinato completo*. Quindi abbiamo definito \mathbb{R} come un corpo ordinato completo.

OSSERVAZIONE 6.10. Fermi tutti, direte voi. Chi ci assicura che esiste un insieme che verifica (1), (2), (3) e (4)? E ammesso che esista, chi ci dice che ne esiste uno solo a cui possiamo dare il nome di \mathbb{R} ? Ne potrebbero esistere molti, magari infiniti. . .

La risposta alla prima domanda è molto facile: per risparmiare tempo e fatica, abbiamo deciso di partire dall’ipotesi che un tale insieme esiste. Il bello del metodo assiomatico è proprio questo: possiamo dedurre tutti i teoremi che vogliamo su \mathbb{R} , indipendentemente dal fatto che \mathbb{R} esista oppure no! Se avessimo più tempo, si potrebbe dimostrare che, partendo dagli assiomi fondamentali della teoria degli insiemi

(gli assiomi ZFC, uno dei quali asserisce che \mathbb{N} esiste e soddisfa il principio di induzione), si possono costruire molto facilmente i razionali \mathbb{Q} e poi, con un po' più di fatica, tutti i reali \mathbb{R} , ossia un insieme che soddisfa gli assiomi (1)–(4). Non avendo tutto questo tempo a disposizione, partiamo direttamente dalle proprietà di \mathbb{R} ; il risultato è completamente equivalente e altrettanto rigoroso.

La risposta alla seconda domanda sull'unicità di \mathbb{R} è interessante: si può dimostrare che di insiemi che verificano gli assiomi (1)–(4) ne esiste “essenzialmente” solo uno. Ad esempio, è abbastanza chiaro che i due modelli costruiti sopra, la retta reale e l'insieme di tutti gli sviluppi decimali, rappresentano in modi diversi lo stesso oggetto. Per essere più precisi, si può dimostrare che due corpi ordinati completi sono isomorfi in un senso opportuno, ma qui non ci dilungheremo su questo fatto.

Una conseguenza dell'assioma precedente è la seguente definizione.

DEFINIZIONE 6.11. Sia E un sottoinsieme di \mathbb{R} .

Se E è limitato superiormente, definiamo **estremo superiore** di E il minimo dei maggioranti:

$$\sup E = \min M(E).$$

Se E è limitato inferiormente, definiamo **estremo inferiore** di E il massimo dei minoranti:

$$\inf E = \max m(E).$$

Grazie al Teorema 6.5, possiamo affermare che se E ammette massimo, allora si ha $\sup E = \max E$, e analogamente, se E ammette minimo, allora $\inf E = \min E$. In altre parole, i concetti di estremo superiore ed inferiore estendono il concetto di massimo e minimo.

Nella Definizione 6.1 avevamo detto che \bar{x} era un massimo per E se 1) \bar{x} era un maggiorante di E che 2) apparteneva ad E . Il seguente teorema modifica la proprietà 2) nel caso dell'estremo superiore (con analoghe modifiche nel caso dell'estremo inferiore).

TEOREMA 6.12. *Sia E un sottoinsieme di \mathbb{R} .*

Sia E limitato superiormente. Allora $S = \sup E$, l'estremo superiore di E , è l'unico numero reale caratterizzato dalle seguenti proprietà:

- a) per ogni y in E si ha $y \leq S$;
- b) per ogni $z < S$ esiste y in E (eventualmente dipendente da z) tale che $z < y \leq S$.

Sia E limitato inferiormente. Allora $I = \inf E$, l'estremo inferiore di E , è l'unico numero reale caratterizzato dalle seguenti proprietà:

- c) per ogni y in E si ha $I \leq y$;
- d) per ogni $z > I$ esiste y in E (eventualmente dipendente da z) tale che $I \leq y < z$.

Dimostrazione. Dimostriamo la prima parte del teorema; la seconda si dimostra in modo analogo.

In primo luogo la a) è equivalente a dire che S è un maggiorante di E , che è esattamente la 4) nella definizione di minimo per $M(E)$. D'altra parte, la b) afferma

che nessun numero reale minore di S è un maggiorante di E , vale a dire che tutti i maggioranti di E sono maggiori od uguali a S , e questo è esattamente ciò che afferma la 3) nella definizione di minimo per $M(E)$. Pertanto, a) e b) sono equivalenti a 3) e 4) nella definizione di minimo per $M(E)$, ovvero S verifica a) e b) se e solo se è il minimo dei maggioranti di E , cioè l'estremo superiore di E . ■

OSSERVAZIONE 6.13. Un modo alternativo di enunciare la b) e la d) è il seguente:

b') per ogni $\varepsilon > 0$ esiste y_ε in E tale che $S - \varepsilon < y_\varepsilon \leq S$;

d') per ogni $\varepsilon > 0$ esiste y_ε in E tale che $I \leq y_\varepsilon < I + \varepsilon$.

La b) e la b') esprimono entrambe la stessa proprietà: non appena ci si sposta un po' a sinistra dall'estremo superiore, si "scavalcano" elementi di E ; analogamente, la d) e la d') affermano che spostandosi a destra dall'estremo inferiore si "scavalcano" elementi di E . Si noti che in entrambi i casi **non** si richiede che z appartenga ad E . Evidentemente, e siccome ci interessa solo quello che accade "vicino" ad S e ad I , per dimostrare la b') o la d') basta verificare l'esistenza di y_ε per ogni ε "sufficientemente piccolo"; anche perché se riusciamo a trovare y_ε tale che la b') è vera per un certo valore di $\varepsilon > 0$, lo stesso y_ε continua a soddisfare la b') per ogni $\varepsilon' > \varepsilon$. Osserviamo esplicitamente che non è affatto detto che l'insieme E sia "pieno" a sinistra dell'estremo superiore, o a destra dell'estremo inferiore. Ad esempio, se $E = \{-2\} \cup \{x \in \mathbb{R} : -1 < x < 1\} \cup \{2\}$, l'estremo superiore di E è 2, ma spostandosi di poco a sinistra troviamo solo il numero 2⁽²²⁾.

ESERCIZIO 6.14. Calcolare estremo superiore ed inferiore (usando $m(E)$ e $M(E)$) dei seguenti insiemi (per l'ultimo si richiede un po' di immaginazione...), e dire se sono rispettivamente massimo e minimo:

$$\begin{aligned} E &= \{x \in \mathbb{R} : 1 < x^2 \leq 9\} \cap \{x \in \mathbb{R} : x > -2\}, \\ E &= \{x \in \mathbb{R} : x^3 > 1\} \cap \{x \in \mathbb{R} : x^5 \leq 32\}, \\ E &= \{x \in \mathbb{Q} : x^2 \leq 4\}, \\ E &= \bigcup_{n=3}^{+\infty} \left\{ x \in \mathbb{R} : \frac{1}{n} \leq x \leq 1 - \frac{1}{n} \right\}. \end{aligned}$$

Una volta calcolati estremo superiore ed inferiore, dimostrare che sono tali usando le a), b) (o b')), c) e d) (o d')).

Risposta 6.14: 4) Dimostriamo che $m(E) = \{x \in \mathbb{R} : x \leq 0\}$; è chiaro che se $x \leq 0$, allora x è minore di tutti gli elementi di E , dato che ogni insieme che contribuisce all'unione è composto da numeri positivi, e quindi appartiene a $m(E)$. Viceversa, se x è un minorante di E , x non può essere positivo. Se, per assurdo, x fosse positivo, consideriamo il numero $y = \frac{1}{x}$, e sia $n = [y] + 1$ (ricordiamo che $[\cdot]$ è la parte intera). Allora $x > \frac{1}{n}$, perché $n > \frac{1}{x}$; se ne deduce che x è più grande di un elemento ($\frac{1}{n}$) appartenente ad E , e quindi non può essere un minorante di E . Concludendo, l'estremo inferiore di E è 0 (e non è un minimo); con ragionamenti analoghi si verifica che l'estremo superiore di E è 1 (e non è un massimo). Per dimostrare che 0 è l'estremo inferiore di E usando c) e d'), basta dimostrare che per ogni

⁽²²⁾Il che spiega per quale motivo nella b) e nella b') si richiede $y \leq S$: se la disuguaglianza fosse stretta saremmo nei guai...

$\varepsilon > 0$ esiste un numero naturale n tale che $\varepsilon > \frac{1}{n}$; come prima, si può scegliere $n = \lceil \frac{1}{\varepsilon} \rceil + 1$. Come esercizio aggiuntivo, verificare che $E = \{x \in \mathbb{R} : 0 < x < 1\}$.

Un'altra proprietà utile soddisfatta da estremo superiore ed inferiore è la “monotonia”.

TEOREMA 6.15. *Sia E un sottoinsieme di \mathbb{R} . Se E è limitato superiormente, e se $F \subseteq E$, allora $\sup F \leq \sup E$. Se E è limitato inferiormente, e se $F \subseteq E$, allora $\inf E \leq \inf F$.*

Dimostrazione. Sia S l'estremo superiore di E . Allora S è un maggiorante di E , ovvero è maggiore o uguale di tutti gli elementi di E . Siccome F è contenuto in E , S è anche più grande di tutti gli elementi di F , ovvero S è in $M(F)$. Essendo in $M(F)$, è maggiore o uguale del minimo di $M(F)$, che è l'estremo superiore di F . La seconda parte della dimostrazione è lasciata per esercizio. ■

Come conseguenza del teorema precedente, se E è limitato, e $F \subseteq E$ (cosicché anche F è limitato), si ha

$$\inf E \leq \inf F \leq \sup F \leq \sup E.$$

L'unica disuguaglianza da dimostrare è la centrale, che discende dal fatto (di dimostrazione evidente) che se x è un maggiorante di F e y è un minorante di F , allora si ha $y \leq x$. In particolare, il massimo dei minoranti è minore del minimo dei maggioranti.

Un altro utile teorema è il seguente.

TEOREMA 6.16. *Siano E ed F due sottoinsiemi di \mathbb{R} tali che per ogni x di E e per ogni y di F si abbia $x \leq y$. Allora*

$$\sup E \leq \inf F.$$

Dimostrazione. Sia y fissato in F ; per ipotesi, ogni x di E è minore o uguale di y , cosicché $M(E)$ non è vuoto. Inoltre, per l'arbitrarietà di y , $M(E)$ contiene F . Per il Teorema 6.15,

$$\sup E = \min M(E) = [\text{Teorema 6.5}] = \inf M(E) \leq \inf F,$$

come volevasi dimostrare. ■

OSSERVAZIONE 6.17. Osserviamo esplicitamente che, anche se E ed F sono tali che $x < y$ per ogni x di E e per ogni y di F , si può solo concludere che $\sup E \leq \inf F$, e non che si ha il minore stretto. Ad esempio, siano $E = \{x \in \mathbb{R} : 0 < x < 1\}$ e $F = \{x \in \mathbb{R} : 1 < x < 2\}$. Se, invece, E ha massimo e F ha minimo, allora $\max E < \min F$.

DEFINIZIONE 6.18. Dati due numeri reali a e b , con $a < b$, definiamo **intervallo** di estremi a e b l'insieme dei numeri reali x compresi tra a e b . Dato che non abbiamo specificato se “compresi” voglia dire maggiori, o maggiori o uguali, di a e minori, o minori o uguali, di b , abbiamo quattro possibilità, che indichiamo come segue:

$$\begin{aligned} [a, b] &= \{x \in \mathbb{R} : a \leq x \leq b\}, \\ [a, b) &= \{x \in \mathbb{R} : a \leq x < b\}, \\ (a, b] &= \{x \in \mathbb{R} : a < x \leq b\}, \\ (a, b) &= \{x \in \mathbb{R} : a < x < b\}. \end{aligned}$$

Il primo si dice intervallo **chiuso** di estremi a e b , mentre l'ultimo si dice intervallo **aperto** di estremi a e b ; il secondo ed il terzo intervallo si dicono **semiaperti** o **semichiusi**, a seconda dei "gusti".

ESERCIZIO 6.19. Calcolare estremo superiore ed inferiore di ognuno dei quattro tipi di intervalli.

Risposta 6.19: a e b .

Consideriamo ora l'insieme

$$E = \{x \in \mathbb{R} : x \geq 0\}.$$

È facile vedere che E è limitato inferiormente e che il suo estremo inferiore è 0 (che è anche minimo); è altrettanto facile rendersi conto che $M(E) = \emptyset$: se supponiamo che x sia un maggiorante di E (fatto questo che implica $x \geq 0$, dato che 0 appartiene ad E), possiamo subito costruire $x + 1$ che appartiene ancora ad E , ma è più grande di x . In altre parole, E non è limitato superiormente.

DEFINIZIONE 6.20. Sia E un sottoinsieme di \mathbb{R} . E si dice **illimitato superiormente** se $M(E)$ è vuoto. E si dice **illimitato inferiormente** se $m(E)$ è vuoto. E si dice **illimitato** se è sia illimitato superiormente che inferiormente.

Equivalentemente:

- i) E è illimitato superiormente se e solo se per ogni $M > 0$ esiste x in E (eventualmente dipendente da M) tale che $x > M$;
- ii) E è illimitato inferiormente se e solo se per ogni $M < 0$ esiste x in E (eventualmente dipendente da M) tale che $x < M$.

Infatti, la i) afferma che nessun numero positivo M appartiene ad $M(E)$ (il che implica automaticamente che nessun numero negativo vi può appartenere), mentre la ii) afferma che nessun numero negativo M appartiene ad $m(E)$ (cosicché nessun numero positivo vi può appartenere).

Come trattare (dal punto di vista di estremo superiore ed inferiore) gli insiemi illimitati? Il risultato chiave è il Teorema 6.15. In altre parole, vogliamo "assegnare" un estremo superiore ad un insieme illimitato superiormente, facendoci guidare dalla monotonia: se F è un qualsiasi sottoinsieme di E limitato superiormente, mentre E non lo è, l'estremo superiore di E (da definire) deve essere maggiore dell'estremo superiore di F . Per capire come fare, riprendiamo l'insieme $E = \{x \in \mathbb{R} : x \geq 0\}$, e, fissato y in \mathbb{R} con $y \geq 0$, consideriamo $F_y = \{x \in \mathbb{R} : 0 \leq x \leq y\} = [0, y]$. Evidentemente, F_y è contenuto in E ; d'altra parte, l'estremo superiore di F_y è proprio y (che è anche massimo), per cui se vogliamo che il Teorema 6.15 valga, deve essere

$$y = \sup F_y \leq \sup E,$$

e quindi il definendo estremo superiore di E deve essere maggiore di ogni numero reale positivo; volendo, deve essere "il più grande" numero reale positivo. Ma un tale numero non esiste! Siccome non esiste, lo "inventiamo", o meglio inventiamo un simbolo che abbia per noi il significato dell'inesistente "più grande numero reale positivo". Tale simbolo è $+\infty$, che si legge "più infinito", ed è quello che scegliamo per definire l'estremo superiore di E e, in generale, l'estremo superiore di un insieme

illimitato superiormente. Ripetiamo — *ad nauseam* — che $+\infty$, ed il suo speculare $-\infty$, che si legge “meno infinito” e viene usato per definire l’estremo inferiore di un insieme illimitato inferiormente, non sono numeri: sono due **simboli**, che usiamo per rappresentare qualcosa di più grande (o più piccolo) di tutti i numeri reali. Pertanto, non ha senso dire che un insieme illimitato superiormente ammette massimo: il suo estremo superiore è $+\infty$, che non è un numero reale.

ESERCIZIO 6.21. Se l’estremo superiore di un insieme illimitato superiormente è $+\infty$, quanto vale il minimo dell’insieme vuoto? E quanto il massimo dell’insieme vuoto? Una volta risposto, la risposta è coerente con il Teorema 6.15?

A questo punto possiamo affermare (usando l’Assioma 6 e la definizione precedente) che ogni sottoinsieme E di \mathbb{R} ammette sia estremo superiore che estremo inferiore. Inoltre, se F è un sottoinsieme di E si ha

$$\inf E \leq \inf F \leq \sup F \leq \sup E.$$

Quest’ultimo fatto è molto utile nel caso in cui si voglia dimostrare che un insieme E è illimitato superiormente (o inferiormente); invece di dimostrarlo direttamente, è sufficiente far vedere che esiste un suo sottoinsieme illimitato superiormente (o inferiormente).

ESERCIZIO 6.22. Sia

$$E = \left\{ x \in \mathbb{R} : x \geq 0 \text{ e } \begin{pmatrix} 2[x] \\ [x] \end{pmatrix} \geq 2^{[x]} \right\}.$$

Calcolare l’estremo superiore e l’estremo inferiore di E .

Risposta 6.22: E contiene \mathbb{N} (si veda l’Esercizio 2.15). Non c’è quindi bisogno di dimostrare che E è l’insieme dei numeri positivi per affermare che $\inf E = 0 = \min E$ e $\sup E = +\infty$.

DEFINIZIONE 6.23. L’insieme dei numeri reali x maggiori (maggiori o uguali) o minori (minori o uguali) di un fissato numero reale a vengono anch’essi detti intervalli, ma illimitati (superiormente o inferiormente a seconda dei casi). Come per gli intervalli limitati, se a è un fissato numero reale abbiamo le seguenti notazioni:

$$\begin{aligned} [a, +\infty) &= \{x \in \mathbb{R} : x \geq a\}, \\ (a, +\infty) &= \{x \in \mathbb{R} : x > a\}, \\ (-\infty, a] &= \{x \in \mathbb{R} : x \leq a\}, \\ (-\infty, a) &= \{x \in \mathbb{R} : x < a\}. \end{aligned}$$

Si noti che il “lato” infinito è sempre aperto, coerentemente con il fatto che $+\infty$ e $-\infty$ non sono numeri reali. In particolare,

$$\mathbb{R} = (-\infty, +\infty).$$

7. Ricostruiamo \mathbb{N} e \mathbb{Q}

Dato che abbiamo scelto come punto di partenza l’insieme \mathbb{R} , e lo abbiamo definito come l’insieme che verifica gli assiomi (1)–(4), è legittimo chiedersi che fine hanno fatto l’insieme \mathbb{N} , con l’importantissimo principio di induzione (o l’equivalente principio di buon ordinamento) e l’insieme \mathbb{Q} . Niente paura: possiamo ricostruire tutti e due questi

insiemi come sottoinsiemi di \mathbb{R} in un modo molto semplice. Basta utilizzare la seguente definizione:

DEFINIZIONE 7.1. Un insieme $E \subseteq \mathbb{R}$ si dice *insieme induttivo* se contiene 0 e ha la proprietà che se x appartiene a E , allora $x + 1$ appartiene a E .

Quindi, per controllare se l'insieme E è induttivo, bisogna dimostrare due cose: (a) che contiene 0; (b) che se contiene un numero x allora contiene anche $x + 1$.

Ovviamente, \mathbb{R} è un insieme induttivo. Ci verrebbe da dire che anche \mathbb{N} è un insieme induttivo, ma purtroppo non lo abbiamo ancora definito. . . Se avessimo a disposizione \mathbb{N} , potremmo costruire subito molti altri insiemi induttivi, ad esempio

$$E = \left\{ 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \dots \right\}$$

e così via. Notiamo che se E è induttivo, dalla definizione segue $0 \in E$, quindi $1 \in E$, quindi $2 \in E$ eccetera. Ci verrebbe da dire che qualunque insieme induttivo contiene \mathbb{N} e quindi che \mathbb{N} è il più piccolo insieme induttivo. E allora diciamolo: **Definizione 7.2.** \mathbb{N} si definisce come l'intersezione di tutti gli insiemi induttivi. Notiamo che se E è induttivo, dalla definizione segue $0 \in E$, quindi $1 \in E$, quindi $2 \in E$ eccetera. Ci verrebbe da dire che qualunque insieme induttivo contiene \mathbb{N} e quindi che \mathbb{N} è *il più piccolo insieme induttivo*. E allora diciamolo:

DEFINIZIONE 7.2. \mathbb{N} si definisce come l'intersezione di tutti gli insiemi induttivi.

L'insieme così definito possiede tutte le proprietà di \mathbb{N} a cui ormai siamo abituati. Ad esempio è facilissimo dimostrare che il principio di induzione (il Teorema 2.8) è ancora valido: se una proprietà $P(n)$ soddisfa le ipotesi del Teorema, chiamiamo E l'insieme degli n per cui $P(n)$ è vera; quindi E contiene 0 e se contiene un numero contiene anche il successivo, ossia E è induttivo; dalla definizione di \mathbb{N} segue allora che E contiene tutto \mathbb{N} , ossia che $P(n)$ è vera per tutti gli $n \in \mathbb{N}$.

Inoltre, ricordando che il principio di induzione è *equivalente* al principio del buon ordinamento, anche quest'ultimo è valido per il nostro insieme \mathbb{N} .

Non perderemo tempo a verificare tutte le altre proprietà di \mathbb{N} ma una in particolare ci servirà fra poco:

PROPOSIZIONE 7.3 (Proprietà archimedea). *Sia $x > 0$ un numero reale e sia E l'insieme di tutti i suoi multipli nx , con $n \in \mathbb{N}$. Allora E non è limitato superiormente.*

Dimostrazione. Supponiamo per assurdo che E sia limitato superiormente, e sia $S = \sup E$. Allora $S - \frac{x}{2}$ non è un maggiorante di E e deve esistere un elemento $nx \in E$ tale che $nx > S - \frac{x}{2}$. Ma allora $(n+1)x > S + \frac{x}{2} > S$, impossibile. ■

Ora \mathbb{Q} si può definire come l'insieme di tutte le frazioni $\pm \frac{p}{q}$ dove $p, q \in \mathbb{N}$, $q > 0$. Anche per \mathbb{Q} è facile verificare che tutte le proprietà abituali sono valide. In particolare possiamo dimostrare che i punti di \mathbb{Q} sono "molto fitti" fra quelli di \mathbb{R} :

TEOREMA 7.4 (Densità di \mathbb{Q} in \mathbb{R}). *Dati x e y numeri reali, con $y < x$, esiste un numero razionale $\frac{p}{q}$ tale che $y < \frac{p}{q} < x$.*

Dimostrazione. Per la proprietà archimedeana, esistono numeri naturali q tali che $q > \frac{1}{x-y}$, ossia tali che $qx > qy + 1$; scegliamone uno qualsiasi. Sempre per la proprietà archimedeana, esistono numeri naturali p tali che $p > qy$; per il principio del buon ordinamento, possiamo prendere il minimo di tali p , quindi si ha $p > qy \geq p - 1$ ossia $qy + 1 \geq p > qy$ e ricordando la disuguaglianza precedente otteniamo $qx > p > qy$. Dividendo per q otteniamo la tesi. ■

8. Radici, potenze. Funzioni elementari.

Nel paragrafo 4 abbiamo “costruito” un numero reale non razionale: la lunghezza della diagonale del quadrato unitario, e abbiamo visto che tale numero, sia esso x , gode della proprietà di avere il quadrato uguale a 2. Come è noto, per semplificare le notazioni tale numero si definisce la radice quadrata di 2, e si indica $x = \sqrt{2}$. Più che sulla notazione (standard), mettiamo l'accento sul fatto che x è “la” radice quadrata di 2: ovvero, l'unico numero positivo il cui quadrato è due. È evidente che di numeri reali il cui quadrato vale due ne esiste un altro (ed uno solo): $-x$, ovvero $-\sqrt{2}$. E questo a motivo del fatto che il quadrato di x ed il quadrato di $-x$ coincidono. Per evitare fraintendimenti, scriviamo $\sqrt{2}$, sottintendendo il segno +, intendendo così che tra i due numeri reali il cui quadrato vale due scegliamo il positivo. Continuando con le nostre costruzioni geometriche, e sempre usando il Teorema di Pitagora, possiamo costruire \sqrt{n} per ogni n naturale⁽²³⁾, ovvero l'unico numero reale positivo il cui quadrato vale n . Ci chiediamo ora se, fissato un qualsiasi numero reale y , esista un unico numero reale positivo x tale che $x^2 = y$. È evidente che se $y < 0$ tale numero x non esiste, dato che $x^2 \geq 0$ per ogni x reale. Limitiamoci pertanto agli $y \geq 0$. Quello che si dimostra è il seguente fatto: dato un qualsiasi $y \geq 0$ esiste un unico numero reale $x \geq 0$ tale che $x^2 = y$.

ESERCIZIO 8.1. Sia $y \geq 0$ e sia $E = \{x \in [0, +\infty) : x^2 \leq y\}$. Dimostrare che E non è vuoto, che ammette estremo superiore S e che $S^2 = y$.

Definiamo \sqrt{y} l'unico numero reale positivo il cui quadrato è y . Che cosa abbiamo ottenuto? Abbiamo ottenuto una **funzione**: vale a dire una legge che, ad ogni numero reale $y \geq 0$ associa l'unico numero reale $x \geq 0$ tale che $x^2 = y$. Scrivendo in maniera compatta:

$$\begin{array}{ccc} \sqrt{\cdot} & : & [0, +\infty) \rightarrow [0, +\infty) \\ & & y \mapsto \sqrt{y}. \end{array}$$

Prima di procedere oltre, definiamo in maniera corretta cosa sia una funzione.

DEFINIZIONE 8.2. Sia A un sottoinsieme di \mathbb{R} . Una funzione f da A in \mathbb{R} è una legge che, ad ogni elemento x di A , associa uno ed un solo elemento y di \mathbb{R} . Tale elemento y si dice “immagine di x tramite f ”, e si indica con $y = f(x)$. In simboli:

$$\begin{array}{ccc} f & : & A \rightarrow \mathbb{R} \\ & & x \mapsto y = f(x). \end{array}$$

L'insieme A si dice **dominio**, o **insieme di definizione** della funzione f , mentre x si dice **argomento** della funzione f . Associato ad una funzione esiste un secondo

⁽²³⁾Come?

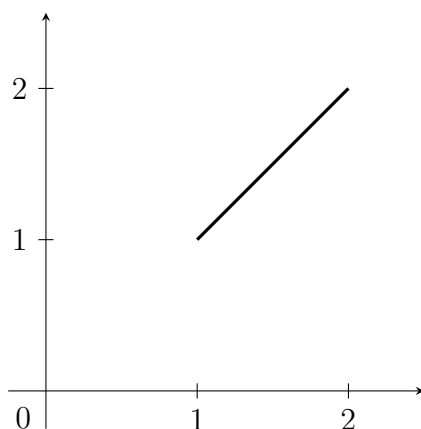
sottoinsieme di \mathbb{R} , detto **immagine** di A tramite f , che è l'insieme dei valori assunti da $f(x)$ al variare di x in A :

$$\text{Im}(f) = f(A) = \{y \in \mathbb{R} : \text{esiste } x \text{ in } A \text{ tale che } f(x) = y\}.$$

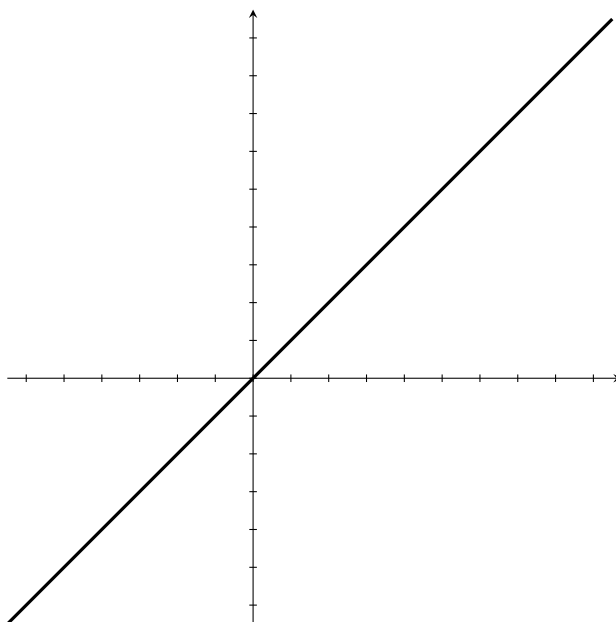
La maniera migliore di “rappresentare” una funzione da A in \mathbb{R} è quella di ricorrere al suo **grafico**, che non è altro che il sottoinsieme del piano cartesiano⁽²⁴⁾ formato dalle coppie $(x, f(x))$ al variare di x in A :

$$G(f) = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x \in A, y = f(x)\}.$$

Ad esempio, se $f : [1, 2] \rightarrow \mathbb{R}$ è la funzione definita da $f(x) = x$ per ogni x in $[1, 2]$, il grafico di f è il seguente:

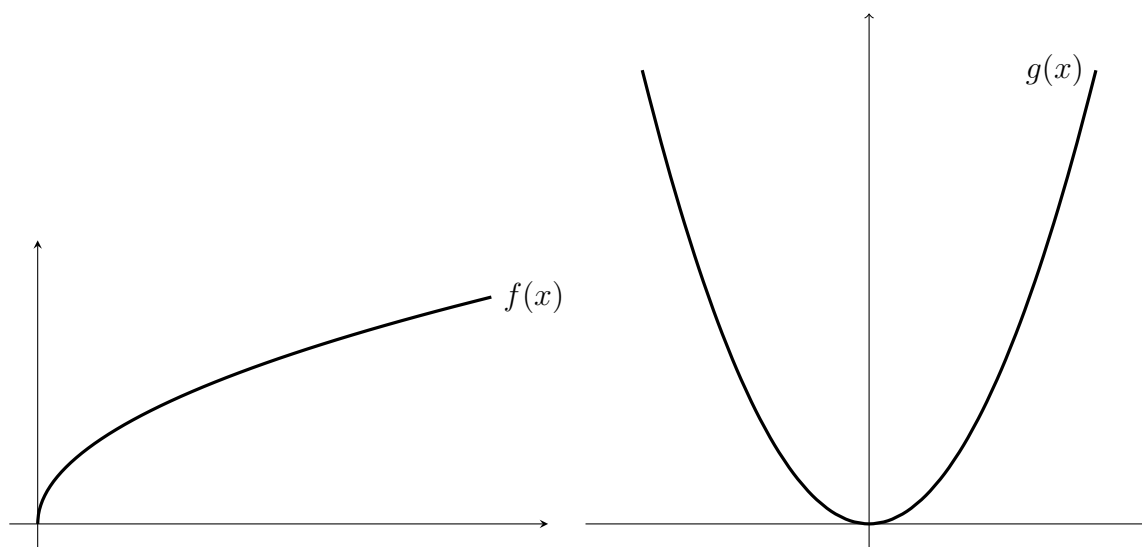


e, più in generale, il grafico (o meglio parte di esso) della funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ definita da $f(x) = x$ per ogni x in \mathbb{R} (una tale funzione si dice “identità” dato che lascia ogni x invariato) è il seguente:



⁽²⁴⁾Qualsiasi cosa esso sia...

Nel caso della funzione $f(x) = \sqrt{x}$, il dominio è $A = [0, +\infty)$, mentre l'immagine è $f(A) = [0, +\infty)$; il grafico è il seguente, affiancato al grafico della funzione $g(x) = x^2$, definita su tutto l'asse reale, e a valori in $[0, +\infty)$.



ESERCIZIO 8.3. Sia x un numero reale qualsiasi: quanto vale $\sqrt{x^2}$?

Svolto l'esercizio precedente? Quale è la risposta? Chi ha risposto $\pm x$? Avendo appena detto che una funzione (come lo è la radice quadrata) associa uno ed un solo numero reale al proprio argomento, è chiaro che $\pm x$ è la risposta sbagliata. Chi ha risposto x ? Dal momento che la funzione radice quadrata ha come immagine l'intervallo $[0, +\infty)$, è chiaro che la risposta è sbagliata nel caso in cui, ad esempio, $x = -1$. In questo caso, infatti, x^2 vale 1, e quindi $\sqrt{x^2} = \sqrt{1} = 1$, dal momento che 1 è l'unico numero reale positivo il cui quadrato vale 1. Se facciamo altri esempi, arriviamo al seguente risultato⁽²⁵⁾:

$$\sqrt{x^2} = \begin{cases} x & \text{se } x \geq 0, \\ -x & \text{se } x < 0. \end{cases}$$

Per aiutare⁽²⁶⁾ i matematici, è stata inventata una notazione apposita per indicare la funzione $\sqrt{x^2}$. Il **valore assoluto**, o **modulo**, indicato con $|\cdot|$, è la funzione da \mathbb{R} in \mathbb{R} definita da

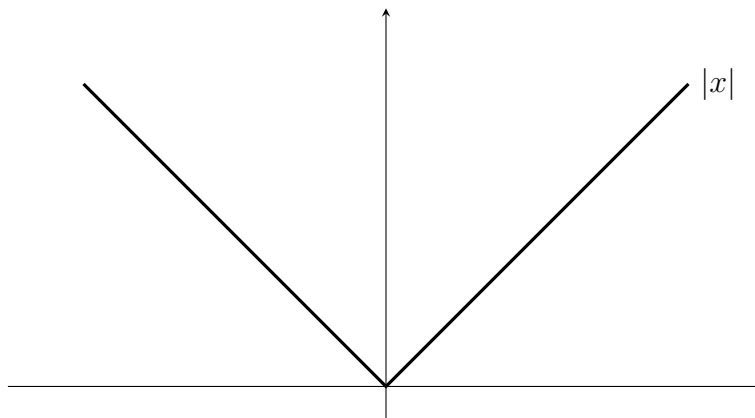
$$(8.1) \quad |x| = \begin{cases} x & \text{se } x \geq 0, \\ -x & \text{se } x < 0. \end{cases}$$

Come si vede abbastanza facilmente, il dominio di $|\cdot|$ è tutto \mathbb{R} , mentre l'immagine è $[0, +\infty)$. **Avvertenza:** il lettore è fortemente invitato a non pensare al modulo di x come “ x privato del segno”. Il valore assoluto è definito in (8.1), ed è la funzione che lascia invariato x se x è non negativo, mentre gli associa il valore $-x$ se x è negativo. Per rendersi conto del perché la frase “ x privato del segno” sia sbagliata, osserviamo che se non diciamo quanto vale x non ha senso parlare del suo segno: quale è il segno di $-x$ se x vale -2 ? In altre parole, se il valore assoluto di x fosse veramente “ x privato

⁽²⁵⁾Altrettanto corretto è lo spostamento dell'uguale: x se $x > 0$ e $-x$ se $x \leq 0$.

⁽²⁶⁾Si fa per dire...

del segno”, allora $|-x|$ sarebbe x , il che è falso se ad esempio $x = -1$. Il grafico della funzione modulo è il seguente:



Notiamo, *en passant*, che $|x - y|$ calcola la “distanza” tra il punto x e il punto y sulla retta reale, indipendentemente dal fatto che x si trovi prima o dopo y ; vale a dire, $|x - y|$ è la lunghezza del segmento che congiunge x a y . La proprietà più importante della funzione modulo è la cosiddetta “disuguaglianza triangolare”: per ogni x e y in \mathbb{R} si ha

$$(8.2) \quad |x + y| \leq |x| + |y|.$$

Per quanto riguarda la differenza, il modulo si comporta in maniera un po’ più complicata: per ogni x e y in \mathbb{R} si ha

$$(8.3) \quad |x - y| \geq \left| |x| - |y| \right|.$$

Infine, osserviamo che se a è non negativo, dire $|x| \leq a$ è equivalente a dire $-a \leq x \leq a$, mentre dire $|x| \geq a$ è equivalente a dire che $x \geq a$ oppure $x \leq -a$.

ESERCIZIO 8.4. Risolvere le seguenti disequazioni:

$$\begin{aligned} |x| &\leq x + 2, \\ |x| &\leq |x - 2|, \\ |x - 4| &< |x - 1|, \\ |2x - 1| &\leq |x - 1| + |x - 2|. \end{aligned}$$

Risposta 8.4: 1) La disequazione è equivalente ai due sistemi

$$\begin{cases} x \leq x + 2 \\ x \geq 0 \end{cases} \quad \begin{cases} -x \leq x + 2 \\ x < 0 \end{cases}$$

Il primo ha come soluzione $x \geq 0$, dato che la prima disequazione è sempre risolta; il secondo ha come soluzione $-1 \leq x < 0$; facendo l’unione, la disequazione è verificata per $x \geq -1$. 2) Questo esercizio si risolve in maniera semplice rispondendo alla domanda “Quali sono i punti della retta reale che distano più da 2 che da 0?”. 3) $x > \frac{5}{2}$. 4) $x \leq -1$.

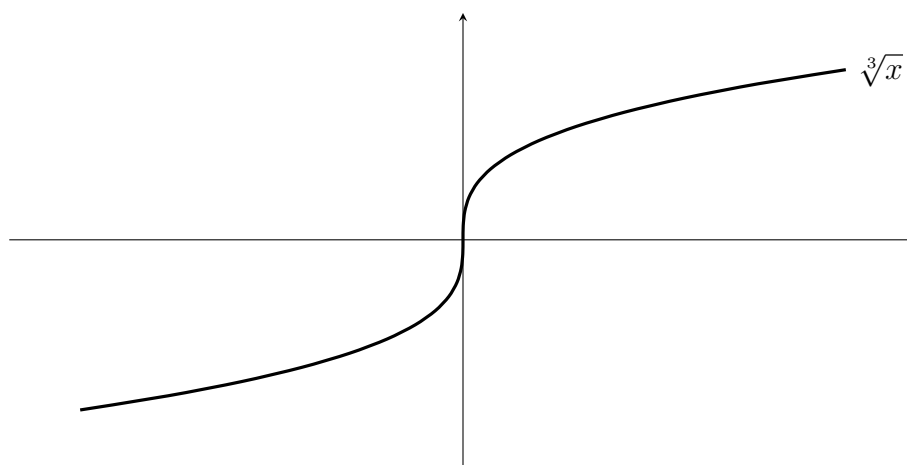
Riprendiamo ora il discorso sulle radici quadrate. Abbiamo visto (o meglio, abbiamo detto) che esiste ed è unica la radice quadrata (positiva) di un numero reale non negativo; in particolare, esiste $\sqrt{2}$, che tra l’altro sappiamo anche “costruire”. L’esigenza di lavorare con le radici quadrate proviene dalla necessità di risolvere equazioni della forma $x^2 = a$, con $a \geq 0$. Ovviamente, il quadrato non è l’unica potenza

che sappiamo fare. Ad esempio, possiamo considerare x^3 , x^4 , ed in generale x^n con n naturale. Così come all'elevazione al quadrato era legata la necessità di risolvere l'equazione $x^2 = 2$, è legittimo chiedersi se si possano risolvere equazioni della forma $x^n = a$, se tali equazioni siano risolubili, e sotto quali condizioni su n e su a . Ad esempio, chiediamoci se sia possibile determinare, se esiste, un numero reale x tale che $x^3 = 2$. In questo caso l'intuizione geometrica non ci aiuta più⁽²⁷⁾, ma fortunatamente abbiamo strumenti analitici sufficientemente potenti (o, equivalentemente, conosciamo sufficientemente bene la retta reale) per poter affermare che tale numero reale esiste, ed è anche unico. A guidarci nella “dimostrazione” dell'esistenza ed unicità di tale x è l'Esercizio 8.1: preso l'insieme

$$E = \{x \in \mathbb{R} : x \geq 0 \text{ e } x^3 \leq 2\},$$

si dimostra facilmente⁽²⁸⁾ che E non è vuoto, che è limitato superiormente, e che il suo estremo superiore S soddisfa l'equazione $S^3 = 2$.

A questo punto, cosa ci impedisce di sostituire 2 con qualsiasi numero reale y ? Niente: ed infatti siamo in grado di dimostrare che per ogni numero reale y esiste un unico numero reale x tale che $x^3 = y$. Osserviamo che siccome la terza potenza di un numero negativo è negativa, non abbiamo più la restrizione sulla positività di y . Chiamando x “radice cubica di y ”, o “radice terza di y ”, indicata in simboli con $x = \sqrt[3]{y}$, abbiamo così definito un'altra funzione: $\sqrt[3]{\cdot} : \mathbb{R} \rightarrow \mathbb{R}$, che a x associa la sua radice cubica. Il grafico di $\sqrt[3]{x}$ è il seguente:



A questo punto possiamo generalizzare: dato un numero naturale $n \geq 1$, e dato un numero reale y (positivo nel caso in cui n sia pari), esiste un unico numero reale x (dello stesso segno di y), tale che $x^n = y$. Definiamo x la “radice n -sima di y ”, e scriviamo $x = \sqrt[n]{y}$. Una notazione alternativa per la radice n -sima, anch'essa molto usata, è quella di elevare x all'esponente frazionario $\frac{1}{n}$. In questa maniera resta valida la legge, soddisfatta dalle potenze intere, secondo la quale $(x^a)^b = x^{ab}$. Infatti, per

⁽²⁷⁾Il problema di costruire x , detto anche “problema della duplicazione del cubo”, è uno dei più antichi della geometria classica; il fatto che non abbia soluzione, ovvero che non esista un modo di costruire x con “riga e compasso”, è stato dimostrato solo agli inizi del secolo diciannovesimo.

⁽²⁸⁾Provateci!

definizione di radice n -sima,

$$x = (x^{\frac{1}{n}})^n = x^{\frac{n}{n}} = x^1.$$

ESERCIZIO 8.5. Per quali x in \mathbb{R} si ha

$$\sqrt[n]{x^n} = x?$$

Un'osservazione: il fatto che per n dispari la radice n -sima sia definita su tutto \mathbb{R} , e non su $[0, +\infty)$ come nel caso di n pari, è un caso fortunato: vedremo tra breve come in generale si potrà parlare di elevamento a potenza solo nel caso in cui la base sia non negativa.

Avendo definito la radice n -sima di x come x elevato alla $\frac{1}{n}$, possiamo ora definire la potenza con esponente razionale di un numero reale: dato $\frac{p}{q}$, numero razionale, definiamo

$$(8.4) \quad x^{\frac{p}{q}} = \sqrt[q]{x^p}.$$

Dobbiamo fare delle restrizioni su p e q , o su x ? Sì, dobbiamo obbligatoriamente chiedere che x sia non negativo, perché non possiamo a priori decidere della parità di q . Inoltre, se $\frac{p}{q}$ è negativo, dobbiamo chiedere che x sia diverso da zero. In definitiva, dato un qualsiasi numero razionale $\frac{p}{q}$ ed un qualsiasi numero reale non negativo x (con $x \neq 0$ se $\frac{p}{q}$ è negativo), è ben definito il numero reale $x^{\frac{p}{q}}$. Alternativamente, se $x \geq 0$, e y è un razionale, è ben definito il numero reale x^y (a patto che x sia diverso da zero se y è negativo). Come numero reale x^y soddisfa tutte le proprietà delle potenze; ad esempio $x^{y+z} = x^y \cdot x^z$, e $(x \cdot y)^z = x^z \cdot y^z$ (nel primo caso $x \geq 0$ e y e z sono razionali, mentre nel secondo x e y sono non negativi e z è razionale). Una delle proprietà più importanti dell'elevamento a potenza razionale è la monotonia: se $x \geq 1$, e se $y > z > 0$ sono due numeri razionali, allora $x^y > x^z$. In altre parole, fissato $x \geq 1$, la funzione che a y razionale positivo associa x^y cresce al crescere di y .

ESERCIZIO 8.6. Verificare che se $0 \leq x \leq 1$, e se $y > z > 0$, allora $x^y < x^z$. Cosa si può dire di x^y e x^z se $y < z < 0$?

Siamo ora pronti a fare il passo successivo: dato x numero reale non negativo (per il momento maggiore di 1), e dato y reale (per il momento positivo), in che modo definire x^y ? Ad esempio, quanto vale (o, meglio, cosa è) $2^{\sqrt{2}}$? Non essendo $\sqrt{2}$ un numero razionale, non possiamo più passare per le radici n -sime, come abbiamo fatto per la potenza ad esponente razionale, ma possiamo usare il fatto che i razionali sono densi in \mathbb{R} . Siccome per ogni $\varepsilon > 0$ esistono due numeri razionali α_ε e β_ε tali che $\sqrt{2} - \varepsilon < \alpha_\varepsilon < \sqrt{2} < \beta_\varepsilon < \sqrt{2} + \varepsilon$, se vogliamo che la monotonia si conservi anche per la potenza ad esponente reale, dovrà essere

$$2^{\alpha_\varepsilon} < 2^{\sqrt{2}} < 2^{\beta_\varepsilon}.$$

Siccome α_ε e β_ε sono arbitrariamente vicini, lo saranno anche 2^{α_ε} e 2^{β_ε} ⁽²⁹⁾, che, in qualche modo, “stringono” tra di loro un unico numero reale, che è quello che definiamo come $2^{\sqrt{2}}$.

⁽²⁹⁾Perché?

Questo procedimento, che può essere reso rigoroso, si può estendere a qualsiasi numero reale $x \geq 0$ ed a qualsiasi numero reale y : l'idea è proprio quella di usare da un lato la monotonia della potenza ad esponente razionale (che garantisce il fatto che, avvicinandosi ad y reale con dei razionali i risultati non "oscillino" troppo), e dall'altro la densità dei razionali nei reali (che permette di approssimare bene quanto si vuole y con dei razionali).

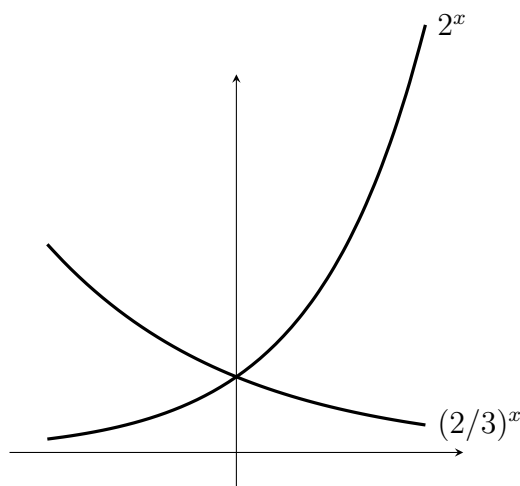
Una maniera alternativa, ma equivalente, di definire $2^{\sqrt{2}}$ è quella di usare in maniera diversa la monotonia e la densità dei razionali: detto

$$E = \{x \in \mathbb{R} : x = 2^q \text{ con } q \text{ razionale e } q \leq \sqrt{2}\},$$

è facile vedere che E è limitato superiormente (ad esempio ogni x di E è minore di 4). Definiamo allora

$$2^{\sqrt{2}} = \sup E.$$

Con lo stesso procedimento, fissato $x \geq 0$ e y reale è possibile definire x^y (sempre con l'avvertenza che se $x = 0$, allora y deve essere non negativo). La prossima figura mostra il grafico di 2^x (la curva crescente) e di $(\frac{2}{3})^x$ (la curva decrescente). Le due curve si incrociano nel punto $x = 0$ e $y = 1$ (dal momento che $z^0 = 1$ per ogni z).



Come è evidente dal grafico⁽³⁰⁾, qualsiasi sia $A > 0$ con $A \neq 1$, la curva A^x è definita su tutto \mathbb{R} e ha come immagine l'intervallo illimitato $(0, +\infty)$.

ESERCIZIO 8.7. Dimostrare che se $f(x) = 2^x$, detto $E = \text{Im}(f) = f(\mathbb{R})$, allora $\inf E = 0$ e $\sup E = +\infty$.

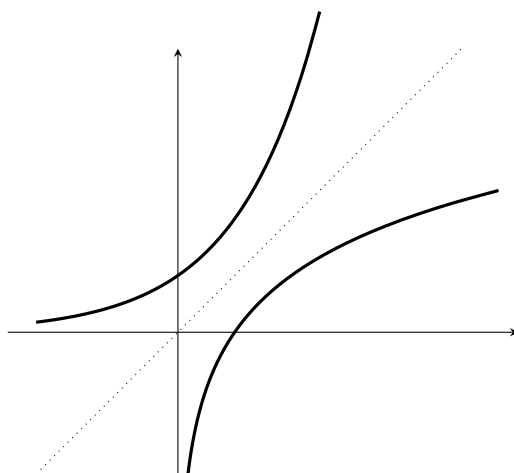
Risposta 8.7: L'insieme E contiene l'insieme dei numeri naturali della forma 2^n , e tale insieme è illimitato superiormente essendo $2^n > n$ per ogni n in \mathbb{N} (dimostrarlo per induzione!). Per mostrare che 0 è l'estremo inferiore, si osservi che E contiene solo numeri positivi (cosicché l'insieme dei minoranti di E contiene $(-\infty, 0]$) e numeri della forma 2^{-n} , che diventano arbitrariamente piccoli al crescere di n .

Sia ora $f : A \rightarrow \mathbb{R}$ una funzione. Per definizione, f associa ad ogni elemento x di A un unico numero reale $y = f(x)$, che ovviamente appartiene ad $f(A)$. Una domanda che ci si può porre è se, fissato y in $f(A)$ (per il quale esiste sicuramente

⁽³⁰⁾Il che vuol dire che potrebbe anche non essere vero...

almeno un x di A tale che $y = f(x)$, esista un **unico** x di A tale che $y = f(x)$. La risposta è, in generale, negativa. Ad esempio, la funzione f definita su \mathbb{R} da $f(x) = 1$ (ovvero la funzione che assume costantemente il valore 1) non verifica questa proprietà, dato che a $y = 1$ corrisponde non un unico x , bensì infiniti. Un altro esempio è la funzione $f(x) = x^2$, definita per x reale. Infatti, se y è un qualsiasi numero reale positivo, esistono esattamente due numeri reali, \sqrt{y} e $-\sqrt{y}$, che hanno come immagine y . Analogamente per $f(x) = |x|$. Le funzioni $f(x) = x$, $f(x) = x^3$ e $f(x) = 2^x$, invece, godono di questa proprietà: ogni numero reale nella loro immagine proviene da uno ed un solo numero del dominio.

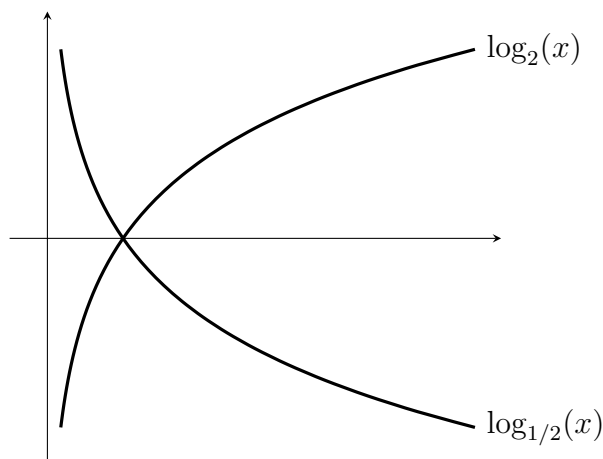
DEFINIZIONE 8.8. Sia $f : A \rightarrow \mathbb{R}$ una funzione. Se per ogni y di $f(A)$ esiste un unico x in A tale che $y = f(x)$, la funzione f si dice **invertibile**. Data una funzione invertibile resta determinata una funzione $g : f(A) \rightarrow A$, detta funzione **inversa**, che è la funzione che a y in $f(A)$ associa l'unico x di A tale che $y = f(x)$. La funzione g è dunque caratterizzata dall'essere tale che, per ogni x di A , $g(f(x)) = x$, mentre per ogni y di $f(A)$ si ha $f(g(y)) = y$. In genere, con una notazione che può trarre in inganno, la funzione inversa viene denotata con f^{-1} . Dal punto di vista del grafico, dato il grafico di una funzione invertibile $y = f(x)$, il grafico della funzione inversa (sugli stessi assi con i nomi cambiati), si ottiene “ribaltando” il grafico rispetto alla retta $y = x$; si veda la figura.



DEFINIZIONE 8.9. Sia $A > 0$, con $A \neq 1$. La funzione A^x , definita su \mathbb{R} a valori in $(0, +\infty)$, è invertibile. Definiamo **logaritmo in base A** la funzione inversa di A^x , e indichiamo tale funzione con \log_A . Come funzione, dunque, $\log_A : (0, +\infty) \rightarrow \mathbb{R}$ è caratterizzata dall'essere

$$\log_A(A^x) = x, \quad \forall x \in \mathbb{R}, \quad A^{\log_A(y)} = y, \quad \forall y \in (0, +\infty).$$

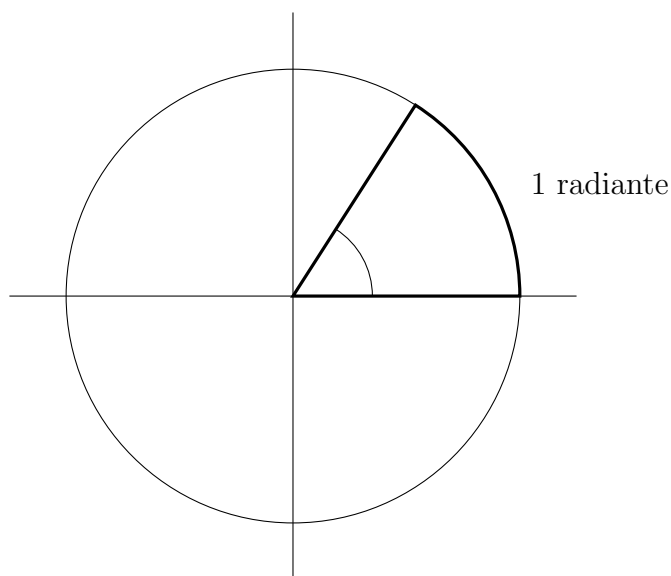
Il grafico che segue rappresenta le curve $\log_2(x)$ (la curva crescente) e $\log_{\frac{1}{2}}(x)$ (la curva decrescente).



Le due curve si intersecano nel punto $x = 1, y = 0$. Una delle proprietà “storicamente” più importanti dei logaritmi, conseguenza del fatto che $A^{x+y} = A^x \cdot A^y$, è il trasformare i prodotti in somme⁽³¹⁾:

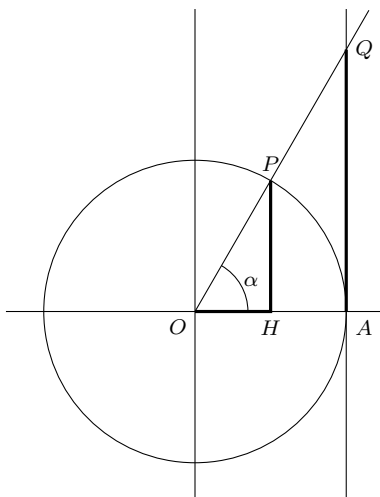
$$\log_A(x \cdot y) = \log_A(x) + \log_A(y), \quad \forall x, y \in (0, +\infty).$$

Nell'introdurre radici n -sime, potenze ad esponente razionale e reale, funzioni esponenziali e logaritmi, siamo partiti dal numero “costruibile” $\sqrt{2}$, la diagonale del quadrato di lato 1. Esistono altri modi di “costruire” esplicitamente funzioni. Uno di questi consiste nell'utilizzare la circonferenza di centro l'origine e raggio 1, e gli angoli misurati in radianti. Ricordiamo che un angolo al centro di una circonferenza di raggio 1 misura un **radiante** se sottende un arco di circonferenza di lunghezza 1. Con questa convenzione, l'angolo retto misura $\frac{\pi}{2}$ radianti, l'angolo piatto π radianti, mentre l'angolo giro misura 2π radianti. Inoltre, gli angoli sono misurati muovendosi lungo la circonferenza in verso antiorario a partire dal punto $x = 1, y = 0$. Nella figura che segue, è mostrato l'angolo che misura un radiante.



⁽³¹⁾E non le somme in prodotti, come talvolta si legge nei compiti scritti. Per esercizio, dimostrare che se il logaritmo della somma fosse il prodotto dei logaritmi, la funzione logaritmo sarebbe identicamente nulla...

Sia ora α un angolo (misurato in radianti a partire dal punto A di coordinate $x = 1$, $y = 0$) compreso tra 0 e 2π , e consideriamo il punto P tale che l'arco di circonferenza AP abbia lunghezza α . A partire da P tracciamo la perpendicolare all'asse x , che incontra tale asse nel punto H . Definiamo **seno** dell'angolo α , indicato con $\text{sen}(\alpha)$, la lunghezza con segno del segmento PH , e **coseno** dell'angolo α , indicato con $\text{cos}(\alpha)$, la lunghezza con segno del segmento OH . Infine, tracciata la tangente alla circonferenza in A , sia Q il punto di intersezione tra la semiretta OP e la tangente. Definiamo **tangente** di α , indicata con $\text{tg}(\alpha)$, la lunghezza con segno del segmento AQ .



Dalla similitudine dei triangoli OPH e OQA si ottiene subito la relazione fondamentale tra seno, coseno e tangente:

$$\text{tg}(\alpha) = \frac{\text{sen}(\alpha)}{\text{cos}(\alpha)}, \quad \forall \alpha \neq \frac{\pi}{2}, \frac{3\pi}{2}.$$

Abbiamo così definito tre funzioni: $\text{sen} : [0, 2\pi] \rightarrow \mathbb{R}$, $\text{cos} : [0, 2\pi] \rightarrow \mathbb{R}$ e $\text{tg} : [0, 2\pi] \setminus \{\frac{\pi}{2}, \frac{3\pi}{2}\} \rightarrow \mathbb{R}$. È facile vedere che $\text{Im}(\text{sen}) = [-1, 1] = \text{Im}(\text{cos})$, cosicché seno e coseno soddisfano le maggiorazioni fondamentali

$$|\text{sen}(\alpha)| \leq 1, \quad |\text{cos}(\alpha)| \leq 1,$$

mentre $\text{Im}(\text{tg}) = \mathbb{R}$. Infine, grazie al Teorema di Pitagora,

$$\text{sen}^2(\alpha) + \text{cos}^2(\alpha) = 1.$$

Che succede se l'angolo α non è compreso tra 0 e 2π ? Ad esempio, se $\alpha = 2\pi + 1$? È chiaro che il punto P che otteniamo è lo stesso che otteniamo per $\alpha = 1$, e quindi $\text{sen}(2\pi + 1) = \text{sen}(1)$, e lo stesso vale per coseno e tangente. Più in generale, se k è un intero relativo, e α è un numero reale qualsiasi, si ha

$$\text{sen}(\alpha + 2k\pi) = \text{sen}(\alpha), \quad \text{cos}(\alpha + 2k\pi) = \text{cos}(\alpha), \quad \text{tg}(\alpha + 2k\pi) = \text{tg}(\alpha).$$

Queste proprietà si esprimono dicendo che le funzioni seno, coseno e tangente sono **periodiche** di **periodo** 2π . Pertanto, le funzioni seno e coseno si possono considerare definite su tutto \mathbb{R} (sempre con la stessa immagine), mentre la tangente è definita su \mathbb{R} privato dei multipli dispari di $\frac{\pi}{2}$. Una volta estese ai reali, le funzioni seno e coseno

soddisfano le relazioni, dette formule di addizione,

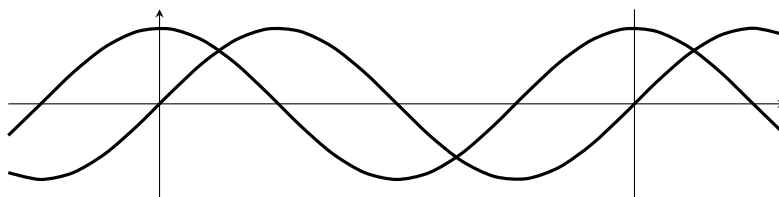
$$\begin{aligned}\operatorname{sen}(\alpha \pm \beta) &= \operatorname{sen}(\alpha) \cos(\beta) \pm \operatorname{sen}(\beta) \cos(\alpha), \\ \cos(\alpha \pm \beta) &= \cos(\alpha) \cos(\beta) \mp \operatorname{sen}(\alpha) \operatorname{sen}(\beta).\end{aligned}$$

Partendo da queste formule, e dalla relazione fondamentale tra seno, coseno e tangente, è facile vedere che

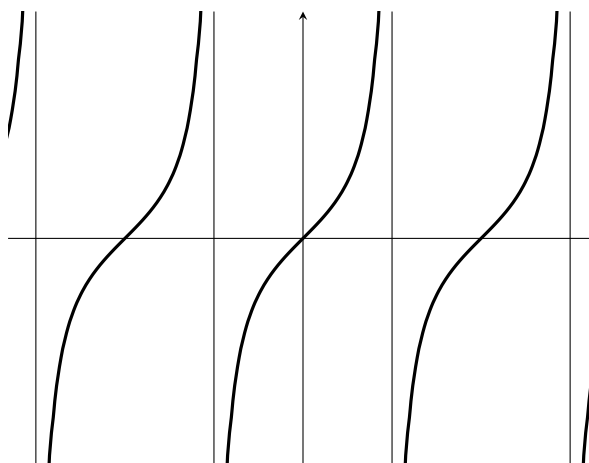
$$\operatorname{tg}(\alpha + k\pi) = \operatorname{tg}(\alpha), \quad \forall k \in \mathbb{Z},$$

e quindi la tangente è periodica di periodo π .

Graficamente abbiamo la seguente situazione⁽³²⁾:

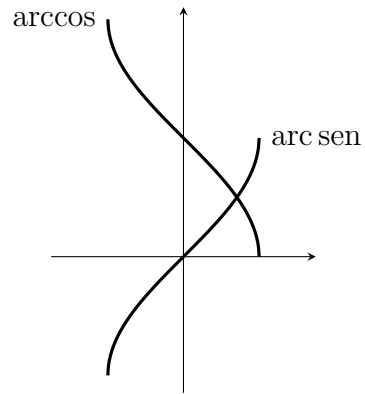


mentre per la tangente abbiamo:

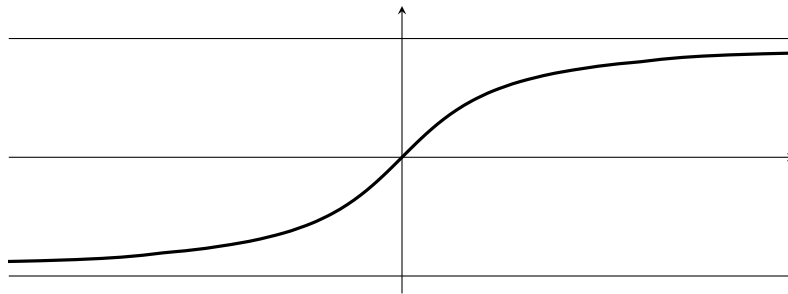


Essendo periodiche (e quindi assumendo infinite volte lo stesso valore), le funzioni seno, coseno e tangente non sono ovviamente invertibili. Se, però, restringiamo nella maniera seguente l'insieme di definizione: $\operatorname{sen} : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-1, 1]$, $\operatorname{cos} : [0, \pi] \rightarrow [-1, 1]$ e $\operatorname{tg} : (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}$, le funzioni risultano essere invertibili. Abbiamo allora l'**arcoseno**, indicato con $\operatorname{arc} \operatorname{sen}$, l'**arcocoseno**, indicato con $\operatorname{arc} \operatorname{cos}$, e l'**arcotangente**, indicato con $\operatorname{arc} \operatorname{tg}$. Le prime due funzioni sono definite su $[-1, 1]$, a valori in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ e $[0, \pi]$ rispettivamente, mentre l'arcotangente è definita su tutto \mathbb{R} a valori in $(-\frac{\pi}{2}, \frac{\pi}{2})$. Graficamente, abbiamo:

⁽³²⁾Al lettore decidere quali delle due curve è il grafico di $\operatorname{sen}(x)$, quale di $\operatorname{cos}(x)$, sapendo che la prima verticale è $x = 0$.



per arcoseno ed arcocoseno, mentre il grafico dell'arcotangente è:



Successioni di numeri reali

1. Definizioni ed esempi. Limite di una successione.

Nell'ultimo paragrafo del capitolo precedente abbiamo introdotto alcune funzioni elementari da (sottoinsiemi di) \mathbb{R} a valori in \mathbb{R} , e abbiamo studiato sommariamente alcune loro proprietà. Facciamo adesso un passo indietro, ed invece di considerare funzioni che associano ad ogni numero reale x un numero reale $y = f(x)$, consideriamo delle “versioni ridotte”: invece di far variare l'argomento della funzione su \mathbb{R} , ci limitiamo a farlo variare su \mathbb{N} . In questo capitolo considereremo dunque funzioni $f : \mathbb{N} \rightarrow \mathbb{R}$. “Funzioni” di questo tipo le incontriamo quotidianamente, ad esempio tutte le volte che ci mettiamo in fila (se occupiamo il decimo posto, noi siamo il valore della funzione per $n = 10$), o quando leggiamo un libro (il contenuto della pagina 105 è il valore della funzione per $n = 105$), o quando guardiamo la televisione (il canale due è il valore della funzione per $n = 2$). L'unica differenza con la vita reale (o naturale...) è che nel nostro caso la fila è infinita, il libro ha infinite pagine, e la televisione ha infiniti canali⁽¹⁾. Per distinguere meglio le “funzioni” da \mathbb{R} in \mathbb{R} dalle “funzioni” da \mathbb{N} in \mathbb{R} , si usano una denominazione diversa ed una notazione particolare.

DEFINIZIONE 1.1. Una **successione** di numeri reali è una funzione $a : \mathbb{N} \rightarrow \mathbb{R}$. Il valore $a(n)$ si indica in genere con a_n ; l'insieme dei valori della successione (ovvero, l'immagine $a(\mathbb{N})$) si indica con $\{a_n, n \in \mathbb{N}\}$, oppure $\{a_n\}_{n \in \mathbb{N}}$, o anche $\{a_n\}$.

ESEMPIO 1.2. Sia $a_n = 1$ per ogni n in \mathbb{N} ; questa successione, e più in generale la successione tale che $a_n = a$ per ogni n in \mathbb{N} , con a numero reale fissato, si dice successione **costante**. Ovviamente, $\{a_n\} = \{1\}$ (o $\{a_n\} = \{a\}$).

Sia $a_n = n$ per ogni n in \mathbb{N} ; questa successione, corrispondente alla funzione identità, non è altro che la successione dei numeri naturali. Pertanto, $\{a_n\} = \mathbb{N}$.

Sia

$$a_n = \begin{cases} \frac{n}{2} & \text{se } n \text{ è pari,} \\ -\frac{n+1}{2} & \text{se } n \text{ è dispari.} \end{cases}$$

Calcolando i primi termini della successione troviamo $a_0 = 0$, $a_1 = -1$, $a_2 = 1$, $a_3 = -2$, $a_4 = 2$, cosicché i valori assunti dalla successione oscillano a sinistra e a destra di zero, e si ottengono man mano tutti gli interi relativi⁽²⁾. Pertanto, $\{a_n\} = \mathbb{Z}$.

Sia $a_n = (-1)^n$. In questo caso la successione è fatta così: $a_0 = 1$, $a_1 = -1$, $a_2 = 1$, $a_3 = -1$, $a_4 = 1$; pertanto la successione oscilla tra 1 e -1. Ovviamente, $\{a_n\} = \{1, -1\}$.

⁽¹⁾In tutti e tre i casi, una catastrofe!

⁽²⁾Sarà vero? Dimostatelo!

Sia $a_n = \frac{1}{n+1}$; in questo caso, i valori assunti dalla successione diventano sempre più piccoli al crescere di n . Sono tutti positivi, nessuno di essi è zero, ma fissato un qualunque $\varepsilon > 0$ esiste un numero naturale $n^{(3)}$, dipendente da ε , a partire dal quale tutti gli a_n sono minori di ε . In questo caso, l'immagine della successione non ha “nulla di particolare” oltre la definizione della successione, nel senso che

$$\{a_n\} = \left\{ \frac{1}{n+1}, n \in \mathbb{N} \right\} = \left\{ 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots \right\}.$$

ESERCIZIO 1.3. Determinare le successioni a_n tali che

$$\begin{aligned} \{a_n\} &= \{\text{numeri pari}\}, \\ \{a_n\} &= \{\text{numeri dispari}\}, \\ \{a_n\} &= \{\text{potenze di due}\}, \\ \{a_n\} &= \{1, 1, 2, 3, 5, 8, 13, 21, 34, \dots\} = \{\text{numeri di Fibonacci}\}, \\ \{a_n\} &= \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \dots\} = \{\text{numeri primi}\}. \end{aligned}$$

Risposta 1.3: 4) L'esercizio è abbastanza complicato. La successione dei numeri di Fibonacci è definita dalla relazione $a_n = a_{n-1} + a_{n-2}$ per $n \geq 2$, con $a_0 = a_1 = 1$. Se cerchiamo una “soluzione” della forma A^n con A numero reale, e sostituiamo, si vede che deve essere $A^2 - A - 1 = 0$, da cui $A_{1,2} = \frac{1 \pm \sqrt{5}}{2}$. Siccome sia A_1^n che A_2^n sono soluzioni, lo è (perché?) ogni loro combinazione lineare:

$$a_n = \alpha \left(\frac{1 + \sqrt{5}}{2} \right)^n + \beta \left(\frac{1 - \sqrt{5}}{2} \right)^n.$$

Assegnando le condizioni per $n = 0$ e $n = 1$ si trova $\alpha + \beta = 1$ e $\alpha - \beta = \frac{\sqrt{5}}{5}$, da cui

$$a_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^{n+1} - \left(\frac{1 - \sqrt{5}}{2} \right)^{n+1} \right].$$

Si osservi il fatto — apparentemente paradossale — che la formula appena scritta assume valori **interi** per ogni n in \mathbb{N} . 5) Non esiste una “formula per i numeri primi”, nonostante sia “facile” dimostrare che sono infiniti. Infatti, supponendo che siano solo in numero finito, siano essi p_1, p_2, \dots, p_n . Se consideriamo il numero intero $p = p_1 \cdot p_2 \cdot \dots \cdot p_n + 1$, allora p è primo, non essendo divisibile per p_1 (la divisione dà 1 come resto), per p_2 (la divisione dà 1 come resto), e così via fino a p_n . Pertanto, supponendo che i numeri primi siano finiti, siamo arrivati ad una contraddizione.

Avendo definito le successioni, come le utilizziamo? La motivazione principale dell'introduzione delle successioni è la seguente: data una successione di numeri reali $\{a_n\}$, sui cui valori abbiamo un controllo totale finché n è piccolo (nel senso che possiamo calcolare esplicitamente i valori che assume la successione), vogliamo determinare — se esiste — un numero reale che ci dia un'indicazione sul comportamento dei valori della successione per n molto grande. In altre parole, vogliamo “distillare” dalla successione un numero reale che ci dia un'idea, approssimata sì, ma con un errore che possiamo rendere piccolo, di cosa accade ai valori di a_n facendo crescere n .

Approssimata in che senso? Esattamente nel senso che abbiamo usato per definire la densità di \mathbb{Q} in \mathbb{R} . Si era detto: per ogni numero reale x , e fissato comunque $\varepsilon > 0$, esiste un numero razionale q_ε tale che $x - \varepsilon < q_\varepsilon < x + \varepsilon$, cioè tale che $|x - q_\varepsilon| < \varepsilon$; in altre

⁽³⁾Quale?

parole, q_ε dista meno di ε da x . Ovvero, a patto di commettere un errore ε che possiamo però scegliere arbitrariamente piccolo, ogni numero reale x può essere approssimato con un razionale q_ε . Potremmo allora dire che un certo numero reale L è il “distillato” della successione $\{a_n\}$ se per ogni $\varepsilon > 0$ esiste un numero naturale n_ε tale che $L - \varepsilon < a_{n_\varepsilon} < L + \varepsilon$; ovvero se, fissato un errore ε arbitrario, esiste un elemento della successione che approssima L a meno di ε . Una tale definizione, però, sarebbe disastrosa; innanzitutto perché nessuno ci garantisce che l’indice n_ε sia “grande” (ricordiamo che vogliamo estrarre dalla successione l’informazione sul suo comportamento per valori di n molto grandi), e poi perché ogni elemento della successione può essere scelto come L : infatti, qualsiasi sia $\varepsilon > 0$, $a_n - \varepsilon < a_n < a_n + \varepsilon$; in altre parole, un qualsiasi elemento della successione è “approssimato” da sé stesso.

Come fare, allora, per “forzare” un determinato numero reale a darci informazioni sui valori della successione per n arbitrariamente grande? Una possibilità sarebbe quella di sostituire la richiesta che **un** elemento della successione sia vicino ad L , con la richiesta che **tutti** gli elementi della successione siano vicini a L a meno di un errore arbitrariamente piccolo. Ovvero che, per ogni $\varepsilon > 0$, si abbia $L - \varepsilon < a_n < L + \varepsilon$ per ogni n in \mathbb{N} . Se la richiesta precedente era troppo debole, questa è però troppo forte; è facile vedere, infatti, che se $L - \varepsilon < a_n < L + \varepsilon$ per ogni n in \mathbb{N} e per ogni $\varepsilon > 0$, allora deve essere obbligatoriamente $a_n = L$ per ogni n in \mathbb{N} ; vale a dire, che sapremmo estrarre informazione sul comportamento di a_n per n grande solo quando tale comportamento fosse già sotto i nostri occhi: se a_n è una successione costante, è sufficiente conoscere il primo elemento per sapere “dove si andrà a parare”.

Né ovviamente, sempre guidati dal nostro interesse per quello che accade per n grande, vale spostare verso destra la soglia di approssimazione; ovvero richiedere che esista un m in \mathbb{N} tale che per ogni $\varepsilon > 0$ si abbia $L - \varepsilon < a_n < L + \varepsilon$ per ogni $n \geq m$, perché questa disuguaglianza implica che a_n è costantemente uguale ad L a partire dall’ n -simo elemento: un’altra volta, l’informazione su quello che accade alla successione per n molto grande, è già contenuta nei valori per n “piccolo”.

A questo punto, non ci resta che una sola cosa da fare: chiedere che la “soglia di ingresso” nell’intervallo $(L - \varepsilon, L + \varepsilon)$ **dipenda dall’errore**, ovvero che, fissato $\varepsilon > 0$, debba esistere n_ε in \mathbb{N} tale che $L - \varepsilon < a_n < L + \varepsilon$, per ogni $n \geq n_\varepsilon$. A partire da un certo punto in poi (ma che dipende dall’errore che accettiamo di fare), tutti gli elementi della successione approssimano L . Siamo allora pronti per dare la definizione chiave di questo capitolo.

DEFINIZIONE 1.4. Sia $\{a_n\}$ una successione di numeri reali. Diremo che a_n **ammette limite per n tendente ad infinito**, ovvero che **converge**, o che **tende**, se esiste un numero reale L tale che

$$(1.1) \quad \forall \varepsilon > 0, \exists n_\varepsilon \in \mathbb{N} : |a_n - L| < \varepsilon, \quad \forall n \geq n_\varepsilon.$$

Se la successione a_n ammette limite L , scriveremo

$$L = \lim_{n \rightarrow +\infty} a_n,$$

che si legge “ L è il limite di a_n per n tendente ad infinito”. Osserviamo che la definizione di limite richiede che la distanza (misurata dal valore assoluto) tra a_n e L diventi arbitrariamente piccola per n sufficientemente grande.

ESEMPIO 1.5. Studiamo le successioni dell'Esempio 1.2.

Sia $a_n = 1$ per ogni n in \mathbb{N} ; pensando all'idea di limite come numero reale approssimato dai valori della successione per n grande, prendiamo $L = 1$; è evidente che qualsiasi sia $\varepsilon > 0$ si ha $|a_n - 1| = |1 - 1| = 0 < \varepsilon$ per ogni n in \mathbb{N} , vale a dire che la (1.1) vale scegliendo $n_\varepsilon = 0$. Quindi, $L = 1$ è un possibile limite della successione a_n . Chiaramente non esistono altri valori per L , perché se L è diverso da 1, allora $|a_n - L| = |1 - L| > \frac{|1-L|}{2}$ per ogni n in \mathbb{N} , e quindi la (1.1) non è verificata per $\varepsilon = \frac{|1-L|}{2}$; vale a dire, l'errore non può essere reso arbitrariamente piccolo. In maniera analoga si dimostra che il limite della successione che vale costantemente a è proprio a .

ESERCIZIO 1.6. Sia $\{a_n\}$ la successione che vale il primo numero estratto nell' n -sima estrazione del lotto sulla ruota di Roma se n è minore di 1000, e 1 se $n \geq 1000$. Calcolare, se esiste, il limite di a_n .

Sia $a_n = n$ per ogni n in \mathbb{N} ; in questo caso, è evidente che non esiste alcun limite L : qualsiasi numero reale fissiamo, e qualsiasi sia $\varepsilon > 0$, la successione cadrà fuori dall'intervallo $(L - \varepsilon, L + \varepsilon)$ non appena $n > [L + \varepsilon]$. Pertanto, per nessun valore di L è soddisfatta la (1.1).

Sia

$$a_n = \begin{cases} \frac{n}{2} & \text{se } n \text{ è pari,} \\ -\frac{n+1}{2} & \text{se } n \text{ è dispari.} \end{cases}$$

Anche in questo caso, non esiste alcun limite. Per dimostrarlo è sufficiente osservare che se $n \neq 0$, allora $|a_n| \geq 1$, e che a_n cambia segno a seconda della parità di n . Se esistesse un valore di L soddisfacente la (1.1), tale valore non potrebbe essere positivo (perché l'intervallo $(L - \varepsilon, L + \varepsilon)$ contiene solo numeri positivi se $\varepsilon < L$, mentre esistono infiniti elementi della successione negativi), non potrebbe essere negativo (per lo stesso motivo), e non potrebbe essere zero, dato che a_n non appartiene all'intervallo $(-\frac{1}{2}, \frac{1}{2})$ per ogni $n \geq 1$.

Sia $a_n = (-1)^n$. Ricordando che in questo caso i valori della successione oscillano tra 1 e -1 , un ragionamento analogo al precedente mostra che non esiste alcun limite.

Sia $a_n = \frac{1}{n+1}$; come abbiamo già detto, i valori assunti dalla successione diventano sempre più piccoli al crescere di n . Essendo tutti positivi un "candidato limite" è $L = 0$. Ed infatti, fissato $\varepsilon > 0$ e scelto $n_\varepsilon = \lceil \frac{1}{\varepsilon} \rceil$ si ha, se $n \geq n_\varepsilon$,

$$|a_n - 0| = \frac{1}{n+1} \leq \frac{1}{n_\varepsilon + 1} < \frac{1}{\frac{1}{\varepsilon}} = \varepsilon,$$

e quindi la (1.1) è soddisfatta. Si dimostra anche facilmente che nessun valore di $L \neq 0$ soddisfa la (1.1), e che quindi $L = 0$ è l'unico limite della successione a_n .

Abbiamo dunque visto che non tutte le successioni ammettono limite, ma nei due esempi nei quali esisteva un limite, tale valore era anche unico. Non era un caso, come mostra il seguente teorema.

TEOREMA 1.7 (Unicità del limite). *Sia a_n una successione di numeri reali che ammette limite. Allora il limite è unico.*

Dimostrazione. Supponiamo per assurdo che la successione a_n ammetta due limiti. Ovvero, che esistano due numeri reali L e M , con $L \neq M$, tali che

$$\lim_{n \rightarrow +\infty} a_n = L, \quad \lim_{n \rightarrow +\infty} a_n = M.$$

Sia ora $\varepsilon > 0$, e scriviamo la (1.1) per L ed M :

$$\begin{aligned} \exists n'_\varepsilon \in \mathbb{N} : |a_n - L| < \varepsilon, & \quad \forall n \geq n'_\varepsilon, \\ \exists n''_\varepsilon \in \mathbb{N} : |a_n - M| < \varepsilon, & \quad \forall n \geq n''_\varepsilon. \end{aligned}$$

Sia ora $n \geq n_\varepsilon = \max(n'_\varepsilon, n''_\varepsilon)$, cosicch  entrambe le disuguaglianze sono vere. Riscrivendole, e ricordando che $|-x| = |x|$ e che $|x| < a$ equivale a $-a < x < a$, si ottiene

$$-\varepsilon < L - a_n < \varepsilon, \quad -\varepsilon < a_n - M < \varepsilon.$$

Sommando le due relazioni, si ottiene

$$-2\varepsilon < L - M < 2\varepsilon \iff |L - M| < 2\varepsilon.$$

Dal momento che la disuguaglianza precedente vale per ogni $\varepsilon > 0$, sar  in particolare vera per $\varepsilon = \frac{|L-M|}{3}$ (che   positivo essendo $L \neq M$). Si ottiene cos 

$$|L - M| < \frac{2|L - M|}{3},$$

che   assurda. ■

Grazie al teorema precedente,   sufficiente mostrare che esiste **un** valore di L che soddisfa la (1.1) per ottenere automaticamente che quello ottenuto   **il** valore del limite della successione. Se il precedente teorema aiuta in una direzione (quella dell'unicit  del limite), non esistono teoremi che aiutino nella direzione della determinazione del valore corretto di L : l'unico aiuto si chiama "esercizio"...

ESERCIZIO 1.8. Date le successioni

$$\begin{aligned} a_n &= \frac{1}{n^2}, \\ a_n &= \frac{(-1)^n}{2n+3}, \\ a_n &= \frac{n}{n+1}, \\ a_n &= \frac{\alpha n + \beta}{\gamma n + \delta}, \quad \gamma \neq 0, \\ a_n &= \operatorname{sen} \left(\frac{1}{n} \right), \\ a_n &= \operatorname{cos} \left(\frac{1}{n} \right), \end{aligned}$$

determinare, se esiste, il valore del limite; successivamente dimostrare che tale valore   effettivamente il limite mostrando che la (1.1)   soddisfatta. Negli ultimi due casi, si usi la definizione geometrica di seno e coseno, ricordando cosa vuol dire che un angolo misura $\frac{1}{n}$ radianti.

Risposta 1.8: 4) Il valore del limite   $\frac{\alpha}{\gamma}$. Infatti, si ha

$$\left| \frac{\alpha n + \beta}{\gamma n + \delta} - \frac{\alpha}{\gamma} \right| = \frac{|\alpha\delta - \beta\gamma|}{|\gamma||\gamma n + \delta|},$$

che si può scrivere nella forma $\frac{A}{Bn+C}$ per opportuni numeri **positivi** A , B e C ⁽⁴⁾. A questo punto si tratta di determinare n_ε tale che $\frac{A}{Bn+C} < \varepsilon$ per ogni $n \geq n_\varepsilon$; è sufficiente scegliere $n \geq \left[\frac{A-C\varepsilon}{B\varepsilon}\right] + 1$.

Come osservazione finale, si noti che la successione a_n ha come limite L se e solo se la successione $b_n = a_n - L$ (o la successione $c_n = L - a_n$, ovvero la successione $d_n = |a_n - L|$) ha come limite 0.

2. Convergenza e limitatezza.

Nel paragrafo dedicato alle proprietà dei numeri reali del precedente capitolo abbiamo introdotto il concetto di estremo superiore ed inferiore di sottoinsiemi di \mathbb{R} , e abbiamo affermato che ogni insieme limitato ha estremo inferiore e superiore finiti. Anche l'insieme $\{a_n\}$ dei valori assunti da una successione a_n è un sottoinsieme di \mathbb{R} e pertanto, qualora fosse limitato, ammetterebbe sia estremo superiore che estremo inferiore. Detta **limitata** una successione a_n il cui insieme dei valori sia limitato, ci chiediamo se dall'esistenza di estremo superiore e inferiore **finiti** si possa ottenere qualche informazione sull'esistenza del limite.

Sfortunatamente, no. Se consideriamo la successione $a_n = (-1)^n$, è facile vedere che a_n è limitata, ma non è convergente. In altre parole, la limitatezza di una successione non basta a garantire l'esistenza del limite. C'è però un legame importante "nell'altro senso": tra convergenza di una successione e limitatezza.

TEOREMA 2.1. *Sia a_n una successione convergente ad L . Allora l'insieme $\{a_n\}$ dei valori della successione è limitato. Ovvero, esiste $M \geq 0$ tale che $|a_n| \leq M$ per ogni n in \mathbb{N} .*

Dimostrazione. Scegliamo $\varepsilon = 1$ nella (1.1). Esiste allora n_ε in \mathbb{N} tale che

$$|a_n - L| < 1, \quad \forall n \geq n_\varepsilon,$$

il che vuol dire che $L - 1 < a_n < L + 1$ per ogni $n \geq n_\varepsilon$. Pertanto, $|a_n| < |L| + 1$ per ogni $n \geq n_\varepsilon$. Sia ora

$$M = \max(|a_1|, |a_2|, \dots, |a_{n_\varepsilon-1}|, |L| + 1).$$

Allora $|a_n| \leq M$ per ogni n in \mathbb{N} . Infatti, se $n < n_\varepsilon$, allora

$$|a_n| \leq \max(|a_1|, |a_2|, \dots, |a_n|, \dots, |a_{n_\varepsilon-1}|, |L| + 1) = M,$$

mentre se $n \geq n_\varepsilon$ si ha

$$|a_n| \leq |L| + 1 \leq \max(|a_1|, |a_2|, \dots, |a_n|, \dots, |a_{n_\varepsilon-1}|, |L| + 1) = M,$$

come volevasi dimostrare. ■

In definitiva, l'esistenza del limite implica la limitatezza, mentre il contrario non è vero. Inoltre, dal teorema precedente segue anche che una successione illimitata non può ammettere come limite un numero reale. Come già nel caso degli insiemi illimitati, cerchiamo di ovviare a questo inconveniente introducendo il concetto di limite "infinito". Vogliamo cioè formalizzare il concetto di successione i cui valori diventano arbitrariamente grandi al crescere di n .

⁽⁴⁾Perché positivi?

DEFINIZIONE 2.2. Sia a_n una successione di numeri reali. Diremo che a_n **diverge a più infinito**, e scriveremo

$$\lim_{n \rightarrow +\infty} a_n = +\infty,$$

se per ogni $M > 0$ esiste n_M in \mathbb{N} tale che $a_n > M$ per ogni $n \geq n_M$. Diremo che a_n **diverge a meno infinito**, e scriveremo

$$\lim_{n \rightarrow +\infty} a_n = -\infty,$$

se per ogni $M < 0$ esiste n_M in \mathbb{N} tale che $a_n < M$ per ogni $n \geq n_M$.

L'esempio più evidente di successione divergente a più infinito (se volete, tendente a più infinito) è la successione dei numeri naturali: fissato un qualsiasi $M > 0$, è sufficiente scegliere $n_M = [M] + 1$. Questo fatto spiega anche la notazione $n \rightarrow +\infty$ che compare nella scrittura “compatta” del limite.

ESERCIZIO 2.3. Date le successioni

$$\begin{aligned} a_n &= n^2 + n, \\ a_n &= 2^n, \\ a_n &= n - n^2, \\ a_n &= \frac{n^2}{n+2}, \\ a_n &= \cos(n\pi) n, \\ a_n &= \operatorname{tg}\left(\frac{\pi}{2} - \frac{1}{n}\right), \quad b_n = \operatorname{tg}\left(\frac{\pi}{2} + \frac{1}{n}\right), \end{aligned}$$

si calcoli il limite (se esiste) per n tendente ad infinito. Per le ultime due successioni si usi un approccio “grafico”.

Come già per la coppia “esistenza del limite–limitatezza”, la divergenza (a più infinito o a meno infinito) della successione a_n implica l'illimitatezza dell'insieme dei suoi valori, mentre il contrario non è vero; se consideriamo la successione $a_n = (-1)^n n$, abbiamo una successione illimitata che non ammette limite dal momento che cambia segno a seconda della parità di \mathbb{N} .

3. Teoremi di confronto.

Supponiamo di avere una successione di numeri reali non negativi a_n , e supponiamo di sapere che a_n ammetta limite. Che possiamo dire del segno del limite?

TEOREMA 3.1. Sia $a_n \geq 0$ una successione di numeri reali convergente ad L . Allora $L \geq 0$.

Dimostrazione. Sia $\varepsilon > 0$, e sia n_ε in \mathbb{N} tale che

$$|a_n - L| < \varepsilon, \quad \forall n \geq n_\varepsilon.$$

Allora si ha

$$0 \leq a_n < L + \varepsilon,$$

e quindi $L + \varepsilon > 0$ per ogni $\varepsilon > 0$. Essendo ε arbitrario, ne segue⁽⁵⁾ che $L \geq 0$, come volevasi dimostrare. ■

⁽⁵⁾Se L fosse negativo, si avrebbe un assurdo per $\varepsilon = -\frac{L}{2} > 0$.

Si noti che l'ipotesi “ $a_n \geq 0$ per ogni n in \mathbb{N} ” può essere sostituita dalla più debole “ $a_n \geq 0$ per ogni n abbastanza grande”, dato che è stata usata solo per $n \geq n_\varepsilon$. Inoltre, anche se $a_n > 0$ per ogni n in \mathbb{N} , la conclusione del teorema precedente è comunque vera solo con il minore od uguale. Ad esempio, $a_n = \frac{1}{n+1}$ converge a zero, e si ha $a_n > 0$ per ogni n .

Si osservi che se abbiamo due successioni a_n e b_n tali che $a_n \leq b_n$ per ogni n (o per n abbastanza grande), e se a_n converge a L e b_n a M , il teorema precedente applicato alla successione non negativa $b_n - a_n$ implica che $L \leq M$.

È vero il contrario del Teorema 3.1? Ovvero, se $L \geq 0$, è vero che $a_n \geq 0$ per ogni n , o almeno per n grande? Ovviamente no se $L = 0$; ad esempio $a_n = \frac{(-1)^n}{n+1}$ converge a zero, ma cambia segno. Se, invece, il limite è positivo, allora il risultato è vero per n grande.

TEOREMA 3.2 (Permanenza del segno). *Sia a_n una successione di numeri reali convergente a $K > 0$. Allora esiste n_0 in \mathbb{N} tale che $a_n > \frac{K}{2}$ per ogni $n \geq n_0$. Sia a_n una successione di numeri reali convergente a $K < 0$. Allora esiste n_0 in \mathbb{N} tale che $a_n < \frac{K}{2}$ per ogni $n \geq n_0$. In particolare, a_n ha lo stesso segno del limite se $n \geq n_0$.*

Dimostrazione. Supponiamo $K > 0$, e prendiamo $\varepsilon = \frac{K}{2}$ nella definizione (1.1). Esiste allora n_K in \mathbb{N} tale che

$$|a_n - K| < \frac{K}{2}, \quad \forall n \geq n_K.$$

Dalla disuguaglianza precedente si ottiene

$$0 < \frac{K}{2} < a_n < \frac{3K}{2}, \quad \forall n \geq n_K,$$

da cui la tesi prendendo $n_0 = n_K$. La seconda parte del teorema si dimostra in maniera analoga scegliendo $\varepsilon = -\frac{K}{2}$. ■

Per ottenere il risultato nel caso di due successioni a_n e b_n con limiti $L < M$ basta applicare il teorema precedente alla differenza $b_n - a_n$, convergente a $K = M - L > 0$.

OSSERVAZIONE 3.3. Osserviamo infine che, se oltre a convergere ad un limite $L \neq 0$, la successione non si annulla mai, allora esiste $K > 0$ tale che

$$(3.1) \quad |a_n| \geq K, \quad \forall n \in \mathbb{N}.$$

Infatti, per il teorema precedente, esiste n_0 in \mathbb{N} tale che $|a_n| > \frac{L}{2}$ per ogni $n \geq n_0$. Sia poi

$$M = \min(|a_1|, |a_2|, \dots, |a_{n_0-1}|) > 0,$$

per le ipotesi su a_n . Scegliendo $K = \min(\frac{L}{2}, M)$ si ha la tesi.

Si noti che sia nel Teorema 3.2 che nella precedente Osservazione l'ipotesi $L \neq 0$ non può essere eliminata; ad esempio, la successione $a_n = \frac{(-1)^n}{n+1}$ converge a zero ma è sia positiva che negativa, ed è diversa da zero per ogni n ma non è (in modulo) maggiore di un numero strettamente positivo.

Abbiamo visto che se $a_n \leq b_n$, allora il limite di a_n è minore od uguale al limite di b_n . Che succede se una successione b_n è “incastrata” tra due successioni che ammettono lo stesso limite?

TEOREMA 3.4 (Teorema dei Carabinieri). *Siano a_n, b_n e c_n tre successioni di numeri reali tali che $a_n \leq b_n \leq c_n$ per ogni n in \mathbb{N} . Se a_n e c_n convergono entrambe allo stesso limite L , allora la successione b_n ammette limite, e tale limite è L .*

Dimostrazione. Sia $\varepsilon > 0$ e siano n'_ε e n''_ε due numeri naturali tali che

$$|a_n - L| < \varepsilon, \quad \forall n \geq n'_\varepsilon, \quad |c_n - L| < \varepsilon, \quad \forall n \geq n''_\varepsilon.$$

Sia $n_\varepsilon = \max(n'_\varepsilon, n''_\varepsilon)$ e sia $n \geq n_\varepsilon$. Allora

$$L - \varepsilon < a_n < L + \varepsilon, \quad L - \varepsilon < c_n < L + \varepsilon,$$

e quindi

$$L - \varepsilon < a_n \leq b_n \leq c_n < L + \varepsilon \iff |b_n - L| < \varepsilon,$$

come volevasi dimostrare. ■

Ancora una volta, non è necessario che b_n sia compresa tra a_n e c_n per ogni n in \mathbb{N} ; affinché il teorema sia vero, è sufficiente che lo sia per n sufficientemente grande (ovvero, da un certo valore di n in poi). Si osservi che il risultato del Teorema dei Carabinieri è differente dagli altri dimostrati in precedenza perché la successione b_n non è a priori convergente.

Nel caso in cui una delle due successioni diverga, è sufficiente che controlli l'altra "dal lato giusto" affinché si possa calcolare il limite.

TEOREMA 3.5. *Siano a_n e b_n due successioni di numeri reali tali che $a_n \geq b_n$ per ogni n in \mathbb{N} . Se b_n diverge a $+\infty$, anche a_n diverge a $+\infty$, mentre se a_n diverge a $-\infty$ anche b_n vi diverge.*

Come prima, alla frase " $a_n \geq b_n$ per ogni n in \mathbb{N} " si può sostituire la frase " $a_n \geq b_n$ per ogni n abbastanza grande". Si noti che, come prima, non si ipotizza né che a_n ammetta limite (nel primo caso), né che lo abbia b_n (nel secondo).

4. Operazioni con i limiti. Forme indeterminate.

Sia $a_n = \frac{1}{n} + \frac{1}{n^3} + \frac{1}{n^5}$. Quanto vale il limite di a_n ? Essendo a_n la somma di tre successioni ognuna delle quali tende a zero, ci aspettiamo che a_n diventi molto piccola al crescere di n , e che quindi tenda a zero anche essa. Il problema è che per dimostrare che a_n è effettivamente tendente a zero, l'unico mezzo a nostra disposizione è la definizione (1.1), il che vuol dire che dobbiamo risolvere una disequazione di quinto grado, operazione questa complicata, e forse impossibile. Fortunatamente, la nostra intuizione (a_n tende a zero come somma di tre successioni che tendono a zero) è in realtà il caso particolare di un teorema più generale.

TEOREMA 4.1. *Siano a_n e b_n due successioni, con a_n convergente ad L e b_n convergente ad M . Allora si ha*

$$\lim_{n \rightarrow +\infty} (a_n \pm b_n) = L \pm M,$$

o, in altre parole, "il limite della somma (della differenza) è la somma (la differenza) dei limiti".

Dimostrazione. Sia $\varepsilon > 0$ fissato, e siano n'_ε e n''_ε i due interi tali che

$$(4.1) \quad |a_n - L| < \frac{\varepsilon}{2}, \quad \forall n \geq n'_\varepsilon, \quad |b_n - M| < \frac{\varepsilon}{2}, \quad \forall n \geq n''_\varepsilon.$$

Sia $n \geq n_\varepsilon = \max(n'_\varepsilon, n''_\varepsilon)$, cosicché entrambe le disuguaglianze sono verificate. Allora, per la disuguaglianza triangolare,

$$(4.2) \quad |(a_n + b_n) - (L + M)| \leq |a_n - L| + |b_n - M| < \varepsilon,$$

e quindi il limite di $a_n + b_n$ è $L + M$. La dimostrazione della formula per la differenza dei limiti è lasciata al lettore. ■

Prima di procedere oltre, torniamo un momento sulla formula (4.1). Abbiamo correttamente usato la definizione di limite (1.1)? Apparentemente, no, dato che in (1.1) compare ε , mentre noi abbiamo scritto $\frac{\varepsilon}{2}$, che è un valore diverso. Ebbene, non abbiamo fatto nessun errore, se “leggiamo” quello che dice la (1.1): fissata comunque una quantità positiva, esiste un intero, dipendente da questa quantità, tale che... Come si vede, la “quantità positiva” è “muta”, il che vuol dire che ci si può permettere di chiamarla come si vuole: ε , $\frac{\varepsilon}{2}$, ε^2 , o in un modo qualsiasi⁽⁶⁾. Nel caso della (4.1) abbiamo deciso (con il senno di poi) di chiamarla $\frac{\varepsilon}{2}$ in modo da ottenere ε nella (4.2), ma avremmo potuto continuare a chiamarla ε , nel qual caso avremmo ottenuto 2ε nella (4.2). In entrambi i casi, data l'arbitrarietà di ε , da cui segue l'arbitrarietà di $\frac{\varepsilon}{2}$, o di 2ε , avremmo scritto correttamente la definizione di limite. Il che vuol dire che, da ora in poi, ci disinteresseremo del “numero” degli ε che troveremo nella disuguaglianza finale; quando dovremo applicare la definizione di limite useremo ε , il che vuol dire che “pagheremo” con un numero eventualmente complicato di ε nella formula finale.

Il primo esempio di questa strategia è il seguente teorema.

TEOREMA 4.2. *Siano a_n e b_n due successioni, con a_n convergente ad L e b_n convergente ad M . Allora si ha*

$$\lim_{n \rightarrow +\infty} a_n \cdot b_n = L \cdot M,$$

o, in altre parole, “il limite del prodotto è il prodotto dei limiti”.

Dimostrazione. Sia $\varepsilon > 0$ fissato, e siano n'_ε e n''_ε i due interi tali che

$$|a_n - L| < \varepsilon, \quad \forall n \geq n'_\varepsilon, \quad |b_n - M| < \varepsilon, \quad \forall n \geq n''_\varepsilon.$$

Sia $n \geq n_\varepsilon = \max(n'_\varepsilon, n''_\varepsilon)$, cosicché entrambe le disuguaglianze sono verificate. Allora

$$\begin{aligned} |a_n \cdot b_n - L \cdot M| &= |a_n \cdot b_n - a_n \cdot M + a_n \cdot M - L \cdot M| \\ &\leq |a_n| \cdot |b_n - M| + |M| \cdot |a_n - L|. \end{aligned}$$

Il secondo addendo è minore di $|M| \cdot \varepsilon$, ma il primo? In esso compare $|a_n|$, che dipende da n . Fortunatamente, per il Teorema 2.1 la successione a_n , essendo convergente, è limitata. Ovvero, esiste $K \geq 0$ tale che $|a_n| \leq K$ per ogni n in \mathbb{N} . Pertanto $|a_n| \cdot |b_n - M| \leq K \cdot |b_n - M| < K \cdot \varepsilon$. In definitiva,

$$|a_n \cdot b_n - L \cdot M| < (K + |M|) \cdot \varepsilon, \quad \forall n \geq n_\varepsilon,$$

e quindi il limite di $a_n \cdot b_n$ è $L \cdot M$. ■

⁽⁶⁾Tranne $-\varepsilon$ (se ε è positivo, ovviamente).

Si osservi che nella dimostrazione precedente il fattore che moltiplica ε è, in un certo senso, “sconosciuto”, nel senso che K , che sappiamo esistere, ci è ignoto, ed è indipendente da L .

Passiamo adesso a trattare il caso del rapporto tra due successioni, a patto che siano soddisfatte delle ipotesi aggiuntive.

TEOREMA 4.3. *Siano a_n e b_n due successioni, con a_n convergente ad L e b_n convergente ad M . Supponiamo che b_n sia diverso da zero per ogni n in \mathbb{N} , e che M sia diverso da zero. Allora*

$$\lim_{n \rightarrow +\infty} \frac{a_n}{b_n} = \frac{L}{M},$$

o, in altre parole, “il limite del rapporto è il rapporto dei limiti”.

Dimostrazione. La dimostrazione segue dal teorema precedente, una volta che si sia dimostrato che se b_n è una successione mai nulla, e convergente ad un limite M diverso da zero, allora la successione $\frac{1}{b_n}$ converge ad $\frac{1}{M}$. Per dimostrare questo fatto, osserviamo che se $\varepsilon > 0$ è fissato, e se $n \geq n_\varepsilon$ (dato da (1.1) per b_n), si ha

$$\left| \frac{1}{b_n} - \frac{1}{M} \right| = \frac{|M - b_n|}{|b_n| \cdot |M|} \leq \frac{\varepsilon}{|b_n| \cdot |M|}.$$

Siccome b_n converge a $M \neq 0$, e b_n è diversa da zero per ogni n in \mathbb{N} , per l'Osservazione 3.3 esiste $K > 0$ tale che $|b_n| \geq K$ per ogni n in \mathbb{N} ; se ne deduce allora che $\frac{1}{|b_n|} \leq \frac{1}{K}$, e quindi

$$\left| \frac{1}{b_n} - \frac{1}{M} \right| \leq \frac{\varepsilon}{K \cdot M},$$

da cui la tesi. ■

La condizione $b_n \neq 0$ del teorema precedente è entrata in gioco due volte; una prima volta affinché sia ben definita la successione $\frac{a_n}{b_n}$, ed una seconda volta per applicare l'Osservazione 3.3. In entrambi i casi, per controllare la stretta positività del modulo di b_n per un numero finito di indici. Per gli infiniti n maggiori di n_0 , infatti, è stato il Teorema della permanenza del segno a far sì che il modulo di b_n fosse maggiore di una costante strettamente positiva. Pertanto, il Teorema 4.3 si può enunciare eliminando la condizione $b_n \neq 0$ per ogni n in \mathbb{N} e lasciando solo la richiesta $M \neq 0$.

Un altro caso molto particolare, nel quale possiamo dedurre l'esistenza del limite del prodotto di due successioni anche nel caso in cui una delle due non ammetta limite, è il seguente.

TEOREMA 4.4. *Sia a_n una successione tendente a zero, e sia b_n una successione limitata. Allora $a_n \cdot b_n$ tende a zero.*

Dimostrazione. Sia $\varepsilon > 0$ e sia n_ε in \mathbb{N} tale che

$$|a_n| < \varepsilon, \quad \forall n \geq n_\varepsilon.$$

Sia ora $M \geq 0$ tale che $|b_n| \leq M$ per ogni n in \mathbb{N} . Se $n \geq n_\varepsilon$ si ha allora

$$|a_n \cdot b_n| = |a_n| \cdot |b_n| < M \cdot \varepsilon,$$

come volevasi dimostrare. ■

Che cosa possiamo dire quando una, o entrambe le successioni divergono? O se la successione b_n nel rapporto $\frac{a_n}{b_n}$ tende a zero? I risultati sono elencati nel seguente teorema.

TEOREMA 4.5. *Siano a_n e b_n sue successioni di numeri reali. Abbiamo allora i seguenti risultati:*

- se a_n e b_n divergono entrambe a $\pm\infty$, allora $a_n + b_n$ diverge a $\pm\infty$;
- se a_n diverge a $\pm\infty$ e b_n diverge a $\mp\infty$, allora $a_n - b_n$ diverge a $\pm\infty$;
- se a_n diverge a $\pm\infty$ e b_n converge a M , allora $a_n \pm b_n$ diverge a $\pm\infty$;
- se a_n converge a L e b_n diverge a $\pm\infty$, allora $a_n + b_n$ diverge a $\pm\infty$;
- se a_n converge a L e b_n diverge a $\pm\infty$, allora $a_n - b_n$ diverge a $\mp\infty$;
- se a_n e b_n divergono entrambe a $\pm\infty$, $a_n \cdot b_n$ diverge a $+\infty$;
- se a_n diverge a $\pm\infty$ e b_n diverge a $\mp\infty$, allora $a_n \cdot b_n$ diverge a $-\infty$;
- se a_n diverge a $\pm\infty$ e b_n converge a $M > 0$, allora $a_n \cdot b_n$ diverge a $\pm\infty$;
- se a_n diverge a $\pm\infty$ e b_n converge a $M < 0$, allora $a_n \cdot b_n$ diverge a $\mp\infty$;
- se a_n converge a L e b_n diverge a $\pm\infty$, allora $\frac{a_n}{b_n}$ converge a 0;
- se a_n diverge a $\pm\infty$ e b_n converge a zero da valori positivi, allora $\frac{a_n}{b_n}$ diverge a $\pm\infty$;
- se a_n diverge a $\pm\infty$ e b_n converge a zero da valori negativi, allora $\frac{a_n}{b_n}$ diverge a $\mp\infty$;
- se a_n converge ad $L > 0$ e b_n converge a zero da valori positivi, allora $\frac{a_n}{b_n}$ diverge a $+\infty$;
- se a_n converge ad $L < 0$ e b_n converge a zero da valori positivi, allora $\frac{a_n}{b_n}$ diverge a $-\infty$;
- se a_n converge ad $L > 0$ e b_n converge a zero da valori negativi, allora $\frac{a_n}{b_n}$ diverge a $-\infty$;
- se a_n converge ad $L < 0$ e b_n converge a zero da valori negativi, allora $\frac{a_n}{b_n}$ diverge a $+\infty$.

Il teorema precedente permette di calcolare diversi casi, ma ne lascia aperti altri; per la precisione

- il limite di $a_n + b_n$ quando a_n diverge a $\pm\infty$ e b_n diverge a $\mp\infty$;
- il limite di $a_n - b_n$ quando a_n e b_n divergono entrambe a $\pm\infty$;
- il limite di $a_n \cdot b_n$ quando a_n diverge a $\pm\infty$ e b_n converge a zero;
- il limite di $\frac{a_n}{b_n}$ quando a_n e b_n divergono entrambe (a $+\infty$ o a $-\infty$);
- il limite di $\frac{a_n}{b_n}$ quando a_n e b_n convergono entrambe a zero.

Questi casi costituiscono le cosiddette “forme indeterminate” che spesso, con una notazione impropria, vengono indicate come $\infty - \infty$, $0 \cdot \infty$, $\frac{\infty}{\infty}$ e $\frac{0}{0}$. Lo “scioglimento” delle forme indeterminate non può essere risolto in maniera generale, in quanto l’esistenza ed il valore del limite dipendono dalle particolari successioni considerate, come nel seguente esercizio.

ESERCIZIO 4.6. Si calcoli (se esiste) il limite di $\frac{a_n}{b_n}$ nei casi che seguono:

$$\begin{aligned} a_n &= \frac{1}{n+1}, & b_n &= \frac{1}{(n+2)^2}, \\ a_n &= \frac{1}{(n+3)^2}, & b_n &= \frac{1}{n+4}, \\ a_n &= \frac{1}{n+1}, & b_n &= \frac{1}{n+2}, \\ a_n &= \frac{(-1)^n}{n+1}, & b_n &= \frac{1}{n+2}, \\ a_n &= \frac{(-1)^n}{n+1}, & b_n &= \frac{1}{n^2+2}. \end{aligned}$$

ESERCIZIO 4.7. Si calcoli, se esiste, il limite delle seguenti successioni:

$$1 + \frac{1}{n}, \quad \cos(\pi n), \quad \arctg\left(\frac{n}{2}\right) + \arctg\left(\frac{3}{n}\right), \quad \left[\operatorname{sen}\left(\frac{1}{n}\right) + \cos\left(\frac{1}{n}\right)\right]^2, \quad \frac{1}{\sqrt{n}} + \frac{4}{\ln(n)}.$$

$$\frac{n^2+1}{n+3}, \quad \frac{n^3+3n^2-1}{n^3+6n+3}, \quad \frac{n^2-2}{n^3-6}, \quad \frac{n^6-n^5}{n^2(n^2+1)(n^2+2)}, \quad \frac{an^p+bn^{p-1}+\dots}{cn^q+dn^{q-1}+\dots}.$$

$$n - \sqrt{n-1}, \quad \sqrt{n^2+n} - \sqrt{n^2+1}, \quad \sqrt[3]{n^2+6n} - \sqrt[3]{n^2+5}, \quad \sqrt[m]{n+1} - \sqrt[m]{n} \quad (m \in \mathbb{N}).$$

$$n \operatorname{sen}\left(\frac{1}{n}\right), \quad \frac{\operatorname{sen}(a_n)}{a_n}, \quad \text{con } a_n \text{ tendente a zero.}$$

Risposta 4.7: Rispondiamo all'ultimo esercizio. Il primo limite è una forma indeterminata del tipo $0 \cdot \infty$. Osserviamo però che, per definizione di seno, tangente, e di angolo misurato in radianti (quest'ultimo fatto è fondamentale!), si ha

$$\operatorname{sen}\left(\frac{1}{n}\right) \leq \frac{1}{n} \leq \operatorname{tg}\left(\frac{1}{n}\right) = \frac{\operatorname{sen}\left(\frac{1}{n}\right)}{\cos\left(\frac{1}{n}\right)},$$

da cui

$$\cos\left(\frac{1}{n}\right) \leq n \operatorname{sen}\left(\frac{1}{n}\right) \leq 1.$$

Dal momento che $\cos\left(\frac{1}{n}\right)$ converge a 1, il Teorema dei Carabinieri permette di concludere: $n \operatorname{sen}\left(\frac{1}{n}\right)$ converge a 1. Un ragionamento analogo permette di affermare che anche il secondo limite vale 1.

Vediamo ora come si comportano i limiti con gli esponenziali. Iniziamo con un caso particolare.

PROPOSIZIONE 4.8. *Le successioni $2^{\frac{1}{n}}$ e $2^{-\frac{1}{n}}$ convergono a 1.*

Dimostrazione. Sia $b_n = 2^{\frac{1}{n}}$. Siccome $2 > 1$, $b_n > 1^{\frac{1}{n}} = 1$, e quindi $b_n = 1 + c_n$, con $c_n > 0$. Se dimostriamo che c_n tende a zero, b_n tende a 1 e la prima parte della proposizione è dimostrata. Ricordando che la radice n -sima è l'operazione inversa della potenza n -sima, si ha $2 = b_n^n = (1 + c_n)^n$. Per la disuguaglianza di Bernoulli (si veda il primo capitolo), si ha

$$2 = (1 + c_n)^n \geq 1 + n c_n,$$

da cui $c_n \leq \frac{1}{n}$. Essendo $c_n > 0$, e $\frac{1}{n}$ tendente a zero, possiamo applicare il Teorema dei Carabinieri e concludere che c_n tende a zero. Una volta dimostrato che $2^{\frac{1}{n}}$ tende a 1, il fatto che $2^{-\frac{1}{n}}$ tenda a 1 segue dal Teorema 4.3 e dal fatto che $2^{-\frac{1}{n}} = \frac{1}{2^{\frac{1}{n}}}$. ■

Si osservi che la dimostrazione continua ad essere valida sostituendo a 2 un qualsiasi numero $A > 1$. Che succede se $0 < A < 1$?

PROPOSIZIONE 4.9. *Sia a_n una successione di numeri reali tendente a zero. Allora la successione 2^{a_n} converge a 1.*

Dimostrazione. Sia $m \geq 1$ un numero naturale fissato. Definendo $\varepsilon = \frac{1}{m}$, e ricordando che a_n tende a zero, esiste n_m in \mathbb{N} tale che

$$|a_n| < \frac{1}{m}, \quad \forall n \geq n_m,$$

ovvero

$$-\frac{1}{m} < a_n < \frac{1}{m}, \quad \forall n \geq n_m.$$

Si ha allora, ricordando che se $x > y$ si ha $2^x > 2^y$,

$$2^{-\frac{1}{m}} < 2^{a_n} < 2^{\frac{1}{m}}, \quad \forall n \geq n_m.$$

Sia ora $\varepsilon > 0$ e siano m'_ε e m''_ε due numeri naturali (che esistono per la Proposizione 4.8) tali che

$$2^{\frac{1}{m}} - 1 < \varepsilon, \quad \forall m \geq m'_\varepsilon, \quad 2^{-\frac{1}{m}} - 1 > -\varepsilon, \quad \forall m \geq m''_\varepsilon.$$

Sia ora $m_\varepsilon = \max(m'_\varepsilon, m''_\varepsilon)$; allora

$$-\varepsilon < 2^{-\frac{1}{m_\varepsilon}} - 1 < 2^{a_n} - 1 < 2^{\frac{1}{m_\varepsilon}} - 1 < \varepsilon, \quad \forall n \geq n_{m_\varepsilon},$$

ovvero

$$|2^{a_n} - 1| < \varepsilon, \quad \forall n \geq n_{m_\varepsilon},$$

e quindi la tesi. ■

PROPOSIZIONE 4.10. *Sia $A > 0$ e sia a_n una successione tendente a zero. Allora A^{a_n} tende ad 1.*

Dimostrazione. Scriviamo

$$A^{a_n} = 2^{\log_2(A^{a_n})} = 2^{a_n \log_2(A)},$$

e applichiamo la dimostrazione precedente, osservando che la successione $a_n \log_2(A)$ tende a zero. ■

TEOREMA 4.11. *Sia $A > 0$ e sia a_n una successione di numeri reali convergente ad L . Allora A^{a_n} converge ad A^L .*

Dimostrazione. Scriviamo

$$A^{a_n} - A^L = A^L (A^{a_n - L} - 1).$$

Dal momento che $a_n - L$ tende a zero, la successione tra parentesi converge a zero per la Proposizione 4.10, e quindi A^{a_n} tende ad A^L , come volevasi dimostrare. ■

Il teorema precedente si può anche invertire.

TEOREMA 4.12. *Sia $A > 0$ e sia a_n una successione di numeri reali tali che A^{a_n} converge ad A^L . Allora a_n converge ad L .*

Dimostrazione. Sia $A > 1$ e supponiamo per assurdo che a_n non converga ad L . Questo vuol dire che esiste un $\bar{\varepsilon} > 0$ tale che $|a_n - L| \geq \bar{\varepsilon}$ per infiniti valori di n . Pertanto, per infiniti n in \mathbb{N} si ha $a_n \geq L + \bar{\varepsilon}$ o $a_n \leq L - \bar{\varepsilon}$. Per gli stessi n si ha $A^{a_n} \geq A^L A^{\bar{\varepsilon}}$, oppure $A^{a_n} \leq A^L A^{-\bar{\varepsilon}}$, e quindi $A^{a_n} - A^L \geq A^L(A^{\bar{\varepsilon}} - 1)$ oppure $A^{a_n} - A^L \leq A^L(A^{-\bar{\varepsilon}} - 1)$. In entrambi i casi, il fatto che A^{a_n} converga ad A^L viene contraddetto. Il caso $0 < A < 1$ si risolve in maniera identica. ■

TEOREMA 4.13. *Sia $A > 0$ e $A \neq 1$, e sia a_n una successione di numeri reali positivi convergente ad $L \neq 0$. Allora $\log_A(a_n)$ converge a $\log_A(L)$.*

Dimostrazione. Sia $b_n = \log_A(a_n)$, cosicché $a_n = A^{b_n}$; per ipotesi, A^{b_n} converge a L , che scriviamo come $A^{\log_A(L)}$. Per il Teorema precedente, b_n converge a $\log_A(L)$. ■

TEOREMA 4.14. *Sia a_n una successione di numeri reali non negativi convergente a $L > 0$. Sia α un numero reale. Allora a_n^α converge a L^α .*

Dimostrazione. È sufficiente scrivere

$$a_n^\alpha = 2^{\log_2(a_n^\alpha)} = 2^{\alpha \log_2(a_n)}.$$

Per il teorema precedente, $\log_2(a_n)$ converge a $\log_2(L)$, e per il Teorema 4.11 $2^{\alpha \log_2(a_n)}$ converge a $2^{\alpha \log_2(L)} = L^\alpha$. ■

TEOREMA 4.15. *Sia a_n una successione di numeri reali non negativi convergente a $L > 0$, e sia b_n una successione di numeri reali convergente a M . Allora $a_n^{b_n}$ converge a L^M .*

Dimostrazione. Come nel teorema precedente, scriviamo

$$a_n^{b_n} = 2^{\log_2(a_n^{b_n})} = 2^{b_n \log_2(a_n)},$$

e applichiamo il Teorema 4.13 ed il Teorema 4.11. ■

Così come nel caso delle operazioni “standard” (somma, prodotto, rapporto), anche nel caso dell’elevamento a potenza sono possibili alcune forme indeterminate, e precisamente quelle del tipo 0^0 , ∞^0 , e 1^∞ .

ESEMPIO 4.16. Abbiamo visto nella Proposizione 4.8 che $\sqrt[n]{2} = 2^{\frac{1}{n}}$ converge ad 1, così come tende ad 1 la radice n -sima di ogni numero reale positivo. Che succede con la radice n -sima di n ? Siamo in presenza di una forma indeterminata del tipo ∞^0 , e quindi è necessario un “trucco” (oltre all’onnipresente disuguaglianza di Bernoulli). Sia $a_n = \sqrt[n]{n}$, e sia $b_n = \sqrt{a_n}$. Dal momento che $a_n > 1$, anche $b_n > 1$, e quindi $b_n = 1 + c_n$, con $c_n > 0$. Si ha

$$1 + c_n = b_n = \sqrt{a_n} = \sqrt{\sqrt[n]{n}} = \sqrt[n]{\sqrt{n}},$$

cosicché

$$\sqrt{n} = (1 + c_n)^n \geq 1 + n c_n \iff 0 < c_n \leq \frac{\sqrt{n} - 1}{n}.$$

Siccome $\frac{\sqrt{n}-1}{n}$ tende a zero, anche c_n tende a zero. Pertanto, b_n tende ad 1, e quindi $\sqrt[n]{n}$ converge ad 1.

ESERCIZIO 4.17. Si calcolino i seguenti limiti

$$\left(1 + \frac{1}{n}\right)^{\frac{3}{n^4}}, \quad (1 + n^2)^{\frac{1}{n}}, \quad \sqrt[n]{2^n + 3^n}, \quad \sqrt[n]{A_1^n + A_2^n + \dots + A_k^n} \quad (A_i > 0).$$

Risposta 4.17: Il penultimo (la cui generalizzazione è l'ultimo):

$$\sqrt[n]{2^n + 3^n} = \sqrt[n]{3^n \left(1 + \frac{2^n}{3^n}\right)} = 3 \sqrt[n]{1 + \frac{2^n}{3^n}} \rightarrow 3.$$

5. Monotonia e convergenza.

Nel paragrafo 2 abbiamo detto che se a_n è una successione convergente ad L , allora è una successione limitata. È facile allora vedere⁽⁷⁾ che L è compreso tra l'estremo inferiore e l'estremo superiore della successione a_n . Cerchiamo, con alcuni esempi, di capire meglio che legami ci possono essere tra il limite e l'estremo inferiore e l'estremo superiore della successione.

ESERCIZIO 5.1. Date le successioni

$$\begin{aligned} a_n &= \frac{1}{n+1}, \\ a_n &= \frac{1}{n^2+1}, \\ a_n &= n^2 - n, \\ a_n &= \frac{n^2 + 3}{n^2 + 1}, \\ a_n &= \frac{(-1)^n}{n+1}, \end{aligned}$$

si calcolino il limite di a_n , $\inf \{a_n\}$ e $\sup \{a_n\}$.

In quattro casi su cinque il limite della successione (finito o infinito) è o l'estremo superiore o l'estremo inferiore dei valori della successione; nell'ultimo caso, il limite è zero, mentre il minimo della successione è $-\frac{1}{2}$, il massimo 1. Esaminiamo la prima successione: possiamo, al crescere di n , "confrontare" fra di loro i valori della successione? Evidentemente sì: essendo $n+1 > n$ si ha $n+2 > n+1$ e quindi $a_{n+1} = \frac{1}{n+2} < \frac{1}{n+1} = a_n$. Pertanto, i valori della successione decrescono al crescere di n . Notiamo che in questo caso il limite, zero, è l'estremo inferiore dei valori della successione; il massimo è evidentemente $a_0 = 1$. Per la seconda successione, si ha $a_n = \frac{1}{n^2+1} = 1 - \frac{1}{n^2+1}$; in questo caso (per quanto visto prima) si ha $a_{n+1} > a_n$, e quindi la successione cresce con n . Il limite di a_n vale 1, che è l'estremo superiore dei valori della successione (il minimo è $a_0 = 0$). Per la terza successione, i primi valori sono 0, 0, 2, 6, 12, ...; ancora una volta la successione cresce con n ⁽⁸⁾, ed il limite della successione, più infinito, coincide con l'estremo superiore.

⁽⁷⁾Esercizio!

⁽⁸⁾Controllando i valori di n fino a 5 possiamo congetturare che l'andamento sia questo. Dimostrare che la disequazione $a_{n+1} \geq a_n$ è soddisfatta per ogni n .

Abbiamo visto negli esempi precedenti che se i valori della successione sono “ordinati” (sia per ogni valore dell’indice n , che per n abbastanza grande), il limite della successione coincide con quello tra estremo superiore o inferiore che si “accorda” con il comportamento della successione (se la successione cresce, l’estremo superiore, altrimenti l’estremo inferiore). Ci chiediamo ora se questo sia un comportamento generale, o sia dovuto ai pochi esempi presentati. Diamo prima una definizione.

DEFINIZIONE 5.2. Sia a_n una successione di numeri reali. La successione si dice **monotona crescente** se si ha $a_{n+1} \geq a_n$ per ogni n in \mathbb{N} . La successione si dice **monotona decrescente** se si ha $a_{n+1} \leq a_n$ per ogni n in \mathbb{N} . Se il segno di \geq (\leq) è sostituito dal segno $>$ ($<$), la successione si dice **strettamente crescente** (**strettamente decrescente**).

Possiamo allora enunciare il risultato principale di questo paragrafo.

TEOREMA 5.3. *Sia a_n una successione di numeri reali monotona crescente. Allora la successione ammette come limite l’estremo superiore di $\{a_n\}$. Sia a_n una successione di numeri reali monotona decrescente. Allora la successione ammette come limite l’estremo inferiore di $\{a_n\}$.*

OSSERVAZIONE 5.4. Si osservi che negli esempi dell’Esercizio 5.1 avevamo verificato che il limite coincideva con l’estremo superiore (o con l’estremo inferiore). Il risultato del teorema appena enunciato è molto più forte: ci dice che la successione 1) ammette limite e 2) che tale limite è l’estremo superiore o inferiore dei valori della successione (a seconda del “verso” della monotonia della successione).

Dimostrazione. Supponiamo che la successione sia monotona crescente e limitata. In questo caso, l’estremo superiore della successione, sia esso L , è finito. Per definizione, L è un numero reale tale che

$$(5.1) \quad a_n \leq L, \quad \forall n \in \mathbb{N},$$

$$(5.2) \quad \forall \varepsilon > 0 \exists n_\varepsilon \in \mathbb{N} : L - \varepsilon < a_{n_\varepsilon} \leq L.$$

La (5.2) non è altro che la b’) nella definizione di estremo superiore, applicata nel nostro caso: la b’) afferma che esiste un elemento x_ε dell’insieme E di cui L è l’estremo superiore tale che $L - \varepsilon < x_\varepsilon \leq L$; essendo gli elementi della successione dipendenti da n , l’elemento x_ε della successione corrisponderà ad un certo indice naturale n_ε . Sia ora $n \geq n_\varepsilon$. Essendo la successione monotona crescente, si ha $a_n \geq a_{n_\varepsilon}$; pertanto

$$L - \varepsilon < [(5.2)] < a_{n_\varepsilon} \leq a_n \leq [(5.1)] \leq L < L + \varepsilon,$$

da cui $|a_n - L| < \varepsilon$, e quindi a_n converge ad L .

Supponiamo ora che la successione sia monotona decrescente e illimitata; in questo caso l’estremo inferiore dei valori della successione è meno infinito, il che vuol dire che per ogni $M < 0$ esiste n_M in \mathbb{N} tale che $a_{n_M} \leq M$ (ancora una volta, stiamo identificando gli elementi della successione con i loro indici). Sia ora $n \geq n_M$; essendo la successione decrescente,

$$a_n \leq a_{n_M} \leq M,$$

e quindi la successione diverge a meno infinito. Gli altri due casi (crescente e illimitata, decrescente e limitata) sono lasciati per esercizio. ■

Che succede se la successione non è monotona, ma “monotona per n grande”? Ovvero, se esiste n_0 in \mathbb{N} tale che, ad esempio, $a_{n+1} \geq a_n$ per ogni $n \geq n_0$? È ancora vero che la successione ammette limite, ma in questo caso il limite non è più l'estremo superiore (o inferiore) di tutti i valori della successione, ma solo dei valori per $n \geq n_0$. Che non sia possibile che il limite sia in generale l'estremo superiore (o inferiore) di tutta la successione, basti pensare all'esempio di una successione che decresce verso zero (ad esempio $\frac{1}{n+1}$) e della quale modifichiamo il primo valore, scegliendo un numero reale negativo (ad esempio, $a_0 = -1$). La successione è decrescente per $n \geq 1$, il limite è sempre zero, ma l'estremo inferiore è (nell'esempio) -1 .

ESERCIZIO 5.5. Sia a_n una successione monotona decrescente e limitata. Sia L il limite della successione (che è dunque l'estremo inferiore dei valori della successione) e supponiamo che $L = \min\{a_n\}$. Come è fatta la successione a_n ?

Risposta 5.5: Esiste n_0 in \mathbb{N} tale che $a_n = L$ per ogni $n \geq n_0$. Infatti, essendo L il minimo dei valori della successione, esiste n_1 in \mathbb{N} tale che $a_{n_1} = L$. Siccome a_n è decrescente, se $n \geq n_1$ si ha $L \leq a_n \leq a_{n_1} = L$, e quindi $a_n = L$ per ogni $n \geq n_1$. A questo punto definiamo

$$n_0 = \min\{n \in \mathbb{N} : a_m = L \ \forall m \geq n\}.$$

Si osservi che il minimo esiste perché l'insieme è un sottoinsieme non vuoto (contiene n_1) dei naturali.

6. Una successione particolare.

Consideriamo la successione

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

La successione a_n “genera” una forma indeterminata del tipo 1^∞ . Cercheremo di “sciogliere” questa forma indeterminata, e di capire se a_n ammetta limite oppure no.

Iniziamo innanzitutto con il dimostrare che a_n è limitata. Infatti, per la disuguaglianza di Bernoulli,

$$a_n = \left(1 + \frac{1}{n}\right)^n \geq 1 + n \frac{1}{n} = 2.$$

Inoltre, per la formula della potenza ennesima del binomio, e ricordando che $m! \geq 2^{m-1}$ per ogni $m \geq 0$ (si veda il primo capitolo),

$$\begin{aligned} a_n &= \left(1 + \frac{1}{n}\right)^n = \sum_{m=0}^n \binom{n}{m} \frac{1}{n^m} \\ &= \sum_{m=0}^n \frac{1}{m!} \frac{n(n-1)\dots(n-m+1)}{n^m} \\ &= \sum_{m=0}^n \frac{1}{m!} \left[\frac{n}{n} \frac{n-1}{n} \dots \frac{n-m+1}{n} \right] \\ \text{[i fattori sono minori di 1]} &\leq \sum_{m=0}^n \frac{1}{m!} = 1 + \sum_{m=1}^n \frac{1}{m!} \leq 1 + \sum_{m=1}^n \frac{1}{2^{m-1}} \\ &= 1 + \left(2 - \frac{1}{2^{n-1}}\right) < 3, \end{aligned}$$

dove nel penultimo passaggio si è usata la formula che dà la somma dei primi termini di una progressione geometrica (si veda il primo capitolo). Possiamo quindi escludere sia la convergenza a zero che il divergere a più infinito, ma ancora non sappiamo se la successione a_n ammetta limite. Dimostriamo allora che la successione è monotona crescente, cosicché esisterà il limite di a_n . Il primo passo consiste nel dimostrare che, per ogni n e m in \mathbb{N} , si ha

$$(6.1) \quad \left(1 - \frac{m}{n+1}\right) \frac{1}{n^m} \leq \frac{1}{(n+1)^m}.$$

La (6.1) segue facilmente dalla disuguaglianza di Bernoulli dal momento che è equivalente alla disuguaglianza

$$\left(\frac{n}{n+1}\right)^m = \left(1 - \frac{1}{n+1}\right)^m \geq 1 - \frac{m}{n+1}.$$

Osserviamo ora che, se $0 \leq m \leq n$,

$$\binom{n+1}{m} = \frac{(n+1)!}{m!(n+1-m)!} = \frac{n+1}{n+1-m} \binom{n}{m},$$

da cui

$$(6.2) \quad \binom{n}{m} = \left(1 - \frac{m}{n+1}\right) \binom{n+1}{m}.$$

Dimostriamo ora che la successione a_n è monotona crescente. Si ha

$$\begin{aligned} a_n &= \left(1 + \frac{1}{n}\right)^n = \sum_{m=0}^n \binom{n}{m} \frac{1}{n^m} \\ [(6.2)] &= \sum_{m=0}^n \binom{n+1}{m} \left(1 - \frac{m}{n+1}\right) \frac{1}{n^m} \\ [(6.1)] &\leq \sum_{m=0}^n \binom{n+1}{m} \frac{1}{(n+1)^m} \\ &\leq \sum_{m=0}^{n+1} \binom{n+1}{m} \frac{1}{(n+1)^m} = \left(1 + \frac{1}{n+1}\right)^{n+1} = a_{n+1}, \end{aligned}$$

che è quello che volevamo dimostrare. Il limite della successione a_n non è un numero “noto”; sappiamo che è compreso tra 2 e 3, ma non corrisponde a nessuna quantità incontrata fino ad ora. Per questo motivo, introduciamo un nuovo simbolo:

$$e = \lim_{n \rightarrow +\infty} a_n = \lim_{n \rightarrow +\infty} \left(1 + \frac{1}{n}\right)^n = 2,7182818284590452353 \dots$$

Il numero reale “e”, detto **numero di Nepero**, non è razionale (e quindi la successione delle sue cifre decimali non è periodica).

Sia ora b_n una successione di numeri **naturali** tendente a più infinito. Allora

$$(6.3) \quad \lim_{n \rightarrow +\infty} \left(1 + \frac{1}{b_n}\right)^{b_n} = e.$$

Per dimostrare la (6.3), sia $\varepsilon > 0$ e sia m_ε in \mathbb{N} tale che

$$\left| \left(1 + \frac{1}{m_\varepsilon}\right)^{m_\varepsilon} - e \right| < \varepsilon.$$

Sia poi n_ε in \mathbb{N} tale che $b_n \geq m_\varepsilon$ per ogni n maggiore di n_ε . Se $n \geq n_\varepsilon$ si ha allora, ricordando che l'applicazione che a m naturale associa $(1 + \frac{1}{m})^m$ è crescente,

$$\left(1 + \frac{1}{m_\varepsilon}\right)^{m_\varepsilon} \leq \left(1 + \frac{1}{b_n}\right)^{b_n} < e,$$

e quindi

$$-\varepsilon < \left(1 + \frac{1}{m_\varepsilon}\right)^{m_\varepsilon} - e \leq \left(1 + \frac{1}{b_n}\right)^{b_n} - e < 0,$$

da cui la (6.3).

Sia ora b_n una successione di numeri reali che diverge a più infinito. Allora

$$(6.4) \quad \lim_{n \rightarrow +\infty} \frac{[b_n]}{[b_n] + 1} = 1.$$

Essendo infatti $b_n \leq [b_n] < b_n + 1$, si ha

$$\frac{b_n}{b_n + 2} \leq \frac{[b_n]}{[b_n] + 1} \leq \frac{b_n + 1}{b_n + 1} = 1,$$

da cui la (6.4) per il Teorema dei Carabinieri, dal momento che

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

Infine, se b_n è una successione di numeri reali tendente a più infinito, si ha

$$(6.5) \quad \lim_{n \rightarrow +\infty} \left(1 + \frac{1}{b_n}\right)^{b_n} = e.$$

Osserviamo esplicitamente che, essendo b_n maggiore di 0 per n sufficientemente grande, la successione $a_n = \left(1 + \frac{1}{b_n}\right)^{b_n}$ è ben definita (al solito, il fatto che possa non essere definita per un numero finito di indici n non ci dà fastidio). Per dimostrare la (6.5), osserviamo che

$$\begin{aligned} \left(1 + \frac{1}{b_n}\right)^{b_n} &\geq \left(1 + \frac{1}{b_n}\right)^{[b_n]} \\ &\geq \left(1 + \frac{1}{[b_n] + 1}\right)^{[b_n]} = \left[\left(1 + \frac{1}{[b_n] + 1}\right)^{[b_n] + 1}\right]^{\frac{[b_n]}{[b_n] + 1}}, \end{aligned}$$

e che l'ultima successione converge a "e" per la (6.3) e la (6.4). D'altra parte,

$$\begin{aligned} \left(1 + \frac{1}{b_n}\right)^{b_n} &\leq \left(1 + \frac{1}{[b_n]}\right)^{b_n} \\ &\leq \left(1 + \frac{1}{[b_n]}\right)^{[b_n] + 1} = \left[\left(1 + \frac{1}{[b_n]}\right)^{[b_n]}\right]^{\frac{[b_n] + 1}{[b_n]}}, \end{aligned}$$

e anche in questo caso l'ultima successione converge a "e" per la (6.3) e la (6.4). La (6.5) segue allora dal Teorema dei Carabinieri.

Sia ora b_n una successione di numeri reali tendente a meno infinito. Allora

$$(6.6) \quad \lim_{n \rightarrow +\infty} \left(1 + \frac{1}{b_n}\right)^{b_n} = e.$$

Per dimostrare la (6.6), definiamo $c_n = -b_n - 1$, cosicché c_n tende a più infinito, ed osserviamo che si ha

$$\left(1 + \frac{1}{b_n}\right)^{b_n} = \frac{1}{\left(1 + \frac{1}{c_n}\right)^{b_n}} = \left[\frac{1}{\left(1 + \frac{1}{c_n}\right)^{c_n}}\right]^{\frac{b_n}{c_n}}.$$

La successione all'interno della parentesi quadra converge a $\frac{1}{e}$, mentre $\frac{b_n}{c_n}$ converge a -1 per definizione di c_n . Pertanto, si ha la (6.6).

Un'importante applicazione delle (6.5) e (6.6) è la seguente: sia L un numero reale e b_n una qualsiasi successione divergente (a più o a meno infinito). Allora

$$(6.7) \quad \lim_{n \rightarrow +\infty} \left(1 + \frac{L}{b_n}\right)^{b_n} = e^L.$$

Per dimostrare la (6.7), è sufficiente scrivere la successione come

$$\left(1 + \frac{L}{b_n}\right)^{b_n} = \left[\left(1 + \frac{L}{b_n}\right)^{\frac{b_n}{L}}\right]^L,$$

e applicare la (6.5) o la (6.6).

Un'altra importante applicazione della (6.5) è la seguente:

$$(6.8) \quad \lim_{n \rightarrow +\infty} \frac{e^{\frac{1}{n}} - 1}{\frac{1}{n}} = 1.$$

Sia $\varepsilon > 0$; allora, dal momento che $\left(1 + \frac{1+\varepsilon}{n}\right)^n$ converge a $e^{1+\varepsilon}$, che è strettamente maggiore di "e", per il Teorema della permanenza del segno esiste n_ε tale che

$$\left(1 + \frac{1+\varepsilon}{n}\right)^n > e, \quad \forall n \geq n_\varepsilon.$$

D'altra parte, essendo "e" maggiore di $\left(1 + \frac{1}{n}\right)^n$ per ogni n in \mathbb{N} , si ha, per ogni $n \geq n_\varepsilon$,

$$1 = \frac{\left(1 + \frac{1}{n}\right) - 1}{\frac{1}{n}} < \frac{e^{\frac{1}{n}} - 1}{\frac{1}{n}} < \frac{\left(1 + \frac{1+\varepsilon}{n}\right) - 1}{\frac{1}{n}} = 1 + \varepsilon,$$

da cui la tesi. A partire dalla (6.8), si dimostra che qualsiasi sia la successione a_n di numeri reali tendente a zero si ha

$$(6.9) \quad \lim_{n \rightarrow +\infty} \frac{e^{a_n} - 1}{a_n} = 1.$$

Una volta definito il numero "e", assume particolare importanza il logaritmo in base "e", detto **logaritmo naturale**, denotato con "ln" (invece che con \log_e). Sfruttando la (6.9) e la definizione di logaritmo naturale, si dimostra che

$$(6.10) \quad \lim_{n \rightarrow +\infty} \frac{\ln(1 + a_n)}{a_n} = 1,$$

per ogni successione a_n di numeri reali tendente a zero. Per dimostrare la (6.10), è sufficiente definire $b_n = \ln(1 + a_n)$, ed osservare che b_n è ben definita (almeno per n sufficientemente grande), che tende a zero e che, per definizione di logaritmo naturale,

$$\frac{\ln(1 + a_n)}{a_n} = \frac{b_n}{e^{b_n} - 1}.$$

Conseguenza della (6.9) è il seguente risultato, valido per ogni successione a_n di numeri reali tendente a zero e per ogni $A > 0$:

$$(6.11) \quad \lim_{n \rightarrow +\infty} \frac{A^{a_n} - 1}{a_n} = \ln(A).$$

Per ottenere la (6.11) è sufficiente scrivere (se $A \neq 1$, essendo il risultato evidentemente vero per $A = 1$)

$$\frac{A^{a_n} - 1}{a_n} = \frac{e^{a_n \ln(A)} - 1}{a_n} = \ln(A) \frac{e^{a_n \ln(A)} - 1}{a_n \ln(A)},$$

e applicare la (6.9).

ESERCIZIO 6.1. Calcolare il limite delle seguenti successioni:

$$\left(1 + \frac{3}{n}\right)^{n^2}, \quad \left(1 - \frac{\pi}{n}\right)^{n\sqrt{n}}, \quad \left(1 + \frac{1}{n^4}\right)^{n^3}, \quad \left(1 + \frac{2^n}{3^n}\right)^{n^2}, \quad \left(1 + \frac{1}{\left(1 + \frac{1}{n}\right)^{n^3}}\right)^n.$$

Risposta 6.1: L'ultima successione può essere riscritta come

$$\left[\left(1 + \frac{1}{\left(1 + \frac{1}{n}\right)^{n^3}}\right)^{\left(1 + \frac{1}{n}\right)^{n^3}} \right]^{\frac{n}{\left(1 + \frac{1}{n}\right)^{n^3}}}.$$

Dal momento che $\left(1 + \frac{1}{n}\right)^{n^3}$ tende a $+\infty$, l'argomento della parentesi quadra converge ad "e". Inoltre, essendo per la disuguaglianza di Bernoulli

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

si ha

$$0 < \frac{n}{\left(1 + \frac{1}{n}\right)^{n^3}} \leq \frac{n}{1 + n^2},$$

e quindi l'esponente "esterno" converge a 0. Il limite vale allora $e^0 = 1$.

7. Altre successioni particolari.

Prendiamo ora in considerazione cinque successioni:

$$\begin{aligned} a_n &= (\ln(n))^\beta, & n \geq 1, \beta > 0, \\ b_n &= n^\alpha, & \alpha > 0, \\ c_n &= A^n, & A > 1, \\ d_n &= n!, & n \geq 0, \\ e_n &= n^n, & n \geq 1. \end{aligned}$$

Ci proponiamo — una volta verificato che divergono tutte e cinque — di studiare i rapporti tra queste successioni. Che a_n diverga si vede usando il fatto che il logaritmo naturale è (essendo la base maggiore di uno) crescente: pertanto $[\ln(n)]^\beta \geq M$ se e solo se $n \geq e^{\sqrt[\beta]{M}}$; quindi è sufficiente scegliere $n_M = [e^{\sqrt[\beta]{M}}] + 1$ affinché sia soddisfatta la Definizione 2.2. Analogamente, per b_n è sufficiente scegliere $n_M = [\sqrt[\alpha]{M}] + 1$. Per c_n , scrivendo $A = 1 + h$ con h maggiore di zero, ed utilizzando la disuguaglianza di Bernoulli, si ha $c_n \geq 1 + nh$, e siccome nh diverge a più infinito, anche c_n diverge a più infinito. Infine, d_n e e_n divergono perché sono entrambe maggiori di n , che diverge.

Dimostriamo ora che

$$(7.1) \quad \lim_{n \rightarrow +\infty} \frac{n^n}{n!} = +\infty.$$

Infatti, per definizione di $n!$,

$$\frac{n^n}{n!} = \frac{n \cdot n \cdot \dots \cdot n \cdot n}{n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1} = \frac{n}{n} \cdot \frac{n}{n-1} \cdot \dots \cdot \frac{n}{2} \cdot \frac{n}{1} \geq n,$$

dal momento che i primi $n-1$ fattori sono maggiori o uguali ad 1.

Il secondo risultato è il seguente:

$$(7.2) \quad \forall A > 1, \quad \lim_{n \rightarrow +\infty} \frac{n!}{A^n} = +\infty.$$

Per dimostrarlo, sia $n > [A]$; allora

$$\frac{n!}{A^n} = \frac{n \cdot (n-1) \cdot \dots \cdot ([A]+1) \cdot [A] \cdot \dots \cdot 2 \cdot 1}{A \cdot A \cdot \dots \cdot A \cdot A \cdot \dots \cdot A \cdot A} \geq \frac{n}{A} \cdot \frac{[A] \cdot \dots \cdot 2 \cdot 1}{A \cdot \dots \cdot A \cdot A},$$

e l'ultima successione diverge.

Abbiamo poi:

$$(7.3) \quad \forall A > 1, \forall \alpha > 0, \quad \lim_{n \rightarrow +\infty} \frac{A^n}{n^\alpha} = +\infty.$$

Per dimostrare la (7.3), scriviamo

$$\frac{A^n}{n^\alpha} = \left(\frac{A^{\frac{n}{2\alpha}}}{n^{\frac{1}{2}}} \right)^{2\alpha} = \left(\frac{(A^{\frac{1}{2\alpha}})^n}{n^{\frac{1}{2}}} \right)^{2\alpha}.$$

Siccome se $A > 1$ allora anche $A^{\frac{1}{2\alpha}} > 1$, possiamo scrivere $A^{\frac{1}{2\alpha}} = 1 + h$, con $h > 0$. Per la disuguaglianza di Bernoulli abbiamo allora

$$\frac{A^n}{n^\alpha} = \left(\frac{(A^{\frac{1}{2\alpha}})^n}{n^{\frac{1}{2}}} \right)^{2\alpha} \geq \left(\frac{1 + nh}{n^{\frac{1}{2}}} \right)^{2\alpha} \geq h^{2\alpha} n^\alpha,$$

e l'ultima successione diverge.

La (7.3) continua a valere se si sostituisce a n l'elemento generico di una qualsiasi successione g_n di numeri positivi divergente a più infinito:

$$(7.4) \quad \forall A > 1, \forall \alpha > 0, \quad \lim_{n \rightarrow +\infty} \frac{A^{g_n}}{g_n^\alpha} = +\infty.$$

Infine,

$$(7.5) \quad \forall \alpha > 0, \forall \beta > 0, \quad \lim_{n \rightarrow +\infty} \frac{n^\alpha}{(\ln(n))^\beta} = +\infty.$$

Iniziamo con l'osservare che, essendo $\beta > 0$, la (7.5) è equivalente a

$$\forall \gamma > 0, \quad \lim_{n \rightarrow +\infty} \frac{n^\gamma}{\ln(n)} = +\infty,$$

e che quest'ultima, essendo n^γ e $\ln(n)$ non negativi, è equivalente a

$$\forall \gamma > 0, \quad \lim_{n \rightarrow +\infty} \frac{\ln(n)}{n^\gamma} = 0.$$

Ricordando le proprietà del logaritmo, in definitiva la (7.5) è equivalente a

$$(7.6) \quad \forall \gamma > 0, \quad \lim_{n \rightarrow +\infty} \ln(n^{\sqrt[\gamma]{n}}) = 0 \iff \lim_{n \rightarrow +\infty} n^{\sqrt[\gamma]{n}} = 1.$$

Definiamo ora

$$\delta_n = \sqrt[n^\gamma]{n^{\frac{\gamma}{2}}},$$

ovvero

$$n^{\frac{\gamma}{2}} = \delta_n^{n^\gamma}.$$

Essendo $\delta_n > 1$, possiamo scrivere $\delta_n = 1 + \varepsilon_n$, con $\varepsilon_n > 0$, e quindi, per la disuguaglianza di Bernoulli,

$$n^{\frac{\gamma}{2}} = \delta_n^{n^\gamma} = (1 + \varepsilon_n)^{n^\gamma} \geq (1 + \varepsilon_n)^{[n^\gamma]} \geq 1 + [n^\gamma] \varepsilon_n.$$

Pertanto,

$$0 < \varepsilon_n \leq \frac{n^{\frac{\gamma}{2}} - 1}{[n^\gamma]} \leq \frac{n^{\frac{\gamma}{2}} - 1}{n^\gamma - 1},$$

e l'ultima successione tende a zero. Per il Teorema dei Carabinieri ε_n converge a 0, e quindi δ_n tende ad 1. La (7.6) è allora dimostrata, essendo

$$\sqrt[n^\gamma]{n} = \left(\sqrt[n^\gamma]{n^{\frac{\gamma}{2}}} \right)^{\frac{2}{\gamma}} = (\delta_n)^{\frac{2}{\gamma}}.$$

Come per la (7.3), anche la (7.5) continua a valere se si sostituisce a n l'elemento generico di una qualsiasi successione g_n di numeri positivi divergente a più infinito:

$$(7.7) \quad \forall \alpha > 0, \forall \beta > 0, \quad \lim_{n \rightarrow +\infty} \frac{g_n^\alpha}{(\ln(g_n))^\beta} = +\infty.$$

Una conseguenza immediata delle (7.1) – (7.5) è che le cinque successioni sono “ordinate”: se $\alpha > 0$, $\beta > 0$ e $A > 1$, si ha

$$(\ln(n))^\beta \prec n^\alpha \prec A^n \prec n! \prec n^n$$

e l'ordine va interpretato nel senso che il rapporto fra due successioni tende a più infinito o a zero a seconda se il numeratore è “più a destra” o “più a sinistra” del denominatore.

ESERCIZIO 7.1. Si calcolino i limiti delle seguenti successioni:

$$\frac{n^2 + \ln^{100}(n)}{2^n - n}, \quad \frac{3^n + n!}{1000^n + n^{50000}}, \quad \frac{n^{\frac{n}{2}} + n^2}{5^n + 8^n}, \quad \frac{n^n}{(2n)!}.$$

Un ultimo risultato, all'apparenza bizzarro, ma utile in molte applicazioni, è la cosiddetta **formula di Stirling**:

$$\lim_{n \rightarrow +\infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1.$$

ESERCIZIO 7.2. Si calcolino i limiti delle seguenti successioni:

$$\sqrt[n]{n!}, \quad \frac{\sqrt[n]{n!}}{n}, \quad \sqrt[n]{\frac{(2n)^n}{(2n)!}}.$$

8. Sottosuccessioni. Il Teorema di Bolzano-Weierstrass.

Nel paragrafo 2 abbiamo visto come una successione convergente sia limitata, mentre il contrario non è necessariamente vero: data la successione $a_n = (-1)^n$, a_n è limitata ma non è convergente, dato che oscilla continuamente tra i due valori ± 1 . Quello che però possiamo fare è non considerare i valori di a_n per ogni n , ma solo per alcuni valori particolari — ad esempio i numeri pari. Se facciamo così, otteniamo la successione di valori $(-1)^0 = 1$, $(-1)^2 = 1$, $(-1)^4 = 1$ e così via, ovvero la successione costantemente uguale ad 1, che sappiamo convergere ad 1. D'altra parte, se invece di considerare gli n pari consideriamo gli n dispari, otteniamo la successione di valori $(-1)^1 = -1$, $(-1)^3 = -1$, $(-1)^5 = -1$ e così via; anche in questo caso abbiamo una successione costante, convergente a -1 . Perché diciamo “una successione”? Per noi una successione è un'applicazione da \mathbb{N} in \mathbb{R} , mentre nei due casi precedenti abbiamo considerato un'applicazione da $\{n \text{ pari}\}$ in \mathbb{R} , o da $\{n \text{ dispari}\}$ in \mathbb{R} : due funzioni alle quali non abbiamo il diritto di dare il nome di “successione”. Se, però, consideriamo l'applicazione k da \mathbb{N} in \mathbb{N} che a n associa $2n$, la nostra applicazione dai numeri pari in \mathbb{R} può essere vista come la composizione della funzione k con la successione a_n : otteniamo così la successione (adesso sì) che a n associa $a_{k(n)} = a_{2n} = (-1)^{2n} = 1$. Analogamente, la seconda applicazione può essere vista come la composizione di $h(n) = 2n + 1$ con a_n .

Dunque, data una successione qualsiasi a_n , possiamo costruire, considerando delle funzioni opportune da \mathbb{N} in \mathbb{N} , delle nuove successioni che assumono solo alcuni dei valori della successione originaria a_n . Perché “opportune”? Perché non possiamo prendere una funzione qualsiasi da \mathbb{N} in \mathbb{N} e sperare di ottenere qualcosa di “significativo”: al solito, a noi non interessa il comportamento della successione per n “piccolo”, ma per n tendente a più infinito. Va da sé che se noi prendiamo come applicazione da \mathbb{N} in \mathbb{N} la funzione k che ad ogni naturale n associa — ad esempio — un fissato naturale n_0 , la composizione di questa funzione con una successione qualsiasi a_n ci dà la successione costante $a_{k(n)} = a_{n_0}$; chiaramente, è uno spreco di energie costruire una successione da un solo elemento di a_n , dato che per capire come è fatto “basta guardarlo”.

In sostanza, dal momento che vogliamo andare a vedere se determinate proprietà (come l'esistenza del limite) possono essere recuperate non considerando la successione nella sua interezza, ma solo una parte di essa, sarà necessario richiedere che la funzione da \mathbb{N} in \mathbb{N} , il “filtro” degli indici, debba almeno divergere al divergere di n . In altre parole, deve essere arbitrariamente grande per n grande. In questa maniera potremo provare a leggere il comportamento al limite di alcuni dei valori della successione a_n attraverso la “lente” della funzione da \mathbb{N} in \mathbb{N} . Per ottenere quello che vogliamo, è sufficiente richiedere una sola proprietà sulla funzione k da \mathbb{N} in \mathbb{N} : k deve essere monotona strettamente crescente.

PROPOSIZIONE 8.1. *Sia $k : \mathbb{N} \rightarrow \mathbb{N}$ una applicazione tale che $k(n + 1) > k(n)$ per ogni n in \mathbb{N} . Allora*

$$\lim_{n \rightarrow +\infty} k(n) = +\infty.$$

Dimostrazione. Si ha $k(0) \geq 0$; inoltre, se $k(n) \geq n$, allora $k(n + 1) > k(n) \geq n$, da cui $k(n + 1) \geq n + 1$. Pertanto, per il principio di induzione, $k(n) \geq n$ per ogni n in \mathbb{N} , e quindi $k(n)$ diverge a più infinito. ■

Come conseguenza del teorema precedente, ogni applicazione monotona strettamente crescente da \mathbb{N} in \mathbb{N} genera una successione divergente a più infinito. Siamo allora pronti a dare la definizione dell'oggetto ottenuto componendo una tale funzione con una successione qualsiasi.

DEFINIZIONE 8.2. Sia a_n una successione di numeri reali, e sia $k : \mathbb{N} \rightarrow \mathbb{N}$ un'applicazione monotona strettamente crescente. La successione $a_{k(n)}$ si dice **sottosuccessione** o **successione estratta** da a_n .

Dal momento che l'applicazione k altro non è che una successione di numeri naturali, al solito si preferisce indicare $k(n)$ con k_n . Parallelamente, la sottosuccessione si indica con a_{k_n} .

Quali proprietà di una successione si conservano prendendo in considerazione le sue sottosuccessioni? Una proprietà è evidentemente la limitatezza, dato che l'insieme dei valori assunti da una sottosuccessione è un sottoinsieme dei valori assunti dalla successione di partenza. Se la successione di partenza ammette limite (finito o infinito), allora ammettono lo stesso limite tutte le sue sottosuccessioni.

TEOREMA 8.3. *Sia a_n una successione convergente ad L , e sia a_{k_n} una qualsiasi sottosuccessione di a_n . Allora a_{k_n} converge a L . Lo stesso vale se a_n diverge a più infinito o a meno infinito.*

Dimostrazione. Sia $\varepsilon > 0$, e sia n_ε in \mathbb{N} tale che $|a_n - L| < \varepsilon$ per ogni $n \geq n_\varepsilon$. Dal momento che $k_n \geq n$ (ricordiamo che la successione k_n è crescente), se $n \geq n_\varepsilon$, allora $k_n \geq k_{n_\varepsilon} \geq n_\varepsilon$, e quindi $|a_{k_n} - L| < \varepsilon$, come volevasi dimostrare. Il caso delle successioni divergenti è lasciato al lettore. ■

Dunque, se a_n converge, considerarne delle sottosuccessioni non modifica le nostre informazioni: il concetto di limite è "stabile" rispetto al passaggio a sottosuccessioni. Che succede se, però, la successione di partenza non ammette limite? È possibile, come nel caso della successione $(-1)^n$, "estrarre" una sottosuccessione convergente ad un numero reale? Se sì, sotto quali ipotesi?

ESERCIZIO 8.4. Si consideri la successione $a_n = (-1)^n n$. Si dimostri che da a_n non si può estrarre alcuna sottosuccessione convergente ad un numero reale.

Risposta 8.4: Dal momento che $|a_n| = n$, la successione $|a_n|$ diverge, e con essa tutte le sue sottosuccessioni. Se esistesse una sottosuccessione a_{k_n} convergente ad un numero reale, a_{k_n} , e quindi $|a_{k_n}|$, sarebbe limitata, cosa che non può essere.

Come si vede dall'esercizio precedente, se la successione non è limitata, la speranza di trovare una sottosuccessione convergente ad un numero reale può essere vana. Il prossimo teorema, di fondamentale importanza⁽⁹⁾, afferma che se la successione è limitata, allora è possibile estrarne una sottosuccessione convergente.

TEOREMA 8.5 (Bolzano-Weierstrass). *Sia a_n una successione limitata. Allora da a_n si può estrarre una sottosuccessione a_{k_n} convergente ad un numero reale L .*

⁽⁹⁾Tanto importante che ne daremo due dimostrazioni...

Prima dimostrazione: il metodo di bisezione.

Dal momento che la successione a_n è limitata, esiste $M > 0$ tale che $|a_n| \leq M$ per ogni n in \mathbb{N} ; vale a dire

$$-M \leq a_n \leq M, \quad \forall n \in \mathbb{N}.$$

Consideriamo ora l'intervallo $I_0 = [-M, M] = [l_0, r_0]$, che sappiamo contenere i valori della successione a_n , e spezziamolo in due sottointervalli disgiunti (ovvero, privi di punti in comune): $I_0 = I_1^+ \cup I_1^-$, con $I_1^+ = [0, M]$ e $I_1^- = [-M, 0]$. Consideriamo ora i due insiemi

$$E_1^+ = \{n \in \mathbb{N} : a_n \in I_1^+\}, \quad E_1^- = \{n \in \mathbb{N} : a_n \in I_1^-\}.$$

Dal momento che a_n è tutta contenuta in I_0 , uno tra E_1^+ e E_1^- è non vuoto. Di più, almeno uno tra E_1^+ e E_1^- contiene infiniti numeri naturali, dato che la loro unione è tutto \mathbb{N} . Se E_1^- contiene infiniti numeri, definiamo $\tilde{I}_1 = I_1^-$, e $E_1 = E_1^-$, altrimenti poniamo $\tilde{I}_1 = I_1^+$ e $E_1 = E_1^+$. Detti l_1 e r_1 gli estremi sinistro e destro di \tilde{I}_1 , definiamo $I_1 = [l_1, r_1]$. Sia poi $n(1)$ il minimo di E_1 (che esiste per il principio del buon ordinamento), cosicché $a_{n(1)}$ appartiene ad I_1 . dopo questo primo passaggio, abbiamo ottenuto un intervallo I_1 , un insieme E_1 (contenente infiniti numeri naturali), un intero $n(1)$ ed un numero reale $a_{n(1)}$ con queste proprietà:

- α_1) l'ampiezza di I_1 è la metà dell'ampiezza di I_0 ; ovvero, $r_1 - l_1 = \frac{r_0 - l_0}{2}$;
- β_1) $l_1 \geq l_0$ e $r_1 \leq r_0$;
- γ_1) $l_1 \leq a_{n(1)} \leq r_1$.

Continuiamo ora il procedimento, spezzando I_1 nell'unione disgiunta di due intervalli I_2^\pm , di ampiezza la metà di I_1 . Definiamo poi

$$E_2^\pm = \{n \in \mathbb{N} : n > n(1) \text{ e } a_n \in I_2^\pm\}.$$

Come nel caso precedente, almeno uno tra E_2^+ e E_2^- contiene infiniti numeri naturali. Se E_2^- è infinito, definiamo $\tilde{I}_2 = I_2^-$ e $E_2 = E_2^-$, altrimenti poniamo $\tilde{I}_2 = I_2^+$ e $E_2 = E_2^+$. Come prima, detti l_2 e r_2 gli estremi di \tilde{I}_2 , definiamo $I_2 = [l_2, r_2]$. Infine, sia $n(2)$ il minimo di E_2 .

Come prima, abbiamo che:

- α_2) l'ampiezza di I_2 è la metà dell'ampiezza di I_1 ; ovvero, $r_2 - l_2 = \frac{r_1 - l_1}{2} = \frac{r_0 - l_0}{4}$;
- β_2) $l_2 \geq l_1 \geq l_0$ e $r_2 \leq r_1 \leq r_0$;
- γ_2) $l_2 \leq a_{n(2)} \leq r_2$;
- δ_2) $n(2) > n(1)$.

Supponiamo ora di aver definito E_k (contenente infiniti numeri naturali), $n(k)$ (il minimo di E_k) e $I_k = [l_k, r_k]$ in modo tale che

- α_k) l'ampiezza di I_k è la metà dell'ampiezza di I_{k-1} ; ovvero, $r_k - l_k = \frac{r_{k-1} - l_{k-1}}{2} = \frac{r_0 - l_0}{2^k}$;
- β_k) $l_k \geq l_{k-1} \geq \dots \geq l_0$ e $r_k \leq r_{k-1} \leq \dots \leq r_0$;
- γ_k) $l_k \leq a_{n(k)} \leq r_k$;
- δ_k) $n(k) > n(k-1) > \dots > n(2) > n(1)$.

Proseguiamo, definendo I_{k+1}^\pm (come le due "metà" di I_k),

$$E_{k+1}^\pm = \{n \in \mathbb{N} : n > n_{k+1} \text{ e } a_n \in I_{k+1}^\pm\},$$

e ponendo come prima $\tilde{I}_{k+1} = I_{k+1}^-$ e $E_{k+1} = E_{k+1}^-$ se E_{k+1}^- contiene infiniti numeri, $\tilde{I}_{k+1} = I_{k+1}^+$ e $E_{k+1} = E_{k+1}^+$ altrimenti. Come prima, $I_{k+1} = [l_{k+1}, r_{k+1}]$ se l_{k+1} e r_{k+1} sono gli estremi sinistro e destro di \tilde{I}_{k+1} . Infine, $n(k+1)$ è il minimo di E_{k+1} .

È allora evidente per costruzione che si ha

$$\begin{aligned} \alpha_{k+1}) \quad & \text{l'ampiezza di } I_{k+1} \text{ è la metà dell'ampiezza di } I_k; \text{ ovvero, } r_{k+1} - l_{k+1} = \frac{r_k - l_k}{2} = \\ & \frac{r_0 - l_0}{2^{k+1}}; \\ \beta_{k+1}) \quad & l_{k+1} \geq l_k \geq \dots \geq l_0 \text{ e } r_{k+1} \leq r_k \leq \dots \leq r_0; \\ \gamma_{k+1}) \quad & l_{k+1} \leq a_{n(k+1)} \leq r_{k+1}; \\ \delta_{k+1}) \quad & n(k+1) > n(k) > \dots > n(2) > n(1). \end{aligned}$$

Dal momento che, per ogni k , l'insieme E_k conterrà infiniti numeri naturali, potremo proseguire indefinitamente il nostro ragionamento. Cosa abbiamo allora trovato? Per ogni k in \mathbb{N} , abbiamo determinato un numero naturale $n(k)$, con la proprietà che $n(k+1) > n(k)$; in altre parole, abbiamo determinato un'applicazione da \mathbb{N} in \mathbb{N} monotona strettamente crescente. Pertanto, $a_{n(k)}$ è una sottosuccessione estratta da a_n . Inoltre, per costruzione, la successione l_k è monotona crescente, e limitata superiormente da $r_0 = M$. Pertanto, l_k converge, per k tendente ad infinito, verso un limite l . Analogamente, r_k è una successione monotona decrescente, limitata inferiormente da $l_0 = -M$. Pertanto, r_k converge verso un limite r . Infine, essendo

$$r_k = l_k + \frac{r_0 - l_0}{2^k},$$

dal momento che $\frac{1}{2^k}$ tende a zero, si ha $r = l$. Applicando il Teorema di Carabinieri alla relazione γ_k):

$$l_k \leq a_{n(k)} \leq r_k,$$

si ha allora la tesi: $a_{n(k)}$ converge a r . ■

ESEMPIO 8.6. Sia $a_n = (-1)^n$, e applichiamo il metodo descritto nel precedente teorema. La successione a_n è limitata, dal momento che $|a_n| \leq 1$, cosicché $I_0 = [-1, 1]$ e $I_1^+ = [0, 1]$ e $I_1^- = [-1, 0)$, mentre

$$\begin{aligned} E_1^+ &= \{n \in \mathbb{N} : a_n \in I_1^+\} = \{\text{pari}\}, \\ E_1^- &= \{n \in \mathbb{N} : a_n \in I_1^-\} = \{\text{dispari}\}. \end{aligned}$$

Dal momento che E_1^- contiene infiniti numeri naturali, definiamo $E_1 = E_1^-$, il che implica $n(1) = 1$ e $I_1 = [-1, 0]$. Nel secondo passaggio, $I_2^+ = [-1, -\frac{1}{2})$ e $I_2^- = [-\frac{1}{2}, 0]$, da cui

$$\begin{aligned} E_2^+ &= \{n \in \mathbb{N} : n > 1 \text{ e } a_n \in I_2^+\} = \emptyset, \\ E_2^- &= \{n \in \mathbb{N} : n > 1 \text{ e } a_n \in I_2^-\} = \{\text{dispari maggiori di } 1\}. \end{aligned}$$

In questo caso, la scelta è obbligata: $E_2 = E_2^-$, da cui $n(2) = 3$ e $I_2 = [-1, -\frac{1}{2}]$. Proseguendo, si avrà

$$I_k^+ = \left[-1, -1 + \frac{1}{2^{k-1}}\right], \quad I_k^- = \left[-1 + \frac{1}{2^{k-1}}, -1 + \frac{1}{2^{k-2}}\right),$$

e quindi

$$\begin{aligned} E_k^+ &= \{n \in \mathbb{N} : n > n(k-1) \text{ e } a_n \in I_k^+\} = \emptyset, \\ E_k^- &= \{n \in \mathbb{N} : n > n(k-1) \text{ e } a_n \in I_k^-\} = \{\text{dispari maggiori di } n(k-1)\}. \end{aligned}$$

Eseguendo i calcoli, si trova $n(k) = 2k - 1$, e $I_k = [-1, -1 + \frac{1}{2^{k-1}}]$. Essendo $a_{n(k)}$ identicamente uguale ad -1 , evidentemente $a_{n(k)}$ converge ad -1 , anche se questo fatto poteva essere dedotto dal Teorema dei Carabinieri e dal fatto che

$$-1 \leq a_{n(k)} \leq -1 + \frac{1}{2^{k-1}}.$$

Seconda dimostrazione: il metodo dell'estremo inferiore.

Data a_n una successione limitata, definiamo

$$E_0 = \{a_k : k \geq 0\} = \{a_0, a_1, a_2, \dots\}.$$

In altri termini, E_0 è l'insieme dei valori della successione. Dal momento che a_n è limitata, anche E_0 sarà limitato. Pertanto, è ben definito il numero reale

$$b_0 = \inf E_0.$$

Dalla definizione equivalente di estremo inferiore sappiamo che, per ogni $\varepsilon > 0$, esiste un elemento in E_0 compreso tra b_0 e $b_0 + \varepsilon$. Siccome gli elementi di E_0 sono identificati da numeri naturali, esiste un numero naturale n_ε tale che

$$b_0 \leq a_{n_\varepsilon} < b_0 + \varepsilon.$$

Scegliamo ora $\varepsilon = 1$, ed in questo modo otteniamo $n(1)$ in \mathbb{N} tale che $b_0 \leq a_{n(1)} < b_0 + 1$. Definiamo ora

$$E_1 = \{a_k, k > n(1)\} = \{a_{n(1)+1}, a_{n(1)+2}, \dots\},$$

ovvero gli elementi della successione da $n(1) + 1$ in poi. Essendo $E_1 \subset E_0$, anche E_1 è limitato, e quindi è ben definito

$$b_1 = \inf E_1.$$

Inoltre, essendo $E_1 \subset E_0$, si ha $b_0 \leq b_1$. Sempre per definizione di estremo inferiore, ma scegliendo questa volta $\varepsilon = \frac{1}{2}$, esiste un indice $n(2) > n(1)$, ed un elemento $a_{n(2)}$ in E_1 , tale che

$$b_1 \leq a_{n(2)} < b_1 + \frac{1}{2}.$$

Proseguiamo nella nostra costruzione; al passo h -simo, si può costruire un insieme limitato

$$E_h = \{a_k, k > n(h)\} \subset E_{h-1},$$

il cui estremo inferiore b_h è maggiore di b_{h-1} , e nel quale possiamo scegliere un elemento $a_{n(h+1)}$, con un indice $n(h+1) > n(h)$, tale che

$$(8.1) \quad b_h \leq a_{n(h+1)} < b_h + \frac{1}{h+1}.$$

Dal momento che l'applicazione $h \mapsto n(h)$ è crescente, $\{a_{n(h)}\}$ è una sottosuccessione di $\{a_n\}$. Essendo b_n crescente, e limitata superiormente dall'estremo superiore dei valori della successione a_n , b_n converge ad un certo limite L . Siccome $b_n + \frac{1}{n+1}$ converge anch'essa a L , il Teorema dei Carabinieri e la (8.1) implicano che $a_{n(h)}$ converge a L . ■

ESEMPIO 8.7. Sia, come nell'esempio precedente, $a_n = (-1)^n$. Allora

$$E_0 = \{a_k, k \geq 0\} = \{-1, 1\},$$

da cui $b_0 = -1$. Fissato $\varepsilon = 1$, possiamo scegliere $n(1)$ un qualsiasi numero dispari; ad esempio, $n(1) = 3$. Abbiamo così

$$E_1 = \{a_k, k \geq 4\} = \{-1, 1\},$$

da cui $b_1 = -1$. Scelto $\varepsilon = \frac{1}{2}$, possiamo scegliere per $n(2)$ un qualsiasi numero dispari maggiore di 4; ad esempio, 5. Continuando, otteniamo per $n(h)$ una successione crescente di numeri dispari (ad esempio, i numeri primi), $b_n = -1$ per ogni n in \mathbb{N} , mentre $a_{n(h)} = -1$ per ogni h (e quindi è banalmente convergente).

OSSERVAZIONE 8.8. Il metodo utilizzato per la prima dimostrazione del Teorema di Bolzano-Weierstrass, che consiste nel dimezzare di volta in volta l'intervallo dentro al quale si trovano i valori della successione, viene detto **metodo di bisezione**. Uno dei vantaggi di tale metodo è che, dopo k passaggi, conosciamo il limite della sottosuccessione con un errore inferiore a $\frac{M}{2^{k-1}}$, dal momento che il limite l si trova tra l_k e r_k e si ha $r_k - l_k = \frac{r_0 - l_0}{2^k} = \frac{M}{2^{k-1}}$. Per ottenere convergenze più rapide, l'intervallo k -simo può essere spezzato in (ad esempio) 10 sottointervalli di ampiezza un decimo dell'intervallo k -simo: il metodo continua a funzionare, dato che almeno uno degli intervalli conterrà infiniti elementi della successione.

Osserviamo infine che il Teorema di Bolzano-Weierstrass dà una condizione sufficiente, ma non necessaria, per l'esistenza di una sottosuccessione convergente. Infatti la successione $a_n = n \operatorname{sen} \left(\frac{n\pi}{2} \right)$ è illimitata (dato che per n dispari vale $\pm n$), ma $a_{2n} = 0$ per ogni n in \mathbb{N} : la sottosuccessione corrispondente agli indici pari è costante, quindi convergente.

Che succede se la successione non è limitata? Come abbiamo visto con l'esempio $a_n = (-1)^n n$ e nell'osservazione precedente, può esistere o non esistere una sottosuccessione convergente ad un numero reale. Quello che è sicuramente vero, però, è che si può estrarre una sottosuccessione divergente.

TEOREMA 8.9. *Sia a_n una successione illimitata superiormente; allora esiste una sottosuccessione divergente a più infinito. Sia a_n una successione illimitata inferiormente; allora esiste una sottosuccessione divergente a meno infinito.*

Dimostrazione. Dal momento che la successione a_n è illimitata superiormente, è non vuoto l'insieme

$$E_1 = \{n \in \mathbb{N} : a_n \geq 1\}.$$

Sia allora $n(1)$ il minimo di E_1 , cosicché $a_{n(1)} \geq 1$. Sempre per l'illimitatezza di a_n , è non vuoto l'insieme

$$E_2 = \{n \in \mathbb{N} : n > n(1) \text{ e } a_n \geq 2\}.$$

Come prima, sia $n(2)$ il minimo di E_2 ; si ha $n(2) > n(1)$ e $a_{n(2)} \geq 2$. Supponiamo ora di aver determinato $n(1), n(2), \dots, n(k)$ tali che $n(k) > n(k-1) > \dots > n(1)$ e $a_{n(j)} \geq j$ per ogni j da 1 a k . Sfruttando l'illimitatezza di a_n , sarà non vuoto l'insieme

$$E_{k+1} = \{n \in \mathbb{N} : n > n(k) \text{ e } a_n \geq k+1\}.$$

Possiamo allora definire $n(k+1)$ il minimo di E_{k+1} , che sarà maggiore di $n(k)$, e sia avrà $a_{n(k)} \geq k$. Continuando in questo modo, possiamo costruire l'applicazione da \mathbb{N} in \mathbb{N} che a k associa $n(k)$; tale applicazione è strettamente crescente, e quindi $a_{n(k)}$ è una sottosuccessione di a_n . Inoltre, essendo $a_{n(k)} \geq k$, la successione $a_{n(k)}$ diverge quando k tende a più infinito. Il lettore completi la dimostrazione nel caso in cui a_n sia illimitata inferiormente. ■

Nel Teorema 8.3 abbiamo visto che se la successione a_n converge, allora ogni sua sottosuccessione converge allo stesso limite. È chiaro che se da una successione si possono estrarre due sottosuccessioni convergenti a due limiti diversi, la successione di partenza non potrà convergere (per l'unicità del limite). Ci chiediamo allora se sia vero che se **ogni** sottosuccessione estratta dalla successione di partenza è convergente allo **stesso** limite, allora tutta la successione è convergente (al limite comune). La risposta è ovviamente affermativa, dato che una delle sottosuccessioni estratte da a_n è a_n stessa (corrispondente all'identità su \mathbb{N} , che è una funzione monotona strettamente crescente). In realtà, possiamo chiedere qualcosa di meno.

TEOREMA 8.10. *Sia a_n una successione di numeri reali. Supponiamo che da ogni sottosuccessione a_{n_k} di a_n si possa estrarre una sotto-sottosuccessione $a_{n_{k_j}}$ convergente ad un limite L indipendente dalla sottosuccessione a_{n_k} . Allora la successione a_n converge ad L .*

Dimostrazione. Supponiamo per assurdo che a_n non converga ad L . Ciò vuol dire — per definizione di limite — che esiste $\bar{\varepsilon} > 0$ tale che a_n dista da L più di $\bar{\varepsilon}$ per infiniti indici n . Pertanto, detti

$$E^+ = \{n \in \mathbb{N} : a_n - L \geq \bar{\varepsilon}\}, \quad E^- = \{n \in \mathbb{N} : a_n - L \leq -\bar{\varepsilon}\},$$

uno tra E^+ e E^- contiene infiniti numeri naturali; supponiamo che sia E^+ . Definiamo $n(1)$ il minimo di E^+ , $n(2)$ il minimo degli elementi di E^+ strettamente maggiori di $n(1)$, $n(3)$ il minimo degli elementi di E^+ strettamente maggiori di $n(2)$ e così via. In questo modo costruiamo una sottosuccessione $a_{n(k)}$ tale che $a_{n(k)} - L \geq \bar{\varepsilon}$ per ogni k in \mathbb{N} . Per ipotesi, da $a_{n(k)}$ si può estrarre una sotto-sottosuccessione $a_{n(k(j))}$ convergente ad L , il che vuol dire che, per j sufficientemente grande, $a_{n(k(j))} - L < \bar{\varepsilon}$. Per come è stata costruita $a_{n(k)}$, ciò non è possibile. ■

9. Le serie numeriche.

Non tutte le successioni “immaginabili” sono suscettibili di una definizione “comoda”, ovvero di una forma analitica che permetta, dato n , di calcolare esplicitamente l' n -simo elemento della successione, come ad esempio $a_n = n$, o $a_n = \sin(n) - n^2$. In alcuni casi, si pensi ad esempio alla successione “ $a_n = n$ -simo numero primo”, non c'è espressione analitica esplicita del valore, il che vuol dire che, per studiare le proprietà della successione, è necessario usare altre tecniche. Ad esempio, una volta dimostrato che esistono infiniti numeri primi, la successione “ $a_n = n$ -simo numero primo” diverge (dato che è monotona strettamente crescente).

Un altro caso — molto importante — di famiglia di successioni per le quali è talvolta necessario conoscere il comportamento al limite senza conoscere gli elementi della successione in maniera analitica, è quello delle cosiddette serie numeriche.

DEFINIZIONE 9.1. Sia a_k una successione di numeri reali. Definiamo **serie numerica** (o **serie**) **associata** ad a_k , o **serie di termine generico** a_k , la successione S_n il cui termine generico è dato da

$$S_n = \sum_{k=0}^n a_k.$$

Se la successione S_n converge ad un numero reale S , la serie si dice **convergente**; se il limite di S_n non è finito la serie si dice **divergente**, mentre si dice **indeterminata** se S_n non ha limite (finito o infinito che sia). Nel caso in cui la successione S_n converga ad S si scrive

$$S = \sum_{k=0}^{+\infty} a_k,$$

e S si definisce la **somma della serie**. Si avverte il lettore che S **non** è in generale la somma di infiniti termini⁽¹⁰⁾, ma il limite della successione S_n , che viene detta **successione delle somme parziali**. In altre parole, le proprietà consuete delle operazioni algebriche tra numeri reali (commutatività della somma, eccetera), non si estendono ai numeri reali ottenuti come somme di una serie.

ESEMPIO 9.2. Sia $a_k = 1$ per ogni n in \mathbb{N} . La serie S_n associata ad a_k è evidentemente

$$(9.1) \quad S_n = \sum_{k=0}^n a_k = \sum_{k=0}^n 1 = n + 1.$$

Pertanto, la serie diverge. Sia $a_k = k$ per ogni n in \mathbb{N} . La serie S_n associata ad a_k è allora

$$S_n = \sum_{k=0}^n a_k = \sum_{k=0}^n k = \frac{n(n+1)}{2}.$$

Anche in questo caso, la serie diverge. Sia $a_k = 2^k$ per ogni n in \mathbb{N} . Ricordando la formula che dà la somma della serie geometrica, la serie (divergente) S_n associata ad a_k è allora

$$S_n = \sum_{k=0}^n a_k = \sum_{k=0}^n 2^k = 2^{n+1} - 1.$$

In generale, se $q \neq 1$, e se $a_k = q^k$, la serie S_n associata ad a_k è data da

$$(9.2) \quad S_n = \sum_{k=0}^n a_k = \sum_{k=0}^n q^k = \frac{q^{n+1} - 1}{q - 1}.$$

Si ha, evidentemente, che S_n diverge se $q \geq 1$, converge a $\frac{1}{1-q}$ se $-1 < q < 1$, mentre è indeterminata se $q \leq -1$.

⁽¹⁰⁾Nessuno ha il tempo sufficiente a sommarli tutti, a meno che la k -sima somma non sia effettuata in un tempo che è la metà di quello necessario per effettuare la $(k-1)$ -sima (esercizio!).

Ricapitolando:

$$(9.3) \quad \text{La serie } \sum_{k=1}^{+\infty} q^k \quad \left\{ \begin{array}{ll} \text{diverge} & \text{se } q \geq 1, \\ \text{converge} & \text{se } -1 < q < 1, \\ \text{non converge} & \text{se } q \leq -1 \end{array} \right.$$

ESEMPIO 9.3. Un caso particolare della serie associata ad una progressione geometrica — e che aiuta a comprendere il motivo del perché la somma di una serie non sia la somma di infiniti termini —, è il caso $q = -1$. Se scriviamo i primi termini della successione S_n definita da

$$S_n = \sum_{k=0}^n (-1)^k,$$

troviamo $S_0 = 1$, $S_1 = 0$, $S_2 = 1$, $S_3 = 0$, e così via, con i valori 0 e 1 che si alternano. Al giorno d'oggi — forti del concetto rigoroso di limite — sappiamo che una successione fatta in questo modo non converge, ma quando in passato si presentò questo esempio, ed il concetto di somma di una serie come limite di una successione non era ancora ben codificato, furono parecchie le dispute attorno al “risultato corretto”:

1) se $S = 1 - 1 + 1 - 1 + 1 - \dots$, allora $S = 0$; infatti

$$S = (1 - 1) + (1 - 1) + (1 - 1) + \dots = 0 + 0 + 0 + \dots = 0;$$

2) se $S = 1 - 1 + 1 - 1 + 1 - \dots$, allora $S = 1$; infatti

$$S = 1 - (1 - 1) + (1 - 1) + (1 - 1) + \dots = 1 + 0 + 0 + 0 + \dots = 1;$$

3) se $S = 1 - 1 + 1 - 1 + 1 - \dots$, allora $S = \frac{1}{2}$; infatti

$$S = 1 - (1 - 1 + 1 - 1 + 1 - \dots = 1 - S \implies 2S = 1 \implies S = \frac{1}{2}.$$

In altre parole, dal momento che la successione delle somme parziali non è convergente (secondo il concetto di convergenza oggi adottato), non è possibile usare l'associatività della somma e sperare di ottenere sempre lo stesso risultato (o di ottenerne uno che sia “migliore” degli altri). Volendo, si potrebbe “dimostrare” che S assume un qualsiasi valore in \mathbb{Z} , riordinando opportunamente i termini: come si vede, pensare che la somma di una serie sia la “somma di infiniti termini” porta rapidamente a delle conclusioni paradossali.

Se, invece, la serie converge “bene”⁽¹¹⁾, si ottiene lo stesso risultato S qualsiasi sia la “regola” scelta per sommare i vari termini della successione. Ad esempio, se $-1 < q < 1$ abbiamo

$$S = \sum_{k=0}^{+\infty} q^k = \lim_{n \rightarrow +\infty} \sum_{k=0}^n q^k = \lim_{n \rightarrow +\infty} \frac{1 - q^{n+1}}{1 - q} = \frac{1}{1 - q},$$

ma anche

$$S = \sum_{k=0}^{+\infty} q^{2k} + \sum_{k=0}^{+\infty} q^{2k+1}.$$

Infatti

$$\sum_{k=0}^{+\infty} q^{2k} = \sum_{k=0}^{+\infty} (q^2)^k = \lim_{n \rightarrow +\infty} \frac{1 - q^{2n+2}}{1 - q^2} = \frac{1}{1 - q^2},$$

⁽¹¹⁾Qualsiasi cosa ciò significhi...

e

$$\sum_{k=0}^{+\infty} q^{2k} = q \sum_{k=0}^{+\infty} (q^2)^k = \lim_{n \rightarrow +\infty} q \frac{1 - q^{2n+2}}{1 - q^2} = \frac{q}{1 - q^2},$$

e si ha

$$\frac{1}{1 - q^2} + \frac{q}{1 - q^2} = \frac{1 + q}{1 - q^2} = \frac{1}{1 - q}.$$

Lo stesso risultato si otterrebbe sommando prima sui multipli di 3, poi sui multipli di 3 più 1, ed infine sui multipli di 3 più 2, o sommando in una maniera “qualsiasi” (purché corretta).

Si noti che nel caso $-1 < q < 0$ (quando cioè i termini della serie sono a segno alterno), spezzando la somma sugli indici pari e dispari si ottengono due serie a termini di segno costante, ed entrambe convergenti.

ESERCIZIO 9.4. Sia q_m una successione convergente ad 1 e tale che $q_m \neq 1$ per ogni m in \mathbb{N} . Sia n un naturale **fissato**. Calcolare, usando le (9.1) e (9.2) e i teoremi sui limiti,

$$\lim_{m \rightarrow +\infty} \frac{q_m^{n+1} - 1}{q_m - 1}.$$

Risposta 9.4: Si ha, per le (9.1) e (9.2),

$$\lim_{m \rightarrow +\infty} \frac{q_m^{n+1} - 1}{q_m - 1} = \lim_{m \rightarrow +\infty} \sum_{k=0}^n q_m^k = \sum_{k=0}^n \left(\lim_{m \rightarrow +\infty} q_m^k \right) = \sum_{k=0}^n 1 = n + 1.$$

ESERCIZIO 9.5. Supponiamo di avere a disposizione un numero infinito di mattoni tutti uguali, omogenei e di lunghezza unitaria. Poggiamo il primo mattone a terra, perfettamente in piano; successivamente sistemiamo un secondo mattone esattamente sotto il primo. Spostiamo ora il secondo mattone verso destra, facendolo scorrere finché il sistema resta in equilibrio. È chiaro che possiamo spostarlo verso destra esattamente di mezza unità. Mettiamo ora un terzo mattone sotto il secondo, e facciamolo scorrere verso destra finché il sistema resta in equilibrio. Quanto possiamo spostarlo? Possiamo arrivare fino al punto in cui l'estremo sinistro del terzo mattone si trova esattamente sotto il baricentro del sistema composto dai primi due mattoni. Adesso mettiamo un quarto mattone, e ripetiamo l'operazione; e così via. Dopo n mattoni, dove si trova l'estremo destro x_n dell' n -simo mattone se l'estremo sinistro del primo mattone è nell'origine?

Risposta 9.5: Come detto, l'estremo sinistro dell' n -simo mattone deve coincidere con il baricentro del sistema formato dai primi $n - 1$ mattoni. Dove si trova quest'ultimo? Il baricentro del primo mattone, b_1 , si trova nel punto $\frac{1}{2}$. Mettendo il secondo mattone e spostandolo verso destra finché il suo estremo destro non si trova sotto b_1 , il baricentro del sistema dei primi 2 mattoni è dato da

$$b_2 = \frac{1}{2} \left[1 \cdot b_1 + \left(b_1 + \frac{1}{2} \right) \right],$$

ovvero, per $n = 2$,

$$\frac{1}{\text{peso degli } n \text{ mattoni}},$$

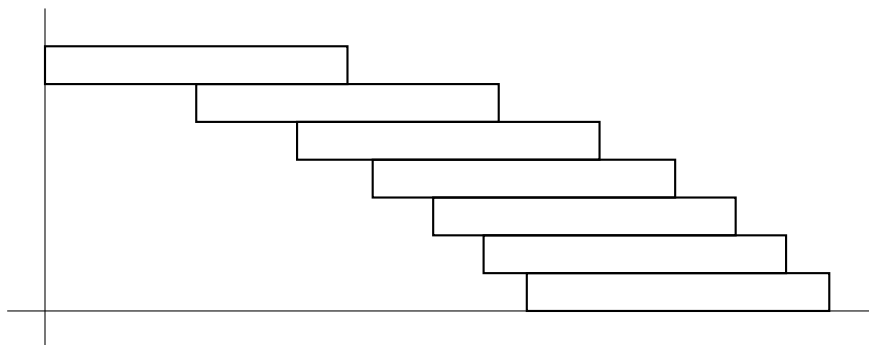
per la somma tra

$$(\text{peso dei primi } n - 1 \text{ mattoni}) \cdot (\text{baricentro dei primi } n - 1 \text{ mattoni}),$$

e

(peso di un mattone) · (baricentro dell' n -simo mattone).Generalizzando, la formula per il calcolo di b_n è

$$b_n = \frac{1}{n} \left[(n-1) \cdot b_{n-1} + \left(b_{n-1} + \frac{1}{2} \right) \right] = b_{n-1} + \frac{1}{2n}.$$

Ovviamente, l'estremo destro dell' n -simo mattone si trova in $b_{n-1} + 1$, cosicché

$$x_n = b_{n-1} + 1 = b_{n-2} + \frac{1}{2(n-1)} + 1 = b_{n-3} + \frac{1}{2(n-2)} + \frac{1}{2(n-1)} + 1 = \dots = 1 + \sum_{k=1}^{n-1} \frac{1}{2k},$$

da cui

$$x_n = 1 + \frac{1}{2} \sum_{k=1}^{n-1} \frac{1}{k}.$$

Una volta svolto l'esercizio precedente, ci troviamo davanti al problema di calcolare esplicitamente x_n . Dalla definizione di serie, è chiaro che x_n è esattamente il termine $(n-1)$ -simo della serie associata alla successione a_k che vale $\frac{1}{2k}$ per ogni $k \geq 1$, e $a_0 = 1$, ma non abbiamo alcuna idea di quanto valga realmente x_n . Quello che possiamo però chiederci è cosa succede alla successione x_n quando n tende all'infinito. Converge? Diverge?

Per semplicità, consideriamo solo

$$S_n = \sum_{k=1}^n \frac{1}{k},$$

che è la cosiddetta **serie armonica**. Iniziamo ad osservare che la successione S_n è monotona strettamente crescente. Infatti $S_{n+1} = S_n + \frac{1}{n+1} > S_n$. Pertanto, per il Teorema 5.3, S_n ammette limite, finito o più infinito che sia. Inoltre, tale limite coincide con quello di una qualsiasi sottosuccessione estratta da S_n (per il Teorema 8.3). Consideriamo allora la sottosuccessione corrispondente a 2^k , ovvero la sottosuccessione S_{2^k} , con $k \geq 0$. Abbiamo

$$\begin{aligned} S_1 &= 1, \\ S_2 &= 1 + \left(\frac{1}{2} \right), \\ S_4 &= 1 + \left(\frac{1}{2} \right) + \left[\frac{1}{3} + \frac{1}{4} \right], \\ \dots \\ S_{2^k} &= 1 + \left(\frac{1}{2} \right) + \left[\frac{1}{3} + \frac{1}{4} \right] + \dots + \left\{ \frac{1}{2^{k-1} + 1} + \dots + \frac{1}{2^k} \right\}. \end{aligned}$$

Notiamo che nelle parentesi tonde c'è un solo termine, uguale a $\frac{1}{2}$, nelle parentesi quadre ci sono due termini, entrambi maggiori di $\frac{1}{4}$, mentre nelle parentesi graffe ci sono esattamente 2^{k-1} termini, tutti maggiori di $\frac{1}{2^k}$. Pertanto,

$$\begin{aligned} S_{2^k} &\geq 1 + \left(\frac{1}{2}\right) + \left[\frac{1}{4} + \frac{1}{4}\right] + \dots + \left\{\frac{1}{2^k} + \dots + \frac{1}{2^k}\right\} \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \dots + \frac{1}{2} = 1 + \frac{k}{2}. \end{aligned}$$

Pertanto, S_{2^k} diverge a più infinito, e quindi S_n diverge a più infinito. Con calcoli analoghi, raggruppando diversamente i termini, si può dimostrare che $S_{2^{k-1}}$ è minore di $1 + k$. Questo significa che S_n è compresa tra $c_1 \log_2(n)$ e $c_2 \log_2(n)$, con c_1 e c_2 costanti opportune. Che l'andamento di S_n fosse di tipo logaritmico (il che vuol dire che diverge sì, ma molto lentamente) lo si poteva dedurre anche dall'immagine della "scala" che definisce S_n .

Come conseguenza del fatto che S_n diverge, accumulando un numero sufficientemente elevato di mattoni è possibile allontanarsi indefinitamente dal primo mattone, a patto però di salire molti gradini: infatti, per spostarsi di 5 unità (ovvero $x_n > 5$), servono 1675 mattoni, essendo

$$x_{1675} = 5,0002427859978895338951523982598487229001.$$

Appare dunque evidente la necessità di usare del cemento per "incollare" i gradini di una scala, in modo da poter arrivare da qualche parte in tempo finito...

Ricapitolando, è possibile che alcune serie (in realtà, la maggior parte di esse) non abbiano un'espressione esplicita in termini analitici del termine generico, ed è quindi necessario ricorrere ad altri strumenti per dimostrarne la convergenza (o la divergenza); nel caso della serie associata alla successione $\frac{1}{k}$, abbiamo concluso che la serie divergeva osservando 1) che ammetteva limite essendo monotona crescente e 2) calcolando tale limite per mezzo di stime su una sottosuccessione (che ha, comunque, lo stesso limite della successione di partenza).

Consideriamo ora la successione $a_k = \frac{1}{k^2}$, definita per $k \geq 1$, e chiediamoci se la serie associata converga o diverga. Come prima, la serie associata è monotona crescente, e pertanto ammette sicuramente limite (finito o infinito). Studiamo allora la sottosuccessione S_{2^k} . Abbiamo

$$\begin{aligned} S_1 &= 1, \\ S_2 &= 1 + \left(\frac{1}{4}\right), \\ S_4 &= 1 + \left(\frac{1}{4}\right) + \left[\frac{1}{9} + \frac{1}{16}\right], \\ \dots & \\ S_{2^k} &= 1 + \left(\frac{1}{4}\right) + \left[\frac{1}{9} + \frac{1}{16}\right] + \dots + \left\{\frac{1}{(2^{k-1} + 1)^2} + \dots + \frac{1}{2^{2k}}\right\}. \end{aligned}$$

Nelle parentesi tonde c'è un unico termine, $\frac{1}{4}$, che è minore di $\frac{1}{2}$, nelle parentesi quadre ci sono due termini, entrambi minori di $\frac{1}{4} = \frac{1}{2^2}$, nelle parentesi graffe ci sono

2^{k-1} termini, tutti minori di $\frac{1}{(2^{k-1})^2}$. Pertanto,

$$\begin{aligned} S_{2^k} &\leq 1 + \left(\frac{1}{(2^0)^2} \right) + \left[\frac{1}{(2^1)^2} + \frac{1}{(2^1)^2} \right] \\ &\quad + \dots + \left\{ \frac{1}{(2^{k-1})^2} + \dots + \frac{1}{(2^{k-1})^2} \right\} \\ &= 1 + \frac{2^0}{(2^0)^2} + \frac{2^1}{(2^1)^2} + \dots + \frac{2^{k-1}}{(2^{k-1})^2} \\ &= 1 + \sum_{m=0}^{k-1} \frac{1}{2^m} = 1 + 2 \left(1 - \frac{1}{2^k} \right) \leq 3. \end{aligned}$$

Essendo S_{2^k} limitata, se ne deduce che anche S_n è limitata (se S_n fosse illimitata, S_n divergerebbe a più infinito, quindi anche S_{2^k} dovrebbe divergere ed essere pertanto illimitata), e quindi S_n ammette limite finito. Non sappiamo quanto valga questo limite⁽¹²⁾, ma sappiamo che esiste ed è finito.

ESERCIZIO 9.6. Dimostrare, ripetendo il ragionamento precedente, che se $\alpha > 1$ allora la serie di termine generico $\frac{1}{n^\alpha}$ è convergente. Trovare una dimostrazione semplice del fatto che se $\alpha < 1$ allora la serie associata alla successione $\frac{1}{n^\alpha}$ è divergente. La serie di termine generico $\frac{1}{n^\alpha}$ si dice **serie armonica generalizzata**.

Risposta 9.6: Se $\alpha < 1$, allora $n^\alpha \leq n$, e quindi $\frac{1}{n^\alpha} \geq \frac{1}{n}$. Pertanto, la serie associata a $\frac{1}{n^\alpha}$ è maggiore della serie associata a $\frac{1}{n}$, che abbiamo visto essere divergente.

ESERCIZIO 9.7. Dimostrare, come nell'esercizio precedente, che la serie di termine generico $\frac{1}{n \ln^2(n)}$ è convergente.

Risposta 9.7: Stimando S_{2^k} , l'addendo tra parentesi graffe è maggiorato da una somma della forma

$$\frac{1}{2^{k-1} \ln^2(2^{k-1})} + \dots + \frac{1}{2^{k-1} \ln^2(2^{k-1})},$$

gli addendi essendo 2^{k-1} , ovvero da un termine come

$$\frac{1}{\ln^2(2^{k-1})} = \frac{1}{(k-1)^2 \ln^2(2)}.$$

Pertanto, S_{2^k} è minore di uno degli elementi della serie di termine generico $\frac{1}{k^2}$, che abbiamo detto essere limitata. Ne consegue che la successione S_{2^k} è limitata, e dunque S_n è convergente (essendo monotona).

Negli esempi precedenti abbiamo sempre usato il fatto che, essendo la successione a_k non negativa, la successione S_n era monotona crescente, e pertanto ammetteva sempre limite. Questo è, ovviamente, un fatto generale.

TEOREMA 9.8. Sia a_k una successione di numeri reali per la quale esiste k_0 in \mathbb{N} tale che $a_k \geq 0$ per ogni $k \geq k_0$. Allora la serie di termine generico a_k ammette limite (finito, o più infinito).

Dimostrazione. Se $n \geq k_0$, allora $S_{n+1} = S_n + a_{n+1} \geq S_n$. Pertanto, S_n è monotona crescente per $n \geq k_0$ e quindi ammette limite. ■

⁽¹²⁾Nel senso che per ora non possiamo calcolarlo; il valore è (magia! magia!) $\frac{\pi^2}{6}$.

Notiamo *en passant* che la condizione di segno su a_k è richiesta da un certo punto in poi: come già per le successioni, modificando un numero finito degli a_k non cambia la convergenza della serie associata (anche se, e diversamente da quanto accade per il limite di una successione, cambia ovviamente il valore della somma della serie).

Esistono delle ipotesi sulla successione a_k che “garantiscono” la convergenza della serie associata? Ahimé, no. Esiste però una condizione che deve essere necessariamente verificata se la serie è convergente.

TEOREMA 9.9. *Sia a_k una successione di numeri reali tale che la serie associata S_n sia convergente. Allora a_k tende a zero.*

Dimostrazione. Sia S la somma della serie, ovvero il limite di S_n . Siccome la sottosuccessione S_{n+1} converge anch'essa ad S , e si ha $S_{n+1} - S_n = a_n$, ne segue che a_n tende a zero come differenza tra due successioni che hanno lo stesso limite finito. ■

Come conseguenza del teorema precedente, se la successione a_k **non** tende a zero, la serie associata non può convergere. Se la successione a_k è non negativa e non tende a zero, allora la serie associata non può che divergere positivamente. Si osservi che la condizione di convergenza a zero di a_k è una condizione necessaria, ma non sufficiente: ad esempio, la serie armonica ha il termine generico tendente a zero, ma non è convergente.

Dal momento che in generale non è possibile calcolare esplicitamente il valore di S_n , l'unica cosa che si può dire di una serie è se essa converga o meno. A tal proposito — unitamente alla conoscenza del comportamento di alcune serie particolari, come le serie associate a progressioni geometriche, o la serie armonica generalizzata — sono utili i seguenti teoremi.

TEOREMA 9.10 (Criterio del confronto). *Siano a_k e b_k due successioni di numeri reali, con $0 \leq a_k \leq b_k$ per ogni k in \mathbb{N} . Se la serie associata ad a_k diverge, la serie associata a b_k diverge; se la serie associata a b_k converge, la serie associata ad a_k converge.*

Dimostrazione. È sufficiente osservare che le serie associate ad a_k e b_k ammettono limite (essendo a termini non negativi), e che la successione delle somme parziali associata ad a_k è maggiorata dalla successione delle somme parziali associata a b_k . Per ottenere la tesi, basta allora applicare i teoremi di confronto tra successioni. ■

Al solito, per ottenere la tesi del precedente teorema è sufficiente che la condizione $0 \leq a_k \leq b_k$ sia soddisfatta per ogni $k \geq k_0$.

TEOREMA 9.11 (Criterio del confronto asintotico – 1). *Siano a_k e b_k due successioni di numeri reali non negativi, e supponiamo che $b_k \neq 0$ per k sufficientemente grande, e che*

$$\lim_{k \rightarrow +\infty} \frac{a_k}{b_k} = L,$$

con L numero reale maggiore di zero. Allora le serie associate ad a_k e b_k o convergono entrambe, o divergono entrambe.

Dimostrazione. Se $\frac{a_k}{b_k}$ converge ad $0 < L < +\infty$, allora

$$\frac{L}{2} \leq \frac{a_k}{b_k} \leq \frac{3L}{2},$$

per ogni k sufficientemente grande. Pertanto (ricordando che b_k è non negativa),

$$\frac{L}{2} b_k \leq a_k \leq \frac{3L}{2} b_k.$$

Per ottenere la tesi, è sufficiente applicare il teorema precedente, osservando che la serie di termine generico $\frac{L}{2} b_k$ (o $\frac{3L}{2} b_k$) ha lo stesso comportamento della serie di termine generico b_k . ■

Se il limite vale 0, o infinito, il teorema precedente è vero “solo a metà”.

TEOREMA 9.12 (Criterio del confronto asintotico – 2). *Siano a_k e b_k due successioni di numeri reali non negativi, e supponiamo che*

$$\lim_{k \rightarrow +\infty} \frac{a_k}{b_k} = 0.$$

Se la serie associata ad a_k diverge, diverge anche la serie associata a b_k ; se quest'ultima converge, converge anche la serie associata ad a_k . Se, invece

$$\lim_{k \rightarrow +\infty} \frac{a_k}{b_k} = +\infty,$$

i ruoli di a_k e b_k sono scambiati.

Dimostrazione. Se $\frac{a_k}{b_k}$ tende a zero, allora per k sufficientemente grande si ha $0 \leq a_k \leq b_k$, e si può applicare il Teorema 9.10. Se $\frac{a_k}{b_k}$ diverge, allora per k sufficientemente grande si ha $a_k \geq b_k \geq 0$, e si può applicare il Teorema 9.10. ■

I criteri esposti precedentemente sono utili quando ci troviamo davanti ad una serie di termine generico a_k , e riusciamo — in qualche modo — a confrontarla con una serie di termine generico b_k della quale sappiamo dire se converge o meno. Come facciamo, però, a sapere se una serie converge o meno? Fino ad ora, abbiamo visto solo alcuni esempi: la serie geometrica di termine generico q^k (che converge se $|q| < 1$, e diverge o non converge altrimenti) e la serie armonica generalizzata di termine generico $1/k^\alpha$, che converge per $\alpha > 1$ e diverge altrimenti.

Supponiamo adesso di avere a che fare con la serie

$$\sum_{k=1}^{+\infty} \frac{2^k}{k!},$$

che soddisfa la condizione necessaria (dato che il termine generico $a_k = \frac{2^k}{k!}$ è infinitesimo) e quindi potrebbe convergere o divergere. Con cosa confrontarla? E la — financo peggiore — serie

$$\sum_{k=1}^{+\infty} \frac{k!}{k^k},$$

che soddisfa sempre la condizione necessaria, con cosa paragonarla? Quale successione — il cui comportamento della serie corrispondente ci è noto — scegliere per applicare il Teorema 9.10 o il Teorema 9.11?

Fortunatamente, abbiamo a disposizione altri due criteri per dire se una serie converge o no.

TEOREMA 9.13 (Criterio del rapporto). *Sia $\{a_k\}$ una successione di numeri positivi, e supponiamo esista*

$$L = \lim_{k \rightarrow +\infty} \frac{a_{k+1}}{a_k}.$$

*Se $L < 1$, allora la serie di termine generico a_k è convergente; se $L > 1$, allora la serie di termine generico a_k è divergente. Se $L = 1$, **non si può dire nulla**.*

Dimostrazione. Supponiamo che sia $L < 1$. Sia allora $\varepsilon > 0$ tale che $L + \varepsilon < 1$, e sia $k_0 > 0$ (che esiste per definizione di limite) tale che

$$-\varepsilon < \frac{a_{k+1}}{a_k} - L < \varepsilon, \quad \forall k \geq k_0.$$

Da questa disuguaglianza otteniamo, sempre per $k \geq k_0$,

$$\frac{a_{k+1}}{a_k} \leq L + \varepsilon \quad \implies \quad a_{k+1} \leq (L + \varepsilon) a_k.$$

Sia ora $m \geq 1$. Si ha allora, applicando ripetutamente l'ultima disuguaglianza, che

$$a_{k_0+m} \leq (L + \varepsilon) a_{k_0+m-1} \leq (L + \varepsilon)^2 a_{k_0+m-2} \leq \dots \leq (L + \varepsilon)^m a_{k_0}.$$

Pertanto,

$$0 \leq \sum_{k=1}^{+\infty} a_k = \sum_{k=1}^{k_0-1} a_k + \sum_{k=k_0}^{+\infty} a_k \leq \sum_{k=1}^{k_0-1} a_k + a_{k_0} \sum_{m=0}^{+\infty} (L + \varepsilon)^m.$$

Essendo $L + \varepsilon < 1$, l'ultima serie è convergente (è una serie geometrica), e quindi la serie di termine generico a_k è convergente.

Se $L > 1$, è sufficiente scegliere $\varepsilon > 0$ tale che $L - \varepsilon > 1$ e usare ripetutamente il fatto che $a_k \geq (L - \varepsilon) a_k$ per arrivare a

$$\sum_{k=1}^{+\infty} a_k = \sum_{k=1}^{k_0-1} a_k + \sum_{k=k_0}^{+\infty} a_k \geq \sum_{k=1}^{k_0-1} a_k + a_{k_0} \sum_{m=0}^{+\infty} (L - \varepsilon)^m,$$

e l'ultima serie è divergente. ■

TEOREMA 9.14 (Criterio della radice). *Sia $\{a_k\}$ una successione di numeri positivi, e supponiamo esista*

$$L = \lim_{k \rightarrow +\infty} \sqrt[k]{a_k}.$$

*Se $L < 1$, allora la serie di termine generico a_k è convergente; se $L > 1$, allora la serie di termine generico a_k è divergente. Se $L = 1$, **non si può dire nulla**.*

Dimostrazione. Supponiamo che sia $L < 1$. Sia allora $\varepsilon > 0$ tale che $L + \varepsilon < 1$, e sia $k_0 > 0$ (che esiste per definizione di limite) tale che

$$-\varepsilon < \sqrt[k]{a_k} - L < \varepsilon, \quad \forall k \geq k_0.$$

Da questa disuguaglianza otteniamo, sempre per $k \geq k_0$,

$$\sqrt[k]{a_k} \leq L + \varepsilon \quad \implies \quad a_k \leq (L + \varepsilon)^k.$$

Dal momento che la serie di termine generico $(L + \varepsilon)^k$ è convergente, la tesi segue dal Teorema 9.10.

Se $L > 1$, è sufficiente scegliere $\varepsilon > 0$ tale che $L - \varepsilon > 1$, ed osservare che a_k risulta essere definitivamente maggiore di $(L - \varepsilon)^k$, che è il termine generico di una serie divergente. ■

OSSERVAZIONE 9.15. Negli enunciati dei due teoremi precedenti, è esplicitamente affermato che se $L = 1$ non si può dire nulla del carattere della serie. Ed infatti, se consideriamo le due successioni $a_k = \frac{1}{k}$ e $b_k = \frac{1}{k^2}$, che sono tali che le serie associate sono rispettivamente divergente e convergente, abbiamo

$$\lim_{k \rightarrow +\infty} \frac{a_{k+1}}{a_k} = \lim_{k \rightarrow +\infty} \sqrt[k]{a_k} = 1 = \lim_{k \rightarrow +\infty} \sqrt[k]{b_k} = \lim_{k \rightarrow +\infty} \frac{b_{k+1}}{b_k}.$$

ESERCIZIO 9.16. Dire se convergono o divergono le serie

$$\sum_{k=1}^{+\infty} \frac{2^k}{k!}, \quad \sum_{k=1}^{+\infty} \frac{k!}{k^k}, \quad \sum_{k=1}^{+\infty} \frac{k^k}{3^k k!}, \quad \sum_{k=1}^{+\infty} \frac{k^k}{e^k k!}.$$

Risposta 9.16: Le prime tre convergono, l'ultima diverge (si usi la formula di Stirling!).

Fin qui, il caso in cui il termine generico a_k della serie sia non negativo (fatto questo che, ricordiamo, implica che la serie o converge o diverge positivamente: *tertium non datur*). Che possiamo dire se la successione a_k cambia segno? Una prima "risposta" a questa domanda è data da una definizione.

DEFINIZIONE 9.17. Data la serie di termine generico a_k , si dice che la serie è **assolutamente convergente** se è convergente la serie

$$\sum_{k=1}^{+\infty} |a_k|.$$

In modo analogo, si definisce **assolutamente divergente** una serie di termine generico a_k tale che la serie di termine generico $|a_k|$ sia divergente.

L'utilità della definizione precedente è resa evidente dal seguente risultato.

TEOREMA 9.18. *Se una serie di termine generico a_k è assolutamente convergente, allora è convergente.*

Dimostrazione. Definiamo $b_k = \max(a_k, 0)$. Chiaramente, si ha $0 \leq b_k \leq |a_k|$; dal momento che la serie di termine generico $|a_k|$ è convergente, è convergente per il Teorema 9.10 la serie di termine generico b_k . Se definiamo $c_k = \max(-a_k, 0)$, abbiamo analogamente che $0 \leq c_k \leq |a_k|$, e quindi che la serie di termine generico c_k è convergente. In altre parole, si ha che esistono L ed M , numeri reali, tali che

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^n b_k = L, \quad \lim_{n \rightarrow +\infty} \sum_{k=1}^n c_k = M.$$

Ma allora, per le operazioni sui limiti, e per la linearità della somma, abbiamo

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^n [b_k - c_k] = L - M \in \mathbb{R}.$$

In altre parole, abbiamo dimostrato che la serie di termine generico $b_k - c_k$ è convergente. D'altra parte, e come si verifica facilmente, $a_k = b_k - c_k$, da cui la tesi. ■

Come conseguenza del risultato precedente, se il termine generico della serie cambia segno, e non sappiamo che peschi pigliare, possiamo passare a studiare la serie dei moduli e — qualora quest'ultima sia convergente — concludere che anche la serie di partenza è convergente. Se, invece, la serie dei moduli diverge, non possiamo dire nulla sulla convergenza della serie di partenza (se non che è assolutamente divergente).

Come fare in questo caso? Ancora una volta, ci viene in aiuto un teorema quando il termine generico a_k ha una forma particolare.

TEOREMA 9.19 (Criterio di Leibnitz). *Sia $a_k = (-1)^k b_k$, con $\{b_k\}$ una successione di numeri reali tali che:*

- 1) $b_k \geq 0$;
- 2) $b_k \geq b_{k+1}$;
- 3) b_k tende a zero.

Allora la serie di termine generico a_k è convergente.

Dimostrazione. Sia

$$S_n = \sum_{k=1}^n a_k = \sum_{k=1}^n (-1)^k b_k.$$

La tesi consiste nel dimostrare che la successione $\{S_n\}$ è convergente. Consideriamo inizialmente la sottosuccessione $\{S_{2n}\}$. Abbiamo, essendo b_k decrescente,

$$S_{2(n+1)} = \sum_{k=1}^{2n+2} (-1)^k b_k = \sum_{k=1}^{2n} (-1)^k b_k - b_{2n+1} + b_{2n+2} \leq \sum_{k=1}^{2n} (-1)^k b_k = S_{2n}.$$

In altre parole, la sottosuccessione $\{S_{2n}\}$ è monotona decrescente. Consideriamo ora la sottosuccessione $\{S_{2n+1}\}$; analogamente a prima, abbiamo

$$S_{2(n+1)+1} = \sum_{k=1}^{2n+1+2} (-1)^k b_k = \sum_{k=1}^{2n+1} (-1)^k b_k + b_{2n+2} - b_{2n+3} \geq \sum_{k=1}^{2n+1} (-1)^k b_k = S_{2n+1},$$

cosicché la sottosuccessione $\{S_{2n+1}\}$ è monotona crescente. Inoltre, essendo b_k non negativa, si ha

$$S_{2n+1} = \sum_{k=1}^{2n+1} (-1)^k b_k = \sum_{k=1}^{2n} (-1)^k b_k - b_{2n+1} \leq S_{2n}.$$

Abbiamo pertanto la seguente catena di disuguaglianze, valida per ogni n in \mathbb{N} :

$$S_1 \leq S_3 \leq \dots \leq S_{2n-1} \leq S_{2n+1} \leq S_{2n} \leq S_{2n-2} \leq \dots \leq S_4 \leq S_2.$$

Se segue che la sottosuccessione $\{S_{2n}\}$ è monotona decrescente e limitata dal basso da S_1 . Pertanto, esiste un numero reale S_p tale che

$$\lim_{n \rightarrow +\infty} S_{2n} = S_p.$$

Inoltre, essendo la sottosuccessione $\{S_{2n+1}\}$ monotona crescente e limitata dall'alto da S_2 , esiste un numero reale S_d tale che

$$\lim_{n \rightarrow +\infty} S_{2n+1} = S_d.$$

Inoltre, si ha

$$S_d - S_p = \lim_{n \rightarrow +\infty} [S_{2n+1} - S_{2n}] = \lim_{n \rightarrow +\infty} -b_{2n+1} = 0,$$

dato che la successione b_k è infinitesima per ipotesi. Pertanto, $S_d = S_p = S$. Dal momento che le due sottosuccessioni $\{S_{2n}\}$ e $\{S_{2n+1}\}$ “esauriscono” tutta la successione $\{S_n\}$, si ha che la successione $\{S_n\}$ converge al numero reale S ⁽¹³⁾. ■

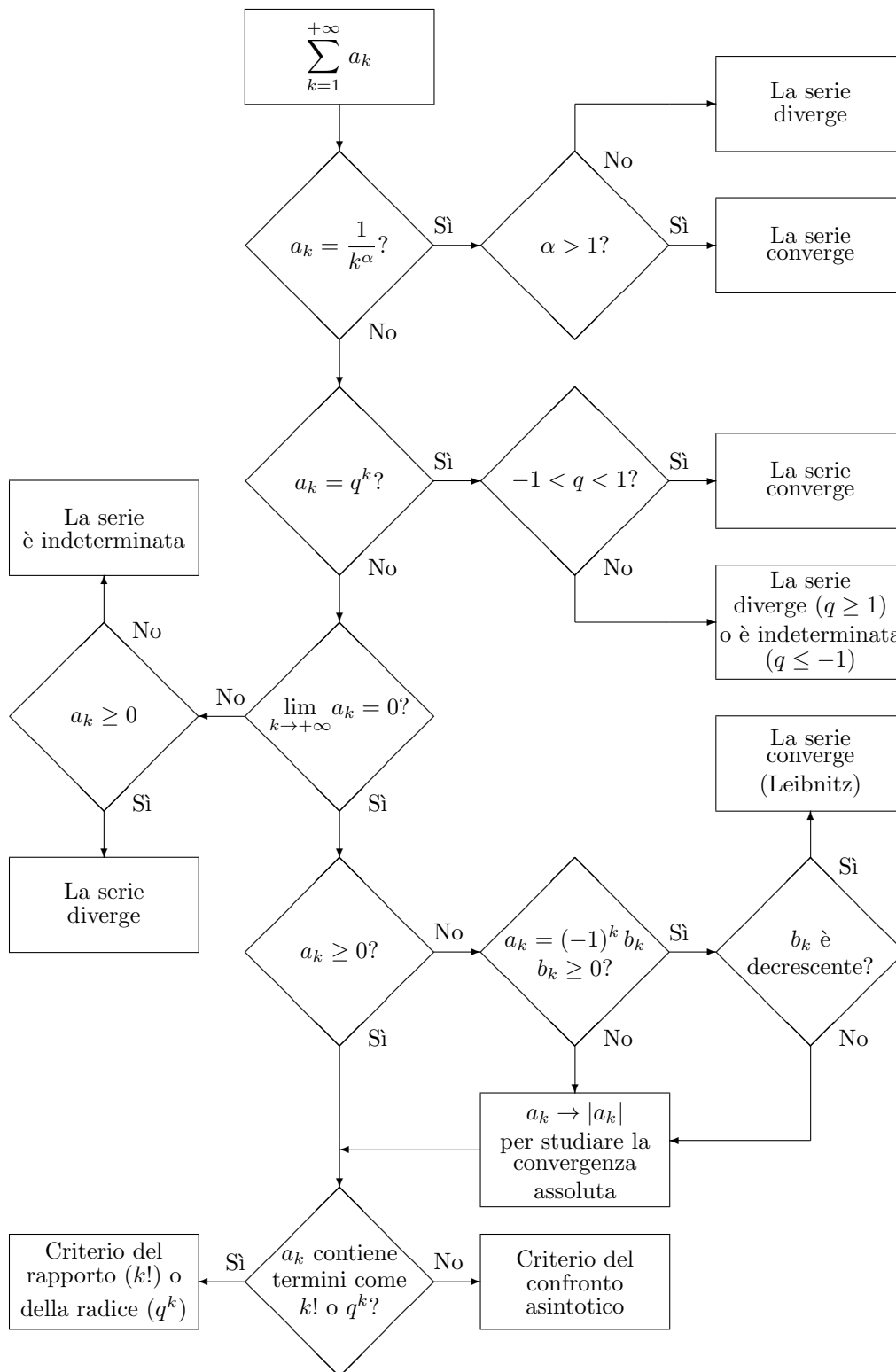
Un esempio di applicazione del precedente risultato è il caso della serie

$$\sum_{k=1}^{+\infty} \frac{(-1)^k}{k}.$$

Se consideriamo la convergenza assoluta, otteniamo la serie armonica, che è divergente. Se applichiamo il Teorema 9.19 (e possiamo farlo perché la successione $b_k = \frac{1}{k}$ è non negativa, decrescente e infinitesima) otteniamo che la serie è convergente.

OSSERVAZIONE 9.20. Rileggendo la dimostrazione del Teorema 9.19, si vede facilmente che se “eliminiamo” la condizione 3) (ovvero il fatto che la successione b_k è infinitesima), vale a dire se eliminiamo la condizione necessaria di convergenza della serie, otteniamo una serie indeterminata. Infatti, detto $B > 0$ il limite della successione b_k (che esiste perché b_k è non negativa e decrescente per le ipotesi 1) e 2)), abbiamo $S_d - S_p = -B$, cosicché la sottosuccessione di indice pari converge ad un limite diverso dalla sottosuccessione di indice dispari, e quindi la successione non converge.

⁽¹³⁾Se n è grande, e pari, S_n è “arbitrariamente vicina” a S ; se n è grande e dispari, S_n è “arbitrariamente vicina” a S . Se n è grande, o è pari (e allora...) o è dispari (e allora...).



FORMULA DI STIRLING

Questo è solo un cenno di dimostrazione della *formula di Stirling*

$$\lim_{n \rightarrow +\infty} \frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} = 1.$$

La dimostrazione non è difficile ma richiede strumenti che ancora non abbiamo sviluppato (e quindi l'abbiamo omessa. Potreste cercare una dimostrazione alternativa...)

Per induzione si ha

$$n! = \int_0^{+\infty} e^{-t} t^{n-1} dt.$$

Ponendo $t = n + s\sqrt{2n}$ otteniamo

$$n! = \sqrt{2nn}^n e^{-n} \int_{-\sqrt{n/2}}^{\infty} e^{-s\sqrt{2n}} (1 + s\sqrt{2/n})^n ds = n^n e^{-n} \sqrt{2n} \int_{-\sqrt{n/2}}^{\infty} e^{-s^2 f(s\sqrt{2/n})} ds$$

dove $f(r) = \frac{2}{r^2}(r - \log(1+r))$. Quando $n \rightarrow +\infty$, l'argomento $r = s\sqrt{2/n}$ di f tende a 0^+ ; dato che $\lim_{r \rightarrow 0^+} f(r) = 1$ e f decresce (notare che $f' \leq 0$), otteniamo $e^{-s^2 f(s\sqrt{2/n})} \uparrow e^{-s^2}$ per ogni s . Passando al limite sotto il segno di integrale concludiamo che l'integrale tende a $\int_{-\infty}^{+\infty} e^{-s^2} ds = \sqrt{\pi}$ da cui la tesi (il passaggio al limite sotto integrale è banale usando il Teorema di Beppo Levi dalla teoria della misura, ma si può anche dimostrare direttamente).

Il teorema di **Riemann-Dini**

DEFINIZIONE 1. Una **permutazione** di \mathbb{N} è una funzione $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ biunivoca.

ESEMPIO 2. La funzione $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ definita da

$$\sigma(\mathbb{N}) = \{1, 3, 2, 4, 5, 7, 6, 8, 9, 11, 10, 12, \dots, 4n+1, 4n+3, 4n+2, 4n+4, \dots\}$$

è una permutazione di \mathbb{N} .

TEOREMA 3 (B. Riemann-U. Dini). Sia a_k una successione di numeri reali tale che

$$(1) \quad \sum_{k=0}^{+\infty} a_k = S, \quad \sum_{k=0}^{+\infty} |a_k| = +\infty.$$

Allora, fissato L in \mathbb{R} esiste una permutazione σ di \mathbb{N} tale che

$$\sum_{k=0}^{+\infty} a_{\sigma(k)} = L.$$

DIMOSTRAZIONE. Per semplicità, supponiamo che sia $a_k = (-1)^k b_k$, con $b_k \geq 0$ (in altre parole, la serie è a segni alterni). Le ipotesi (1) diventano quindi

$$(2) \quad \sum_{k=0}^{+\infty} (-1)^k b_k < +\infty, \quad \sum_{k=0}^{+\infty} b_k = +\infty.$$

Iniziamo con osservare che si ha

$$(3) \quad \lim_{k \rightarrow +\infty} b_k = 0.$$

Infatti, per la condizione necessaria di convergenza di una serie si ha che $(-1)^k b_k$ tende a zero; quindi $|(-1)^k b_k|$ tende a zero, ma $|(-1)^k b_k| = |(-1)^k| |b_k| = b_k$, e quindi la (3).

Siano ora

$$E_n = \sum_{k=0}^n b_{2k}, \quad O_n = \sum_{k=0}^n b_{2k+1},$$

e notiamo che si ha, per ogni n in \mathbb{N} ,

$$S_{2n+1} = \sum_{k=0}^{2n+1} (-1)^k b_k = \sum_{\substack{k=0 \\ k \text{ pari}}}^{2n+1} (-1)^k b_k + \sum_{\substack{k=0 \\ k \text{ dispari}}}^{2n+1} (-1)^k b_k = E_n - O_n,$$

mentre

$$T_{2n+1} = \sum_{k=0}^{2n+1} |(-1)^k b_k| = \sum_{k=0}^{2n+1} b_k = \sum_{k=0}^n b_{2k} + \sum_{k=0}^n b_{2k+1} = E_n + O_n.$$

Dal momento che E_n e O_n sono successioni monotone crescenti (dato che sono definite come somme di quantità positive), esistono i limiti di E_n e di O_n , che possono essere o numeri reali, ovvero più infinito. Se però una delle due successioni ha limite finito, si ha allora una contraddizione con le ipotesi. Infatti, se — per esempio — O_n convergesse ad un numero reale M , si avrebbe

$$\begin{aligned} \lim_{n \rightarrow +\infty} E_n &= \lim_{n \rightarrow +\infty} [S_{2n+1} + O_n] = S + M \in \mathbb{R}, \\ \lim_{n \rightarrow +\infty} E_n &= \lim_{n \rightarrow +\infty} [T_{2n+1} - O_n] = +\infty - M = +\infty, \end{aligned}$$

che è chiaramente una contraddizione (per l'unicità del limite). Si ha pertanto, come conseguenza delle ipotesi sulla successione $a_k = (-1)^k b_k$, che

$$(4) \quad \sum_{\substack{k=0 \\ k \text{ pari}}}^{+\infty} b_k = \lim_{n \rightarrow +\infty} \sum_{k=0}^n b_{2k} = \lim_{n \rightarrow +\infty} E_n = +\infty,$$

e

$$(5) \quad \sum_{\substack{k=0 \\ k \text{ dispari}}}^{+\infty} b_{2k+1} = \lim_{n \rightarrow +\infty} \sum_{k=0}^n b_{2k+1} = \lim_{n \rightarrow +\infty} O_n = +\infty.$$

Sia ora $h \geq 1$ un numero naturale fissato. Allora

$$\begin{aligned} \sum_{k=h}^n b_{2k} &= \sum_{k=0}^n b_{2k} - \sum_{k=0}^{h-1} b_{2k} = E_n - \sum_{k=0}^{h-1} b_{2k}, \\ \sum_{k=h}^n b_{2k+1} &= \sum_{k=0}^n b_{2k+1} - \sum_{k=0}^{h-1} b_{2k+1} = O_n - \sum_{k=0}^{h-1} b_{2k+1}. \end{aligned}$$

Dato che le due somme da 0 a $h-1$ sono due quantità fissate (perché h è fissato), da (4) e (5) segue che, qualsiasi sia $h \geq 0$, si ha

$$(6) \quad \lim_{n \rightarrow +\infty} \sum_{k=h}^n b_{2k} = +\infty, \quad \lim_{n \rightarrow +\infty} \sum_{k=h}^n b_{2k+1} = +\infty.$$

Dalla prima delle (6) segue che se M è un numero reale positivo, e h è un numero intero, esiste \bar{n} in \mathbb{N} (dipendente sia da M che da h) tale che

$$\sum_{k=h}^n b_{2k} > M, \quad \forall n \geq \bar{n}.$$

In altre parole, non è vuoto l'insieme

$$E = \left\{ n \in \mathbb{N} : \sum_{k=h}^n b_{2k} > M \right\}.$$

Per il principio del buon ordinamento, esiste il minimo di E :

$$\rho = \min(E) = \min \left(\left\{ n \in \mathbb{N} : \sum_{k=h}^n b_{2k} > M \right\} \right).$$

Dalla definizione di minimo, segue che ρ appartiene a E , mentre $\rho-1$ non vi appartiene, e quindi

$$\sum_{k=h}^{\rho-1} b_{2k} \leq M < \sum_{k=h}^{\rho} b_{2k}.$$

Ma allora, dato che $b_{2\rho} \geq 0$,

$$\left| \sum_{k=h}^{\rho} b_{2k} - M \right| \leq \left| \sum_{k=h}^{\rho} b_{2k} - \sum_{k=h}^{\rho-1} b_{2k} \right| = b_{2\rho}.$$

Abbiamo pertanto dimostrato la seguente proprietà:

PROPRIETÀ 4. Per ogni $M > 0$ e per ogni $h \geq 0$ esiste un intero ρ (dipendente da M e da h) tale che

$$(7) \quad \sum_{k=h}^{\rho-1} b_{2k} \leq M < \sum_{k=h}^{\rho} b_{2k},$$

con la convenzione che se $h = 0$ e $\rho = 0$ la somma a sinistra vale 0. Se ne deduce che ρ è tale che

$$(8) \quad \left| \sum_{k=h}^{\rho} b_{2k} - M \right| \leq b_{2\rho}.$$

Ripetendo lo stesso ragionamento per le somme con indici dispari, usando la seconda delle (6), possiamo dedurre la seguente proprietà:

PROPRIETÀ 5. Per ogni $M > 0$ e per ogni $j \geq 0$ esiste un intero η (dipendente da M e da h) tale che

$$(9) \quad \sum_{k=j}^{\eta-1} b_{2k+1} \leq M < \sum_{k=j}^{\eta} b_{2k+1},$$

con la convenzione che se $j = 0$ e $\eta = 0$ la somma a sinistra vale 0. Se ne deduce che η è tale che

$$(10) \quad \left| \sum_{k=j}^{\eta} b_{2k} - M \right| \leq b_{2\eta+1}.$$

A questo punto possiamo iniziare la dimostrazione vera e propria. Supponiamo, per semplicità, $L > 0$. L'idea è la seguente: sommiamo alcuni termini di indice pari della successione, finché la somma non supera L . A questo punto, sottraiamo termini di indice dispari, finché il totale non scende sotto L ; continuiamo sommando termini di posto pari (non sommati precedentemente) finché non si supera di nuovo L , per poi sottrarre elementi di posto dispari (non sottratti precedentemente) finché non si scende sotto L , e così via. Che questo procedimento si possa fare, e che in questo modo si ottenga una successione di somme che converge a L discende dalle Proprietà 4 e 5.

Vediamo perché.

Applichiamo la Proprietà 4 con $h_0 = 0$ e $M = L$. Esiste quindi un numero intero ρ tale che (si vedano (7) e (8))

$$(11) \quad \sum_{k=0}^{\rho-1} b_{2k} \leq L < \sum_{k=0}^{\rho} b_{2k}, \quad \left| \sum_{k=0}^{\rho} b_{2k} - L \right| \leq b_{2\rho}.$$

Definiamo $p_0 = 2\rho$, $h_1 = \rho + 1$ e

$$\Pi_0 = \{0, 2, 4, \dots, p_0 - 2, p_0\}, \quad P_0 = \sum_{k \in \Pi_0} a_k = \sum_{k \in \Pi_0} b_k, \quad T_0 = P_0.$$

Grazie a queste definizioni, le (11) diventano

$$(12) \quad T_0 > L, \quad |T_0 - L| \leq b_{p_0}.$$

Avendo “superato” L , applichiamo la Proprietà 5 con $j_0 = 0$ e $M = T_0 - L$. Troviamo quindi un numero intero η tale che (si vedano (9) e (10))

$$(13) \quad \sum_{k=0}^{\eta-1} b_{2k+1} \leq T_0 - L < \sum_{k=0}^{\eta} b_{2k+1}, \quad \left| \sum_{k=0}^{\eta} b_{2k+1} - (T_0 - L) \right| \leq b_{2\eta+1}.$$

Definiamo $d_0 = 2\eta + 1$, $j_1 = \eta + 1$ e

$$\Omega_0 = \{1, 3, 5, \dots, d_0 - 2, d_0\}, \quad D_0 = \sum_{k \in \Omega_0} a_k = - \sum_{k \in \Omega_0} b_k, \quad T_1 = T_0 + D_0.$$

Osserviamo che la prima delle (13) si può riscrivere come $T_0 - L < -D_0$, mentre all'interno del modulo, nella seconda, c'è la quantità $-D_0 - T_0 + L$; in altre parole, dalle due formule di (13) segue che

$$(14) \quad T_1 < L, \quad |T_1 - L| \leq b_{d_0}.$$

Continuiamo, applicando la Proprietà 4 con $h = h_1$ e $M = L - T_1$. Troviamo un numero intero ρ tale che

$$(15) \quad \sum_{k=h_1}^{\rho-1} b_{2k} \leq L - T_1 < \sum_{k=h_1}^{\rho} b_{2k}, \quad \left| \sum_{k=h_1}^{\rho} b_{2k} - (L - T_1) \right| \leq b_{2\rho}.$$

Definiamo $p_1 = 2\rho$, $h_2 = \rho + 1$ e

$$\Pi_1 = \{p_0 + 2, \dots, p_1 - 2, p_1\}, \quad P_1 = \sum_{k \in \Pi_1} a_k = \sum_{k \in \Pi_1} b_k, \quad T_2 = T_1 + P_1.$$

Con queste definizioni abbiamo (usando (15)) che

$$T_2 > L, \quad |T_2 - L| \leq b_{p_1}.$$

Proseguiamo ulteriormente, usando la Proprietà 5 con $j = j_1$ e $M = T_2 - L$; troviamo η intero tale che

$$\sum_{k=j_1}^{\eta-1} b_{2k+1} \leq T_2 - L < \sum_{k=j_1}^{\eta} b_{2k+1}, \quad \left| \sum_{k=j_1}^{\eta} b_{2k+1} - (T_2 - L) \right| \leq b_{2\eta+1}.$$

Definiamo $d_1 = 2\eta + 1$, $j_2 = \eta + 1$ e

$$\Omega_1 = \{d_0 + 2, \dots, d_1 - 2, d_1\}, \quad D_1 = \sum_{k \in \Omega_1} a_k = - \sum_{k \in \Omega_1} b_k, \quad T_3 = T_2 + D_1.$$

Con queste definizioni, si ha che

$$T_3 < L, \quad |T_3 - L| \leq b_{d_1}.$$

A questo punto concludiamo la dimostrazione procedendo per induzione.

Definiamo $p_{-1} = -2$, $d_{-1} = -1$ e $\mathcal{D}_0 = \emptyset$. Vogliamo far vedere che per ogni m in \mathbb{N} esistono due numeri p_m e d_m (uno pari e uno dispari) e due insiemi Π_m e Ω_m , tali che:

- Π_m è costituito dai numeri pari compresi tra $p_{m-1} + 2$ e p_m ;
- Ω_m è costituito dai numeri dispari compresi tra $d_{m-1} + 2$ e d_m ;
- si ha $p_m > p_{m-1}$ e $d_m > d_{m-1}$; definiamo $h_m = p_m/2 + 1$ e $j_m = (d_m - 1)/2 + 1$;

- definendo

$$\mathcal{P}_m = \bigcup_{k=0}^m \Pi_k, \quad \mathcal{D}_m = \bigcup_{k=0}^m \Omega_k,$$

e

$$T_{2m} = \sum_{k \in \mathcal{P}_m \cup \mathcal{D}_{m-1}} a_k, \quad T_{2m+1} = \sum_{k \in \mathcal{P}_m \cup \mathcal{D}_m} a_k,$$

si ha

$$(16) \quad T_{2m} > L, \quad |T_{2m} - L| \leq b_{p_m}, \quad T_{2m+1} < L, \quad |T_{2m+1} - L| \leq b_{d_m}.$$

Chiaramente, se $m = 0$ gli insiemi Π_0 e Ω_0 esistono (li abbiamo costruiti prima) e T_0 e T_1 (definiti prima) soddisfano le disuguaglianze (16) (si vedano (12) e (14)): il passo iniziale dell'induzione è stato compiuto!

Supponiamo ora di aver già costruito $\{\Pi_\ell\}$ e $\{\Omega_\ell\}$ per ogni ℓ tra 1 e m (l'ipotesi induttiva), nonché $\{T_\ell\}$ con ℓ tra 0 e $2m + 1$, e dimostriamo che siamo in grado di costruire Π_{m+1} , Ω_{m+1} , T_{2m+2} e T_{2m+3} in modo tale che tutte le proprietà elencate sopra siano soddisfatte.

Iniziamo¹ con applicare la Proprietà 4 con $h = h_m$ e $M = L - T_{2m+1}$. Troviamo quindi un numero intero ρ tale che

$$\sum_{k=h_m}^{\rho-1} b_{2k} \leq L - T_{2m+1} < \sum_{k=h_m}^{\rho} b_{2k}, \quad \left| \sum_{k=h_m}^{\rho} b_{2k} - (L - T_{2m+1}) \right| \leq b_{2\rho}.$$

Definiamo $p_{m+1} = 2\rho > p_m$, $h_{m+1} = \rho + 1$,

$$\Pi_{m+1} = \{p_m + 2, \dots, p_{m+1} - 2, p_{m+1}\}, \quad P_{m+1} = \sum_{k \in \Pi_{m+1}} a_k = \sum_{k \in \Pi_{m+1}} b_k,$$

e

$$T_{2m+2} = T_{2m+1} + P_{m+1}.$$

Con queste definizioni si ha

$$T_{2m+2} > L, \quad |T_{2m+2} - L| \leq b_{p_{m+1}}.$$

Infine (siamo quasi arrivati!) applichiamo la Proprietà 5 con $j = j_m$ e $M = T_{2m+2} - L$, trovando un numero intero η tale che

$$\sum_{k=j_m}^{\eta-1} b_{2k+1} \leq T_{2m+2} - L < \sum_{k=j_m}^{\eta} b_{2k+1}, \quad \left| \sum_{k=j_m}^{\eta} b_{2k+1} - (T_{2m+2} - L) \right| \leq b_{2\eta+1}.$$

Definiamo $d_{m+1} = 2\eta + 1 > d_m$, $j_{m+1} = \eta + 1$ e

$$\Omega_{m+1} = \{d_m + 2, \dots, d_{m+1} - 2, d_{m+1}\}, \quad D_{m+1} = \sum_{k \in \Omega_{m+1}} a_k = - \sum_{k \in \Omega_{m+1}} b_k,$$

e

$$T_{2m+3} = T_{2m+2} + D_{m+1}.$$

Con queste definizioni si ha

$$T_{2m+3} < L, \quad |T_{2m+3} - L| \leq b_{d_{m+1}}.$$

Raccogliendo tutte le informazioni, abbiamo che Π_{m+1} e Ω_{m+1} sono formati da soli numeri pari o numeri dispari, che $p_{m+1} > p_m$ e che $d_{m+1} > d_m$ (e quindi le due

¹Qui seguiamo il ragionamento che ci ha portato alla costruzione di Π_1 e di T_2 avendo già costruito Π_0 , Ω_0 , T_0 e T_1 .

successioni sono crescenti), e che T_{2m+2} e T_{2m+3} soddisfano le (16) per $m + 1$. Il secondo passo dell'induzione è completo!

Per il principio di induzione, abbiamo quindi che esistono due famiglie $\{\Pi_m\}$ e $\{\Omega_m\}$ di insiemi, con m appartenente a \mathbb{N} , con le proprietà elencate sopra. Definiamo

$$\mathcal{P}_m = \bigcup_{k=0}^m \Pi_k = \{0, 2, 4, \dots, p_m\}, \quad \mathcal{D}_m = \bigcup_{k=0}^m \Omega_k = \{1, 3, 5, \dots, d_m\},$$

ed osserviamo che, dato che p_m e d_m sono due successioni (di interi!) strettamente crescenti, si ha

$$\lim_{m \rightarrow +\infty} p_m = \lim_{m \rightarrow +\infty} d_m = +\infty,$$

e quindi

$$\mathcal{P} = \bigcup_{k=0}^{+\infty} \Pi_k = \{\text{numeri pari}\}, \quad \mathcal{D} = \bigcup_{k=0}^{+\infty} \Omega_k = \{\text{numeri dispari}\}.$$

Pertanto, l'unione di tutti i Π_m e di tutti gli Ω_m è l'insieme \mathbb{N} dei numeri naturali, che può essere "riordinato" (costruendo così una permutazione di \mathbb{N}) in questa maniera:

$$\mathbb{N} = \sigma(\mathbb{N}) = \Pi_0 \cup \Omega_0 \cup \Pi_1 \cup \Omega_1 \cup \dots \cup \Pi_m \cup \Omega_m \cup \dots$$

Inoltre, dato che sia p_m che d_m divergono, si ha per la (3) che

$$\lim_{m \rightarrow +\infty} b_{p_m} = 0 = \lim_{m \rightarrow +\infty} b_{d_m}.$$

Ma allora, per il teorema dei carabinieri, da (16) segue che

$$0 \leq \lim_{m \rightarrow +\infty} |T_{2m} - L| \leq \lim_{m \rightarrow +\infty} b_{p_m} = 0, \quad 0 \leq \lim_{m \rightarrow +\infty} |T_{2m+1} - L| \leq \lim_{m \rightarrow +\infty} b_{d_m} = 0,$$

e quindi sia T_{2m} che T_{2m+1} convergono a L ; in definitiva (dato che le due sottosuccessioni "indici pari" e "indici dispari" esauriscono \mathbb{N}) si ha

$$\lim_{m \rightarrow +\infty} T_m = L.$$

A questo punto osserviamo che, per definizione,

$$\lim_{m \rightarrow +\infty} T_m = \sum_{k \in \sigma(\mathbb{N})} a_k = \sum_{k=0}^{+\infty} a_{\sigma(k)} = L,$$

che è quello che si voleva dimostrare. □

ESEMPIO 6. Consideriamo la successione

$$a_k = \frac{(-1)^k}{k+1}.$$

Per il criterio di Leibnitz (la successione $b_k = \frac{1}{k+1}$ è positiva, decrescente e infinitesima) si ha che

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

mentre

$$\sum_{k=0}^{+\infty} |a_k| = \sum_{k=0}^{+\infty} \frac{1}{k+1} = +\infty,$$

dato che si tratta della serie armonica. Per il Teorema di Riemann-Dini, esiste una permutazione σ di \mathbb{N} tale che, ad esempio,

$$\sum_{k=0}^{+\infty} a_{\sigma(k)} = \sum_{k=0}^{+\infty} \frac{(-1)^{\sigma(k)}}{\sigma(k) + 1} = \frac{\pi}{2} \approx 1.5707963267948966.$$

Per trovare σ , seguiamo la dimostrazione del Teorema. Il primo passo è quello di trovare n tale che

$$\sum_{k=0}^n a_{2k} = \sum_{k=0}^n \frac{1}{2k+1} = 1 + \frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{2n+1} > \frac{\pi}{2},$$

e che tale n sia il primo valore per cui la somma “supera” $\frac{\pi}{2}$. Dopo lunghi e faticosi calcoli², si ha che

$$1.5333333333333333 \approx \sum_{k=0}^2 \frac{1}{2k+1} < \frac{\pi}{2} < \sum_{k=0}^3 \frac{1}{2k+1} \approx 1.676190476190476.$$

Per quanto visto nella dimostrazione, si ha $p_0 = 6$, $h_1 = 4$, $\Pi_0 = \{0, 2, 4, 6\}$ e

$$T_0 = P_0 = \sum_{k=0}^3 \frac{1}{2k+1}.$$

Dobbiamo ora iniziare a sottrarre gli elementi di posto dispari, ovvero le frazioni della forma $\frac{1}{2k+2}$, con $k \geq 0$. Si vede facilmente che se sottraiamo la frazione per $k = 0$ (che vale $\frac{1}{2} = 0.5$), “scendiamo” sotto $\frac{\pi}{2}$, dato che $1.676190476190476 - 0.5 = 1.176190476190476 < \frac{\pi}{2}$. Definiamo dunque $d_0 = 1$, $j_0 = 1$, $\Omega_0 = \{1\}$ e

$$T_1 = T_0 - \sum_{k=0}^0 \frac{1}{2k+2} < \frac{\pi}{2}.$$

Proseguendo, si ha che

$$T_1 + \sum_{k=4}^8 \frac{1}{2k+1} = 1.5806239512121862 > \frac{\pi}{2},$$

da cui segue che $p_1 = 16$, $h_2 = 9$, $\Pi_1 = \{8, 10, 12, 14, 16\}$, e

$$T_2 = T_1 + \sum_{k=4}^8 \frac{1}{2k+1} > \frac{\pi}{2}.$$

Per il momento, la somma è data da

$$T_2 = \left(1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7}\right) - \left(\frac{1}{2}\right) + \left(\frac{1}{9} + \frac{1}{11} + \frac{1}{13} + \frac{1}{15} + \frac{1}{17}\right).$$

Proseguendo, si trovano i seguenti valori:

$$d_1 = 3, p_2 = 28, d_2 = 5, p_3 = 40, d_3 = 7, p_4 = 52, \dots$$

Si noti che il fatto che sia sufficiente sottrarre un unico termine per tornare “sotto” L è essenzialmente dovuto alla scelta di $L = \frac{\pi}{2}$. Siccome al primo passaggio servono 4 termini per “scavalcare” L , la somma parziale supera L di poco (di meno di $\frac{1}{7}$); il che vuol dire che basta sottrarre $\frac{1}{2}$ (che è maggiore di $\frac{1}{7}$) per passare sotto a L ; e questa cosa si ripete ad ogni passaggio, perché l’ultima frazione che aggiungiamo è sempre

²Non tanto lunghi, e non tanto faticosi, in realtà...

molto più piccola dell'ultima frazione che avevamo sottratto, e quindi basta sottrarre un unico valore per scendere sotto L .

Tanto per chiarire: se ripetiamo l'algoritmo per 1000 passaggi, abbiamo $p_{1000} = 11576$ e $d_{1000} = 1999$ (che è il millesimo numero dispari) e

$$\sum_{k=0}^{11576} \frac{1}{2k+1} - \sum_{k=0}^{1999} \frac{1}{2k+2} = 1.5703775048971924,$$

non molto lontano da $\frac{\pi}{2} = 1.5707963267948966$ (ma nemmeno vicinissimo, la convergenza è molto lenta).

Se vogliamo ottenere come limite π , invece, abbiamo $p_{1000} = 267610$ e $d_{1000} = 1999$ (essenzialmente perché $p_0 = 150$: bisogna sommare molti termini per salire per la prima volta sopra π) e

$$\sum_{k=0}^{267610} \frac{1}{2k+1} - \sum_{k=0}^{1999} \frac{1}{2k+2} = 3.1410927170968543.$$

Essendo $\pi = 3.141592653589793$ siamo parecchio lontani... Va un po' meglio con

$$\sum_{k=0}^{267610} \frac{1}{2k+1} - \sum_{k=0}^{1998} \frac{1}{2k+2} = 3.1415927170968545,$$

ma, di nuovo, siamo ancora abbastanza lontani.

ESERCIZIO 7. ³ Data la successione

$$a_k = \frac{(-1)^k}{k+1},$$

calcolare p_m e d_m , con m da 0 a 10, corrispondenti al valore $L = 0.6931471805599453$.

ESERCIZIO 8. ⁴ Dimostrare che è possibile trovare una permutazione σ di \mathbb{N} tale che

$$\sum_{k \in \sigma(\mathbb{N})} a_{\sigma(k)} = +\infty.$$

ESEMPIO 9. Qui sotto un programma python (non molto efficiente...) che calcola gli insiemi \mathcal{P}_k e \mathcal{D}_k .

```
import math

# elemento generico della serie
# sarà chiamata con esp = 1, e quindi è 1/(val + 1)
def function(val, esp):
    return (val + 1)**(-esp)

# somma elementi della serie finché non si supera il valore 'target'
# l'indice aumenta di 2: passiamo da un pari/dispari al successivo
# ritorna l'indice successivo a quello che ha generato il sorpasso
# e il valore raggiunto
def calc_sum(start_idx, start_val, target):
    val = start_val
    idx = start_idx
    while val <= target:
        val += function(idx, esp)
        idx += 2
```

³Per i coraggiosi che siano arrivati fino a qui.

⁴Per i coraggiosissimi.

```

    return idx, val

# il valore limite
L = math.pi / 2
# l'esponente della funzione (1/(k+1), in questo caso)
esp = 1
# quante iterazioni vogliamo fare?
iterations = 1000

# inizializzazione
total_sum = 0
# indici di base per i pari e per i dispari
idx_eve_old, idx_odd_old = 0, 1
# lista con la permutazione trovata dei naturali
indices = []
for idx in range(iterations):
    # nuovo indice dei pari, e nuovo valore
    # la soglia è L - il valore raggiunto nell'iterazione precedente
    # (vedere la dimostrazione per capire il perché)
    idx_eve_new, temp_val = calc_sum(idx_eve_old, 0, L - total_sum)
    # aggiorniamo il valore raggiunto
    total_sum += temp_val
    # aggiungiamo tutti gli indici pari (l'insieme  $Pi_{\{k\}}$ )
    indices += [i for i in range(idx_eve_old, idx_eve_new, 2)]
    # vecchio = nuovo :)
    idx_eve_old = idx_eve_new
    # ripetiamo per gli indici dispari, sottraendo
    idx_odd_new, temp_val = calc_sum(idx_odd_old, 0, total_sum - L)
    total_sum -= temp_val
    indices += [i for i in range(idx_odd_old, idx_odd_new, 2)]
    idx_odd_old = idx_odd_new

# lista con le somme parziali
partials = []
total = 0
for idx in indices:
    # pari o dispari
    rem = idx % 2
    # 1 - 2*rem = 1 se rem = 0, -1 altrimenti
    total += (1 - (2 * rem)) * function(idx, esp)
    # aggiungiamo la coppia indice, valore
    partials.append([idx, total])
'''
In uscita, indices contiene la permutazione trovata
Ad esempio, per calcolare pi/2:
[
0, 2, 4, 6, 1, 8, 10, 12, 14, 16, 3, 18, 20, 22, 24, 26, 28,
5, 30, 32, 34, 36, 38, 40, 7, 42, 44, 46, 48, 50, 52, 9, 54,
56, 58, 60, 62, 11, 64, 66, 68, 70, 72, 74, 13, 76, 78, 80,
82, 84, 86, 15, 88, 90, 92, 94, 96, 98, 17, 100, 102, 104,
106, 108, 19, 110, 112, 114, 116, 118, 120, 21, 122, 124,
126, 128, 130, 132, 23, 134, 136, 138, 140, 142, 144, 25,
146, 148, 150, 152, 154, 156, 27, 158, 160, 162, 164, 166,
29, 168, 170, 172, 174, 176, 178, 31, 180, 182, 184, 186,
188, 190, 33, 192, 194, 196, 198, 200, 202, 35, 204, 206,
208, 210, 212, 214, 37, 216, 218, 220, 222, 224, 39

```

```
]
mentre partials contiene le coppie (indice, valore)
Ad esempio
[
[0, 1], [2, 1.3333333], [4, 1.5333333] eccetera
]
,,,
```

CAPITOLO 3

Limiti di funzioni

1. Il concetto di funzione.

Alla fine del primo capitolo abbiamo introdotto il concetto di funzione e abbiamo studiato rapidamente alcuni esempi elementari: la potenza n -esima, la radice n -esima, il modulo, le funzioni esponenziali, i logaritmi, e alcune funzioni trigonometriche con le loro inverse.

Il concetto di funzione che stiamo utilizzando è semplice e comprensibile, rivediamolo in dettaglio:

DEFINIZIONE 1.1. Siano A, B due sottoinsiemi di \mathbb{R} . Una funzione f da A in B è una legge che, ad ogni elemento x di A , associa uno ed un solo elemento $y \in B$. Tale elemento y si dice **immagine** di x **tramite** f , o **valore** di f in x , e si indica con $y = f(x)$. La formula

$$f : A \rightarrow B$$

dice sinteticamente che f è una funzione definita sul dominio A a valori in B ; spesso scriveremo anche

$$x \mapsto y = f(x)$$

per indicare la legge che ci permette di calcolare f . L'insieme A si dice **dominio** (o anche **insieme di definizione**) della funzione f , l'insieme B si dice **codominio**, e x si dice **argomento** della funzione f . Associato ad una funzione esiste un secondo sottoinsieme di \mathbb{R} , detto **immagine** di A tramite f , che è l'insieme di tutti i valori assunti da $f(x)$ al variare di x in A :

$$\text{Im}(f) = f(A) = \{y \in \mathbb{R} : \text{esiste } x \text{ in } A \text{ tale che } f(x) = y\}.$$

Un ottimo strumento per visualizzare una funzione è il suo **grafico**, ossia il sottoinsieme del piano cartesiano formato dalle coppie $(x, f(x))$ al variare di x in A :

$$G(f) = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x \in A, y = f(x)\}.$$

Notiamo tre fatti importanti:

- Per assegnare una funzione, non basta scrivere la legge $x \mapsto f(x)$ che ci permette di calcolarla, ma si deve specificare anche il dominio A su cui facciamo variare x . Una funzione è la coppia di un dominio e una legge definita su tale dominio. Per essere ancora più chiari, la funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ di legge $f(x) = x^2$ e la funzione $g : [0, 1] \rightarrow \mathbb{R}$ di legge $g(x) = x^2$ saranno per noi due funzioni diverse (potete pensare a f come “tutta la parabola” e g come “un pezzo di parabola”).
- Se abbiamo una funzione $f : A \rightarrow B$, e C è un sottoinsieme di A , possiamo definire una nuova funzione $g : C \rightarrow B$ riutilizzando la stessa legge $g(x) = f(x)$; diremo che g è la **restrizione** di f all'insieme C e scriveremo $g = f|_C$.
- Se assegnamo soltanto una legge $x \mapsto f(x)$ tramite una combinazione di funzioni elementari (ad esempio $f(x) = \sqrt{x^2 - 3}$), possiamo chiederci per quali valori di x abbia senso calcolare

l'espressione data (nel nostro esempio, tutti gli $x \geq \sqrt{3}$ e tutti gli $x \leq -\sqrt{3}$). Chiaramente possiamo prendere come dominio della funzione l'insieme di tali x , detto il **dominio naturale** o **massimale** della funzione f .

Quando è nato il concetto di funzione? Non è chiaro: Ipparco di Nicea compilò delle tavole della funzione $\sin x$ intorno al 150 a.C.! Tuttavia, di funzioni in un senso generale si cominciò a parlare verso la metà del seicento. Per molto tempo i matematici non si posero il problema di cosa fosse esattamente una funzione, lavoravano utilizzando due concetti vaghi collegati fra di loro: un concetto geometrico (una "curva" nel piano cartesiano: Descartes 1637) e uno analitico (Newton-Leibniz, fine seicento; per loro una funzione era sostanzialmente quella che noi chiamiamo una combinazione di funzioni elementari, somma, prodotto, rapporto, potenza, esponenziale, logaritmo, funzioni trigonometriche e loro inverse. Il termine "funzione" apparve per la prima volta in uno scritto di Leibniz, 1692). Si noti che l'idea di rappresentare graficamente i risultati di misurazioni sperimentali di grandezze fisiche è molto più recente, e risale essenzialmente all'ottocento. I due concetti entrarono in collisione nel settecento, nello studio del moto della corda vibrante: se si pizzica una corda in modo che essa assuma la forma di una retta spezzata, a questa forma corrisponde una funzione o no? Secondo i geometrici ovviamente sì (cosa c'è di più semplice di una spezzata?), secondo gli analitici ovviamente no (come si fa a scrivere una spezzata come combinazione di funzioni elementari?). Dopo un secolo di aspri dibattiti si arrivò all'unificazione e al concetto moderno di funzione; la definizione di Dirichlet (1837) è identica a quella che abbiamo dato noi (anche se lui si sentì in obbligo di aggiungere che la "legge" poteva essere specificata in "qualunque modo"...). In effetti i matematici erano ancora insoddisfatti, che vuol dire esattamente una "legge"? quindi continuarono a discutere sul concetto di funzione ancora per parecchio tempo. La versione definitiva fu elaborata dal gruppo Bourbaki negli anni 1920–30: dati due insiemi A e B , una *relazione* da A in B è un qualunque sottoinsieme del prodotto cartesiano $A \times B$, e una relazione si dice *funzione* se per ogni $x \in A$ c'è al più un $y \in B$ tale che la coppia (x, y) appartiene alla relazione. In pratica, la nostra definizione di grafico $G(f)$ è precisamente la definizione rigorosa di funzione f !

ESERCIZIO 1.2. Determinare il dominio naturale delle seguenti funzioni:

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

$$|x|, \quad \sqrt{|x|-1}, \quad \sqrt{\frac{|x-1|}{|x-2|(3-x)}}, \quad \frac{1}{(1-|x|)(2-|x|)}.$$

Ricordiamo anche la definizione di funzione inversa:

DEFINIZIONE 1.3. Siano $A, B \subseteq \mathbb{R}$ e sia $f : A \rightarrow B$ una funzione. Se per ogni y di $f(A)$ esiste un unico x in A tale che $y = f(x)$, la funzione f si dice **iniettiva** (mentre si dice **suriettiva** se $f(A) = B$). Data una funzione iniettiva resta determinata una funzione $f^{-1} : f(A) \rightarrow \mathbb{R}$, detta funzione **inversa**, che è la funzione che a y in $f(A)$ associa l'unico x di A tale che $y = f(x)$. Notare che se $g = f^{-1}$ è l'inversa di f , allora il dominio di g è esattamente l'immagine di f , e viceversa.

Il grafico $G(f^{-1})$ di una funzione inversa si ottiene semplicemente dal grafico di f scambiando i posti di tutte le coppie: $G(f^{-1}) = \{(y, x) : (x, y) \in G(f)\}$.

Con le funzioni possiamo fare delle operazioni, proprio come con i numeri reali:

DEFINIZIONE 1.4. Siano $f : A \rightarrow \mathbb{R}$ e $g : A \rightarrow \mathbb{R}$ due funzioni. La **somma** $f + g$, la **differenza** $f - g$, il **prodotto** fg , e il prodotto cf di f per un numero reale c sono delle funzioni da A in \mathbb{R} definite nel modo seguente:

$$(f \pm g)(x) = f(x) \pm g(x), \quad (fg)(x) = f(x) \cdot g(x), \quad (cf)(x) = c \cdot f(x).$$

Naturalmente si può definire anche il **rapporto** f/g

$$\frac{f}{g}(x) = \frac{f(x)}{g(x)}$$

ma si dovranno escludere i punti in cui il denominatore si annulla, e la **potenza** $g^f(x) = g(x)^{f(x)}$, nei punti in cui $g > 0$.

Infine, date due funzioni $f : A \rightarrow \mathbb{R}$ e $g : B \rightarrow \mathbb{R}$ tali che $f(A) \subseteq B$, si definisce la **composizione** di g con f , indicata con $g \circ f : A \rightarrow \mathbb{R}$ o anche con $g(f)$, nel modo seguente:

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

Ad esempio, consideriamo le due funzioni $f(x) = x^2$ e $g(x) = \sin x$ definite su tutto \mathbb{R} . Allora,

- la somma $f + g$ ha legge $x^2 + \sin x$ (e dominio naturale \mathbb{R})
- la differenza $f - g$ ha legge $x^2 - \sin x$ (dominio \mathbb{R})
- il prodotto fg ha legge $x^2 \sin x$ (dominio \mathbb{R})
- il rapporto f/g ha legge $x^2/\sin x$ ma il dominio naturale è $\mathbb{R} \setminus \mathbb{Z}\pi$ (l'insieme degli $x \neq k\pi$, $k \in \mathbb{Z}$). E la funzione g/f ?
- la composizione $g \circ f$ ha legge $g(f(x)) = \sin(x^2)$ mentre la composizione $f \circ g$ ha legge $f(g(x)) = \sin^2 x = (\sin x)^2$ (dominio \mathbb{R} in entrambi i casi).

ESERCIZIO 1.5. 1) Sia c una costante, sia $f(x) = cx$. Dimostrare che $f(x) + f(1-x) = f(1)$ per qualunque x . È vero che $f(x^2) = f(x)^2$ per ogni x ? Quanto vale $f(x+4) - f(x^3)$?

b) Sia $f : A \rightarrow \mathbb{R}$ data da $f(x) = x^2 - 2x - 4$, dove $A = [1, 4]$. Qual è l'immagine $f(A)$?

c) Siano $f(x) = x^3 - x - 1$, $g(x) = \frac{1}{x+1}$. Calcolare le funzioni $f + g$, $f - g$, fg , f/g , $f \circ g$ e $g \circ f$. È vero che $f \circ g = g \circ f$?

Ricordiamo infine alcune definizioni che ci permettono di classificare il comportamento delle funzioni:

DEFINIZIONE 1.6. Sia $f : A \rightarrow \mathbb{R}$ una funzione. Diciamo che f è **crescente** (o **non decrescente**) se per tutti i punti $x, y \in A$ tali che $x \leq y$ si ha $f(x) \leq f(y)$. Diciamo che f è **strettamente crescente** se per tutti i punti $x, y \in A$ tali che $x < y$ si ha $f(x) < f(y)$.

In modo simile, diciamo che f è **decrescente** (o **non crescente**) se per tutti gli $x, y \in A$ tali che $x \leq y$ si ha $f(x) \geq f(y)$. Diciamo che f è **strettamente decrescente** se per tutti gli $x, y \in A$ tali che $x < y$ si ha $f(x) > f(y)$.

Una funzione crescente o decrescente si dice **monotona**.

Piccolo quiz: una funzione strettamente crescente è crescente? Esiste una funzione sia crescente che decrescente? Esiste una funzione sia crescente che strettamente decrescente? E se una funzione soddisfa la proprietà $x > y \implies f(x) < f(y)$ che si può dire? E se soddisfa la proprietà $x < y \implies f(x) \leq f(y)$, è crescente o strettamente decrescente?

Una osservazione semplice ma importante:

TEOREMA 1.7. Sia A un sottoinsieme di \mathbb{R} , e sia $f : A \rightarrow \mathbb{R}$ una funzione strettamente crescente (oppure strettamente decrescente). Allora f è **iniettiva**, e **l'inversa è ancora strettamente crescente (strettamente decrescente)**.

Dimostrazione. Consideriamo solo il caso di f crescente (l'altro è analogo). Prendiamo un punto y di $f(A)$, e supponiamo che esistano x e x' in A tali che $f(x) = f(x') = y$. Allora non può essere $x < x'$ perché questo implicherebbe $f(x) < f(x')$ e analogamente non può essere $x > x'$; concludiamo che $x = x'$ ossia f è iniettiva.

Infine, l'inversa f^{-1} è strettamente crescente: infatti, se $y < y'$ sono due punti di $f(A)$ e $x = f^{-1}(y)$, $x' = f^{-1}(y')$, allora deve essere necessariamente $x < x'$ perché se fosse $x \geq x'$ ne seguirebbe $y = f(x) \geq f(x') = y'$.

DEFINIZIONE 1.8. Una funzione $f : A \rightarrow \mathbb{R}$ si dice **limitata superiormente** se la sua immagine $f(A)$ è limitata superiormente; si dice **limitata inferiormente** se $f(A)$ è limitata inferiormente; e

si dice **limitata** se la sua immagine $f(A)$ è un insieme limitato. L'**estremo superiore** e l'**estremo inferiore** di f su A sono definiti allora come l'estremo superiore e inferiore dell'immagine $f(A)$:

$$\sup_A f = \sup f(A), \quad \inf_A f = \inf f(A).$$

Qualora esista il massimo di $f(A)$ esso si chiama **massimo (assoluto)** di f su A e si indica con $\max_A f$; analoga definizione del **minimo (assoluto)** di f su A , indicato con $\min_A f$.

Riepiloghiamo rapidamente la lista delle *funzioni elementari* già introdotte in precedenza, le cui combinazioni (somma, prodotto e composizione) sono alla base dell'analisi:

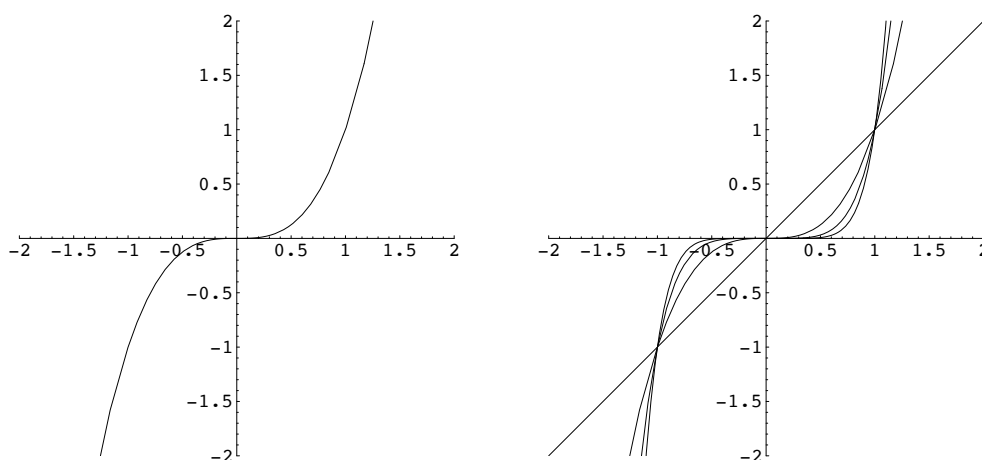
1) **POTENZE** $f(x) = x^\alpha$, dove α è un numero reale fissato. In generale f è definita solo per $x > 0$; se $\alpha \geq 0$, f è definita per $x \geq 0$; se $\alpha \in \mathbb{N}$, f è definita su tutto \mathbb{R} ; e infine se α è un intero negativo, f è definita per tutti gli $x \neq 0$. Limitandoci al dominio $x > 0$, le funzioni x^α sono *strettamente crescenti* se $\alpha > 0$ e *strettamente decrescenti* se $\alpha < 0$ (e naturalmente $x^0 = 1$; sorvoliamo sul problema di definizione per $x = 0$). Che si può dire per $x < 0$? Ricordiamo qui anche le proprietà basilari della potenza con esponente reale

$$x^{\alpha+\beta} = x^\alpha \cdot x^\beta, \quad (x^\alpha)^\beta = x^{\alpha\beta}, \quad x^{-\alpha} = \frac{1}{x^\alpha}$$

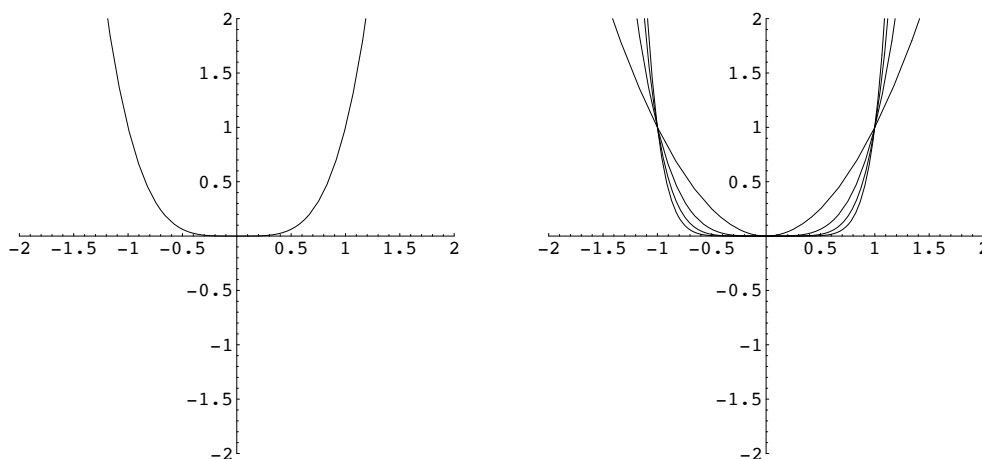
e le notazioni alternative per $n \in \mathbb{N}$, $n \geq 1$

$$x^{1/n} = \sqrt[n]{x}, \quad x^{\alpha/n} = \sqrt[n]{x^\alpha} = (\sqrt[n]{x})^\alpha.$$

Vediamo qui di seguito nella prima figura il grafico della funzione $f(x) = x^3$, nella seconda quello delle funzioni x, x^3, x^5, x^7 sovrapposte.



Analogamente, ecco qui di seguito nella prima figura il grafico della funzione $f(x) = x^4$, nella seconda quello delle funzioni x^2, x^4, x^6, x^8 sovrapposte.



2) **POLINOMI**: un **polinomio** è una funzione del tipo

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n;$$

i numeri reali a_0, \dots, a_n sono fissati e sono detti i coefficienti del polinomio, e se $a_n \neq 0$ chiamiamo n il grado del polinomio. Il dominio è tutto \mathbb{R} . L'immagine è tutto \mathbb{R} se n è dispari; invece se $n \geq 2$ è pari l'immagine di f è un intervallo del tipo $[a, +\infty)$ se $a_n > 0$, ed è un intervallo del tipo $(-\infty, a]$ se $a_n < 0$. Che si può dire quando $n = 0$?

3) **FUNZIONE RAZIONALE**: è semplicemente il rapporto di due polinomi $P(x)/Q(x)$ (definito dove non si annulla il denominatore).

4) **ESPONENZIALE E LOGARITMO**. Se $A > 0$ è un numero reale fissato, la funzione $f(x) = A^x$ definita per $x > 0$ si dice **funzione esponenziale**. La sua immagine è sempre $]0, +\infty[$ (con l'esclusione del caso $A = 1$ perché allora f è la funzione costante uguale a 1), A^x è strettamente crescente se $A > 1$ e strettamente decrescente se $A < 1$. Se $A \neq 1$ allora A^x è invertibile e l'inversa si chiama \log_A , **logaritmo in base A**, definita su $]0, +\infty[$ e crescente per $A > 1$, decrescente per $A < 1$.

Quando la base A è uguale al numero di Nepero

$$e = 2,7182818284590452353602874713526624977572470936999\dots$$

(che abbiamo definito come il limite della successione $(1 + 1/n)^n$), otteniamo la funzione strettamente crescente $f : \mathbb{R} \rightarrow \mathbb{R}$

$$f(x) = e^x$$

detta anche semplicemente l'**esponenziale**; questo è il caso più importante, e il motivo si vedrà dopo l'introduzione delle derivate nei capitoli successivi. Il \log_e si indica anche semplicemente con \log o \ln e si chiama **logaritmo naturale**; è una funzione strettamente crescente $\ln :]0, +\infty[\rightarrow \mathbb{R}$. Alcune proprietà base:

$$A^{x+y} = A^x A^y, \quad (A^x)^y = A^{xy};$$

queste proprietà equivalgono alle seguenti:

$$\log_A x + \log_A y = \log_A(xy), \quad y \cdot \log_A x = \log_A(x^y).$$

5) **FUNZIONI TRIGONOMETRICHE**. Si tratta delle funzioni \sin , \cos , che sono definite su tutto \mathbb{R} con immagine $[-1, 1]$ e periodicità 2π :

$$\forall k \in \mathbb{Z}, x \in \mathbb{R} \quad \sin(x + 2k\pi) = \sin x, \quad \cos(x + 2k\pi) = \cos x$$

e della funzione

$$\operatorname{tg} x = \frac{\sin x}{\cos x}$$

definita per $x \neq k\pi + \pi/2$, $k \in \mathbb{Z}$, con immagine \mathbb{R} e periodicità π :

$$\forall k \in \mathbb{Z}, x \in \mathbb{R} \quad \text{tg}(x + k\pi) = \text{tg } x.$$

Alcune proprietà che useremo:

$$\text{sen}^2 x + \text{cos}^2 x = 1,$$

$$\text{sen}(-x) = -\text{sen } x, \quad \text{cos}(-x) = \text{cos } x, \quad \text{tg}(-x) = -\text{tg } x;$$

inoltre cos è uguale a sen composta con una traslazione:

$$\text{cos}(x) = \text{sen}(x + \pi/2)$$

e valgono le formule di addizione

$$\text{sen}(x \pm y) = \text{sen } x \text{cos } y \pm \text{cos } x \text{sen } y,$$

$$\text{cos}(x \pm y) = \text{cos } x \text{cos } y \mp \text{sen } x \text{sen } y.$$

Un esercizio facile: usando le formule precedenti, dimostrare le formule di prostaferesi

$$\text{sen } x + \text{sen } y = 2 \text{sen} \left(\frac{x+y}{2} \right) \text{cos} \left(\frac{x-y}{2} \right),$$

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

(basta scrivere $x = \frac{x+y}{2} + \frac{x-y}{2}$, $y = \frac{x+y}{2} - \frac{x-y}{2}$ e applicare le formule di addizione). Ricordiamo poi la disuguaglianza

$$\forall 0 < x < \frac{\pi}{2} : \quad 0 < \text{sen } x < x < \text{tg } x$$

da cui segue anche (calcolando la precedente in $-x$)

$$\forall 0 > x > -\frac{\pi}{2} : \quad 0 > \text{sen } x > x > \text{tg } x.$$

Più sinteticamente possiamo scrivere

$$0 \leq |\text{sen } x| \leq |x| \leq |\text{tg } x| \quad \text{per} \quad 0 \leq |x| < \frac{\pi}{2}$$

da cui segue anche $|\text{sen } x| \leq |x|$ per *tutti gli* x (perché?).

Ricordiamo anche che la restrizione di sen all'intervallo $[-\frac{\pi}{2}, \frac{\pi}{2}]$

$$\text{sen} : \left[-\frac{\pi}{2}, \frac{\pi}{2} \right] \rightarrow \mathbb{R}$$

è una funzione strettamente crescente, dunque iniettiva, la cui inversa è strettamente crescente e si indica con arcsin , definita su $[-1, 1]$ e con immagine $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Analogamente

$$\text{cos} : [0, \pi] \rightarrow \mathbb{R}$$

è strettamente decrescente ed ha per inversa la funzione arccos , strettamente decrescente, definita su $[-1, 1]$ con immagine $[0, \pi]$. Infine

$$\text{tg} : \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \rightarrow \mathbb{R}$$

ha per inversa la funzione arctan definita su tutto \mathbb{R} e con immagine $]-\frac{\pi}{2}, \frac{\pi}{2}[$ e strettamente crescente.

ESERCIZIO 1.9. a) Disegnare il grafico delle funzioni x^2 , $x^2 - 3x - 2$, $2x - 4x^2$, $1/x$, $1/(x - 1)$, $\sin(x - 2)$, $\operatorname{tg}(x - 3)$, $-e^x$, $|x|$, $|x - 1|$, $|2 - 3x|$, $|e^x|$, $|1 - e^x|$, trovare il loro dominio (naturale) e la loro immagine, e stabilire se sono crescenti, decrescenti, e iniettive.

b) Stesso esercizio per le funzioni $[x]$ (parte intera di x), $x - [x]$, $|\sin x|$, $|x| + |2x - 1|$.

c) Se componiamo una funzione crescente con una funzione crescente, cosa otteniamo? e una funzione crescente con una decrescente?

d) Disegnare (a piacere) il grafico di una funzione $f : [a, b] \rightarrow \mathbb{R}$. Sapreste disegnare allora i grafici di $-f(x)$, $f(-x)$, $-f(-x)$, $|f(x)|$, $f(x)^2$, $\sqrt{f(x)}$; e quelli di $f(x - 3)$, $f(x + 3)$? Notare che alcune di queste funzioni hanno un insieme di definizione diverso da $[a, b]$.

e) Determinare il dominio naturale delle seguenti funzioni:

$$\frac{1}{e^{2x} - 1}, \quad \operatorname{tg}(e^x), \quad \sqrt{\sin x}, \quad \frac{1}{\operatorname{tg} x}, \quad \frac{1}{\operatorname{tg}(\sqrt{x})}, \quad \ln(x - 5),$$

$$\ln(-x), \quad \ln(x^2 - 2x - 3), \quad \frac{1}{\ln(x^2 - 1)}, \quad \sqrt{\ln x}, \quad \ln\left(\frac{1}{x}\right),$$

$$\arcsin(x^2 - 4), \quad \frac{1}{1 - \operatorname{arc} \operatorname{tg} x}, \quad \arccos e^x.$$

2. Limiti di funzioni.

In questo paragrafo daremo un senso alla formula

$$(2.1) \quad \lim_{x \rightarrow c} f(x) = L.$$

Il concetto che vogliamo definire è più facile da comprendere se si rappresenta graficamente la funzione f , e si immagina di “muovere” il punto x lungo l’ascissa (sembra che questa metafora di movimento sia intimamente legata alla nozione di limite). In corrispondenza del movimento di x sull’asse x , il punto del piano cartesiano $(x, f(x))$ traccia la curva grafico di f . La formula (2.1) allora vuol dire: quando x si avvicina al punto c , il corrispondente valore $f(x)$ si avvicina al numero L , ossia il punto $(x, f(x))$ si muove verso il punto (c, L) .

Quest’idea di movimento è molto intuitiva, ma non si può esprimere precisamente in formule. Invece si può dare una versione “statica” della stessa idea in modo molto preciso; in un certo senso immaginiamo di “fotografare” il movimento in istanti successivi; nelle immagini successive il punto x è sempre più vicino a c e $f(x)$ è sempre più vicino a L . La definizione esatta è la seguente (inventata da Cauchy, ma scritta rigorosamente solo da Weierstrass):

DEFINIZIONE 2.1. Sia A un sottoinsieme di \mathbb{R} contenente due intervalli aperti $]a, c[$ e $]c, b[$ (notare che c può anche non appartenere ad A). Sia $f : A \rightarrow \mathbb{R}$. Diciamo che f **ha limite L nel punto c** , o che f **tende a L per x che tende a c** , e scriviamo

$$(2.2) \quad \lim_{x \rightarrow c} f(x) = L$$

se: per ogni $\varepsilon > 0$ esiste un $\delta = \delta(\varepsilon)$ tale che

$$(2.3) \quad |f(x) - L| < \varepsilon \quad \text{per tutti gli } x \text{ tali che } 0 < |x - c| < \delta.$$

Si scrive anche:

$$f(x) \rightarrow L \quad \text{per } x \rightarrow c.$$

Notiamo che il punto c può anche non appartenere al dominio A , ossia la funzione può anche non essere definita nel punto c ; in ogni caso, anche quando accade che $c \in A$ e quindi conosciamo il valore $f(c)$, tale valore (per il momento!) non ci interessa e non interviene nella definizione di limite.

In questa definizione, δ esprime la distanza di x da c , e ε quella di $f(x)$ da L . Se vogliamo che i valori di $f(x)$ siano “sempre più vicini” ad L , cioè che la differenza $|f(x) - L|$ sia sempre più piccola, e quindi riduciamo il numero ε , dobbiamo prendere dei valori di δ “sempre più piccoli”.

Vediamo alcuni casi tipici che si possono verificare.

1) Il caso più semplice e importante è il seguente: c appartiene al dominio A , e si ha

$$f(x) \rightarrow f(c) \quad \text{per } x \rightarrow c.$$

In altri termini, il limite di f per x che tende a c esiste, ed è uguale esattamente al valore della funzione nel punto c . (Torneremo più avanti su questa situazione: si tratta delle funzioni continue). La quasi totalità delle funzioni elementari rientra in questo caso. L'esempio più ovvio è quello di una funzione costante $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = C$ per ogni x ; allora in qualunque punto $c \in \mathbb{R}$ si ha $f \rightarrow C$ per $x \rightarrow c$. Infatti

$$|f(x) - C| = |C - C| = 0 < \varepsilon$$

è vera comunque si scelgano $\varepsilon > 0$ e $x \in \mathbb{R}$, e quindi la definizione di limite si applica banalmente.

Un esempio più interessante: consideriamo la funzione $f(x) = x^2$ e proviamo a dimostrare che

$$x^2 \rightarrow c^2 \quad \text{per } x \rightarrow c.$$

Seguendo la definizione, fissiamo un $\varepsilon > 0$ e cerchiamo in corrispondenza un $\delta > 0$ con la proprietà (2.3). Ossia vogliamo scegliere δ in modo che sia $|x^2 - c^2| < \varepsilon$ se $|x - c| < \delta$. Notiamo anzitutto che possiamo limitarci ai valori di δ minori di 1, quindi nel seguito sarà $|x - c| < \delta < 1$ da cui segue $|x| < |c| + 1$. Ora possiamo scrivere

$$|x^2 - c^2| = |x + c| \cdot |x - c| \leq (|x| + |c|)|x - c| < (2|c| + 1) \cdot |x - c|.$$

Per esempio, se scegliamo

$$\delta = \frac{\varepsilon}{2|c| + 1}$$

otteniamo subito

$$|x - c| < \delta \quad \implies \quad |x^2 - c^2| < (2|c| + 1) \cdot |x - c| < \varepsilon.$$

2) Un caso più delicato è il seguente: c appartiene al dominio A , e si ha

$$f(x) \rightarrow L \neq f(c) \quad \text{per } x \rightarrow c.$$

Ossia il limite L nel punto c esiste, possiamo anche calcolare il valore di f nel punto c , ma L e $f(c)$ sono diversi. Un esempio (un po' artificiale, ma abbiamo completa libertà nella scelta della legge che definisce f): consideriamo la funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ definita come segue

$$(2.4) \quad f(x) = \begin{cases} 0 & \text{se } x \neq 0, \\ 1 & \text{se } x = 0. \end{cases}$$

In questo caso si ha subito

$$f \rightarrow 0 \quad \text{per } x \rightarrow 0$$

e chiaramente $f(0) \neq 0$.

3) Altro caso interessante: il punto c non appartiene al dominio A , quindi non è possibile calcolare il valore di f in c , e tuttavia quando $x \rightarrow c$ esiste il limite $f(x) \rightarrow L$. In questo caso è naturale *aggiungere* il punto c al dominio A e definire $f(c) = L$. In questo modo diciamo che *abbiamo esteso per continuità* la funzione f al punto c . Questo tipo di comportamento, detto anche di *singolarità eliminabile*, si presenta in vari esempi importanti, e vi torneremo nel seguito. (Talvolta si usa questo termine anche nel caso 2: invece di aggiungere il punto c al dominio e porre $f(c) = L$, in questo caso si può modificare il valore di f in c ponendo $f(c) = L$).

4) Infine, può darsi che il limite di f nel punto c non esista, e in questo caso non importa molto se c appartiene ad A o no.

ESEMPIO 2.2. Un esempio di funzione che non ha limite: sia $f(x) = [x]$ la parte intera di x , che è definita su tutto \mathbb{R} , e consideriamo il suo limite nel punto $x = 1$. Vediamo che $f = 0$ per $0 < x < 1$ mentre $f = 1$ per $1 \leq x < 2$; allora è chiaro che per $x \rightarrow 1$ la funzione f non può tendere ad alcun valore L . Infatti se così fosse avremmo $|f(x) - L| < \varepsilon$ per $0 < |x - 1| < \delta(\varepsilon)$, qualunque sia $\varepsilon > 0$. Ma ad esempio in $x = 1 + \delta/2$ la funzione f vale 1, quindi $|1 - L| < \varepsilon$; invece in $x = 1 - \delta/2$ la f vale 0, quindi $|0 - L| = |L| < \varepsilon$; e in conclusione possiamo scrivere

$$\varepsilon > |1 - L| \geq 1 - |L| > 1 - \varepsilon \implies 2\varepsilon > 1$$

e questo è assurdo perché possiamo scegliere ε piccolo a piacere.

Dimostriamo anzitutto un risultato di confronto molto importante per limiti di funzioni: se una funzione tende ad un limite strettamente positivo in un punto c allora essa deve essere positiva “vicino” a c ; e se una funzione è positiva “vicino” al punto c allora il suo limite nel punto c deve essere positivo.

TEOREMA 2.3. (*Permanenza del segno*). Sia $A \subseteq \mathbb{R}$ contenente $]a, c[\cup]c, b[$, e sia $f : A \rightarrow \mathbb{R}$ una funzione tale che $f \rightarrow L$ per $x \rightarrow c$. Allora:

(i) Supponiamo che $L > 0$. Allora esiste un $\delta > 0$ tale che $f(x) > 0$ per tutti gli x con $0 < |x - c| < \delta$.

(ii) Supponiamo che per un certo $\delta > 0$ si abbia $f(x) \geq 0$ per gli x tali che $0 < |x - c| < \delta$. Allora anche $L \geq 0$.

Proprietà analoghe valgono quando L ed f sono negativi.

Dimostrazione. Per ipotesi, sappiamo che per ogni $\varepsilon > 0$ possiamo trovare δ dipendente da ε tale che

$$L - \varepsilon < f(x) < L + \varepsilon \quad \text{per} \quad 0 < |x - c| < \delta.$$

Per dimostrare (i) basta scegliere $\varepsilon = L/2 > 0$ e osservare che in particolare

$$f(x) > L - \varepsilon = L/2 > 0$$

quando $0 < |x - c| < \delta$. Discorso analogo quando $L < 0$ (provate a scriverlo).

Per dimostrare (ii) procediamo per assurdo: se f fosse positiva ma $L < 0$, allora applicando (i) nel caso L negativo vediamo che f dovrebbe essere negativa vicino a c e questo è impossibile.

ESEMPIO 2.4. Supponiamo di sapere che due funzioni f e g tendono rispettivamente ai limiti L e M nel punto c , e che si ha sempre $f(x) \geq g(x)$ (basterebbe anche: per tutti gli x tali che $0 < |x - c| < \delta$). Allora dev'essere necessariamente $L \geq M$. Infatti la funzione $f - g$ è positiva, e per la permanenza del segno il suo limite $L - M$ deve essere un numero positivo.

Nel capitolo precedente abbiamo lavorato moltissimo sulle successioni e sui loro limiti. Dimostriamo adesso un risultato di collegamento che ci permetterà di riciclare buona parte di quei risultati per calcolare limiti di funzioni:

TEOREMA 2.5. (*Teorema ponte*). Sia $f : A \rightarrow \mathbb{R}$ con $]a, c[\cup]c, b[\subseteq A$. Allora si ha

$$(2.5) \quad \lim_{x \rightarrow c} f(x) = L$$

se e solo se, per qualunque successione di numeri reali $x_n \in A$ che tende a c , con $x_n \neq c$, si ha

$$(2.6) \quad \lim_{n \rightarrow \infty} f(x_n) = L.$$

Dimostrazione. (2.5) \implies (2.6): sappiamo per ipotesi che, comunque si scelga $\varepsilon > 0$, esiste $\delta = \delta(\varepsilon) > 0$ con la proprietà

$$|f(x) - L| < \varepsilon \quad \text{per} \quad 0 < |x - c| < \delta.$$

Se inoltre $x_n \rightarrow c$, $x_n \in A$, con $x_n \neq c$, sappiamo che per ogni $\delta > 0$ esiste un indice $n(\delta)$ a partire dal quale

$$n \geq n(\delta) \implies 0 < |x_n - c| < \delta.$$

Quindi è evidente che, posto $n_\varepsilon = n(\delta(\varepsilon))$,

$$n \geq n_\varepsilon \implies |f(x_n) - L| < \varepsilon$$

ossia $f(x_n) \rightarrow L$.

(2.6) \implies (2.5): procediamo “per assurdo” ossia dimostriamo che se la (2.5) è falsa, allora si può costruire una successione che non soddisfa la condizione (2.6). Dire che la (2.5) è falsa vuol dire: esiste un $\varepsilon_0 > 0$ tale che per ogni $\delta > 0$ si può trovare un punto $x(\delta) \in A$ con le proprietà

$$0 < |x(\delta) - c| < \delta \quad \text{e} \quad |f(x(\delta)) - L| \geq \varepsilon_0.$$

A questo punto è semplicissimo costruire una successione che viola la (2.6): basta porre $x_n = x(1/n)$, cioè scegliamo successivamente $\delta = 1, 1/2, 1/3, 1/4, \dots$ e chiamiamo x_n i punti ottenuti. È chiaro che la successione x_n converge a c perché $0 < |x_n - c| < 1/n$, e d'altra parte non è vero che $f(x_n)$ converge a L perché $|f(x_n) - L| \geq \varepsilon_0$ per tutti gli n .

Il teorema precedente ci consente di dimostrare immediatamente alcune proprietà fondamentali dei limiti di funzioni, come conseguenza delle corrispondenti proprietà delle successioni.

TEOREMA 2.6. (*Operazioni sui limiti*). Siano $A, B \subseteq \mathbb{R}$ contenenti $]a, c[\cup]c, b[$, siano $f : A \rightarrow \mathbb{R}$ e $g : B \rightarrow \mathbb{R}$ due funzioni, e supponiamo che per $x \rightarrow c$ si abbia $f(x) \rightarrow L$ e $g(x) \rightarrow M$. Allora valgono le proprietà: per $x \rightarrow c$,

$$f \pm g \rightarrow L \pm M, \quad f \cdot g \rightarrow L \cdot M, \quad \alpha f \rightarrow \alpha L$$

dove α è un qualunque numero reale. Inoltre se $M \neq 0$ si ha anche

$$\frac{f}{g} \rightarrow \frac{L}{M}.$$

Infine, se $M > 0$ si ha

$$g^f \rightarrow M^L \quad \text{e} \quad \log g \rightarrow \log M.$$

(Nelle ultime tre proprietà, le funzioni f/g , g^f e $\log g$ sono definite solo dove $g \neq 0$ e $g > 0$ rispettivamente).

Dimostrazione. Se $x_n \neq c$ è una qualunque successione in A convergente a c , per il teorema precedente sappiamo che $f(x_n) \rightarrow L$ e $g(x_n) \rightarrow M$. Ma allora per le proprietà delle successioni abbiamo subito che $f(x_n) + g(x_n) \rightarrow L + M$, e applicando di nuovo il teorema precedente otteniamo subito che $f + g \rightarrow L + M$ dato che x_n è una successione qualunque. Lo stesso ragionamento mostra che $f - g \rightarrow L - M$ e $fg \rightarrow LM$.

Le altre proprietà si dimostrano allo stesso modo; nel caso del rapporto è necessario osservare in aggiunta che, essendo $M \neq 0$, ad esempio $M > 0$ (il caso $M < 0$ è identico), applicando il teorema della permanenza del segno si ottiene che $g(x) > 0$ per $0 < |x - c| < \delta$, δ opportuno, e quindi il rapporto f/g è definito per tali x . Una precisazione simile serve per il limite di g^f e $\log g$.

TEOREMA 2.7. (*Carabinieri*). Sia $A \subseteq \mathbb{R}$ contenente $]a, c[\cup]c, b[$, siano $f, g, h : A \rightarrow \mathbb{R}$ tre funzioni, e supponiamo che

$$f(x) \leq g(x) \leq h(x)$$

per tutti gli x . Se $f \rightarrow L$ e $h \rightarrow L$ per $x \rightarrow c$, allora anche $g \rightarrow L$ per $x \rightarrow c$.

Dimostrazione. Consideriamo una qualunque successione $x_n \rightarrow c$, $x_n \neq c$ di punti di A . Sappiamo che $f(x_n) \rightarrow L$ e $g(x_n) \rightarrow L$, e inoltre che $f(x_n) \leq g(x_n) \leq h(x_n)$. Dal Teorema dei Carabinieri per la successione otteniamo subito che $g(x_n) \rightarrow L$, ed essendo x_n qualunque, questo conclude la dimostrazione.

ESERCIZIO 2.8. Sia f una funzione e $g(x) = |f(x)|$. Dimostrare che se $g \rightarrow 0$ per $x \rightarrow c$, allora anche $f \rightarrow 0$ per $x \rightarrow c$.

ESEMPIO 2.9. Provate a “dimostrare” che la funzione $f(x) = x$ ha la proprietà $\lim_{x \rightarrow c} x = c$; abbiamo messo le virgolette perché in effetti c'è ben poco da dimostrare (basta prendere $\delta = \varepsilon$ nella definizione di limite).

A questo punto il Teorema sulle operazioni fra limiti garantisce che $x^2 \rightarrow c^2$ (fare il prodotto $f \cdot f$); analogamente $x^n \rightarrow c^n$; e più in generale, se $P(x)$ è un qualunque polinomio, $P(x) \rightarrow P(c)$; e se $Q(x)$ è un altro polinomio con $Q(c) \neq 0$, si ha anche $P/Q \rightarrow P(c)/Q(c)$.

Inoltre, sempre dal Teorema 2.6, se $x \rightarrow c$ si ha $e^x \rightarrow e^c$, e, se $c \geq 0$, si ha anche $\sqrt{x} \rightarrow \sqrt{c}$; più in generale, se $c > 0$ si ha $x^\alpha \rightarrow c^\alpha$ per ogni numero reale α e anche $\log x \rightarrow \log c$.

ESEMPIO 2.10. Abbiamo dimostrato che, per ogni successione $a_n \rightarrow 0$, si ha

$$\frac{e^{a_n} - 1}{a_n} \rightarrow 1.$$

Se ora applichiamo il Teorema 2.5 otteniamo subito il seguente risultato molto importante:

$$(2.7) \quad \lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1.$$

Allo stesso modo, riutilizzando il corrispondente risultato dimostrato per le successioni, abbiamo

$$(2.8) \quad \lim_{x \rightarrow 0} \frac{\ln(1+x)}{x} = 1.$$

ESEMPIO 2.11. Con il metodo dell'esempio precedente otteniamo senza difficoltà i limiti di $\sin x$, $\cos x$, $\operatorname{tg} x$ e $\frac{\sin x}{x}$ per $x \rightarrow 0$. Proviamo a calcolarli senza ricorrere al Teorema Ponte (per cambiare). Anzitutto si ha

$$(2.9) \quad \lim_{x \rightarrow 0} \sin x = 0.$$

Per calcolare questo limite partiamo dalla disuguaglianza

$$(2.10) \quad 0 \leq |\sin x| \leq |x| \leq |\operatorname{tg} x| \quad \text{per} \quad |x| \leq \frac{\pi}{2}.$$

Si noti che le disuguaglianze sono strette se $x \neq 0$, e che si ha anche

$$(2.11) \quad |\sin x| \leq |x| \quad \text{per tutti gli } x \in \mathbb{R}.$$

Se scriviamo la (2.11) sotto la forma $-x \leq \sin x \leq x$, otteniamo subito la (2.9) applicando il Teorema dei Carabinieri.

Come conseguenze immediate abbiamo

$$(2.12) \quad \lim_{x \rightarrow 0} \cos x = 1, \quad \lim_{x \rightarrow 0} \operatorname{tg} x = \lim_{x \rightarrow 0} \frac{\sin x}{\cos x} = 0.$$

Basta scrivere $\cos x = \sqrt{1 - \sin^2 x}$ e applicare il Teorema 2.6.

Infine studiamo la funzione $\frac{\sin x}{x}$; si noti che la funzione non è definita per $x = 0$. Dato che $\operatorname{tg} x = \frac{\sin x}{\cos x}$, da (2.10) otteniamo subito (dividendo per $\sin x$ e prendendo l'inverso)

$$\cos x < \frac{\sin x}{x} < 1 \quad \text{per} \quad 0 < x < \frac{\pi}{2}.$$

Se invece x è negativo, compreso fra 0 e $-\pi/2$, otteniamo nello stesso modo

$$\cos x < \frac{\sin x}{x} < 1 \quad \text{per} \quad 0 > x > -\frac{\pi}{2}.$$

In conclusione,

$$(2.13) \quad \cos x < \frac{\sin x}{x} < 1 \quad \text{per} \quad -\frac{\pi}{2} < x < \frac{\pi}{2}, \quad x \neq 0.$$

Dal Teorema dei Carabinieri e dalla (2.12) otteniamo subito

$$(2.14) \quad \lim_{x \rightarrow 0} \frac{\operatorname{sen} x}{x} = 1.$$

ESEMPIO 2.12. Quanto fa il limite

$$\lim_{x \rightarrow 0} \frac{\operatorname{sen}(x^2)}{x^2}?$$

Una idea naturale è “cambiare variabile”, ossia porre $y = x^2$; infatti quando $x \rightarrow 0$ si ha $y = x^2 \rightarrow 0$ e quindi si dovrebbe avere

$$\lim_{x \rightarrow 0} \frac{\operatorname{sen}(x^2)}{x^2} = \lim_{y \rightarrow 0} \frac{\operatorname{sen} y}{y} = 1.$$

Il risultato è corretto, ma come si giustifica questo procedimento? Se usiamo il Teorema 2.5 e consideriamo $x_n \rightarrow 0$ arbitraria con $x_n \neq 0$, abbiamo $y_n = x_n^2 \rightarrow 0$ con $y_n \neq 0$ e quindi $\operatorname{sen}(y_n)/y_n \rightarrow 1$, ossia $\operatorname{sen}(x_n^2)/x_n^2 \rightarrow 1$ per qualunque successione, e questo dimostra la validità del calcolo precedente.

È chiaro che questo metodo funziona anche in generale: se vogliamo calcolare un limite del tipo

$$\lim_{x \rightarrow c} g[f(x)]$$

sapendo che $f(x) \rightarrow L$ per $x \rightarrow c$, siamo autorizzati a cambiare variabile e scrivere

$$\lim_{x \rightarrow c} g[f(x)] = \lim_{y \rightarrow L} g(y)$$

purché si abbia $f(x) \neq L$ per $x \neq c$ (e anzi è sufficiente che questa condizione sia soddisfatta per i punti x vicini a c). Provate a scrivere la dimostrazione di questo risultato usando il Teorema 2.5.

ESERCIZIO 2.13. Calcolare i seguenti limiti: sviluppando,

$$\lim_{x \rightarrow 1} \frac{x^3 + 1}{x^2 + 1}, \quad \lim_{y \rightarrow 2} \frac{y^2 - 4}{y^2 - 3y + 2}, \quad \lim_{h \rightarrow 0} \frac{(x+h)^3 - x^3}{h},$$

$$\lim_{x \rightarrow 1} \frac{x^2 - 1}{1 - x}, \quad \lim_{y \rightarrow 1} \frac{\sqrt{y} - 1}{y - 1}, \quad \lim_{h \rightarrow 0} \frac{(x+h)^3 - x^3}{h},$$

razionalizzando,

$$\lim_{x \rightarrow 2} \frac{\sqrt{x} - \sqrt{2}}{x - 2}, \quad \lim_{y \rightarrow 0} \frac{\sqrt{1+y} - \sqrt{1-y}}{y}, \quad \lim_{h \rightarrow 0} \frac{\sqrt{x+h} - \sqrt{x}}{h},$$

utilizzando astutamente qualcuno dei limiti notevoli precedenti (ma non sempre!),

$$\lim_{x \rightarrow 1} \frac{\operatorname{sen} x}{x}, \quad \lim_{x \rightarrow 0} \frac{\operatorname{sen}(5x)}{x}, \quad \lim_{x \rightarrow -2} \frac{\operatorname{tg}(\pi x)}{x + 2}, \quad \lim_{x \rightarrow 0} \frac{\operatorname{sen}(5x)}{\operatorname{sen}(3x)},$$

$$\lim_{x \rightarrow 0} x \operatorname{sen} \frac{1}{x}, \quad \lim_{x \rightarrow 0} \frac{\arcsin(5x)}{x}, \quad \lim_{x \rightarrow 0} \frac{\operatorname{sen}^2 x}{x}, \quad \lim_{x \rightarrow 1} \frac{\operatorname{sen}(\pi x)}{\operatorname{sen}(3\pi x)},$$

$$\lim_{x \rightarrow 0} \frac{\operatorname{sen} x}{\sqrt{x}}, \quad \lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2}, \quad \lim_{x \rightarrow 1} (1 - x) \operatorname{tg} \frac{\pi x}{2}.$$

Concludiamo con qualche variazione sul concetto di limite. Anzitutto, possiamo immaginare che il “movimento” di x verso il punto c avvenga in una sola direzione, cioè possiamo limitarci a considerare i valori $x < c$ (movimento da sinistra) o quelli $x > c$ (movimento da destra).

DEFINIZIONE 2.14. Supponiamo che $A \subseteq \mathbb{R}$ contenga un intervallo $]a, c[$ e sia $f : A \rightarrow \mathbb{R}$ una funzione. Diciamo che f **ha limite sinistro** L in c , o che f **tende a L per x che tende a c da sinistra**, e scriviamo

$$(2.15) \quad \lim_{x \rightarrow c^-} f(x) = L$$

se: per ogni $\varepsilon > 0$ esiste un $\delta = \delta(\varepsilon)$ tale che

$$(2.16) \quad |f(x) - L| < \varepsilon \quad \text{per tutti gli } x \text{ tali che } c - \delta < x < c.$$

Si scrive anche:

$$f(x) \rightarrow L \quad \text{per } x \rightarrow c^-.$$

La definizione del limite destro è completamente analoga: supponiamo che $A \subseteq \mathbb{R}$ contenga un intervallo $]c, b[$ e sia $f : A \rightarrow \mathbb{R}$ una funzione. Diciamo che f **ha limite destro** L in c , o che f **tende a L per x che tende a c da destra**, e scriviamo

$$(2.17) \quad \lim_{x \rightarrow c^+} f(x) = L$$

se: per ogni $\varepsilon > 0$ esiste un $\delta = \delta(\varepsilon)$ tale che

$$(2.18) \quad |f(x) - L| < \varepsilon \quad \text{per tutti gli } x \text{ tali che } c < x < c + \delta.$$

Si scrive anche:

$$f(x) \rightarrow L \quad \text{per } x \rightarrow c^+.$$

Può succedere che una funzione abbia limite destro e limite sinistro in un punto c , ma che questi limiti siano diversi. Allora chiaramente la funzione non ha limite per $x \rightarrow c$.

ESERCIZIO 2.15. Dimostrare che se $f \rightarrow L$ per $x \rightarrow c$, allora in particolare si ha anche $f \rightarrow L$ per $x \rightarrow c^-$ e $x \rightarrow c^+$. Dimostrare che vale anche il viceversa: ossia se f ha limite sinistro e limite destro uguali allo stesso numero L nel punto c , allora $f \rightarrow L$ per $x \rightarrow c$.

Dimostrare l'analogo del Teorema 2.5 per limite destro e sinistro. Ad esempio: $f \rightarrow L$ per $x \rightarrow c^+$ se e solo se, per ogni successione x_n tendente a c con $x_n > c$ si ha $f(x_n) \rightarrow L$. È sufficiente considerare successioni x_n *decrecenti* a c ?

ESERCIZIO 2.16. Calcolare i seguenti limiti:

$$\begin{array}{cccc} \lim_{x \rightarrow 0^+} \frac{|\operatorname{sen} x|}{x}, & \lim_{x \rightarrow 0^-} \frac{|\operatorname{sen} x|}{x}, & \lim_{x \rightarrow 1^+} \frac{|x-1|}{x-1}, & \lim_{x \rightarrow 1^-} \frac{|x-1|}{x-1}, \\ \lim_{x \rightarrow -2^+} \frac{x}{x+2}, & \lim_{x \rightarrow -2^-} \frac{1-x}{2+x}, & \lim_{x \rightarrow 5^+} \frac{x - [x]}{x}, & \lim_{x \rightarrow \pi^-} \frac{[5-x]}{x-1}. \end{array}$$

ESERCIZIO 2.17. Esiste il limite per $x \rightarrow 0$ della funzione seguente?

$$f(x) = \begin{cases} x^2 + x & \text{se } x \geq 0 \\ 1 - \frac{\operatorname{sen} x}{x} & \text{se } x < 0. \end{cases}$$

3. Limiti senza limiti.

Come per le successioni, anche per le funzioni può succedere che al tendere di x a c i valori della funzione siano sempre più grandi. Per esempio, proviamo a calcolare il limite in 0 di $f(x) = 1/x^2$; scegliendo valori di $x > 0$ sempre più vicini a 0 otteniamo valori di f sempre più grandi, e anzi se prendiamo una successione $x_n \rightarrow 0$, $x_n > 0$, otteniamo una successione di valori $f(x_n)$ che tende a $+\infty$. In questo caso diremo che $f \rightarrow +\infty$ per $x \rightarrow 0$. Cerchiamo di dare una definizione più precisa. Notiamo che nelle definizioni precedenti esprimevamo la vicinanza di f ad un certo numero L tramite la quantità $|f(x) - L|$, richiedendo che essa fosse “piccola”; ora invece vogliamo esprimere il fatto che $f(x)$ è “grande” quando x è vicino a c . Ricorriamo ad una definizione simile a quella usata per le successioni divergenti:

DEFINIZIONE 3.1. Sia $A \subseteq \mathbb{R}$ contenente due intervalli $]a, c[$ e $]c, b[$. Diciamo che f **diverge a $+\infty$ nel punto c** , o che f **tende a $+\infty$ per x che tende a c** , e scriviamo

$$\lim_{x \rightarrow c} f(x) = +\infty$$

oppure $f \rightarrow +\infty$ per $x \rightarrow c$, se per ogni $M > 0$ esiste $\delta = \delta(M) > 0$ tale che

$$f(x) > M \quad \text{per tutti gli } x \text{ tali che } 0 < |x - c| < \delta.$$

Diciamo che f **diverge a $-\infty$ nel punto c** , o che f **tende a $-\infty$ per x che tende a c** , e scriviamo

$$\lim_{x \rightarrow c} f(x) = -\infty$$

oppure $f \rightarrow -\infty$ per $x \rightarrow c$, se per ogni $M < 0$ esiste $\delta = \delta(M) > 0$ tale che

$$f(x) < M \quad \text{per tutti gli } x \text{ tali che } 0 < |x - c| < \delta.$$

Se al posto della disuguaglianza $0 < |x - c| < \delta$ sostituiamo la disuguaglianza $c < x < c + \delta$ (con A contenente $]c, b[$), otteniamo le corrispondenti definizioni di limite destro, e se invece usiamo la disuguaglianza $c - \delta < x < c$ (con A contenente $]a, c[$) otteniamo le corrispondenti definizioni di limite sinistro.

Quando si verifica una delle condizioni della definizione precedente, si dice anche che la funzione (o meglio il suo grafico) *ha un asintoto verticale nel punto $x = c$* .

ESERCIZIO 3.2. a) Sia $f(x) = 1/x$. Mostriamo che: $f \rightarrow +\infty$ per $x \rightarrow 0^+$, $f \rightarrow -\infty$ per $x \rightarrow 0^-$, e che il limite per $x \rightarrow 0$ di f non esiste. Per dimostrare il primo fatto fissiamo $M > 0$ e cerchiamo $\delta > 0$ tale che $1/x > M$ per $0 < x < 0 + \delta$: chiaramente, basta prendere $\delta = 1/M$. Per l'altro limite prendiamo $M < 0$ e scegliendo $\delta = -1/M$ (positivo) abbiamo $1/x < M$ per $0 - \delta < x < 0$; la terza affermazione segue subito dall'esercizio 2.15 perché limite destro e sinistro sono diversi.

b) Dimostrare che $1/x^2 \rightarrow +\infty$ per $x \rightarrow 0$.

c) Dimostrare che se $f \rightarrow 0$ per $x \rightarrow c$ e f è strettamente positiva per $x \neq c$, allora

$$\frac{1}{f(x)} \rightarrow +\infty \quad \text{per } x \rightarrow c.$$

Che si può dire se f è strettamente negativa? e se è strettamente positiva solo per x vicino a c ? e se è strettamente positiva solo per $x > c$?

d) Dire se esistono, e calcolare i limiti

$$\begin{aligned} & \lim_{x \rightarrow 0^+} \frac{1}{\operatorname{sen} x}, \quad \lim_{x \rightarrow 0^-} \frac{1}{\operatorname{sen} x}, \quad \lim_{x \rightarrow 0} \frac{1}{\operatorname{sen} x}; \\ & \lim_{x \rightarrow 2^+} \frac{x^2 - 3x^2}{2 - x}, \quad \lim_{x \rightarrow 2^-} \frac{x^2 - 3x^2}{2 - x}, \quad \lim_{x \rightarrow 2} \frac{x^2 - 3x^2}{2 - x}; \\ & \lim_{x \rightarrow 2^+} \frac{x^3 - 2x^2}{2 - x}, \quad \lim_{x \rightarrow 2^-} \frac{x^3 - 2x^2}{2 - x}, \quad \lim_{x \rightarrow 2} \frac{x^3 - 2x^2}{2 - x}. \end{aligned}$$

C'è un ultimo caso interessante di limite di funzioni, il comportamento per valori di x "grandi". Prendiamo ad esempio le funzioni $f(x) = x^2$ e $g(x) = 1/x$. Calcoliamole per valori di x sempre più grandi: vediamo che f e g hanno un andamento completamente diverso: i valori di f sono sempre più grandi, quelli di g sono sempre più vicini a 0. Nel primo caso diremo che x^2 tende a $+\infty$ quando x tende a $+\infty$, nel secondo che $1/x$ tende a 0 per $x \rightarrow +\infty$. Non è difficile definire in modo preciso queste nozioni:

DEFINIZIONE 3.3. Sia $A \subseteq \mathbb{R}$ contenente un intervallo $]a, +\infty[$, e sia $f : A \rightarrow \mathbb{R}$. Diciamo che f **tende a L per x che tende a $+\infty$** , e scriviamo

$$\lim_{x \rightarrow +\infty} f(x) = L$$

o $f \rightarrow L$ per $x \rightarrow +\infty$, se per ogni $\varepsilon > 0$ esiste $K = K(\varepsilon)$ tale che

$$|f(x) - L| < \varepsilon \quad \text{per tutti gli } x \text{ tali che } x > K.$$

La definizione di $f \rightarrow L$ per $x \rightarrow -\infty$ è simile: si suppone che A contenga un intervallo $] -\infty, b[$ e si richiede che la condizione $|f(x) - L| < \varepsilon$ valga per $x < K$.

Infine diciamo che f **tende a $+\infty$ per $x \rightarrow +\infty$** , e scriviamo

$$\lim_{x \rightarrow +\infty} f(x) = +\infty$$

oppure $f \rightarrow +\infty$ per $x \rightarrow +\infty$, se per ogni $M > 0$ esiste $K = K(M)$ tale che

$$f(x) > M \quad \text{per tutti gli } x \text{ tali che } x > K.$$

Non è difficile dare le definizioni negli altri casi possibili $f \rightarrow \pm\infty$ per $x \rightarrow \pm\infty$ (quattro casi in tutto).

Quando $f \rightarrow L$ per $x \rightarrow +\infty$ si dice anche che f e il suo grafico *hanno un asintoto orizzontale per $x \rightarrow +\infty$* , e analogamente nel caso $x \rightarrow -\infty$.

ESERCIZIO 3.4. Anzitutto scrivete per esteso le definizioni precedenti in tutti i casi possibili. Poi, provate a modificare l'enunciato del Teorema Ponte 2.5 in modo da coprire ciascuno dei casi. Le dimostrazioni sono praticamente identiche (farle in dettaglio sarebbe inutile crudeltà).

Ad esempio: $f \rightarrow +\infty$ per $x \rightarrow +\infty$ se e solo se, per ogni successione $a_n \rightarrow +\infty$ si ha $f(a_n) \rightarrow +\infty$. Eccetera (l'esercizio è facile!).

ESEMPIO 3.5. Come al solito da risultati già dimostrati per le successioni otteniamo subito i risultati corrispondenti per le funzioni. Ad esempio,

$$\lim_{x \rightarrow \pm\infty} \left(1 + \frac{1}{x}\right)^x = e;$$

infatti sappiamo che $(1 + 1/a_n)^{a_n} \rightarrow e$ per qualunque successione $a_n \rightarrow \pm\infty$. Più in generale,

$$\lim_{x \rightarrow \pm\infty} \left(1 + \frac{L}{x}\right)^x = e^L$$

(ragionamento identico).

Un caso particolarmente interessante è quello delle funzioni monotone:

TEOREMA 3.6. Sia $f :]a, c[\rightarrow \mathbb{R}$ una funzione crescente. Se f è limitata superiormente, allora esiste finito il limite

$$\lim_{x \rightarrow c^-} f(x) = L$$

e L coincide con l'estremo superiore di f su $]a, c[$. Se invece f non è limitata superiormente, allora

$$\lim_{x \rightarrow c^-} f(x) = +\infty$$

(e anche in questo caso il limite coincide con l'estremo superiore).

Un risultato analogo vale con l'estremo inferiore se f è decrescente.

Dimostrazione. Nel primo caso, sia L l'estremo superiore di f ; sappiamo che $f(x) \leq L$ per ogni $x \in]a, c[$, e che comunque si prenda $\varepsilon > 0$ si può trovare un punto y_ε dell'immagine di f tale che $L - \varepsilon < y_\varepsilon \leq L$. Quindi si deve avere $y_\varepsilon = f(x_\varepsilon)$ per qualche punto x_ε del dominio $]a, c[$, e dato che f è crescente si ha

$$L - \varepsilon < f(x_\varepsilon) \leq f(x) \leq L \quad \text{per tutti gli } x \text{ tali che } x_\varepsilon < x < c.$$

Se poniamo $\delta = \delta(\varepsilon) = c - x_\varepsilon$ possiamo riscrivere la disuguaglianza ottenuta come

$$L - \varepsilon < f(x) \leq L \quad \text{per tutti gli } x \text{ tali che } c - \delta < x < c,$$

e pertanto (dato che $\varepsilon > 0$ è arbitrario, e che $L < L + \varepsilon$) abbiamo dimostrato la tesi.

Se f non è limitata superiormente, procediamo in modo analogo: sappiamo che per ogni $M > 0$ c'è un punto $y_M = f(x_M)$ nell'immagine tale che $f(x) > M$ per tutti gli x tali che $x_M < x < c$, e ponendo $\delta = \delta(M) = c - x_M$ otteniamo anche in questo caso la tesi.

OSSERVAZIONE 3.7. Naturalmente risultati analoghi valgono per il limite destro: se $f :]c, b[$ è crescente e limitata inferiormente, allora esiste il limite per $x \rightarrow c^+$ di $f(x)$ ed esso coincide con l'estremo inferiore della funzione su $]c, b[$ (attenzione, stiamo "andando verso sinistra!"); se $f(x)$ non è limitata inferiormente allora $f \rightarrow -\infty$ per $x \rightarrow c^+$. Risultato analogo con l'estremo superiore se f è decrescente.

Infine, analoghi risultati si hanno per $x \rightarrow +\infty$ o per $x \rightarrow -\infty$ (provare a scriverne qualcuno!).

ESERCIZIO 3.8. Se $f : [a, b] \rightarrow \mathbb{R}$ è una funzione monotona definita su un intervallo chiuso, allora esistono finiti i limiti destro e sinistro in ogni punto interno dell'intervallo, e inoltre esistono finiti il limite sinistro in b e il limite destro in a .

Se $f :]a, b[\rightarrow \mathbb{R}$ è una funzione monotona definita su un intervallo aperto, allora esistono finiti i limiti destro e sinistro in ogni punto dell'intervallo; inoltre esistono il limite sinistro in b e il limite destro in a , ma possono essere rispettivamente $+\infty$ e $-\infty$. Sapreste costruire degli esempi di ciascuna di queste situazioni?

ESEMPIO 3.9. Siano $A > 1$, $\alpha > 0$, $\beta > 0$. Le funzioni

$$(\ln x)^\alpha, \quad x^\beta, \quad A^x, \quad x^x$$

tendono a $+\infty$ per $x \rightarrow +\infty$. Infatti sono crescenti (perché è crescente x^x ?) e non sono limitate superiormente (basta ad esempio calcolarle per $x = n$ e otteniamo delle successioni che divergono a $+\infty$).

Se invece $0 < A < 1$, si ha

$$A^x \rightarrow 0 \quad \text{per } x \rightarrow +\infty.$$

Perché?

Anche per le funzioni, come per le successioni, si possono estendere le operazioni fra i limiti in modo da coprire i casi di limiti $\pm\infty$, ma non tutti. Il Teorema seguente elenca i vari casi possibili:

TEOREMA 3.10. Siano f e g due funzioni aventi limite per $x \rightarrow c$ (o per $x \rightarrow c^\pm$). Abbiamo allora che, per $x \rightarrow c$, (o per $x \rightarrow c^\pm$)

- se f e g divergono entrambe a $\pm\infty$, allora $f + g$ diverge a $\pm\infty$;
- se f diverge a $\pm\infty$ e g diverge a $\mp\infty$, allora $f - g$ diverge a $\pm\infty$;
- se f diverge a $\pm\infty$ e g tende a M , allora $f \pm g$ diverge a $\pm\infty$;
- se f tende a L e g diverge a $\pm\infty$, allora $f + g$ diverge a $\pm\infty$;
- se f tende a L e g diverge a $\pm\infty$, allora $f - g$ diverge a $\mp\infty$;
- se f e g divergono entrambe a $\pm\infty$, $f \cdot g$ diverge a $+\infty$;
- se f diverge a $\pm\infty$ e g diverge a $\mp\infty$, allora $f \cdot g$ diverge a $-\infty$;
- se f diverge a $\pm\infty$ e g tende a $M > 0$, allora $f \cdot g$ diverge a $\pm\infty$;

- se f diverge a $\pm\infty$ e g tende a $M < 0$, allora $f \cdot g$ diverge a $\mp\infty$;
- se f tende a L e g diverge a $\pm\infty$, allora $\frac{f}{g}$ tende a 0;
- se f diverge a $\pm\infty$ e g tende a zero da valori positivi, allora $\frac{f}{g}$ diverge a $\pm\infty$;
- se f diverge a $\pm\infty$ e g tende a zero da valori negativi, allora $\frac{f}{g}$ diverge a $\mp\infty$;
- se f tende ad $L > 0$ e g tende a zero da valori positivi, allora $\frac{f}{g}$ diverge a $+\infty$;
- se f tende ad $L < 0$ e g tende a zero da valori positivi, allora $\frac{f}{g}$ diverge a $-\infty$;
- se f tende ad $L > 0$ e g tende a zero da valori negativi, allora $\frac{f}{g}$ diverge a $-\infty$;
- se f tende ad $L < 0$ e g tende a zero da valori negativi, allora $\frac{f}{g}$ diverge a $+\infty$;
- se f tende a L da valori positivi con $0 \leq L < 1$ e $g \rightarrow +\infty$ allora $f^g \rightarrow 0$;
- se f tende a L da valori positivi con $0 \leq L < 1$ e $g \rightarrow -\infty$ allora $f^g \rightarrow +\infty$;
- se f tende a L da valori positivi con $L > 1$ e $g \rightarrow +\infty$ allora $f^g \rightarrow +\infty$;
- se f tende a L da valori positivi con $L > 1$ e $g \rightarrow -\infty$ allora $f^g \rightarrow 0$.

Proprio come per le successioni, vi sono alcuni casi di “forme indeterminate”:

- il limite di $f + g$ quando f diverge a $\pm\infty$ e g diverge a $\mp\infty$;
- il limite di $f - g$ quando f e g divergono entrambe a $\pm\infty$;
- il limite di $f \cdot g$ quando f diverge a $\pm\infty$ e g converge a zero;
- il limite di $\frac{f}{g}$ quando f e g divergono entrambe (a $+\infty$ o a $-\infty$);
- il limite di $\frac{f}{g}$ quando f e g convergono entrambe a zero.

Per calcolare limiti di questo tipo (“ $\infty - \infty$, $0 \cdot \infty$, ∞/∞ , $0/0$ ”) non vale una regola generale ma bisogna procedere caso per caso, e talvolta, come già per le successioni, il limite non esiste. Stesso discorso per le forme indeterminate del tipo “ $(+\infty)^0$, 1^∞ ”.

ESEMPIO 3.11. Sappiamo che

$$\lim_{x \rightarrow +\infty} e^x = +\infty$$

(perché crescente e non limitata) e quindi

$$\lim_{x \rightarrow +\infty} e^{-x} = 0$$

perché $e^{-x} = 1/e^x$ (stiamo applicando il Teorema precedente, caso L/∞). Da questo segue anche che

$$\lim_{x \rightarrow -\infty} e^x = 0$$

perché ponendo $z = -x$ otteniamo esattamente il limite precedente.

Se invece proviamo a calcolare il limite di e^x/x per $x \rightarrow \infty$ abbiamo una forma indeterminata ∞/∞ . Fortunatamente abbiamo già dimostrato che per ogni successione a_n divergente a $+\infty$ si ha

$$\forall A > 1, \quad \forall \alpha > 0 \quad \lim_{n \rightarrow +\infty} \frac{A^n}{a_n^\alpha} = +\infty$$

e quindi otteniamo subito

$$(3.1) \quad \forall A > 1, \quad \forall \alpha > 0 \quad \lim_{x \rightarrow +\infty} \frac{A^x}{x^\alpha} = +\infty.$$

In modo simile si dimostra che

$$(3.2) \quad \forall \alpha > 0, \quad \forall \beta > 0, \quad \lim_{x \rightarrow \infty} \frac{x^\alpha}{(\ln(x))^\beta} = +\infty.$$

ESEMPIO 3.12. Anche per calcolare limiti infiniti o per $x \rightarrow \pm\infty$ si può usare il metodo del cambiamento di variabile, con le stesse cautele. Vediamo qualche esempio:

$$\lim_{x \rightarrow 0^+} e^{1/x} = \lim_{y \rightarrow +\infty} e^y = +\infty$$

perché $y = 1/x \rightarrow +\infty$ per $x \rightarrow 0^+$;

$$\lim_{x \rightarrow 0^+} e^{1/x} = \lim_{y \rightarrow +\infty} e^y = +\infty$$

perché $y = 1/x \rightarrow -\infty$ per $x \rightarrow 0^-$; notare che il limite per $x \rightarrow 0$ non esiste. Altro esempio: dal limite

$$\lim_{x \rightarrow +\infty} \frac{e^x}{x} = +\infty$$

con il cambiamento di variabile $y = e^x$ (e quindi $x = \ln y$) segue subito che

$$\lim_{y \rightarrow +\infty} \frac{y}{\ln y} = +\infty$$

(e più in generale, il limite (3.2) segue subito dal limite (3.1)). Ultimo esempio:

$$\lim_{x \rightarrow 0^+} x e^{1/x} = \lim_{y \rightarrow +\infty} \frac{e^y}{y} = +\infty$$

ponendo $y = 1/x$.

ESERCIZIO 3.13. Calcolare i seguenti limiti:

$$\begin{aligned} & \lim_{x \rightarrow +\infty} (\sqrt{x+a} - \sqrt{x}), & \lim_{x \rightarrow +\infty} (\sqrt{x(x+a)} - x), & \lim_{x \rightarrow +\infty} \frac{x^2 - 5x}{3 - x}, \\ & \lim_{x \rightarrow -\infty} \frac{3 - x^2}{x^3 + x^2}, & \lim_{x \rightarrow +\infty} \frac{\sqrt{x^3 - x}}{3 - x^2}, & \lim_{x \rightarrow +\infty} \frac{\sqrt{x}}{\sqrt{x + \sqrt{x}}}, \\ & \lim_{x \rightarrow +\infty} x \operatorname{sen} \frac{1}{x}, & \lim_{x \rightarrow +\infty} \left(1 + \frac{3}{x}\right)^x, & \lim_{x \rightarrow 0^+} (1+x)^{1/x}, \\ & \lim_{x \rightarrow 0^+} x^x, & \lim_{x \rightarrow 0^+} (\operatorname{sen} x)^x, & \lim_{x \rightarrow +\infty} (2 + \operatorname{sen} x)^x, & \lim_{x \rightarrow 0^+} x e^{-1/x^2}, \\ & \lim_{x \rightarrow 0^-} x e^{-1/x^2}, & \lim_{x \rightarrow 0} \frac{1 - e^{-2x}}{x}, & \lim_{x \rightarrow 0^+} \frac{1 - e^x}{\sqrt{x}}, & \lim_{x \rightarrow 0} \frac{e^{2x} - e^{3x}}{x}, \\ & \lim_{x \rightarrow 0} \left(\frac{2+x}{3-x}\right)^x, & \lim_{x \rightarrow +\infty} \left(\frac{x+1}{x-1}\right)^x, & \lim_{x \rightarrow 0^+} \frac{1}{x} e^{-1/x}, & \lim_{x \rightarrow 0^-} \frac{1}{x} e^{-1/x}. \end{aligned}$$

4. Funzioni continue.

Siamo arrivati a una delle definizioni centrali dell'Analisi Matematica:

DEFINIZIONE 4.1. Sia $A \subseteq \mathbb{R}$ un insieme contenente l'intervallo $]a, b[$, e sia $c \in]a, b[$. Una funzione $f :]a, b[\rightarrow \mathbb{R}$ si dice **continua nel punto** c se

$$\lim_{x \rightarrow c} f(x) = f(c).$$

Se $f : [a, b] \rightarrow \mathbb{R}$, estendiamo la definizione precedente dicendo che f è continua in a se il limite sinistro di f in a è uguale ad $f(a)$, e analogamente nel punto b usando il limite destro. Se questa condizione non è soddisfatta (ossia il limite non esiste, oppure esiste ma è diverso dal valore della funzione nel punto), allora f si dice **discontinua** in c .

La funzione si dice **continua** se è continua in ogni punto del suo dominio.

Notare che adesso il punto c *deve* appartenere al dominio della funzione: infatti qui richiediamo che il limite in c esista e sia proprio uguale al valore $f(c)$ della funzione in quel punto.

Il seguente risultato è una conseguenza immediata del Teorema sulle operazioni fra limiti:

TEOREMA 4.2. *Se due funzioni f, g sono continue nello stesso punto c , allora $f + g, f - g, f \cdot g, \alpha \cdot f$ ($\alpha \in \mathbb{R}$) sono continue in c , e il rapporto f/g è continuo in c se $g(c) \neq 0$. Inoltre g^f è continua in c se $g(c) > 0$.*

ESEMPIO 4.3. 1) Le potenze x^n con $n \in \mathbb{N}$ e più in generale tutti i polinomi sono funzioni continue. Infatti abbiamo visto che per qualunque c si ha $\lim_{x \rightarrow c} P(x) = P(c)$.

La funzione e^x è continua perché, come abbiamo già dimostrato, $e^x \rightarrow e^c$ per $x \rightarrow c$. Analogamente, A^x è continua per ogni $A > 0$, e $\ln x$ è continua (su $]0, +\infty[$).

Anche $\sin x$ e $\cos x$ sono continue: sappiamo già che $\sin x$ è continua in zero, ossia che

$$\lim_{x \rightarrow 0} \sin x = 0 = \sin 0.$$

Quindi anche $\cos x$ è continua in 0, infatti

$$\lim_{x \rightarrow 0} \cos x = \lim_{x \rightarrow 0} \sqrt{1 - \sin^2 x} = 1$$

applicando le regole di calcolo che conosciamo. Inoltre scrivendo

$$\sin x = \sin(x - c + c)$$

e applicando la formula di addizione otteniamo subito la continuità in ogni punto:

$$\lim_{x \rightarrow c} \sin x = \sin c$$

(scrivere i dettagli!). Discorso simile per $\cos x$. Ne segue che la funzione $\operatorname{tg} x = \sin x / \cos x$ è continua in tutti i punti $x \neq \frac{\pi}{2} + k\pi$.

2) La funzione $|x|$ è continua. Basta osservare che

$$||x| - |c|| \leq |x - c|$$

e da questa disuguaglianza segue subito che $|x| \rightarrow |c|$ per $x \rightarrow c$.

3) Supponiamo che f sia una funzione continua, e sia g un'altra funzione con la proprietà $f(x) = g(x)$ per $x \in]a, b[$ mentre potrebbe essere $f \neq g$ negli altri punti. Allora g è continua in tutti i punti $c \in]a, b[$. Infatti quando si calcola il limite di g per $x \rightarrow c$ ci si può limitare a considerare i valori di x vicini a c (vedi la definizione di limite).

Non tutte le funzioni sono continue, ovviamente; studiamo qualche esempio tipico di discontinuità.

ESEMPIO 4.4. 1) La funzione parte intera di x , $f(x) = [x]$, è discontinua in tutti i punti $x \in \mathbb{Z}$, e continua negli altri punti. Che sia continua in ogni $c \notin \mathbb{Z}$ è evidente perché vicino a tali punti f è una funzione costante (detto meglio: la funzione ristretta a $(c - \epsilon, c + \epsilon)$ è costante se prendiamo ϵ abbastanza piccolo). D'altra parte, fissato $k \in \mathbb{Z}$, si ha

$$\lim_{x \rightarrow k^+} = k, \quad \lim_{x \rightarrow k^-} = k - 1,$$

e quindi il limite in k non esiste.

2) La funzione segno di x , definita come

$$f(x) = \begin{cases} +1 & \text{se } x > 0 \\ 0 & \text{se } x = 0 \\ -1 & \text{se } x < 0 \end{cases}$$

(o in modo più rapido ma meno preciso, $f(x) = x/|x|$) è continua in ogni punto tranne 0.

DEFINIZIONE 4.5. Qualche volta si usa la seguente terminologia: se in un punto c esistono sia il limite destro che il limite sinistro di f , ma essi non coincidono, oppure coincidono ma sono diversi dal valore $f(c)$, si dice che f ha una **discontinuità di prima specie** in c . In particolare quando i limiti destro e sinistro coincidono fra loro ma sono diversi da $f(c)$, è ovvio che basta cambiare il valore di f

in c per ottenere una funzione continua; per questo si parla anche di **discontinuità eliminabile**. (E talvolta si usa questo termine anche per funzioni che non sono definite in c , ma hanno limite destro e sinistro in c uguali).

In tutti gli altri casi si parla di **discontinuità di seconda specie**.

ESEMPIO 4.6. Un esempio più sofisticato. La *funzione di Dirichlet* si definisce come

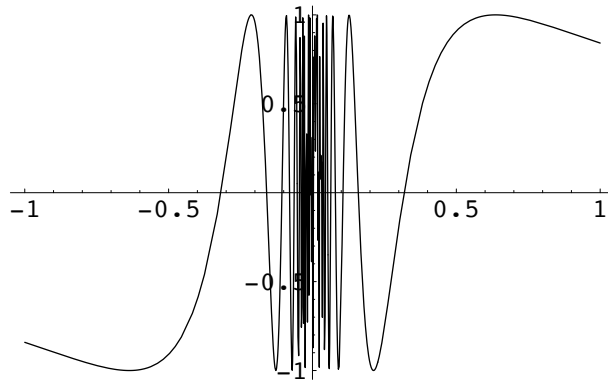
$$f(x) = \begin{cases} 1 & \text{se } x \in \mathbb{Q} \\ 0 & \text{se } x \notin \mathbb{Q}. \end{cases}$$

Questa funzione è discontinua in tutti i punti, e anzi ha solo discontinuità di seconda specie. Sapreste dimostrarlo?

ESEMPIO 4.7. Un altro esempio interessante è dato dalla funzione

$$f(x) = \operatorname{sen} \frac{1}{x}.$$

Il suo grafico è il seguente:



Questa funzione è continua in tutti i punti tranne $x = 0$. Sapete dimostrarlo?

Provate anche a disegnare il grafico della funzione

$$f(x) = x \cdot \operatorname{sen} \frac{1}{x}.$$

ESEMPIO 4.8. Supponiamo che una funzione $f(x)$ abbia la proprietà seguente: esiste una costante $K \geq 0$ tale che

$$|f(x) - f(y)| \leq K|x - y|$$

per tutti i punti x, y nel dominio di A . Una tale funzione si dice **lipschitziana**. È evidente che una funzione lipschitziana è anche continua: infatti se $x \rightarrow c$ dalla disuguaglianza

$$|f(x) - f(c)| \leq K|x - c|$$

segue subito che $f(x) \rightarrow f(c)$.

ESERCIZIO 4.9 (L.Orsina). Pronunciare “lipschitziana”.

Anche la composizione di funzioni continue è continua:

TEOREMA 4.10. Siano $f :]a, b[\rightarrow \mathbb{R}$ e $g :]\alpha, \beta[\rightarrow \mathbb{R}$ due funzioni con la proprietà $f(]a, b[) \subseteq]\alpha, \beta[$; si può quindi considerare la funzione composta $h = g \circ f$. Se f è continua in un punto c e g è continua nel punto $f(c)$ allora h è continua in c . Quindi, la composizione di funzioni continue è una funzione continua.

Dimostrazione. Presa una successione $a_n \rightarrow c$ qualunque, la successione $\alpha_n = f(a_n)$ converge a $f(c)$ dato che f è continua in c , e quindi $g(\alpha_n) = h(a_n)$ converge a $g(f(c))$ dato che g è continua in $f(c)$. Applicando il Teorema Ponte, ne segue la continuità di h .

ESERCIZIO 4.11. a) Determinare l'insieme dei punti di continuità per le funzioni seguenti. (Esempio: la funzione $f(x) = 1/x$ è discontinua in 0 ma continua in tutti gli altri punti in quanto rapporto di due funzioni continue, e pertanto il suo insieme di continuità è $\mathbb{R} \setminus \{0\}$. Naturalmente l'insieme di continuità è contenuto nel dominio).

$$\frac{1}{1-3x}, \quad xe^{1/x}, \quad xe^{-1/x^2}, \quad \operatorname{sen} \frac{1}{x}, \quad \frac{1}{\operatorname{sen} \frac{1}{x}}.$$

b) Per quali valori dei parametri α, β la funzione seguente è continua?

$$f(x) = \begin{cases} \frac{2x}{\operatorname{sen} x} & \text{per } x > 0, \\ \alpha & \text{per } x = 0, \\ \frac{e^{-\beta x} - 1}{x} & \text{per } x < 0. \end{cases}$$

Stesso problema per la funzione

$$f(x) = \begin{cases} \frac{e^{2x} - e^{3x}}{\sqrt{x}} & \text{per } x > 0, \\ \alpha & \text{per } x = 0, \\ \frac{e^{-\beta x} - 1}{\operatorname{sen} x} & \text{per } x < 0. \end{cases}$$

c) Sia f continua. Dimostrare che le funzioni $f^+(x) = \max\{f(x), 0\}$ e $f^-(x) = \max\{-f(x), 0\}$ sono continue. Inoltre data una funzione $f(x)$ qualunque, sapreste disegnare i grafici di f^+ e f^- ? Infine, delle seguenti quattro identità quali sono vere e quali no?

$$f = f^+ + f^-, \quad f = f^+ - f^-, \quad |f| = f^+ + f^-, \quad |f| = f^+ f^-.$$

5. Alcune proprietà delle funzioni continue.

TEOREMA 5.1. (*Permanenza del segno per funzioni continue*). Sia $f :]a, b[\rightarrow \mathbb{R}$ continua in c e supponiamo che $f(c) > 0$. Allora esiste $\delta > 0$ tale che $f(x) > 0$ per tutti gli x tali che $|x - c| < \delta$. Proprietà analoga se $f(c) < 0$.

Dimostrazione. Basta applicare il Teorema 2.3, caso (i), ricordando che $L = f(c)$.

Dimostriamo ora una proprietà fondamentale delle funzioni continue:

TEOREMA 5.2. (*Teorema degli zeri*). Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua e supponiamo che $f(a) < 0$ e $f(b) > 0$ (o viceversa). Allora esiste un punto $z \in]a, b[$ in cui f si annulla: $f(z) = 0$.

Dimostrazione. Supponiamo ad esempio di essere nella situazione $f(a) < 0 < f(b)$. Consideriamo il punto $c = (a+b)/2$, ossia il punto di mezzo dell'intervallo. Se $f(c) = 0$ possiamo prendere $z = c$ e il teorema è dimostrato. Altrimenti definiamo un nuovo intervallo più piccolo $[a_1, b_1]$ distinguendo due casi:

- $f(c) > 0$, e allora poniamo $a_1 = a$, $b_1 = c$;
- $f(c) < 0$, e allora poniamo $a_1 = c$, $b_1 = b$.

In altri termini, abbiamo costruito un intervallo largo la metà di quello di partenza in cui si verifica ancora la situazione originaria: $f(a_1) < 0 < f(b_1)$. Notare anche che il nuovo intervallo è contenuto nel precedente: detti $a_0 = a$ e $b_0 = b$ si ha $a_0 \leq a_1 < b_1 \leq b_0$. Chiaramente possiamo ripetere il procedimento, prendendo il punto medio dell'intervallo $[a_1, b_1]$ e distinguendo i vari casi $f(c_1) = 0, < 0, > 0$.

Andiamo avanti cosí: se abbiamo ripetuto l'operazione n volte siamo arrivati a costruire i punti $a_n < b_n$ con le proprietà

$$a_0 \leq a_1 \leq \dots \leq a_n \quad \text{e} \quad b_n \leq \dots \leq b_1 \leq b_0$$

e inoltre la larghezza dell'intervallo $[a_n, b_n]$ è quella dell'intervallo di partenza divisa per 2^n volte, ossia

$$b_n - a_n = \frac{b_{n-1} - a_{n-1}}{2} = \dots = \frac{b_0 - a_0}{2^n}.$$

Per fare il passo successivo consideriamo il punto medio dell'intervallo $[a_n, b_n]$ ossia $c_n = (a_n + b_n)/2$; come al solito, se $f(c_n) = 0$ abbiamo terminato, altrimenti abbiamo i due casi

- $f(c_n) > 0$, e allora poniamo $a_{n+1} = a_n$, $b_{n+1} = c_n$;
- $f(c_n) < 0$, e allora poniamo $a_{n+1} = c_n$, $b_{n+1} = b_n$.

Se ad un certo passo il punto medio dell'intervallo è uno zero di f , il procedimento termina; altrimenti esso produce due successioni a_n , b_n con le seguenti proprietà: a_n è crescente, b_n è decrescente, inoltre $a_n < b_n$ e anzi

$$0 < b_n - a_n = \frac{b - a}{2^n}.$$

Tutte e due queste successioni convergono ad un limite finito, infatti sono monotone e limitate (in quanto appartengono all'intervallo $[a, b]$). Sia L il limite della prima e M il limite della seconda: chiaramente $L = M$ perché passando al limite nella disuguaglianza precedente per il Teorema dei Carabinieri otteniamo

$$0 \leq L - M \leq 0.$$

Poniamo allora $z = L = M$: è facile vedere che $f(z) = 0$. Infatti da $f(a_n) < 0$ e $a_n \rightarrow L$ segue subito $f(L) \leq 0$, e d'altra parte da $f(b_n) > 0$ e $b_n \rightarrow M$ segue subito $f(M) = f(L) \geq 0$.

Dal Teorema degli zeri deduciamo con facilità un risultato molto utile:

TEOREMA 5.3. (*Teorema dei valori intermedi*). *Sia f una funzione continua su un intervallo $[a, b]$. Allora essa assume tutti i valori intermedi fra $f(a)$ e $f(b)$.*

Dimostrazione. Consideriamo il caso $f(a) < f(b)$, il caso $f(a) > f(b)$ è identico (e se $f(a) = f(b)$ non c'è nulla da dimostrare). Preso un qualunque valore intermedio $f(a) < \gamma < f(b)$, vogliamo dimostrare che esiste un punto z in cui $f(z) = \gamma$.

Basta considerare la funzione $g : [a, b] \rightarrow \mathbb{R}$ definita da

$$g(x) = f(x) - \gamma.$$

Si ha infatti $g(a) = f(a) - \gamma < 0$ e $g(b) = f(b) - \gamma > 0$, e inoltre g è una funzione continua. Per il Teorema degli zeri allora deve esistere un punto z in cui $0 = g(z) = f(z) - \gamma$ da cui la tesi.

OSSERVAZIONE 5.4. Il Teorema dei valori intermedi stabilisce quindi che l'immagine di una funzione continua $f : [a, b] \rightarrow \mathbb{R}$ contiene tutto l'intervallo di estremi $f(a)$ e $f(b)$. Con un piccolo sforzo in più possiamo verificare che *l'immagine di f è un intervallo*. Infatti se l'immagine di f contiene due valori $c < d$, ne segue che esistono $x_1, x_2 \in [a, b]$ tali che $f(x_1) = c$ e $f(x_2) = d$; applicando il Teorema a f definita sull'intervallo tra x_1 e x_2 , otteniamo che l'immagine di f deve contenere tutto l'intervallo $[c, d]$, e questo vuol dire che $\text{Im}(f)$ è un intervallo.

E se la funzione è definita su un intervallo aperto, o illimitato? Il risultato vale ancora (provate a dimostrarlo per esercizio). Possiamo quindi enunciare il fatto generale che *una funzione continua definita su un intervallo di qualunque tipo (aperto, chiuso, semiaperto, finito o infinito) ha per immagine un intervallo*.

ESEMPIO 5.5. Determiniamo l'immagine $E = \text{Im}(f)$ della funzione $f(x) = e^x$. In effetti abbiamo già osservato che $E =]0, +\infty[$ ma senza dimostrarlo. Ora abbiamo gli strumenti per farlo. Anzitutto per il teorema dei valori intermedi sappiamo che E deve essere un intervallo, dato che f è definita su \mathbb{R} ed è continua; inoltre $E \subseteq]0, +\infty[$ perché $e^x > 0$ per qualunque x . Infine sappiamo che $\lim_{x \rightarrow -\infty} e^x = 0$, quindi l'estremo inferiore di E è 0, e e^x non è limitata superiormente, quindi la tesi.

Altra proprietà fondamentale delle funzioni continue è l'esistenza di massimi e minimi.

DEFINIZIONE 5.6. Sia $A \subseteq \mathbb{R}$. Diciamo che la funzione $f : A \rightarrow \mathbb{R}$ **ammette massimo** su A se esiste un punto $c \in A$ tale che $f(c) \geq f(x)$ per tutti gli $x \in A$, e tale punto è detto **punto di massimo** per f su A . Analoga la definizione per il minimo.

TEOREMA 5.7. (Teorema di Weierstrass). Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua sull'intervallo chiuso e limitato $[a, b]$. Allora f ammette massimo e minimo su $[a, b]$.

Dimostrazione. Dimostriamo l'esistenza del massimo, quella del minimo è analoga. Distinguiamo anzitutto due casi.

Caso A: se f non è limitata superiormente, possiamo costruire una successione di punti a_n tali che $f(a_n) \rightarrow +\infty$. Basta osservare che per ogni intero n devono esistere dei punti in cui f vale più di n , scegliere uno di questi punti e chiamarlo a_n .

Caso B: se invece f è limitata superiormente, allora il suo estremo superiore $M = \sup f([a, b])$ è un numero reale. Dalla definizione di estremo superiore abbiamo che:

- (1) $M \geq f(x)$ per tutti gli $x \in [a, b]$;
- (2) dato qualunque $\varepsilon > 0$ esiste un punto $a_\varepsilon \in [a, b]$ tale che $f(a_\varepsilon) > M - \varepsilon$.

Allora per ogni $n \in \mathbb{N}$ possiamo prendere $\varepsilon = 1/n$ e definire $a_n = a_\varepsilon$; abbiamo costruito una successione $a_n \in [a, b]$ tale che

$$M - \frac{1}{n} < f(a_n) \leq M.$$

Da questa disuguaglianza, passando al limite segue subito che $f(a_n) \rightarrow M$.

In conclusione, sia nel Caso A che nel Caso B siamo riusciti a costruire una successione $a_n \in [a, b]$ con la proprietà $f(a_n) \rightarrow \sup f$. Per il Teorema di Bolzano–Weierstrass, da a_n si può estrarre una sottosuccessione a_{n_k} convergente ad un punto $c \in [a, b]$. Essendo f continua si ha $f(a_{n_k}) \rightarrow f(c)$.

Ma allora il Caso A è impossibile: infatti se fosse $f(a_n) \rightarrow +\infty$ si avrebbe anche $f(a_{n_k}) \rightarrow +\infty$ ma sappiamo che $f(a_{n_k}) \rightarrow f(c) \neq +\infty$. Resta solo il caso B, e otteniamo allora che

$$f(a_{n_k}) \rightarrow M = f(c)$$

ossia il valore di f nel punto c è uguale all'estremo superiore di f . Ne segue che c è il punto di massimo cercato.

OSSERVAZIONE 5.8. Le ipotesi del teorema precedente sono: la continuità di f ; il fatto che il dominio è un intervallo chiuso; e il fatto che il dominio è un intervallo limitato. Si tratta veramente di ipotesi essenziali?

1) Per esempio, proviamo ad abolire l'ipotesi di continuità. Allora vediamo che è molto facile trovare dei casi in cui la funzione non ha massimo, anche se la funzione è discontinua solo in un punto. Ad esempio si può prendere la funzione $f : [0, 1] \rightarrow \mathbb{R}$ definita come

$$f(x) = \begin{cases} x & \text{se } 0 \leq x < 1, \\ 0 & \text{se } x = 1. \end{cases}$$

Allora l'estremo superiore della funzione è 1 (perché?), ma la funzione non assume mai il valore 1 e quindi non ha massimo su $[0, 1]$. È chiaro che la continuità è un'ipotesi essenziale.

2) Proviamo allora a cambiare le proprietà del dominio. Consideriamo la funzione $f : [0, 1[\rightarrow \mathbb{R}$ data semplicemente da $f(x) = x$. Anche in questo caso si ha subito che l'estremo superiore di f è 1; e anche in questo caso la funzione non assume mai il valore 1, questa volta perché il dominio di f non è chiuso.

3) E se si abolisce l'altra proprietà del dominio, la limitatezza? In questo caso è chiaro che si possono considerare funzioni il cui estremo superiore è $+\infty$. Ad esempio $f : \mathbb{R} \rightarrow \mathbb{R}$ data da $f(x) = x$ non ha massimo su \mathbb{R} perché non è limitata superiormente (e non ha neanche minimo).

Anche se si considerano solo funzioni limitate la situazione non migliora: per esempio $f(x) = \arctg x$, $f : \mathbb{R} \rightarrow \mathbb{R}$, è sempre compresa fra $-\pi/2$ e $+\pi/2$, il suo estremo superiore è proprio $\pi/2$, ma non assume mai tale valore e quindi non ha massimo (né minimo).

OSSERVAZIONE 5.9. Però va anche detto che in certe situazioni si può dimostrare che esiste il massimo (o il minimo) anche se le ipotesi del Teorema di Weierstrass non sono soddisfatte.

Ad esempio, supponiamo che il dominio sia l'unione di due intervalli chiusi disgiunti come

$$A = [a, b] \cup [c, d]$$

con $a < b < c < d$. È chiaro che una funzione continua $f : A \rightarrow \mathbb{R}$ deve avere massimo; infatti ha sicuramente un massimo sul primo intervallo, ha un massimo sul secondo, e basta confrontare i due valori e scegliere quello più grande per trovare un massimo su tutto A . È abbastanza ovvio quindi che non è essenziale prendere un intervallo nel Teorema di Weierstrass; ciò che conta è una proprietà più generale detta *compattezza*: $K \subseteq \mathbb{R}$ si dice **compatto** se da ogni successione in K si può estrarre una sottosuccessione convergente ad un punto di K . In altri termini, K è compatto se soddisfa il Teorema di Bolzano–Weierstrass. È molto facile vedere che la dimostrazione del Teorema di Weierstrass funziona perfettamente se invece di $[a, b]$ si considera una funzione definita su un dominio compatto.

Ci sono casi in cui anche se il dominio della funzione continua non è compatto, è possibile ricondursi ad una situazione compatta, e quindi si riesce ancora a trovare il massimo, o minimo, come nei seguenti esercizi.

ESERCIZIO 5.10. 1) Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ continua con la proprietà

$$\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow -\infty} f(x) = 0.$$

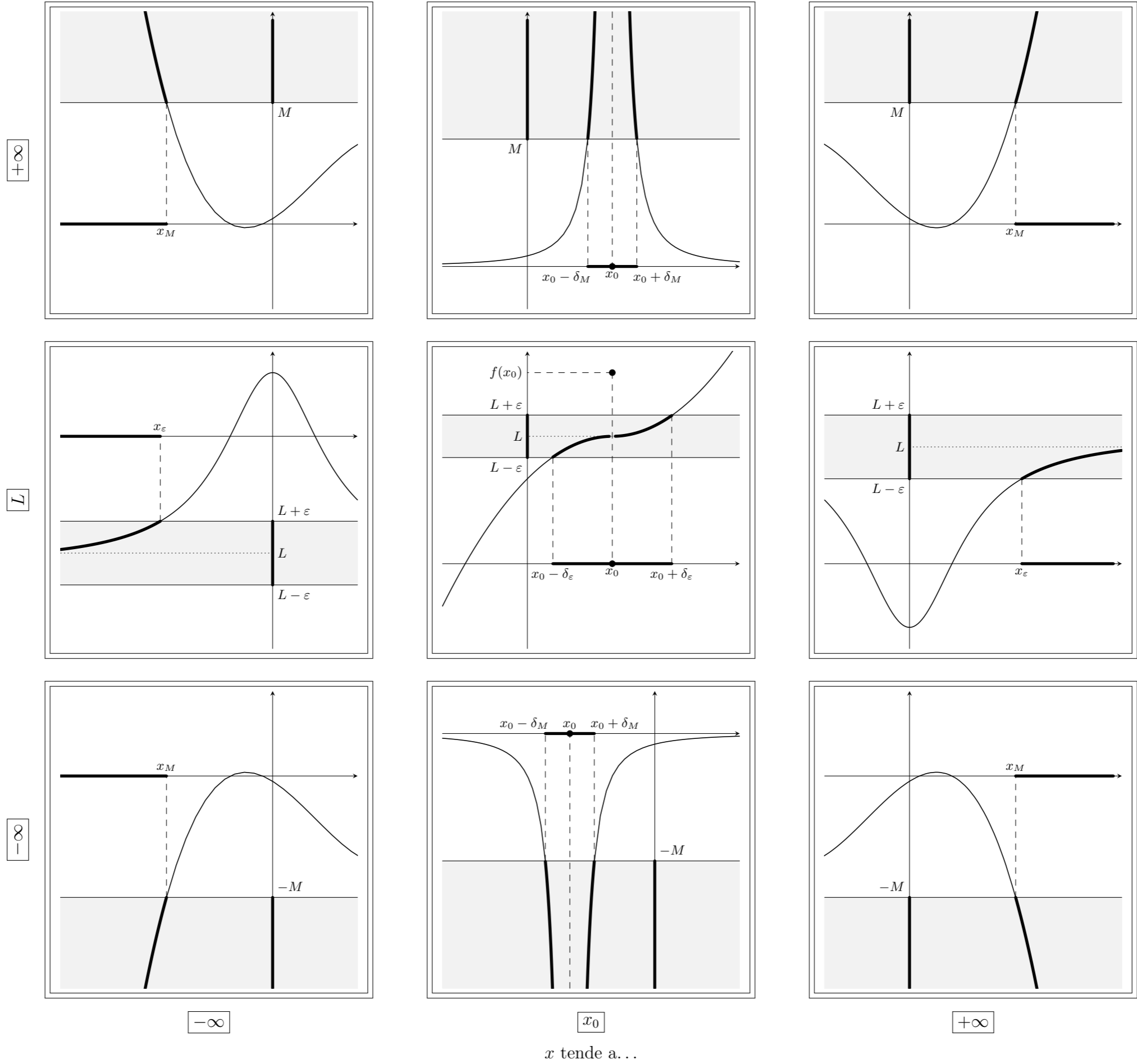
Dimostrare che f ammette o massimo o minimo su \mathbb{R} . Mostrare con un esempio che f può avere massimo e non minimo, o viceversa.

2) Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ continua con la proprietà

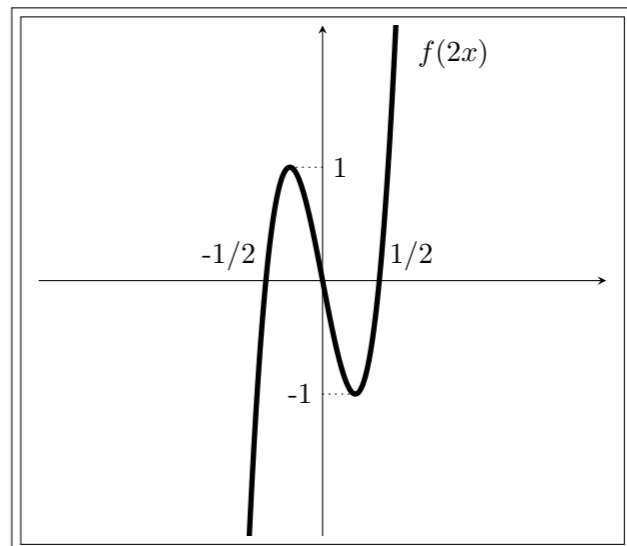
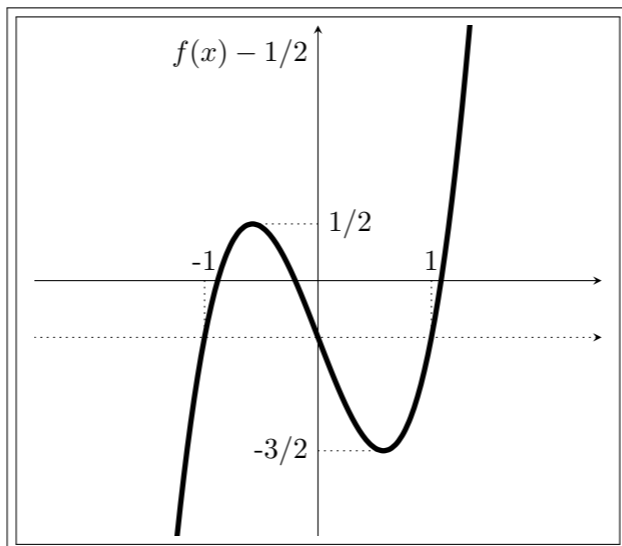
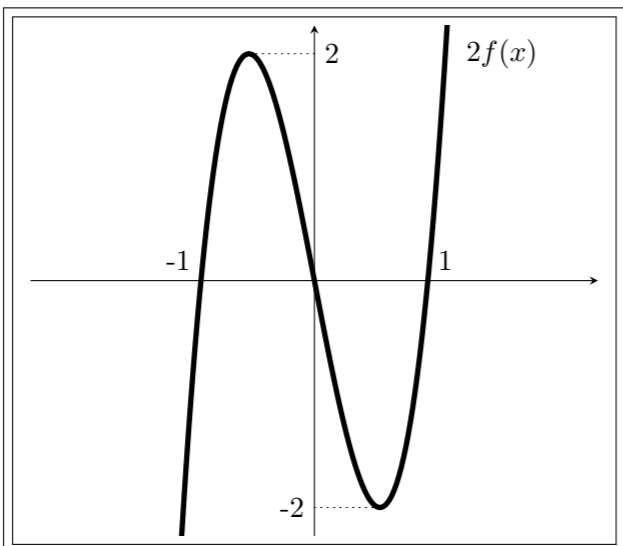
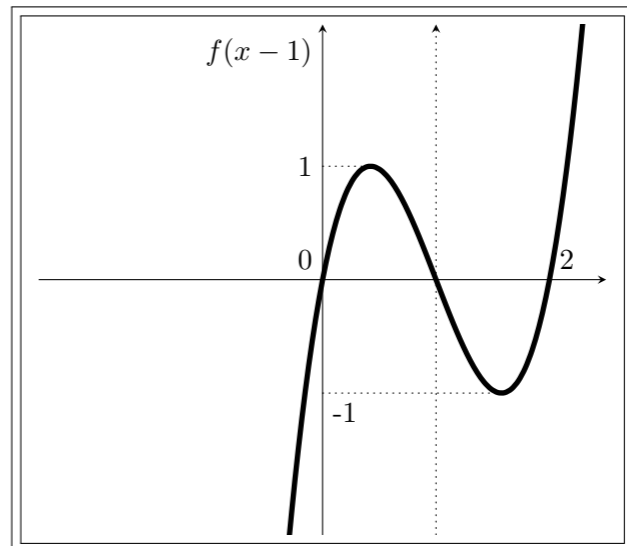
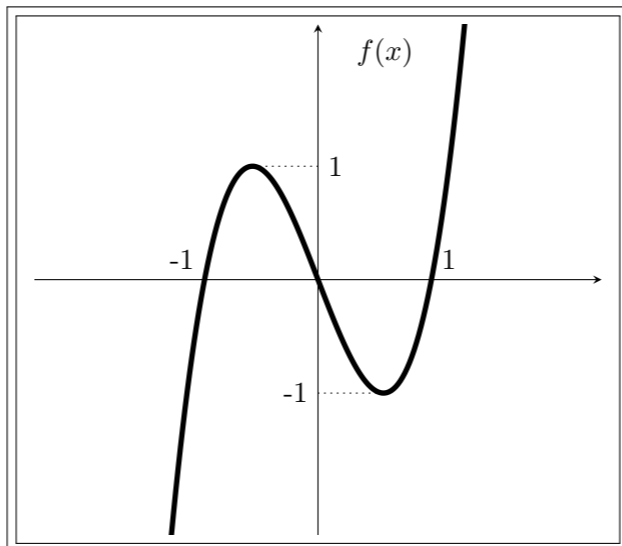
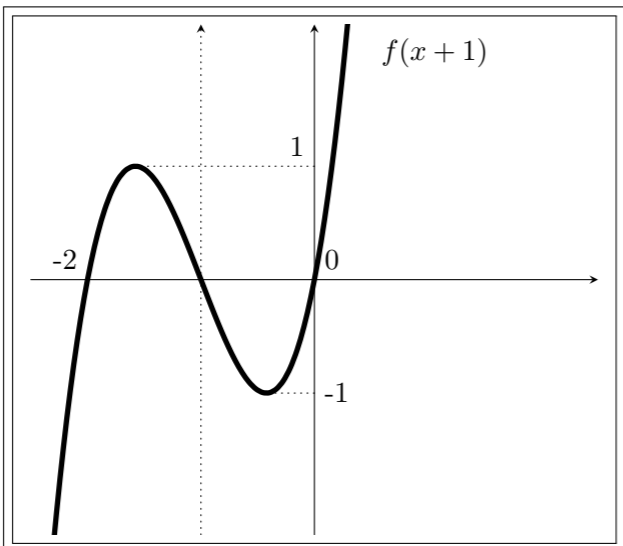
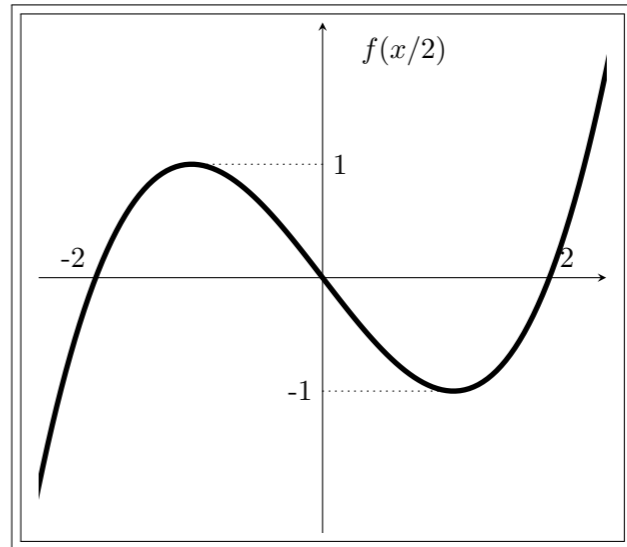
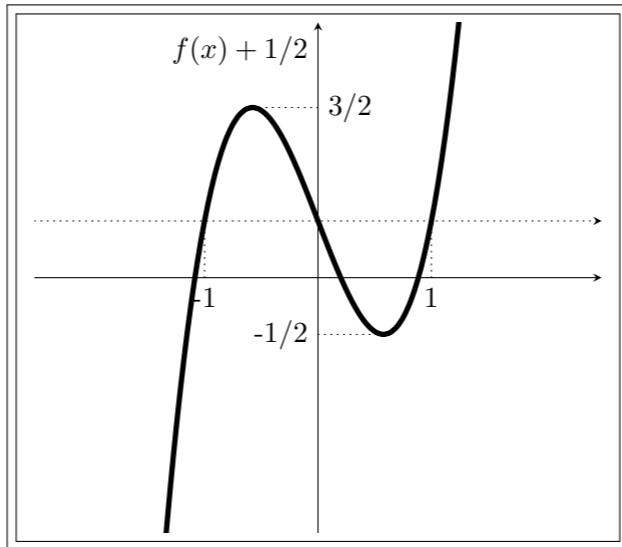
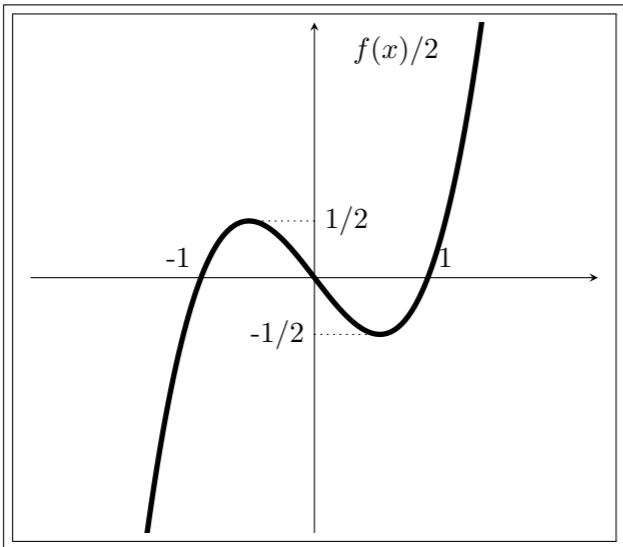
$$\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow -\infty} f(x) = +\infty.$$

Allora f ammette minimo.

$f(x)$ tende a...



x tende a...



Il teorema di Weierstrass e il teorema dei valori intermedi

Ricordiamo il teorema di Weierstrass:

TEOREMA 1 (Weierstrass). *Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua. Allora esistono x_m e x_M in $[a, b]$ tali che*

$$m = f(x_m) \leq f(x) \leq f(x_M) = M, \quad \forall x \in [a, b].$$

In altre parole, esistono il minimo m e il massimo M di $f(x)$ su $[a, b]$:

$$m = \min\{f(x), x \in [a, b]\}, \quad M = \max\{f(x), x \in [a, b]\}.$$

Conseguenza del teorema di Weierstrass è il teorema dei valori intermedi per funzioni continue su un intervallo chiuso e limitato:

TEOREMA 2 (Valori intermedi). *Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua. Allora f assume tutti (e soli) i valori compresi tra massimo e minimo, vale a dire si ha*

$$f([a, b]) = \{f(x), x \in [a, b]\} = [m, M].$$

La dimostrazione di questo risultato si basa su un principio generale, valido per le funzioni continue su intervalli qualsiasi: *se, dato un numero reale t , esistono due punti x_0 e x_1 nell'intervallo di definizione di f tali che $f(x_0) < t < f(x_1)$, allora esiste x_t appartenente all'intervallo di estremi x_0 e x_1 tale che $f(x_t) = t$.*

Nel caso del teorema dei valori intermedi, fissato t in (m, M) , si ha $m = f(x_m) < t < f(x_M) = M$, e quindi la funzione f assume il valore t .

Il teorema di Weierstrass (e di conseguenza il teorema dei valori intermedi) hanno bisogno di una funzione continua definita su un intervallo chiuso e limitato (come si vede tramite opportuni controesempi). Cosa si può dire se l'intervallo non è chiuso, o non è limitato?

Primo caso: f è definita su tutto \mathbb{R} .

In questo caso, è possibile che si perdano sia il massimo che il minimo (ad esempio: se $f(x) = x$, non esistono né il minimo, né il massimo). È però possibile recuperare uno dei due con un'ipotesi aggiuntiva sul comportamento della funzione a più infinito e a meno infinito (le uniche due "direzioni" che possono essere responsabili della perdita del minimo e del massimo). Più precisamente, abbiamo il seguente:

TEOREMA 3 (Weierstrass generalizzato - 1). *Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua.*

a) Se

$$\lim_{x \rightarrow -\infty} f(x) = +\infty \quad \lim_{x \rightarrow +\infty} f(x) = +\infty$$



allora esiste il minimo m di $f(x)$ su \mathbb{R} .

b) Se

$$\lim_{x \rightarrow -\infty} f(x) = -\infty \quad \lim_{x \rightarrow +\infty} f(x) = -\infty$$



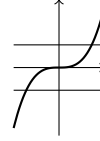
allora esiste il massimo M di $f(x)$ su \mathbb{R} .

Se siamo nelle ipotesi del teorema precedente, possiamo generalizzare anche il teorema dei valori intermedi:

TEOREMA 4 (Valori intermedi generalizzato - 1). Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua.

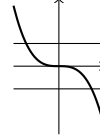
a) Se

$$\lim_{x \rightarrow -\infty} f(x) = -\infty \quad \lim_{x \rightarrow +\infty} f(x) = +\infty$$



ovvero se

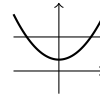
$$\lim_{x \rightarrow -\infty} f(x) = +\infty \quad \lim_{x \rightarrow +\infty} f(x) = -\infty$$



allora $f(\mathbb{R}) = \mathbb{R}$.

b) Se

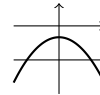
$$\lim_{x \rightarrow -\infty} f(x) = +\infty \quad \lim_{x \rightarrow +\infty} f(x) = +\infty$$



allora $f(\mathbb{R}) = [m, +\infty)$ dove m è il minimo di f su \mathbb{R} .

c) Se

$$\lim_{x \rightarrow -\infty} f(x) = -\infty \quad \lim_{x \rightarrow +\infty} f(x) = -\infty$$



allora $f(\mathbb{R}) = (-\infty, M]$ dove M è il massimo di f su \mathbb{R} .

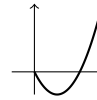
Secondo caso: f è definita su una semiretta chiusa $[a, +\infty)$ o $(-\infty, b]$.

In questo secondo caso, se la funzione $f(x)$ diverge (positivamente o negativamente) quando x diverge (positivamente o negativamente), uno tra massimo e minimo si “perde”, ma l’altro esiste. Questo è quanto afferma il seguente risultato:

TEOREMA 5 (Weierstrass generalizzato - 2). Sia $f : [a, +\infty) \rightarrow \mathbb{R}$ una funzione continua.

a) Se

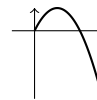
$$\lim_{x \rightarrow +\infty} f(x) = +\infty$$



allora esiste il minimo m di $f(x)$ su $[a, +\infty)$.

b) Se

$$\lim_{x \rightarrow +\infty} f(x) = -\infty$$



allora esiste il massimo M di $f(x)$ su $[a, +\infty)$.

Sia $f : (-\infty, b] \rightarrow \mathbb{R}$ una funzione continua.

a) Se

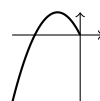
$$\lim_{x \rightarrow -\infty} f(x) = +\infty$$



allora esiste il minimo m di $f(x)$ su $(-\infty, b]$.

b) Se

$$\lim_{x \rightarrow -\infty} f(x) = -\infty$$



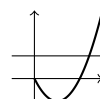
allora esiste il massimo M di $f(x)$ su $(-\infty, b]$.

Conseguentemente, il teorema dei valori intermedi si generalizza così:

TEOREMA 6 (Valori intermedi generalizzato - 2). Sia $f : [a, +\infty) \rightarrow \mathbb{R}$ una funzione continua.

a) Se

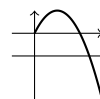
$$\lim_{x \rightarrow +\infty} f(x) = +\infty$$



allora $f([a, +\infty)) = [m, +\infty)$, dove m è il minimo di $f(x)$ su $[a, +\infty)$.

b) Se

$$\lim_{x \rightarrow +\infty} f(x) = -\infty$$

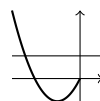


allora $f([a, +\infty)) = (-\infty, M]$, dove M è il massimo di $f(x)$ su $[a, +\infty)$.

Sia $f : (-\infty, b] \rightarrow \mathbb{R}$ una funzione continua.

a) Se

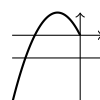
$$\lim_{x \rightarrow -\infty} f(x) = +\infty$$



allora $f((-\infty, b]) = [m, +\infty)$, dove m è il minimo di $f(x)$ su $(-\infty, b]$.

b) Se

$$\lim_{x \rightarrow -\infty} f(x) = -\infty$$



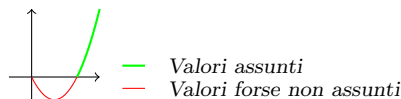
allora $f((-\infty, b]) = (-\infty, M]$, dove M è il massimo di $f(x)$ su $(-\infty, b]$.

Se, per un motivo qualsiasi, non si è in grado di calcolare il minimo m , o il massimo M delle funzioni f , si possono comunque ricavare delle informazioni parziali sull'immagine di f .

TEOREMA 7 (Valori intermedi generalizzato - 4). Sia $f : [a, \infty) \rightarrow \mathbb{R}$ una funzione continua.

a) Se

$$\lim_{x \rightarrow +\infty} f(x) = +\infty$$

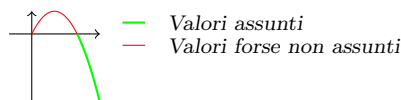


— Valori assunti
— Valori forse non assunti

allora $[f(a), +\infty) \subseteq f([a, +\infty))$, ovvero f assume tutti i valori maggiori o uguali a $f(a)$.

b) Se

$$\lim_{x \rightarrow +\infty} f(x) = -\infty$$



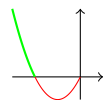
— Valori assunti
— Valori forse non assunti

allora $(-\infty, f(a)] \subseteq f([a, +\infty))$, ovvero f assume tutti i valori minori o uguali a $f(a)$.

Sia $f : (-\infty, b] \rightarrow \mathbb{R}$ una funzione continua.

a) Se

$$\lim_{x \rightarrow -\infty} f(x) = +\infty$$

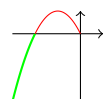


— Valori assunti
— Valori forse non assunti

allora $[f(b), +\infty) \subseteq f((-\infty, b])$, ovvero f assume tutti i valori maggiori o uguali a $f(b)$.

b) Se

$$\lim_{x \rightarrow -\infty} f(x) = -\infty$$



— Valori assunti
— Valori forse non assunti

allora $(-\infty, f(b)] \subseteq f((-\infty, b])$, ovvero f assume tutti i valori minori o uguali a $f(b)$.

Terzo caso: f è definita su un intervallo aperto (limitato o meno).

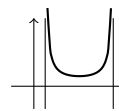
Anche in questo caso, come nei precedenti, entrano in gioco i limiti della funzione “agli estremi dell’intervallo”. Per semplicità, supponiamo che la funzione f sia definita su un intervallo limitato (a, b) ; l’estensione al caso di un intervallo illimitato (in una direzione) è lasciata al lettore...

TEOREMA 8 (Weierstrass generalizzato – 3). Sia $f : (a, b) \rightarrow \mathbb{R}$ una funzione continua.

a) Se

$$\lim_{\substack{x \rightarrow a \\ x > a}} f(x) = +\infty$$

$$\lim_{\substack{x \rightarrow b \\ x < b}} f(x) = +\infty$$

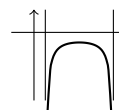


allora esiste il minimo m di $f(x)$ su (a, b) .

b) Se

$$\lim_{\substack{x \rightarrow a \\ x > a}} f(x) = -\infty$$

$$\lim_{\substack{x \rightarrow b \\ x < b}} f(x) = -\infty$$



allora esiste il massimo M di $f(x)$ su (a, b) .

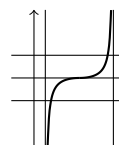
Analogamente a quanto detto in precedenza, il teorema dei valori intermedi si trasforma così:

TEOREMA 9 (Valori intermedi generalizzato – 5). Sia $f : (a, b)$ una funzione continua.

a) Se

$$\lim_{\substack{x \rightarrow a \\ x > a}} f(x) = -\infty$$

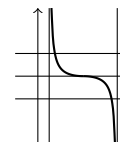
$$\lim_{\substack{x \rightarrow b \\ x < b}} f(x) = +\infty$$



o se

$$\lim_{\substack{x \rightarrow a \\ x > a}} f(x) = +\infty$$

$$\lim_{\substack{x \rightarrow b \\ x < b}} f(x) = -\infty$$

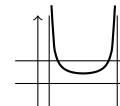


allora $f((a, b)) = \mathbb{R}$.

b) Se

$$\lim_{\substack{x \rightarrow a \\ x > a}} f(x) = +\infty$$

$$\lim_{\substack{x \rightarrow b \\ x < b}} f(x) = +\infty$$

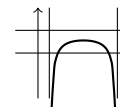


allora $f((a, b)) = [m, +\infty)$, dove m è il minimo di $f(x)$ su (a, b) .

c) Se

$$\lim_{\substack{x \rightarrow a \\ x > a}} f(x) = -\infty$$

$$\lim_{\substack{x \rightarrow b \\ x < b}} f(x) = -\infty$$



allora $f((a, b)) = (-\infty, M]$, dove M è il massimo di $f(x)$ su (a, b) .

CAPITOLO 1

Derivazione

1. Derivata.

Che cos'è la velocità? L'idea di velocità fa parte dell'esperienza comune di tutti. Non solo: sappiamo bene che la velocità si può misurare, e abbiamo sviluppato una certa sensibilità al riguardo. Sappiamo che essa ha a che fare con quanto spazio si riesce a percorrere in un certo tempo, cioè si tratta di un rapporto. Sappiamo anche che ha perfettamente senso parlare della velocità in un certo istante (pensate al tachimetro di una macchina): non c'è bisogno di percorrere una distanza, misurarla e calcolare un rapporto, si può misurare la “velocità istantanea”. Ma di che si tratta esattamente?

Riuscire ad afferrare questo concetto e utilizzarlo in un procedimento di calcolo è stata un'impresa intellettuale importante, compiuta tra la fine del seicento e l'inizio del settecento sostanzialmente da Newton (la priorità è discussa: certamente Galileo fu un precursore, e Leibniz ebbe le stesse idee negli stessi anni, ma Newton oltre ad inventare l'analisi fu anche in grado di usarla per risolvere vari problemi di meccanica celeste con successo spettacolare). È difficile immaginare oggi l'impatto che queste idee poterono avere nel settecento. Oggetti che sembravano al limite del soprannaturale, le comete e i pianeti, si rivelavano invece obbedire alle stesse leggi degli oggetti comuni, e addirittura se ne potevano prevedere i movimenti da terra. Il successo di Newton segnò l'inizio della scienza moderna.

DEFINIZIONE 1.1. Sia $f :]a, b[\rightarrow \mathbb{R}$ una funzione e $x_0 \in]a, b[$. Diciamo che f è **derivabile in x_0** se esiste il limite

$$(1.1) \quad \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

detto **derivata** di f nel punto x e indicato con varie notazioni:

$$f'(x_0) = Df(x_0) = \frac{df}{dx}(x_0)$$

(notazioni attribuite a Lagrange, Leibniz e Louis Arbogeste). Se f è derivabile in tutti i punti del dominio si dice **derivabile**, e la funzione $x \mapsto f'(x)$ si indica con f' (e si chiama naturalmente la derivata di f).

Si può anche considerare in (1.1) il limite destro per $h \rightarrow 0^+$, detto **derivata destra**, e il limite sinistro per $h \rightarrow 0^-$, detto **derivata sinistra**. In particolare per una funzione $f : [a, b] \rightarrow \mathbb{R}$ si possono definire la derivata sinistra nel punto b e la derivata destra nel punto a . Ovviamente, una funzione è derivabile in un punto se e solo se derivata destra e quella sinistra in quel punto esistono e sono uguali.

OSSERVAZIONE 1.2. Notare che se si pone $y = x_0 + h$ nella definizione di derivata si ottiene la formula equivalente

$$f'(x_0) = \lim_{y \rightarrow x_0} \frac{f(y) - f(x_0)}{y - x_0}$$

perché evidentemente $y = x_0 + h \rightarrow x_0$ quando $h \rightarrow 0$.

OSSERVAZIONE 1.3 (Notazione degli o piccoli). Introduciamo una notazione che avrà molta importanza nel seguito. Consideriamo una funzione $f(x)$ definita in un

intervallo aperto contenente 0. Diciamo che f è un **o piccolo di x** , per x che tende a 0, e scriviamo $f(x) \in o(x)$, se f ha la seguente proprietà:

$$f(x) \in o(x) \quad \text{significa} \quad \lim_{x \rightarrow 0} \frac{f(x)}{x} = 0.$$

(si dice anche: f è un **infinitesimo di ordine superiore a 1** per x che tende a 0). Quindi $o(x)$ è l'insieme di tutte le funzioni che hanno la proprietà $f(x)/x \rightarrow 0$ per $x \rightarrow 0$. Di solito si fa un piccolo abuso di notazione e si scrive semplicemente

$$f(x) = o(x) \quad \Longleftrightarrow \quad \frac{f(x)}{x} \rightarrow 0$$

e questa notazione è molto comoda ma bisogna stare un po' attenti ad usarla. Ad esempio, la differenza $o(x) - o(x)$ non si cancella! cioè, se $f(x) = o(x)$ e $g(x) = o(x)$, si ha sicuramente $f(x) - g(x) = o(x) - o(x) = o(x)$, ma $f - g$ è uguale a zero solo se le due funzioni coincidono. Alcuni esempi:

- (1) La funzione $f(x) = x^2$ è un $o(x)$: infatti $\frac{f(x)}{x} = \frac{x^2}{x} \rightarrow 0$ per $x \rightarrow 0$.
- (2) La funzione $f(x) = x$ non è un $o(x)$ per $x \rightarrow 0$
- (3) In generale $x^k = o(x)$ se e solo se $k > 1$.
- (4) La funzione $f(x) = \sin x$ non è un $o(x)$, infatti $\frac{\sin x}{x}$ non tende a 0 per $x \rightarrow 0$.
- (5) $\sin(x^2) = o(x)$ per $x \rightarrow 0$.
- (6) $\ln(1 + x^3) = o(x)$ per $x \rightarrow 0$.

OSSERVAZIONE 1.4. Vediamo come si può riformulare la definizione di derivata usando la notazione degli o piccoli. Una funzione f è derivabile nel punto x con derivata $f'(x) = L$ se e solo se il rapporto incrementale $\frac{f(x+h)-f(x)}{h} \rightarrow L$ per $h \rightarrow 0$, cioè se e solo si ha

$$\frac{f(x+h) - f(x) - Lh}{h} = \frac{f(x+h) - f(x)}{h} - L \rightarrow 0.$$

Quindi vediamo che $f(x)$ è derivabile in x con derivata L se e solo se $f(x+h) - f(x) - Lh = o(h)$, ossia

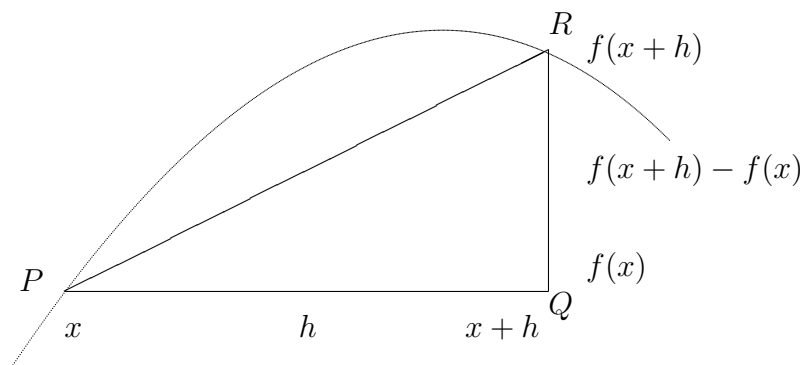
$$(1.2) \quad f(x+h) = f(x) + Lh + o(h).$$

OSSERVAZIONE 1.5. Il rapporto

$$(1.3) \quad \frac{f(x+h) - f(x)}{h}$$

si chiama **rapporto incrementale** di f in x e allora h si dice anche l'**incremento**. Notare che l'incremento può essere qualunque numero positivo o negativo diverso da 0, e quindi $x+h$ può essere anche a sinistra di x ; purché naturalmente si abbia sempre $x+h \in]a, b[$.

È facile visualizzare geometricamente il rapporto incrementale: nella figura,



se il punto P del grafico ha coordinate $(x, f(x))$ e ad esempio prendiamo $h > 0$, il numeratore del rapporto incrementale rappresenta la lunghezza del segmento QR parallelo all'asse delle ordinate, il denominatore h è la lunghezza del segmento PQ parallelo all'asse delle ascisse, e quindi il rapporto incrementale esprime la pendenza del segmento PR : rapporto 1 vuol dire pendenza di $45^\circ = \frac{\pi}{4}$, e così via. Allora è naturale interpretare il limite del rapporto incrementale, ossia la derivata $f'(x)$, come la pendenza limite delle rette PR quando il punto P tende al punto R ; ossia la pendenza della retta tangente al grafico di f nel punto P .

Possiamo utilizzare quest'osservazione per scrivere l'equazione della retta tangente al grafico di una funzione (quando tale retta esiste!). Una retta di pendenza a ha la forma $ax + b$ (ossia il suo coefficiente angolare è a), quindi la retta tangente ad f in un punto fissato c deve avere pendenza $a = f'(c)$ ossia è del tipo

$$f'(c)x + b.$$

A questo punto basta scegliere il coefficiente b in modo che la retta passi per il punto del grafico $(c, f(c))$, ossia

$$f'(c) \cdot c + b = f(c) \implies b = f(c) - cf'(c)$$

e quindi la retta tangente al grafico di f nel punto c ha la forma

$$f'(c)(x - c) + f(c).$$

Anche la formulazione equivalente dell'Osservazione 1.4 si può visualizzare geometricamente: se x è un punto vicino a c , ossia $x = c + h$, allora la (1.2) si può scrivere, dato che $L = f'(c)$,

$$f(x) = f(c) + f'(c)(x - c) + o(x - c).$$

Quindi la funzione $o(x - c)$ si può interpretare come l'errore che si commette approssimando $f(x)$ con la sua tangente; vediamo che l'esistenza della derivata implica che tale errore è di ordine superiore al primo, in quanto

$$\frac{o(h)}{h^1} \rightarrow 0 \quad \text{per} \quad h \rightarrow 0.$$

ESEMPIO 1.6. Sia $f(x) = C$ una funzione costante (ossia una funzione f il cui valore è C in tutti i punti). Il suo grafico naturalmente è una retta orizzontale ad altezza C . Se calcoliamo il rapporto incrementale in un qualunque punto x abbiamo

$$\frac{f(x+h) - f(x)}{h} = \frac{C - C}{h} = 0$$

(si ricordi che $h \neq 0$) e quindi otteniamo che *la derivata di una funzione costante è 0*; per ricordare questa regola scriviamo in modo sintetico (e leggermente impreciso)

$$(C)' = 0.$$

Notare che la “pendenza” della curva è zero, cioè la curva è orizzontale in tutti i punti.

Sia $f(x) = x$: il rapporto incrementale in qualunque punto x è

$$\frac{f(x+h) - f(x)}{h} = \frac{x+h-x}{h} = 1$$

e quindi otteniamo la regola

$$(x)' = 1$$

ossia la derivata della funzione x è la funzione costante 1.

Se invece consideriamo $f(x) = cx + d$ dove c, d sono numeri reali fissati, il rapporto incrementale è

$$\frac{f(x+h) - f(x)}{h} = \frac{c(x+h) - cx}{h} = c$$

da cui $f' = c$, ossia sinteticamente

$$(cx + d)' = c.$$

Notiamo che la pendenza della retta $cx + d$ è espressa proprio dal coefficiente angolare c (positivo, negativo o nullo), indipendentemente dal valore di d .

Altra derivata facile da calcolare è quella di $f(x) = x^2$:

$$\frac{f(x+h) - f(x)}{h} = \frac{(x+h)^2 - x^2}{h} = 2x + h$$

e se calcoliamo il limite per $h \rightarrow 0$ otteniamo subito $f'(x) = 2x$. Abbiamo quindi la regola sintetica

$$(x^2)' = 2x.$$

Proviamo a calcolare la derivata di $f(x) = e^x$:

$$\frac{e^{x+h} - e^x}{h} = e^x \cdot \frac{e^h - 1}{h};$$

il limite per $h \rightarrow 0$ si calcola facilmente ricordando che

$$\frac{e^h - 1}{h} \rightarrow 1 \quad \text{per } h \rightarrow 0$$

e dà $f' = e^x$:

$$(e^x)' = e^x.$$

Invece per $f(x) = \text{sen } x$ il calcolo è un po' più complicato; il modo più rapido è utilizzare la formula di *prostaferesi*

$$\text{sen } a - \text{sen } b = 2\text{sen} \left(\frac{a-b}{2} \right) \cos \left(\frac{a+b}{2} \right)$$

da cui si ha subito

$$\frac{\text{sen}(x+h) - \text{sen } x}{h} = \frac{2}{h} \text{sen} \left(\frac{h}{2} \right) \cdot \cos \left(x + \frac{h}{2} \right).$$

Ricordando che

$$\frac{\text{sen } x}{x} \rightarrow 1 \quad \text{per } x \rightarrow 0$$

otteniamo subito che il limite per $h \rightarrow 0$ è $f'(x) = \cos x$:

$$(\text{sen } x)' = \cos x.$$

Un procedimento simile per $f(x) = \cos x$ dà $f'(x) = -\text{sen } x$:

$$(\cos x)' = -\text{sen } x$$

(verificare!).

ESERCIZIO 1.7. Quanto fa la derivata di $f(t) = \sin t$ nel punto t ? E la derivata di $f(z) = e^z$ nel punto $z = 4$? E la derivata di $f(z) = e^z$ nel punto $-z$? E la derivata di $f(t) = t^2$ nel punto x ?

OSSERVAZIONE 1.8. Come dicevamo all'inizio, la derivata è un concetto nato dall'idea intuitiva di velocità. Se $f(t)$ esprime la posizione di un oggetto in moto, in funzione del tempo, allora il rapporto incrementale di f è il rapporto tra lo spazio percorso nel tempo h , e il tempo h impiegato a percorrerlo, cioè esprime la velocità media nell'intervallo di tempo h . Allora il limite per $h \rightarrow 0$, cioè la derivata $f'(t)$, esprime la velocità all'istante t : è la rapidità con cui varia la posizione all'istante t .

Supponiamo di avere calcolato la velocità di un oggetto in ogni istante; abbiamo allora la funzione $v(t)$, velocità in funzione del tempo. Se ripetiamo il procedimento precedente e deriviamo v , otteniamo una nuova quantità che misura la rapidità con cui varia la velocità all'istante t : è detta *accelerazione* dell'oggetto all'istante t . Si tratta della *derivata seconda* della posizione; torneremo su questo concetto in seguito.

Quando $f(t)$ rappresenta una posizione in funzione del tempo, si usa talvolta la notazione $\dot{f}(t)$ invece di $f'(t)$; si tratta della notazione originale di Newton. Allora l'accelerazione si indica con $\ddot{f}(t)$.

ESERCIZIO 1.9. a) Determinare la derivata delle funzioni seguenti calcolando il limite del rapporto incrementale:

$$\frac{1}{2}x^2, \quad 1 - 2x, \quad \frac{1}{x}, \quad x + \frac{1}{x}, \quad (x + 3)^2, \quad A^x \ (A > 0), \quad \sqrt{x}, \quad \sqrt{1 + x^2}.$$

b) Una particella P si muove di moto uniformemente accelerato. Questo vuol dire che lo spazio percorso all'istante t è dato dalla formula

$$s = \frac{1}{2}at^2 + v_0t + s_0$$

dove s_0 è la posizione iniziale e v_0 la velocità iniziale. Calcolare le espressioni della velocità e dell'accelerazione di P all'istante t .

ESEMPIO 1.10. Nel seguito saremo interessati quasi esclusivamente a funzioni derivabili; ma naturalmente vi sono moltissime funzioni non derivabili (anzi, in un senso ben preciso, sono molte di più). Un esempio semplice è dato dal modulo $f(x) = |x|$: f non è derivabile in 0. Infatti il rapporto incrementale in $x = 0$

$$\frac{f(0 + h) - f(0)}{h} = \frac{|h|}{h}$$

vale $+1$ per $h > 0$, e quindi il limite destro è $+1$, mentre vale -1 per $h < 0$, e quindi il limite sinistro è -1 . Pertanto il limite per $h \rightarrow 0$ non esiste (notare che esistono la derivata sinistra e quella destra in 0, ma hanno valori diversi).

In questo esempio la funzione non è derivabile soltanto in un punto; non è difficile costruire funzioni non derivabili in un numero infinito di punti (ad esempio la funzione di Dirichlet vista nel Capitolo precedente). Si può fare di meglio e costruire una funzione continua che non è derivabile in nessun punto del suo dominio, ma questo è un po' più difficile (Weierstrass 1861).

Anche se non tutte le funzioni continue sono derivabili, è però vero il contrario:

TEOREMA 1.11. *Una funzione derivabile in un punto è anche continua in quel punto.*

Dimostrazione. Possiamo scrivere

$$f(c+h) = f(c) + \frac{f(c+h) - f(c)}{h} \cdot h$$

e se prendiamo il limite per $h \rightarrow 0$ otteniamo

$$\lim_{h \rightarrow 0} f(c+h) = f(c) + f'(x) \cdot 0 = f(c)$$

e quindi la funzione è continua nel punto c .

ESERCIZIO 1.12. Sia $f(x) = x^{1/3} = \sqrt[3]{x}$. Esiste la derivata di f in 0? Stessa domanda per la funzione $f(x) = x|x|$ (attenzione!).

2. Regole di derivazione.

Calcolare la derivata di una combinazione di funzioni elementari non è difficile, perché il procedimento si può ridurre all'applicazione a catena di una serie di semplici regole.

TEOREMA 2.1. (*Regole di derivazione*). Siano $f, g :]a, b[\rightarrow \mathbb{R}$ due funzioni derivabili nel punto x , e sia $\alpha \in \mathbb{R}$. Allora $f \pm g$, αf e fg sono derivabili in x e valgono le regole di derivazione

$$(f \pm g)'(x) = f'(x) \pm g'(x), \quad (\alpha f)'(x) = \alpha f'(x),$$

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x).$$

Inoltre, se $g(x) \neq 0$, anche il rapporto f/g è (definito e) derivabile in x e si ha

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

In particolare,

$$\left(\frac{1}{g}\right)'(x) = -\frac{g'(x)}{g(x)^2}.$$

Dimostrazione. Il rapporto incrementale di $f + g$ nel punto x è

$$\frac{(f(x+h) + g(x+h)) - (f(x) + g(x))}{h} = \frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h}$$

e prendendo il limite per $h \rightarrow 0$ otteniamo subito che $f + g$ è derivabile in x e vale la formula $(f + g)'(x) = f'(x) + g'(x)$. Dimostrazione identica per $f - g$ e analoga per αf .

Il rapporto incrementale di fg nel punto x si può scrivere

$$\frac{f(x+h)g(x+h) - f(x)g(x)}{h} = \frac{f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x)}{h}$$

che è uguale a

$$= \frac{f(x+h) - f(x)}{h} \cdot g(x+h) + f(x) \frac{g(x+h) - g(x)}{h}.$$

Prendendo il limite per $h \rightarrow 0$, e osservando che $g(x+h) \rightarrow g(x)$ per il Teorema 1.11, otteniamo che fg è derivabile in x e vale la formula $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$.

Infine, se $g(x) \neq 0$, dato che g è derivabile e quindi continua in x , per la permanenza del segno si ha $g(x+h) \neq 0$ per h vicino a 0, pertanto il rapporto $1/g(x+h)$ è ben

definito. Per calcolare la derivata di $1/g$ utilizziamo direttamente la definizione: il rapporto incrementale della funzione $1/g$ è dato da

$$\frac{1}{h} \left(\frac{1}{g(x+h)} - \frac{1}{g(x)} \right) = -\frac{g(x+h) - g(x)}{h} \cdot \frac{1}{g(x+h)g(x)};$$

se ora prendiamo il limite per $h \rightarrow 0$, e ricordiamo che $g(x+h) \rightarrow g(x)$ per il Teorema 1.11, otteniamo subito

$$\lim_{h \rightarrow 0} \frac{1}{h} \left(\frac{1}{g(x+h)} - \frac{1}{g(x)} \right) = -\frac{g'(x)}{g(x)^2}.$$

La regola per la derivata di f/g segue immediatamente scrivendo

$$\frac{f}{g} = f \cdot \frac{1}{g}$$

e utilizzando le regole precedenti per il prodotto ed il reciproco di funzioni.

ESEMPIO 2.2. Abbiamo già a disposizione le regole di calcolo $(x)' = 1$ e $(x^2)' = 2x$. Per calcolare la derivata di una qualunque potenza intera $f(x) = x^n$ possiamo procedere per induzione e dimostrare la formula

$$f'(x) = (x^n)' = nx^{n-1}.$$

Infatti se la formula è vera per un certo n , possiamo applicare la regola per la derivata di un prodotto a $x^{n+1} = x \cdot x^n$:

$$(x^{n+1})' = (x \cdot x^n)' = 1 \cdot x^n + x \cdot nx^{n-1} = (n+1)x^n$$

e otteniamo la formula al passo $n+1$.

In realtà la formula vale per qualunque potenza reale $a \in \mathbb{R}$:

$$(x^a)' = ax^{a-1}.$$

Dimostreremo questa regola più sotto (vedi l'Esempio 2.11) utilizzando i teoremi seguenti. Ad esempio per $f(x) = \sqrt{x} = x^{1/2}$ si ha

$$(\sqrt{x})' = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}},$$

per $f(x) = \frac{1}{x} = x^{-1}$ si ha

$$\left(\frac{1}{x}\right)' = -\frac{1}{x^2}$$

e per $f(x) = 1/\sqrt{x} = x^{-1/2}$

$$\left(\frac{1}{\sqrt{x}}\right)' = -\frac{1}{2} \frac{1}{\sqrt{x^3}}$$

(verificare).

Calcoliamo la derivata di $f(x) = \operatorname{tg} x$: usando la regola per la derivata di un rapporto otteniamo

$$(\operatorname{tg} x)' = \left(\frac{\operatorname{sen} x}{\operatorname{cos} x}\right)' = \frac{\operatorname{cos}^2 x + \operatorname{sen}^2 x}{\operatorname{cos}^2 x} = \frac{1}{\operatorname{cos}^2 x}.$$

ESERCIZIO 2.3. a) Calcolare la derivata delle funzioni seguenti:

$$6x^7 - \frac{1}{x}, \quad \frac{1}{x^2}, \quad xe^x, \quad x(e^x - 2), \quad \operatorname{sen} x \operatorname{cos} x.$$

TEOREMA 2.4. (*Derivata delle funzioni composte*). Sia $f :]a, b[\rightarrow \mathbb{R}$ derivabile nel punto x , e sia $g :]\alpha, \beta[\rightarrow \mathbb{R}$ con $f(]a, b[) \subseteq]\alpha, \beta[$ derivabile nel punto $y = f(x)$. Allora la funzione composta $g \circ f$ è derivabile nel punto x e vale la regola

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x).$$

Dimostrazione. Poniamo $y = f(x)$ e definiamo per k piccolo la funzione

$$\phi(k) = \begin{cases} g'(y) & \text{se } k = 0, \\ \frac{g(y+k) - g(y)}{k} & \text{se } k \neq 0 \end{cases}$$

(per la precisione, questa funzione è definita per tutti i k tali che $y + k \in]\alpha, \beta[$ ossia sull'intervallo $]y - \epsilon, y + \epsilon[$). Chiaramente ϕ è continua in $k = 0$ dato che la funzione g è derivabile nel punto $y = f(x)$. Inoltre vale l'identità (anche per $k = 0$)

$$g(y + k) - g(y) = k \cdot \phi(k).$$

Ora consideriamo il rapporto incrementale della funzione composta, di cui dobbiamo calcolare il limite per $h \rightarrow 0$:

$$\frac{g(f(x+h)) - g(f(x))}{h}.$$

Se poniamo $k = f(x+h) - f(x)$ e usiamo l'identità precedente abbiamo

$$\frac{g(f(x+h)) - g(f(x))}{h} = \frac{g(y+k) - g(y)}{h} = \frac{k}{h} \phi(k) = \frac{f(x+h) - f(x)}{h} \phi(k).$$

Adesso calcoliamo il limite per $h \rightarrow 0$; allora si ha anche $k = f(x+h) - f(x) \rightarrow 0$ (dato che f è continua in 0), quindi $\phi(k) \rightarrow g'(y)$, e in conclusione

$$\lim_{h \rightarrow 0} \frac{g(f(x+h)) - g(f(x))}{h} = f'(x)g'(y) = f'(x)g'(f(x)).$$

ESEMPIO 2.5. Calcoliamo la derivata di 2^x : basta scrivere $2^x = e^{x \cdot \ln 2}$ come funzione composta di $g(x) = e^x$, che sappiamo derivare: $g'(x) = e^x$, e $f(x) = x \cdot \ln 2$ che è del tipo cx e quindi sappiamo derivare: $f'(x) = c = \ln 2$. Allora si ha

$$(2^x)' = (e^{x \ln 2})' = g'(f(x))f'(x) = e^{x \ln 2} \ln 2 = 2^x \ln 2.$$

In generale per ogni esponenziale A^x abbiamo

$$(A^x)' = A^x \ln A$$

che naturalmente per $A = e$ dà di nuovo la nota formula.

Molto spesso si incontrano funzioni del tipo $f(x) = \operatorname{tg}(2x)$, $g(x) = \operatorname{sen}(3x - 4)$ eccetera: funzioni composte con una espressione lineare, cioè del tipo $f(ax + b)$. In questi casi la regola precedente dà immediatamente

$$(f(ax + b))' = f'(ax + b)(ax + b)' = a \cdot f'(ax + b).$$

Ad esempio

$$(\operatorname{tg}(2x))' = \frac{2}{\cos^2(2x)}, \quad (\operatorname{sen}(3x - 4))' = 3 \cos(3x - 4).$$

ESERCIZIO 2.6. a) Calcolare la derivata delle funzioni seguenti:

$$\operatorname{sen}(e^x), \quad \operatorname{sen}(x^2), \quad \operatorname{sen}^2(x), \quad \operatorname{sen}(x^2) \cos(e^x), \quad e^{x^2},$$

$$\frac{1}{\sqrt{1+x^2}}, \quad \sqrt{1+x^2}, \quad \sqrt{1-5x^2}, \quad \frac{1}{3+x^2}, \quad \frac{1}{\sqrt{1-2x}}$$

b) Data una funzione $f(x)$, calcolare la derivata di

$$(f(x))^2, \quad \frac{1}{f(x)}, \quad \text{sen}(f(x)), \quad f(e^x), \quad e^{f(x)}.$$

TEOREMA 2.7. (*Derivata delle funzioni inverse*). Sia $f : [a, b] \rightarrow \mathbb{R}$ continua e strettamente crescente (o decrescente), e supponiamo che essa sia derivabile in $x \in]a, b[$ con $f'(x) \neq 0$. Allora la funzione inversa f^{-1} è derivabile nel punto $y = f(x)$ e vale la regola

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

Dimostrazione. Basta considerare il caso f crescente. Indichiamo con $g = f^{-1}$ l'inversa di f ; vogliamo scrivere il suo rapporto incrementale nel punto $y = f(x)$. Dal Teorema 2.8 del Capitolo 2 sappiamo che anche g è strettamente crescente. Se chiamiamo k l'incremento rispetto a y , la quantità

$$h = g(y + k) - g(y)$$

sarà diversa da zero se k è diverso da zero, e vale la relazione

$$k = f(x + h) - f(x)$$

(basta ricavare k dalla relazione precedente). Quindi si può scrivere

$$\frac{g(y + k) - g(y)}{k} = \frac{h}{f(x + h) - f(x)};$$

osserviamo adesso che al tendere di $k \rightarrow 0$ anche $h \rightarrow 0$ (per continuità) e quindi facendo il limite per $k \rightarrow 0$ nell'identità precedente otteniamo la tesi.

ESEMPIO 2.8. La funzione $\ln x$ è definita come l'inversa della funzione esponenziale $f(x) = e^x$. Proviamo a calcolarne la derivata usando il Teorema precedente; usiamo le stesse notazioni dell'enunciato per maggiore chiarezza, fissiamo un punto x , poniamo $y = f(x) = e^x$, e calcoliamo la derivata di $f^{-1} = \ln$ nel punto y . Dato che $(e^x)' = e^x$, si ha

$$(\ln)'(y) = \frac{1}{e^x} = \frac{1}{y}.$$

Abbiamo cioè dimostrato la regola di derivazione

$$(\ln x)' = \frac{1}{x}.$$

Consideriamo adesso la funzione $\text{arctg } x$ che è l'inversa di $f(x) = \text{tg } x$. Procedendo allo stesso modo (e ricordando la regola $(\text{tg } x)' = 1/\cos^2 x$), nel punto $y = \text{tg } x$ si ha

$$(\text{arctg } y)' = \frac{1}{(\text{tg } x)'} = \cos^2 x = \frac{1}{1 + \text{tg}^2 x} = \frac{1}{1 + y^2}$$

dove abbiamo usato l'identità $1 + \text{tg}^2 x = 1/\cos^2 x$. Abbiamo così la regola

$$(\text{arctg } x)' = \frac{1}{1 + x^2}.$$

Provate a dimostrare le regole

$$(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}}, \quad (\arccos x)' = -\frac{1}{\sqrt{1 - x^2}}.$$

ESEMPIO 2.9. L'insieme di definizione di $f(x) = \ln x$ è $\{x > 0\}$. La derivata di $\ln x$ è $1/x$, e l'insieme di definizione di $1/x$ è $\{x \neq 0\}$. Sembra che la derivata di f sia definita su un dominio più grande del dominio di f . Possibile?

La contraddizione è solo apparente: la funzione $g(x) = 1/x$ coincide con la derivata di f sul dominio di f , e quindi sul dominio possiamo dire che $f' = g$; invece, fuori dal dominio di f , la funzione g non è la derivata di f semplicemente perché in quei punti f non è definita e neanche derivabile. In altri termini può accadere che f' si possa *estendere* ad un dominio più grande di quello di f .

Possiamo anche osservare che la funzione $h(x) = \ln |x|$ è definita su tutto $\mathbb{R} \setminus \{0\}$ ed estende la funzione $f(x) = \ln x$ ai valori negativi di x ; abbiamo subito

$$(\ln |x|)' = \frac{1}{x} \quad \text{per tutti gli } x \neq 0.$$

ESEMPIO 2.10. Qualche volta abbiamo a che fare con funzioni la cui definizione non è esplicita ma *implicita*. Ad esempio, consideriamo la curva nel piano definita dalla relazione

$$x^2 + y^2 = 1.$$

Sappiamo che si tratta di una circonferenza di raggio 1 centrata nell'origine. Si può considerare questa curva come il grafico di una funzione? Sì e no. Possiamo ad esempio "ricavare y " ottenendo

$$y = \pm\sqrt{1 - x^2}$$

ma chiaramente abbiamo due possibilità. Allora potremmo dire ad esempio: spezziamo la curva in due semicirconferenze, quella superiore è il grafico di $f(x) = \sqrt{1 - x^2}$, quella inferiore è il grafico di $g(x) = -\sqrt{1 - x^2}$. Date queste espressioni esplicite, possiamo calcolare le derivate f' e g' e procedere con lo studio delle funzioni.

E se non siamo in grado di "ricavare" le funzioni? C'è un modo di fare qualche calcolo anche se f è data in modo implicito. Nell'esempio precedente la funzione $y = f(x) = \sqrt{1 - x^2}$ verifica

$$x^2 + [f(x)]^2 = 1.$$

Se deriviamo questa identità otteniamo la relazione

$$2x + 2f(x)f'(x) = 0$$

da cui

$$f'(x) = -\frac{x}{f(x)}.$$

Ad esempio nel punto della circonferenza di coordinate $(x, f(x)) = (1/\sqrt{2}, 1/\sqrt{2})$ la derivata della funzione f vale

$$f'(1/\sqrt{2}) = -\frac{1/\sqrt{2}}{1/\sqrt{2}} = -1.$$

Vediamo quindi che si può calcolare f' senza esplicitare f .

Un esempio più complesso: consideriamo la curva $y = f(x)$ definita implicitamente dalla relazione

$$x^3 + 2y^3 = 3xy$$

(un *folium di Cartesio*) passante per il punto $(1, 1)$. In questo caso non possiamo esplicitare f (e anzi per essere sicuri che la relazione definisca davvero una funzione serve uno strumento più avanzato: il Teorema del Dini o delle Funzioni Implicite. Per ora diamo questo fatto per dimostrato). Quanto vale $f'(1)$?

ESEMPIO 2.11. Usando i teoremi precedenti è facile dimostrare la formula generale

$$(x^a)' = ax^{a-1}$$

per qualunque esponente reale a . Basta scrivere

$$x^a = (e^{\ln x})^a = e^{a \ln x}$$

e applicare le regole per la derivata dell'esponenziale, di una funzione composta e del logaritmo:

$$(x^a)' = (e^{a \ln x})' = e^{a \ln x} \cdot a(\ln x)' = e^{a \ln x} \cdot \frac{a}{x} = ax^{a-1}.$$

ESERCIZIO 2.12. a) Calcolare la derivata delle funzioni seguenti:

$$\ln(3x+2), \quad \ln(e^x+1), \quad \ln(\sin x), \quad x \ln x, \quad \frac{x}{\ln x},$$

$$\operatorname{arctg}(x^2-1), \quad x \operatorname{arctg} x, \quad \arcsin(2x), \quad \arccos(3x),$$

$$x^x, \quad (\sin x)^x, \quad \frac{x}{\sqrt{1+\ln x}}, \quad \frac{\ln x}{x}.$$

b) Data una funzione $f(x)$, calcolare la derivata di

$$\ln(f(x)), \quad A^{f(x)}, \quad \operatorname{arctg}(f(x)), \quad f(\operatorname{arctg} x), \quad \arcsin(f(x)).$$

c) Sapreste scrivere una regola generale per la derivata di $f(x)^{g(x)}$? (Suggerimento: scrivere $f^g = e^{f \cdot \ln g}$).

OSSERVAZIONE 2.13. Consideriamo la funzione

$$f(x) = \begin{cases} x^4 & \text{se } x \geq 0, \\ x^2 + ax & \text{se } x < 0 \end{cases}$$

e chiediamoci per quali valori del parametro a la funzione è continua e per quali è derivabile. Vediamo che:

$$\lim_{x \rightarrow 0^+} f(x) = 0, \quad \lim_{x \rightarrow 0^-} f(x) = 0$$

qualunque sia il valore di a , quindi la risposta è: la funzione f è continua per qualunque $a \in \mathbb{R}$.

Se invece vogliamo determinare per quali valori di a la funzione è derivabile, osserviamo subito che la domanda riguarda solo l'esistenza di $f'(0)$ dato che negli altri punti f è derivabile. Scriviamo il rapporto incrementale di f in 0; per incrementi h positivi abbiamo subito

$$\frac{f(0+h) - f(0)}{h} = \frac{h^4 - 0}{h} = h^3 \rightarrow 0$$

per $h \rightarrow 0^+$. In altri termini, la derivata destra in 0 esiste e vale 0. Per incrementi negativi si ha invece

$$\frac{f(0+h) - f(0)}{h} = \frac{h^2 + ah - 0}{h} = h^2 + a \rightarrow a$$

per $h \rightarrow 0^-$. La derivata sinistra in 0 quindi vale a , e la funzione è derivabile se e solo se $a = 0$.

ESEMPIO 2.14. ATTENZIONE: questo esempio contiene un errore. Consideriamo la funzione

$$f(x) = \begin{cases} x + 1 & \text{se } x \geq 0, \\ x & \text{se } x < 0. \end{cases}$$

Calcolare la derivata di f è molto semplice: dalle regole elementari di derivazione si ottiene subito

$$f'(x) = \begin{cases} 1 & \text{se } x \geq 0, \\ 1 & \text{se } x < 0 \end{cases}$$

e quindi f è derivabile in tutti i punti con derivata costante uguale a 1. Dov'è l'errore?

ESERCIZIO 2.15. Per ciascuna delle seguenti funzioni, determinare per quali valori dei parametri esse sono continue e per quali valori sono derivabili:

$$f(x) = \begin{cases} e^{x-1} & \text{se } x \geq 0, \\ ax + b & \text{se } x < 0 \end{cases}$$

$$f(x) = \begin{cases} ax^2 \ln x & \text{se } x > 0, \\ b & \text{se } x = 0, \\ ce^{-1/x^2} & \text{se } x < 0 \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{x} \text{sen}(x^2) & \text{se } x > 0, \\ ax^2 + b & \text{se } x \leq 0 \end{cases}$$

$$f(x) = \begin{cases} x^3 - ax & \text{se } x \geq 0, \\ x^2 + bx & \text{se } x < 0 \end{cases}$$

$$f(x) = \begin{cases} x^3 - ax & \text{se } x \geq 1, \\ x^2 + bx & \text{se } x < 1. \end{cases}$$

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

Ricordare che derivabile \implies continua! Quindi è più comodo verificare la continuità prima della derivabilità; se in un punto una funzione è discontinua, allora non c'è bisogno di verificare la derivabilità in quel punto, perché la funzione sicuramente non è neanche derivabile in quel punto.

3. Applicazioni delle derivate.

Le derivate sono uno strumento molto efficace per lo studio del grafico di funzioni. Data un'espressione analitica, vogliamo capire con il minimo sforzo l'*andamento* del grafico corrispondente, ossia la forma della curva che la rappresenta. Quindi ci interessa stabilire i tratti in cui la curva è crescente e decrescente, e i punti in cui il comportamento della curva cambia (le "creste" e le "valli"), in modo da poterla disegnare almeno approssimativamente. Se la funzione che stiamo studiando modella un fenomeno che ci interessa, conoscere il grafico ci permette di capire l'andamento del fenomeno. Questi punti si definiscono come segue:

DEFINIZIONE 3.1. Data una funzione $f : [a, b] \rightarrow \mathbb{R}$ diciamo che un punto $c \in [a, b]$ è di **massimo (relativo)** per f se esiste un $\delta > 0$ tale che

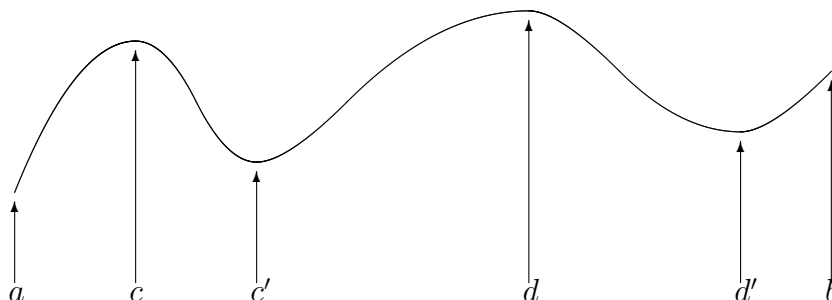
$$f(c) \geq f(x) \quad \text{per tutti gli } x \text{ di } [a, b] \text{ tali che } |x - c| < \delta,$$

ossia il valore $f(c)$ è più grande dei valori $f(x)$ per x vicino a c . I **minimi relativi** di f si definiscono in modo analogo: esiste $\delta > 0$ tale che

$$f(c) \leq f(x) \quad \text{per tutti gli } x \text{ di } [a, b] \text{ tali che } |x - c| < \delta.$$

ESEMPIO 3.2. In particolare, fra i punti studiati nella definizione precedente compaiono anche i punti di massimo e di minimo la cui esistenza è garantita dal Teorema di Weierstrass. Per distinguerli dagli altri li denominiamo massimo e minimo **assoluti** di f su $[a, b]$; e ciò che li contraddistingue è che f vi assume il valore più grande e il più piccolo in confronto a *tutti* gli altri punti dell'intervallo. Notare che in un intervallo $[a, b]$ vi possono essere vari massimi e minimi assoluti, anche infiniti (ad esempio?).

ESEMPIO 3.3. Bisogna fare un po' di attenzione per rendersi conto dei vari casi possibili. Osserviamo il grafico della funzione $f : [a, b] \rightarrow \mathbb{R}$, rappresentato nella figura seguente.



Vediamo che i punti c e d soddisfano la definizione di massimo relativo; come ci aspettiamo, abbiamo due “creste” nel grafico. Analogamente, i punti c' e d' sono minimi relativi e vediamo due “valli”. Ma notiamo che anche il punto a è un minimo relativo, mentre il punto b è un massimo relativo: e in questo caso la curva non ha lo stesso andamento che negli altri casi.

Allora è importante distinguere i massimi o minimi “interni”, cioè che cadono all'interno dell'intervallo, dagli estremi a, b dell'intervallo di definizione.

Il risultato seguente mostra come i potenziali candidati nella nostra ricerca dei massimi e minimi relativi (interni) siano esattamente i punti in cui si annulla la derivata (ossia la retta tangente in quei punti deve essere orizzontale, vedi più avanti):

TEOREMA 3.4. (Teorema di Fermat). Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione, sia c un punto di massimo (o minimo) relativo interno ad $[a, b]$, e supponiamo che f sia derivabile in c . Allora $f'(c) = 0$.

Dimostrazione. Se ad esempio c è un minimo relativo, abbiamo la disuguaglianza

$$f(c + h) \geq f(c)$$

valida per tutti gli h positivi o negativi abbastanza piccoli (vedi definizione). Studiamo il segno del rapporto incrementale in c : se $h > 0$ abbiamo

$$\frac{f(c + h) - f(c)}{h} = \frac{+}{+} \geq 0,$$

ossia il rapporto incrementale è positivo per incrementi positivi; mentre se $h < 0$ abbiamo

$$\frac{f(c+h) - f(c)}{h} = \frac{+}{-} \leq 0,$$

ossia il rapporto incrementale è negativo per incrementi negativi. Passiamo al limite per $h \rightarrow 0$ (il limite esiste per ipotesi). Il limite destro è un numero positivo perché il rapporto è positivo, il limite destro è negativo, ma i due limiti devono coincidere, quindi il limite fa 0.

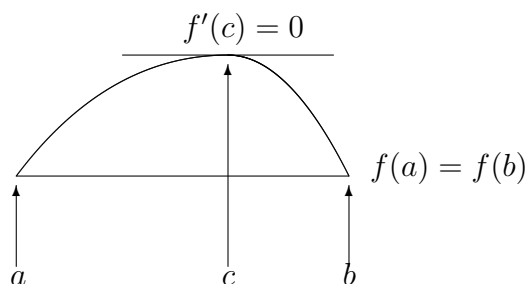
ESERCIZIO 3.5. Consideriamo una funzione con un andamento simile a quello della figura precedente. L'estremo b è un massimo relativo, ma è evidente che la derivata in b non si annulla. Quindi è chiaro che la dimostrazione del Teorema di Fermat non funziona in questo caso. Perché? (Qual è il segno della derivata sinistra in b ?)

OSSERVAZIONE 3.6. Riassumendo: se un punto (interno) è di massimo o minimo relativo, allora la derivata si annulla in quel punto; il viceversa non è vero in generale (ad esempio la funzione $f(x) = x^3$ è strettamente crescente ma ha derivata nulla in $x = 0$). In particolare, per trovare i massimi e minimi relativi di una funzione basta calcolare gli zeri della derivata, e poi esaminarli uno ad uno per distinguere i massimi dai minimi (e quelli che non sono né massimi né minimi).

Ancora un po' di terminologia: un punto c in cui la derivata $f'(c)$ si annulla si dice anche **punto critico** per la funzione f . Un punto critico c tale che $f(x) \geq f(c)$ per $x \geq c$ e $f(x) \leq f(c)$ per $x \leq c$ (o viceversa) si chiama anche **punto di flesso** (a tangente orizzontale). Ad esempio, $f(x) = x^3$ ha un flesso in $x = 0$.

Il Teorema di Fermat afferma semplicemente che un massimo o minimo relativo interno all'intervallo di definizione è necessariamente un punto critico. Viceversa, un punto critico può essere un punto di massimo, o di minimo, o di flesso. Ma vi sono anche altri casi possibili. Sapreste dare un esempio di punto critico che non è né massimo né minimo né flesso? (Basta un disegno).

TEOREMA 3.7. (Teorema di Rolle). Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua, e supponiamo che essa sia derivabile sull'intervallo aperto $]a, b[$. Se $f(a) = f(b)$, allora esiste un punto $c \in]a, b[$ nel quale $f'(c) = 0$.



Dimostrazione. Il Teorema di Weierstrass ci assicura che f assume massimo e minimo su $[a, b]$, quindi esistono due punti c e d in $[a, b]$ tali che

$$f(c) \leq f(x) \leq f(d)$$

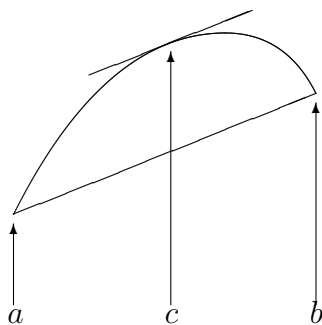
per ogni $x \in [a, b]$. Se almeno uno dei due punti è interno all'intervallo, applicando il Teorema di Fermat otteniamo che la derivata vi si annulla. Se invece sia c che d cadono agli estremi, abbiamo

$$f(c) = f(a) = f(b) = f(d)$$

e questo significa che la funzione f è costante. Allora $f'(x) = 0$ in ogni punto e possiamo scegliere come c un qualunque punto di $]a, b[$.

TEOREMA 3.8. (*Teorema di Lagrange o del Valor Medio*). *Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua, e supponiamo che essa sia derivabile sull'intervallo aperto $]a, b[$. Allora esiste un punto $c \in]a, b[$ nel quale*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$



Dimostrazione. Sia $g(x)$ la funzione definita nel modo seguente:

$$g(x) = f(x) - \frac{x - a}{b - a} \cdot [f(b) - f(a)].$$

Allora g verifica le ipotesi del Teorema di Rolle: è continua sull'intervallo chiuso, derivabile su quello aperto, e inoltre $g(a) = g(b)$ (verificare!). Quindi, per il teorema, esiste un punto c in cui $g'(c) = 0$. Calcoliamo la derivata di g : si ha

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

e quindi se $g'(c) = 0$ si ottiene la tesi.

Il Teorema di Lagrange ci fornisce un utile criterio per stabilire se una funzione è crescente o decrescente, uno degli strumenti più importanti nello studio di funzioni:

TEOREMA 3.9. *Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua, e supponiamo che essa sia derivabile sull'intervallo aperto $]a, b[$. Valgono le seguenti proprietà:*

$$f'(x) \geq 0 \text{ per ogni } x \in]a, b[\text{ se e solo se } f \text{ è crescente su } [a, b];$$

$$f'(x) \leq 0 \text{ per ogni } x \in]a, b[\text{ se e solo se } f \text{ è decrescente su } [a, b];$$

$$f'(x) = 0 \text{ per ogni } x \in]a, b[\text{ se e solo se } f \text{ è costante su } [a, b].$$

Dimostrazione. Consideriamo il primo caso. Vogliamo dimostrare anzitutto che, se $f' \geq 0$, presi due punti qualunque $c < d$ interni ad $[a, b]$, si ha $f(c) \leq f(d)$. Appliciamo il Teorema di Lagrange a f sull'intervallo $[c, d]$; otteniamo che esiste un punto p interno all'intervallo $[c, d]$ nel quale

$$f'(p) = \frac{f(d) - f(c)}{d - c}.$$

Sappiamo che $f'(p) \geq 0$, quindi il rapporto al secondo membro è positivo. Moltiplicando per $d - c (> 0)$ otteniamo

$$f(d) - f(c) \geq 0$$

cioè la tesi.

Mostriamo il viceversa: sia f crescente, consideriamo il suo rapporto incrementale in un qualunque punto x

$$\frac{f(x+h) - f(x)}{h}$$

e notiamo che numeratore e denominatore hanno lo stesso segno per qualunque valore di h . In altri termini, il rapporto incrementale è sempre positivo e ne segue che anche il suo limite, cioè $f'(x)$, deve essere positivo.

La dimostrazione nel caso $f' \leq 0$ è identica. Nel terzo caso, se $f'(x) = 0$ per ogni x allora dal primo caso otteniamo che f è crescente mentre dal secondo caso abbiamo che f è decrescente, e in conclusione f deve essere costante. Viceversa, se f è costante il rapporto incrementale è identicamente nullo e otteniamo $f' = 0$ in ogni punto.

OSSERVAZIONE 3.10. Notiamo che se f è derivabile e strettamente crescente, non è detto che sia $f'(x) > 0$ in tutti i punti. Ad esempio $f(x) = x^3$ è strettamente crescente (verificare) ma $f'(0) = 0$. Però in questo esempio f' si annulla solo in un punto, e in un certo senso questa è la situazione tipica. Più precisamente, consideriamo una funzione $f :]a, b[\rightarrow \mathbb{R}$ derivabile; allora f è *strettamente crescente* $\iff f'(x) \geq 0$ in ogni punto, e non succede mai che f' ristretta ad un intervallo sia la funzione nulla. Risultato analogo per funzioni decrescenti.

Sapreste dimostrarlo? (Sia in un verso che nell'altro sappiamo già che f è crescente, quindi dire che f non è strettamente crescente equivale a dire che è costante su un certo tratto....)

OSSERVAZIONE 3.11. Supponiamo che $f : [a, b] \rightarrow \mathbb{R}$ sia continua su $[a, b]$, derivabile su (a, b) e che la derivata di f sia limitata, cioè esista una costante C tale che $|f'(x)| \leq C$ per tutti gli x . Allora f è lipschitziana di costante C . Infatti presi due punti x, y qualsiasi il Teorema di Lagrange applicato all'intervallo di estremi x, y garantisce l'esistenza di un punto c tale che

$$\frac{f(x) - f(y)}{x - y} = f'(c);$$

ne segue

$$\left| \frac{f(x) - f(y)}{x - y} \right| = |f'(c)| \leq C$$

da cui

$$|f(x) - f(y)| \leq C|x - y|.$$

ESEMPIO 3.12. Consideriamo una funzione $f : A \rightarrow \mathbb{R}$ definita come segue: A è l'unione dei due intervalli $(1, 2)$ e $(3, 4)$, e f è data da

$$f(x) = \begin{cases} 1 & \text{se } x \in (1, 2), \\ 2 & \text{se } x \in (3, 4). \end{cases}$$

La derivata di f è nulla in tutti i punti, come si verifica immediatamente. Ma f non è costante. Perché non c'è contraddizione con il Teorema precedente?

ESEMPIO 3.13. Considerate con attenzione i due esempi seguenti.

La derivata di $f(x) = \arcsin x + \arccos x$ è

$$f' = \frac{1}{\sqrt{1-x^2}} - \frac{1}{\sqrt{1-x^2}} = 0,$$

qualunque sia x . Ne segue che f è costante: $f = C$. Quanto vale C ? (Calcolare ad esempio $f(0)$).

La derivata di $g(x) = \operatorname{arctg} x + \operatorname{arctg} \left(\frac{1}{x}\right)$ è

$$g' = \frac{1}{1+x^2} - \frac{1}{x^2} \frac{1}{1+\frac{1}{x^2}} = 0.$$

Possiamo dire anche in questo caso che g è costante? Provate ad esempio a calcolare $g(1)$ e $g(-1)$. Che succede?

ESEMPIO 3.14. Possiamo finalmente verificare in modo semplice alcune proprietà delle funzioni elementari: ad esempio e^x , $\ln x$, $\arcsin x$, $\operatorname{arctg} x$ sono strettamente crescenti (derivata strettamente positiva), $\arccos x$ è strettamente decrescente, e così via.

Ad esempio studiamo $f(x) = A^x$, $A > 0$:

$$f'(x) = A^x \ln A$$

quindi A^x è strettamente crescente se $\ln A > 0$ cioè $A > 1$, strettamente decrescente se $A < 1$.

ESEMPIO 3.15. Più in generale lo studio degli intervalli di crescita e decrescenza di una funzione è, assieme allo studio dei massimi e minimi relativi, uno dei mezzi più efficaci per tracciarne il grafico. Studiamo ad esempio l'andamento della funzione

$$f(x) = 3x^3 - 4x.$$

La funzione è ben definita e continua per tutti gli $x \in \mathbb{R}$. La sua immagine è un intervallo, e dato che

$$\lim_{x \rightarrow \pm\infty} f(x) = \pm\infty,$$

abbiamo subito che l'immagine di f è tutto \mathbb{R} . La derivata

$$f' = 9x^2 - 4$$

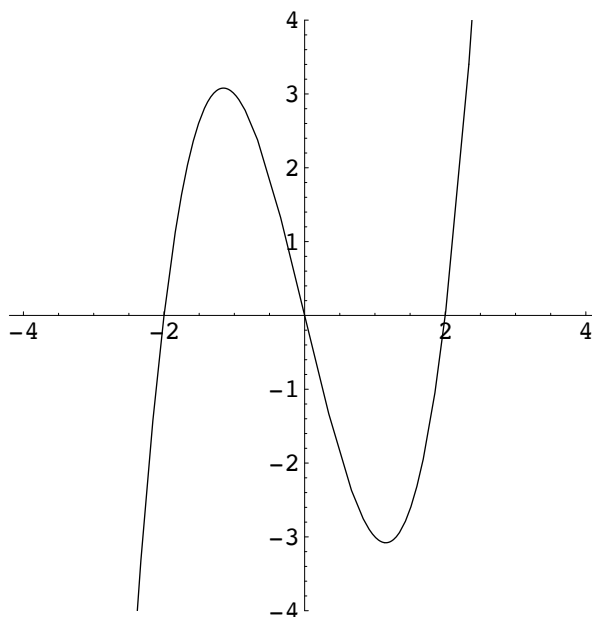
si annulla per $x = 3/2$ e $x = -3/2$. Questi sono gli unici punti che possono essere massimi o minimi relativi di f (tutti i punti del dominio sono interni al dominio). Come fare a distinguere i massimi dai minimi? Il modo più generale (e spesso anche il più semplice) è determinare gli intervalli di crescita e decrescenza di f . Ad esempio se in c si ha $f'(c) = 0$, e la funzione f cresce per $x < c$ e decresce per $x > c$, necessariamente c è un punto di massimo. Nel nostro esempio

$$f'(x) = 9x^2 - 4 \geq 0 \iff x \geq \frac{3}{2} \text{ oppure } x \leq -\frac{3}{2},$$

mentre

$$f'(x) = 9x^2 - 4 \leq 0 \iff -\frac{3}{2} \leq x \leq \frac{3}{2}.$$

Pertanto, $x = 3/2$ è un punto di minimo e $x = -3/2$ un punto di massimo. Abbiamo elementi sufficienti per disegnare il grafico (dopo aver calcolato quanto vale f in $x = \pm 3/2$ in modo da collocare nel grafico il punto di minimo e di massimo).



OSSERVAZIONE 3.16. (Studio di funzioni). Riassumiamo i passi principali per impostare lo studio del grafico di una funzione $f(x)$:

1) Stabilire l'insieme di definizione. Nella quasi totalità dei casi si tratta di una unione di intervalli. Stabilire l'insieme dei punti in cui f è continua (l'**insieme di continuità** di f). Per una funzione ottenuta come combinazione di funzioni elementari, vi saranno solo alcuni punti singolari, di solito in numero finito, e in tutti gli altri punti f sarà continua.

2) Stabilire eventuali simmetrie di $f(x)$: se è **pari**, ossia

$$f(-x) = f(x)$$

e in questo caso il grafico è simmetrico rispetto all'asse delle ordinate; oppure se è **dispari**, ossia

$$f(-x) = -f(x)$$

e in tal caso il grafico è simmetrico rispetto all'origine degli assi. Può essere inoltre molto utile (ma non sempre facile) stabilire gli eventuali zeri di f e gli insiemi dove $f > 0$ e $f < 0$.

3) Calcolare i limiti di f agli estremi degli intervalli di definizione (che possono essere anche i limiti per $x \rightarrow \pm\infty$). Questo consente di determinare gli eventuali **asintoti verticali**, ossia i punti $c \in \mathbb{R}$ in cui f tende a $\pm\infty$, o se la funzione abbia un **asintoto orizzontale**, termine usato per indicare che esiste finito il limite di f per $x \rightarrow +\infty$ (o $x \rightarrow -\infty$).

4) Determinare l'insieme dei punti in cui la funzione è derivabile (l'**insieme di derivabilità** di f) e naturalmente calcolare la derivata f' . Valgono le stesse osservazioni che per l'insieme di continuità.

5) Determinare i punti critici $f'(c) = 0$. Tra questi si nascondono i massimi e minimi relativi di f ; per determinarli il modo più semplice è proseguire lo studio del segno di f' e determinare gli intervalli su cui $f' > 0$ (intervalli di crescita) e quelli su cui $f' < 0$ (intervalli di decrescenza). Ora sia c un punto critico: se f è crescente per $x < c$ e decrescente per $x > c$, il punto deve essere un punto di massimo; se f è decrescente per $x < c$ e crescente per $x > c$, il punto deve essere un punto di minimo; se cresce (o decresce) sia prima che dopo, c è un flesso.

5) Gli elementi precedenti consentono quasi sempre di individuare l'andamento del grafico. Se fosse necessario uno studio più approfondito si può ricorrere alle nozioni del paragrafo seguente e determinare gli intervalli di convessità e concavità di f , ossia gli intervalli in cui f'' è positiva o negativa. Un altro elemento utile è la presenza di **asintoti obliqui**: diciamo che $f(x)$ per $x \rightarrow +\infty$ è asintotica alla retta $ax + b$ se si ha

$$\lim_{x \rightarrow +\infty} [f(x) - (ax + b)] = 0$$

e una definizione analoga si ha per $x \rightarrow -\infty$. È facile ricavare i coefficienti a e b : dividendo per x si ottiene subito

$$\lim_{x \rightarrow +\infty} \left[\frac{f(x)}{x} - \left(a + \frac{b}{x} \right) \right] = 0$$

da cui

$$a = \lim_{x \rightarrow +\infty} \frac{f(x)}{x}.$$

Chiaramente, dopo aver ricavato a , il valore di b si ottiene subito come

$$b = \lim_{x \rightarrow +\infty} [f(x) - ax].$$

Notare che per dimostrare l'esistenza di un asintoto obliquo non è sufficiente dimostrare che $\frac{f(x)}{x}$ tende ad un limite finito a (ad esempio $f(x) = \ln x$ soddisfa $f/x \rightarrow 0$ ma non ha un asintoto); bisogna dimostrare che $f - (ax + b) \rightarrow 0$, e quindi, seguendo il metodo precedente, si devono calcolare sia a che b ottenendo due limiti finiti.

ESEMPIO 3.17. Studio della funzione

$$f(x) = \begin{cases} \frac{xe^{1/x}}{1 + e^{1/x}} & \text{se } x \neq 0, \\ 0 & \text{se } x = 0. \end{cases}$$

La funzione è definita su tutto \mathbb{R} e non ha simmetrie. Vediamo subito che $f > 0$ per $x > 0$ e $f < 0$ per $x < 0$, e l'unico zero di f è $x = 0$. Sicuramente continua e derivabile per $x \neq 0$; controlliamo in 0. Scrivendo

$$\frac{xe^{1/x}}{1 + e^{1/x}} = \frac{x}{e^{-1/x} + 1}$$

otteniamo facilmente

$$\lim_{x \rightarrow 0^+} f(x) = \lim_{x \rightarrow 0^-} f(x) = 0$$

e quindi la funzione è continua anche in 0. Per la derivabilità in 0 scriviamo il rapporto incrementale

$$\frac{f(h+0) - f(0)}{h} = \frac{e^{1/h}}{1 + e^{1/h}}$$

e otteniamo che il limite destro in 0 vale 1, il limite sinistro vale 0, quindi f ha derivata destra e sinistra in 0 diverse: non è derivabile in 0. Per $x \neq 0$ invece possiamo scrivere

$$f' = e^{1/x} \frac{1 + e^{1/x} - 1/x}{(1 + e^{1/x})^2}$$

che è sempre positiva (basta usare la disuguaglianza $e^x \geq 1 + x$) quindi f è sempre crescente. I limiti all'infinito sono

$$\lim_{x \rightarrow \pm\infty} f(x) = \pm\infty.$$

Quindi non ci sono asintoti orizzontali o verticali. Osservando che

$$f(x) = \frac{e^{1/x}}{1 + e^{1/x}} \cdot x \sim \frac{1}{2} \cdot x$$

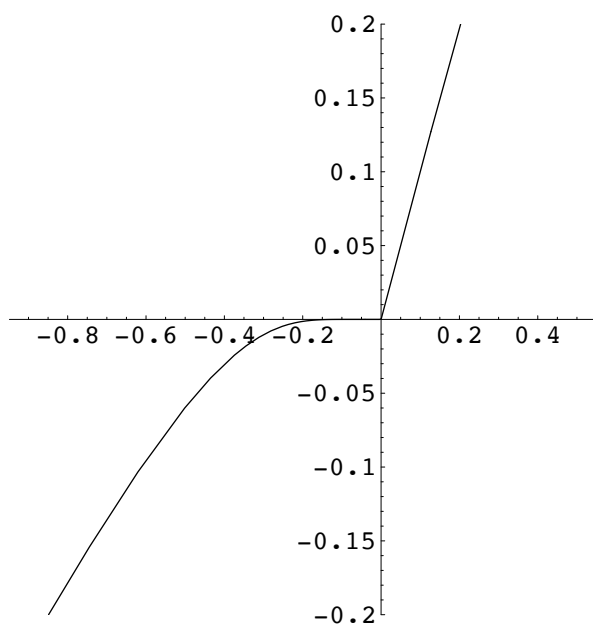
per x grande, sorge il sospetto che vi sia un asintoto obliquo $ax + b$ per $x \rightarrow +\infty$. Infatti per $x \rightarrow +\infty$

$$\frac{f(x)}{x} = \frac{e^{1/x}}{1 + e^{1/x}} \rightarrow \frac{1}{2}$$

da cui $a = 1/2$, e

$$f(x) - ax = \frac{x e^{1/x} - 1}{2 e^{1/x} + 1} = \frac{1}{2} \frac{e^{1/x} - 1}{1/x} \frac{1}{e^{1/x} + 1} \rightarrow \frac{1}{4}$$

sempre per $x \rightarrow +\infty$, da cui $b = 1/4$. Anche per $x \rightarrow -\infty$ c'è un asintoto obliquo, determinarlo per esercizio.



ESERCIZIO 3.18. Studiare le seguenti funzioni determinando quanti più elementi possibile e tracciandone un grafico approssimativo. Più precisamente: determinare gli insiemi di definizione, di continuità e derivabilità; il segno della funzione; i limiti agli estremi degli intervalli che compongono l'insieme di definizione; l'eventuale presenza di asintoti orizzontali, verticali o obliqui; calcolare la derivata; determinare gli zeri della derivata, gli intervalli di crescita e decrescenza, e i massimi e minimi relativi (e assoluti). Con questi elementi si può tracciare il grafico della funzione; in qualche caso informazioni più precise si ottengono calcolando la derivata seconda e trovando gli intervalli di convessità (derivata seconda positiva) e concavità (derivata seconda negativa); per quest'ultimo punto si veda il paragrafo successivo.

$$\begin{aligned} & \frac{(x-3)^2}{4(x-1)}, \quad \frac{x}{2x^2-1}, \quad x + \frac{1}{x}, \quad \frac{x^2-2x-3}{x}, \quad \frac{1}{x} + \frac{1}{x+1} + \frac{1}{x+2}, \\ & x \ln x, \quad x \ln |x|, \quad x + \ln(\sqrt{|x|} + 3), \quad \frac{x}{\ln x}, \quad \frac{\ln x}{x}, \quad \frac{(\ln x)^2}{x} \\ & x e^{-1/x}, \quad (2-x)e^{1/(x-2)}, \quad \frac{e^{-1/x}}{x}, \quad e^{\frac{x+1}{x-2}}, \quad x e^{-x^2}, \quad \frac{x}{e^{|x|}}, \quad \frac{e^{1/x} - 1}{e^{1/x} + 1}, \\ & x^x, \quad |x|^x, \quad x^{\ln x}, \quad x(\ln |x| - 1), \quad x^2(\ln |x| - 1), \quad x^{\sqrt{x}}, \quad x^{(\ln x)^2} \end{aligned}$$

$$\begin{aligned} &\sqrt{2-3x^2}, \quad x - \sqrt{5-x}, \quad x - \sqrt{2x-x^2}, \quad \sqrt{\frac{x}{x-1}}, \quad \sqrt{x} - \sqrt{3-x}, \\ &\operatorname{sen} x - \cos x, \quad \operatorname{tg} x + \cos x, \quad \operatorname{arctg} \left(\frac{1}{x} \right), \quad \operatorname{arctg} \left(x - \frac{1}{x} \right), \\ &x - \operatorname{arctg} \sqrt{1-x}, \quad \operatorname{arctg} (1 - e^{-x}), \quad \operatorname{arctg} \left(\frac{\ln |x|}{x} \right), \quad \frac{1}{\operatorname{arctg} x}, \\ &x^2 - 4 - \ln(2|x|), \quad \frac{e^{1-x^2}}{2-x}, \quad (1-x-x^2)e^{|x-1|}, \quad |x-1| - |x-4|, \\ &x^{1/3}, \quad x^{-2/3}, \quad 2x + 3x^{2/3}, \quad \sqrt[3]{x^2 - 6x}, \quad x - [x], \quad \frac{1}{[x]}, \quad \left[\frac{1}{x} \right], \quad [|x|]. \end{aligned}$$

ESERCIZIO 3.19. Una applicazione molto utile delle derivate è per la dimostrazione di alcune disuguaglianze. Ad esempio, per dimostrare che $e^x \geq 1 + x$ per qualunque x , è sufficiente dimostrare che la funzione

$$f(x) = e^x - (1 + x)$$

è sempre positiva. Abbiamo

$$f'(x) = e^x - 1,$$

quindi f è crescente per $x \geq 0$; dato che $f(0) = 0$, otteniamo subito $f(x) \geq 0$ per ogni $x \geq 0$. Inoltre vediamo che f è decrescente per $x \leq 0$. In altri termini, $x = 0$ è un punto di minimo della funzione f . Dato che $f(0) = 0$, abbiamo dimostrato che f è sempre positiva.

Usando idee simili, dimostrare le disuguaglianze

$$\begin{aligned} \ln x &\leq x - 1 \quad \forall x > 0, \\ \operatorname{sen} x &\leq x \quad \forall x \in [0, \pi/2], \\ \operatorname{sen} x &\geq \frac{2}{\pi}x \quad \forall x \in [0, \pi/2], \\ \cos x &\geq 1 - \frac{x^2}{2} \quad \forall x \in [-\pi/2, \pi/2]. \end{aligned}$$

ESERCIZIO 3.20. Studiare la funzione

$$f(x) = \frac{\operatorname{sen} x}{x}$$

e determinarne il grafico. Notare che f non è definita su tutto \mathbb{R} . È possibile estendere f a una funzione continua definita su tutto \mathbb{R} ? E a una funzione derivabile definita su tutto \mathbb{R} ?

Stesse domande per la funzione

$$f(x) = \begin{cases} x \ln x & \text{per } x > 0, \\ x^2 - x & \text{per } x < 0 \end{cases}$$

e per la funzione

$$f(x) = \begin{cases} x^2 \ln x & \text{per } x > 0, \\ e^{1/x} & \text{per } x < 0. \end{cases}$$

OSSERVAZIONE 3.21. Quando si studiano funzioni definite a tratti, il seguente criterio è molto utile per stabilire se una funzione è derivabile in un punto “cattivo” circondato da punti “buoni”:

Sia $x_0 \in (a, b)$ e $f : (a, b) \rightarrow \mathbb{R}$ una funzione continua su (a, b) e derivabile su (a, x_0) e su (x_0, b) . Supponiamo che esista il limite

$$\lim_{x \rightarrow x_0} f'(x) = L.$$

Se L è finito, allora f è derivabile anche in x_0 e $f'(x_0) = L$. Se invece L è uguale a $+\infty$ o $-\infty$, allora $f(x)$ non è derivabile in x_0 .

La dimostrazione è semplicissima: scriviamo il rapporto incrementale di f e applichiamo Lagrange sull'intervallo fra x_0 e $x_0 + h$:

$$\frac{f(x_0 + h) - f(x_0)}{h} = f'(y)$$

dove y è un certo punto compreso fra x_0 e $x_0 + h$. Dato che quando $h \rightarrow 0$ si ha anche $y \rightarrow x$, possiamo scrivere

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{y \rightarrow x} f'(y) = L$$

e da qui otteniamo subito la tesi.

Lo stesso argomento ci dice che se esiste finito il limite destro (o sinistro) di f' in x_0 allora esso coincide con la derivata destra (o sinistra) di f in x_0 .

Esempio: la funzione

$$f(x) = \begin{cases} e^x - 1 & x > 0, \\ 0 & x = 0, \\ x - x^2 & x < 0. \end{cases}$$

è continua su tutto \mathbb{R} (perché?), inoltre è chiaro che

$$f'(x) = e^x \text{ per } x > 0, \quad f'(x) = 1 - 2x \text{ per } x < 0.$$

Vediamo che limite destro e sinistro di f' in 0 esistono e sono uguali a 1. Ma allora per il criterio appena visto la funzione è derivabile anche in $x = 0$ e la derivata è uguale al limite cioè $f'(0) = 1$.

4. Convessità e derivate di ordine superiore.

Quasi tutte le funzioni che abbiamo studiato hanno una proprietà molto forte, che finora non abbiamo sfruttato: sono *convesse* o *concave* a tratti. Introduciamo il concetto di convessità.

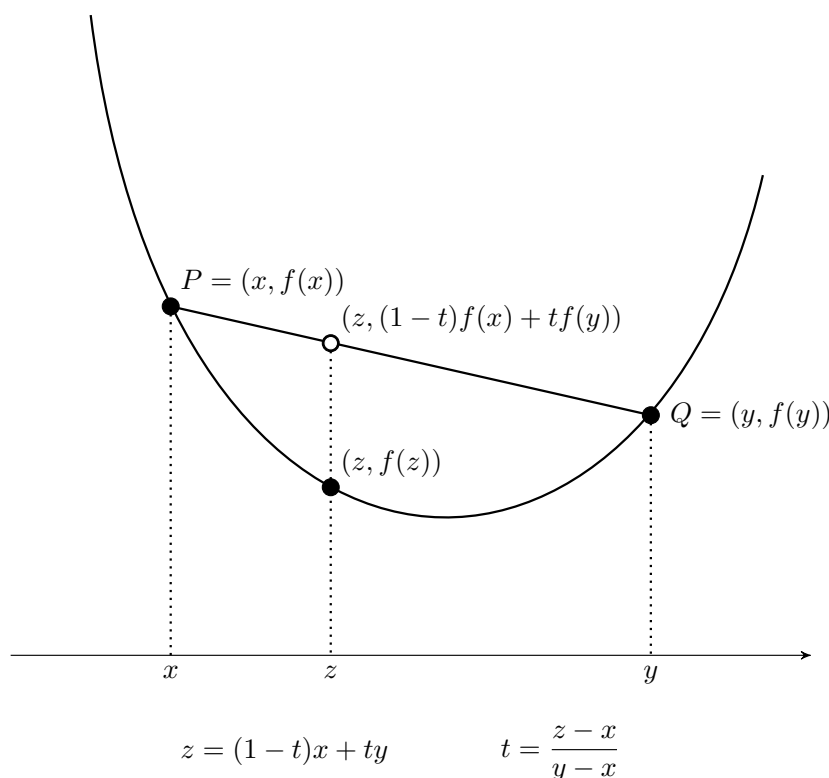
DEFINIZIONE 4.1. Una funzione $f : [a, b] \rightarrow \mathbb{R}$ si dice **convessa** se per ogni coppia di punti $x, y \in [a, b]$ e ogni $t \in [0, 1]$ vale la disuguaglianza

$$(4.1) \quad f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

f si dice **concava** se $-f$ è convessa.

OSSERVAZIONE 4.2. Per interpretare geometricamente la definizione di convessità, consideriamo due punti $x < y$ nel dominio di f e i corrispondenti punti del grafico di f ossia $P = (x, f(x))$ e $Q = (y, f(y))$. Il segmento che unisce P e Q si chiama anche la *corda del grafico di f fra x e y* . È facile descriverla analiticamente: infatti la retta passante per P e Q ha equazione parametrica $\phi(t) = (1-t)(x, f(x)) + t(y, f(y))$ ossia

$$\phi(t) = ((1-t)x + ty, (1-t)f(x) + tf(y)) \quad t \in \mathbb{R},$$



e quindi i punti della corda si ottengono per $0 \leq t \leq 1$. Il numero

$$z = (1-t)x + ty, \quad 0 \leq t \leq 1$$

è un qualunque numero compreso fra x e y , e la definizione (4.1) dice semplicemente che: dato un qualunque punto z intermedio fra x e y , il punto $(z, f(z))$ del grafico di f si trova sempre *al di sotto* della corda fra x e y .

OSSERVAZIONE 4.3. La definizione di convessità si può enunciare anche in un altro modo equivalente. Ricavando il valore di t da $z = (1-t)x + ty$ si ottiene subito

$$t = \frac{z-x}{y-x}, \quad 1-t = \frac{y-z}{y-x}$$

e quindi (4.1) equivale a: per tutti i punti $x < z < y$ nel dominio di f si ha

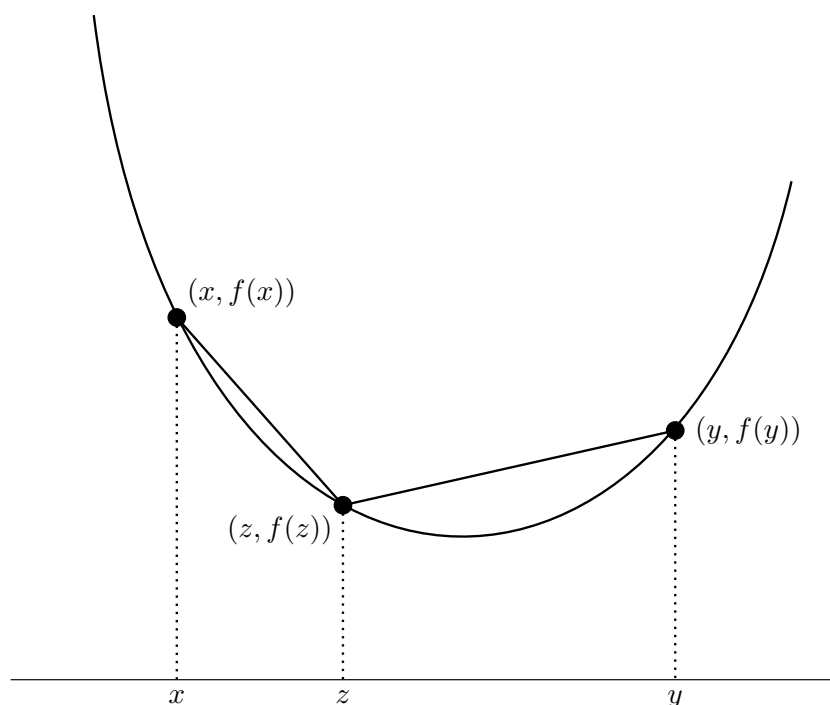
$$f(z) \leq \frac{y-z}{y-x}f(x) + \frac{z-x}{y-x}f(y).$$

Moltiplicando per $y-x$ e semplificando, questa disuguaglianza si scrive anche così: per tutti i punti $x < z < y$ nel dominio di f si ha

$$(4.2) \quad \frac{f(z) - f(x)}{z-x} \leq \frac{f(y) - f(z)}{y-z}.$$

Dato che questi due rapporti esprimono la pendenza (coefficiente angolare) della corda fra x e z e della corda fra z e y , un altro modo geometrico di definire la convessità è il seguente: *una funzione $f : (a, b) \rightarrow \mathbb{R}$ è convessa se e solo se, dati tre punti $x < z < y$ qualunque in (a, b) , la pendenza della corda fra x e z è minore o uguale alla pendenza della corda fra z e y .*

Se f è derivabile, si può dare una caratterizzazione alternativa molto semplice delle funzioni convesse:



$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(z)}{y - z}$$

TEOREMA 4.4. Sia $f : [a, b] \rightarrow \mathbb{R}$ continua su $[a, b]$ e derivabile su $]a, b[$. Allora f è convessa se e solo se f' è crescente.

Dimostrazione. Sia f convessa. Prendiamo x, y tali che $a < x < y < b$, e notiamo che se $h > 0$ è abbastanza piccolo si ha anche $a < x < x + h < y < y + h < b$. Consideriamo le tre corde del grafico di f da x a $x + h$, da $x + h$ a y , e da y a $y + h$; per quanto visto sopra, la pendenza della prima corda è minore o uguale alla pendenza (della seconda corda e quindi) della terza corda:

$$\frac{f(x + h) - f(x)}{h} \leq \frac{f(y + h) - f(y)}{h}.$$

Se ora mandiamo $h \rightarrow 0^+$ in questa disuguaglianza otteniamo $f'(x) \leq f'(y)$ ossia f' è crescente.

Viceversa, sia f' crescente. Presi x, z, y con $a \leq x < z < y \leq b$, vogliamo dimostrare la disuguaglianza (4.2) ossia

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(z)}{y - z}.$$

Per il Teorema di Lagrange, esiste un punto $c \in (x, z)$ tale che il primo rapporto incrementale è uguale a $f'(c)$, ed esiste anche un punto $d \in (z, y)$ tale che il secondo rapporto incrementale è uguale a $f'(d)$; ma essendo $c < d$ si ha $f'(c) \leq f'(d)$ e quindi la disuguaglianza è dimostrata.

Una ulteriore caratterizzazione si ha utilizzando la derivata seconda di f . Introduciamo il concetto di derivate di ordine superiore:

DEFINIZIONE 4.5. Sia $f :]a, b[\rightarrow \mathbb{R}$ derivabile, e consideriamo la funzione derivata $f' :]a, b[\rightarrow \mathbb{R}$. Se f' è a sua volta derivabile, diciamo che f è **due volte derivabile**, e

chiamiamo la derivata di f' la **derivata seconda** di f , indicata con f'' oppure D^2f o anche con $\frac{d^2f}{dx^2}$.

È chiaro che si può procedere allo stesso modo e, quando possibile, definire la derivata terza f''' , la derivata quarta $f^{(4)}$, e in generale la derivata n -esima $f^{(n)}$ (notare il cambiamento di notazione a partire dalla derivata quarta).

TEOREMA 4.6. *Sia $f : [a, b] \rightarrow \mathbb{R}$ continua; supponiamo inoltre che essa sia due volte derivabile su $]a, b[$. Allora f è convessa se e solo se $f''(x) \geq 0$ per ogni $x \in]a, b[$.*

Dimostrazione. Dal Teorema precedente abbiamo subito: f convessa se e solo se f' crescente se e solo se $(f')' = f'' \geq 0$.

OSSERVAZIONE 4.7. Per una funzione f derivabile, esiste una seconda proprietà equivalente alla convessità: f è convessa se e solo se il grafico di f giace tutto al di sopra delle rette tangenti al grafico. In formule, se c è un punto qualunque del dominio, si deve avere per tutti gli x

$$f(x) \geq f(c) + f'(c)(x - c).$$

Geometricamente la proprietà è evidente; ma sapreste dimostrarla?

ESERCIZIO 4.8. a) Calcolare la derivata seconda e terza delle funzioni

$$\begin{aligned} &x^2, \quad x^3 + x^5, \quad x^n; \\ &\text{sen } x, \quad \cos x, \quad \text{tg } x; \\ &\frac{1}{x}, \quad \ln x, \quad e^x, \quad x^{2e^x}. \end{aligned}$$

Calcolare la derivata centesima di x^{100} . E la derivata n -esima di x^n ? e quella $(n + 1)$ -esima di x^n ? E la derivata centesima di $\text{sen } x$?

b) Trovare gli intervalli di convessità e concavità delle funzioni

$$x^3 - 2x, \quad x \ln x, \quad x^4 - x^3 - 3x^2, \quad e^{-x^2}.$$

c) Sia f due volte derivabile con f, f', f'' continue. Se in un punto c si ha $f'(c) = 0$, $f''(c) > 0$ allora c è un punto di minimo. (Per la permanenza del segno, $f''(x) > 0$ vicino a c , ossia f è convessa; ma allora per l'Osservazione 4.7...). Trovare una condizione che garantisca l'esistenza di un massimo.

d) Sia $f : [a, b] \rightarrow \mathbb{R}$ continua e convessa. Dov'è il suo punto di massimo?

e) Imitando la dimostrazione della formula di Newton per la potenza di un binomio, dimostrare la *formula di Leibniz* per la derivata n -esima di un prodotto:

$$(fg)^{(n)} = \sum_{k=0}^n \binom{n}{k} f^{(k)} g^{(n-k)}.$$

5. Il Teorema di de l'Hospital-Bernoulli

Uno strumento molto potente per calcolare i limiti di forme indeterminate del tipo $0/0$ e ∞/∞ (e anche di altro tipo) è il Teorema di de l'Hospital. Premettiamo un risultato che ci servirà per dimostrarlo:

TEOREMA 5.1. (Cauchy) *Siano $f, g : [a, b] \rightarrow \mathbb{R}$ due funzioni continue su $[a, b]$ e derivabili su (a, b) . Supponiamo inoltre che $g'(x) \neq 0$ per ogni $x \in (a, b)$. Allora esiste un punto $c \in (a, b)$ tale che*

$$(5.1) \quad \frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}.$$

OSSERVAZIONE 5.2. Notare che nella formula (5.1) il secondo membro sicuramente è ben definito, infatti $g'(x) \neq 0$ per ipotesi e la divisione si può fare. E a primo membro, chi ci assicura che il denominatore sia diverso da 0? Facile: se fosse $g(b) = g(a)$, per il Teorema di Rolle dovrebbe esistere un punto dentro l'intervallo (a, b) in cui g' si annulla, ma per ipotesi sappiamo che questo non accade, e quindi possiamo essere sicuri che $g(b) - g(a) \neq 0$.

Dimostrazione. Consideriamo la funzione

$$h(x) = f(x)[g(b) - g(a)] - g(x)[f(b) - f(a)].$$

Verifichiamo subito che h soddisfa tutte le ipotesi di Rolle: è continua su $[a, b]$ (somma di funzioni continue moltiplicate per costanti), è derivabile su (a, b) (stesso motivo), e inoltre $h(a) = h(b)$: infatti

$$h(a) = f(a)[g(b) - g(a)] - g(a)[f(b) - f(a)] = f(a)g(b) - g(a)f(b),$$

$$h(b) = f(b)[g(b) - g(a)] - g(b)[f(b) - f(a)] = f(a)g(b) - g(a)f(b).$$

Ma allora sappiamo che c'è un punto $c \in (a, b)$ tale che $h'(c) = 0$, ossia

$$h'(c) = f'(c)[g(b) - g(a)] - g'(c)[f(b) - f(a)] = 0$$

da cui segue subito la formula (5.1).

Passiamo ora al risultato principale. Supponiamo di dover calcolare il limite di una forma indeterminata del tipo $0/0$ (vedi più avanti per il caso ∞/∞), ossia un limite di un rapporto $f(x)/g(x)$ quando sia f sia g tendono a 0: ad esempio,

$$\frac{f(x)}{g(x)} = \frac{e^x - 1}{\sin x}$$

per $x \rightarrow 0$. Consideriamo invece il rapporto delle *derivate* di f e g :

$$\frac{f'(x)}{g'(x)} = \frac{(e^x - 1)'}{(\sin x)'} = \frac{e^x}{\cos x}.$$

Il Teorema di de l'Hospital afferma che, *se questo secondo limite esiste, allora esiste anche quello di partenza f/g e il risultato è lo stesso*:

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{\sin x} = \lim_{x \rightarrow 0} \frac{(e^x - 1)'}{(\sin x)'} = \lim_{x \rightarrow 0} \frac{e^x}{\cos x} = 1.$$

Fate molta attenzione: questa tecnica si può applicare esclusivamente nel caso di forme indeterminate $0/0$ (o ∞/∞), altrimenti si ottengono risultati assurdi (e sbagliati): considerate ad esempio i limiti

$$\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow 0} \frac{\sin x}{e^x} = 0, \quad \lim_{x \rightarrow 0} \frac{f'(x)}{g'(x)} = \lim_{x \rightarrow 0} \frac{\cos x}{e^x} = 1.$$

TEOREMA 5.3. (*de l'Hospital*) Siano $f, g: [c, b] \rightarrow \mathbb{R}$ due funzioni continue tali che

$$f(c) = g(c) = 0.$$

Inoltre supponiamo che f e g siano derivabili su (c, b) e che $g(x) \neq 0$, $g'(x) \neq 0$ per ogni $x \in (c, b)$. Infine supponiamo che esista il limite (finito o infinito)

$$\lim_{x \rightarrow c^+} \frac{f'(x)}{g'(x)}.$$

Allora esiste anche il limite di f/g per $x \rightarrow c^+$, e i due limiti coincidono:

$$\lim_{x \rightarrow c^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c^+} \frac{f'(x)}{g'(x)}.$$

Dimostrazione. Basta applicare in modo astuto il teorema precedente. Sia a_n una qualunque successione dentro (c, b) che converge a c . Dato che $f(c) = g(c) = 0$ possiamo scrivere

$$\frac{f(a_n)}{g(a_n)} = \frac{f(a_n) - f(c)}{g(a_n) - g(c)}$$

e poi possiamo applicare il Teorema di Cauchy sull'intervallo $[c, a_n]$; otteniamo che esiste un punto in (c, a_n) , che chiamiamo b_n (dipende da n), con la proprietà

$$(5.2) \quad \frac{f(a_n)}{g(a_n)} = \frac{f(a_n) - f(c)}{g(a_n) - g(c)} = \frac{f'(b_n)}{g'(b_n)}.$$

Notare che la successione b_n è compresa fra c e a_n , quindi anche $b_n \rightarrow c^+$.

Ora usiamo l'ipotesi che f'/g' converge ad un certo limite L (oppure $\pm\infty$) per $x \rightarrow c^+$; per il Teorema ponte, anche $f'(b_n)/g'(b_n)$ converge allo stesso limite, e dalla formula (5.2) vediamo che anche $f(a_n)/g(a_n)$ converge allo stesso limite. Dato che la successione a_n è qualunque, sempre per il Teorema ponte possiamo concludere che anche f/g tende allo stesso limite per $x \rightarrow c^+$.

OSSERVAZIONE 5.4. È ovvio che il teorema precedente vale anche se $x \rightarrow c^-$, e quindi anche se studiamo limiti per $x \rightarrow c$. Ma vi sono numerose altre varianti, tutte piuttosto utili per calcolare i limiti di forme indeterminate di vario tipo. Vediamone in dettaglio una nella quale si studiano limiti del tipo $x \rightarrow +\infty$ (o $x \rightarrow -\infty$, è lo stesso).

TEOREMA 5.5. (de l'Hospital, versione 2) Siano $f, g : [a, +\infty) \rightarrow \mathbb{R}$ due funzioni continue tali che

$$\lim_{x \rightarrow +\infty} f(x) = \lim_{x \rightarrow +\infty} g(x) = 0.$$

Inoltre supponiamo che f e g siano derivabili su $(a, +\infty)$ e che $g(x) \neq 0$, $g'(x) \neq 0$ per ogni $x \in (a, +\infty)$. Infine supponiamo che esista il limite (finito o infinito)

$$\lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)}.$$

Allora esiste anche il limite di f/g per $x \rightarrow +\infty$, e i due limiti coincidono:

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)}.$$

Dimostrazione. La maniera più semplice di dimostrare il teorema è ricondursi alla versione precedente. Anzitutto notiamo che per calcolare i limiti quando $x \rightarrow +\infty$, è sufficiente utilizzare i valori di f e g per $x > 0$; quindi possiamo tranquillamente supporre $a > 0$. Poniamo $A = 1/a$ e definiamo due nuove funzioni $F, G : [0, A] \rightarrow \mathbb{R}$ come segue:

$$F(x) = \begin{cases} f\left(\frac{1}{x}\right) & \text{se } 0 < x \leq A, \\ 0 & \text{se } x = 0, \end{cases} \quad G(x) = \begin{cases} g\left(\frac{1}{x}\right) & \text{se } 0 < x \leq A, \\ 0 & \text{se } x = 0. \end{cases}$$

Notare che F è continua su $[0, A]$; questo è evidente nei punti $x > 0$ (composizione di funzioni continue) mentre per $x = 0$ si ha $F(0) = 0$ e

$$\lim_{x \rightarrow 0^+} F(x) = \lim_{x \rightarrow 0^+} f\left(\frac{1}{x}\right) = \lim_{y \rightarrow +\infty} f(y) = 0$$

per ipotesi. Stesso discorso per G . Inoltre F, G sono derivabili su $(0, A)$ (composizione di funzioni derivabili) e si ha

$$F'(x) = \left(f\left(\frac{1}{x}\right) \right)' = f'\left(\frac{1}{x}\right) \cdot \left(-\frac{1}{x^2}\right)$$

e analogamente per G ; in particolare $G' \neq 0$. Notiamo anche che

$$\frac{F'(y)}{G'(y)} = \frac{f'(1/y)}{g'(1/y)}$$

(i fattori $-1/y^2$ si semplificano!) e quindi il limite

$$\lim_{y \rightarrow 0^+} \frac{F'(y)}{G'(y)} = \lim_{y \rightarrow 0^+} \frac{f'(1/y)}{g'(1/y)} = \lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)}$$

esiste, per ipotesi. Ma allora possiamo applicare il Teorema di de l'Hospital nella prima versione e concludere che anche il limite di F/G esiste e coincide con il precedente:

$$\lim_{y \rightarrow 0^+} \frac{F(y)}{G(y)} = \lim_{y \rightarrow 0^+} \frac{F'(y)}{G'(y)}$$

il che vuol dire precisamente

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow +\infty} \frac{f'(x)}{g'(x)}.$$

Un'altra variante utilissima del Teorema di de l'Hospital riguarda le forme indeterminate del tipo ∞/∞ :

TEOREMA 5.6. (*de l'Hospital, versione 3*) Siano $f, g : (c, b] \rightarrow \mathbb{R}$ due funzioni continue tali che

$$f(x) \rightarrow +\infty, \quad g(x) \rightarrow +\infty \quad \text{per } x \rightarrow c^+.$$

Inoltre supponiamo che f e g siano derivabili su (c, b) e che $g(x) \neq 0$, $g'(x) \neq 0$ per ogni $x \in (c, b)$. Infine supponiamo che esista il limite (finito o infinito)

$$\lim_{x \rightarrow c^+} \frac{f'(x)}{g'(x)}.$$

Allora esiste anche il limite di f/g per $x \rightarrow c^+$, e i due limiti coincidono:

$$\lim_{x \rightarrow c^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c^+} \frac{f'(x)}{g'(x)}.$$

Omettiamo la dimostrazione (ci si riconduce ai casi precedenti con astute trasformazioni). Notiamo però che il teorema vale anche se una o tutte e due le funzioni tendono a $-\infty$, se i limiti sono da sinistra ossia $x \rightarrow c^-$, e se i limiti sono per $x \rightarrow +\infty$ oppure $x \rightarrow -\infty$.

ESEMPIO 5.1. Vediamo che succede applicando l'Hospital allo studio di alcuni limiti notevoli:

$$\lim_{x \rightarrow 0} \frac{\text{sen } x}{x} = \lim_{x \rightarrow 0} \frac{(\text{sen } x)'}{(x)'} = \lim_{x \rightarrow 0} \frac{\cos x}{1} = 1$$

oppure

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{x} = \lim_{x \rightarrow 0} \frac{(e^x - 1)'}{(x)'} = \lim_{x \rightarrow 0} \frac{e^x}{1} = 1$$

o anche, usando de l'Hospital due volte,

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \lim_{x \rightarrow 0} \frac{(1 - \cos x)'}{(x^2)'} = \lim_{x \rightarrow 0} \frac{\sin x}{2x} = \lim_{x \rightarrow 0} \frac{(\sin x)'}{(2x)'} = \lim_{x \rightarrow 0} \frac{\cos x}{2} = \frac{1}{2}.$$

Nel caso ∞/∞ abbiamo ad esempio

$$\lim_{x \rightarrow +\infty} \frac{e^x}{x} = \lim_{x \rightarrow +\infty} \frac{(e^x)'}{(x)'} = \lim_{x \rightarrow +\infty} \frac{e^x}{1} = +\infty$$

eccetera.

Un dubbio frequente: perchè nel capitolo sui limiti abbiamo perso tempo a calcolare i limiti notevoli di $\frac{\sin x}{x}$ e di $\frac{e^x-1}{x}$, quando bastava aspettare un po' e usare il Teorema di de l'Hospital che rende le dimostrazioni così semplici?

Risposta: per dimostrare che

$$(5.3) \quad (e^x)' = e^x$$

abbiamo utilizzato proprio il limite notevole

$$(5.4) \quad \lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1.$$

Se ora usassimo l'Hospital per dimostrare il limite notevole, avremmo solo dimostrato che (5.3) e (5.4) sono *equivalenti*, ma in effetti non avremmo dimostrato nessuna delle due proprietà (potrebbero essere equivalenti ma tutte e due false...). Stesso discorso per $\frac{\sin x}{x}$ e altri limiti calcolati "a mano" in precedenza.

ESEMPIO 5.2. Possiamo usare il metodo anche per forme indeterminate del tipo $0 \cdot \infty$: ad esempio scriviamo

$$\lim_{x \rightarrow 0^+} x \cdot \ln x = \lim_{x \rightarrow 0^+} \frac{\ln x}{\frac{1}{x}}$$

e quindi ora dobbiamo studiare una forma del tipo ∞/∞ ; applicando de l'Hospital

$$= \lim_{x \rightarrow 0^+} \frac{(\ln x)'}{(\frac{1}{x})'} = \lim_{x \rightarrow 0^+} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = - \lim_{x \rightarrow 0^+} \frac{x^2}{x} = 0.$$

Simbolicamente, il trucco consiste nello scrivere

$$0 \cdot \infty = \frac{0}{\frac{1}{\infty}} \quad \text{oppure} \quad = \frac{\infty}{\frac{1}{0}}.$$

Vi sono parecchie altre applicazioni di questa tecnica. Ne menzioniamo solo una: per studiare una forma del tipo ∞^0 scriviamo (simbolicamente!)

$$\infty^0 = e^{0 \cdot \ln(\infty)} = e^{\frac{\ln(\infty)}{\frac{1}{0}}}.$$

ESEMPIO 5.3. Calcolare il seguente limite.

$$(5.5) \quad \lim_{x \rightarrow 0} \left(\frac{e^x - 1 - x}{x^2} \right).$$

Usiamo il Teorema di de l'Hospital in versione base. Ovviamente $f(x) = e^x - 1 - x$, $g(x) = x^2$. Quindi per $x \neq 0$

$$\frac{f'(x)}{g'(x)} = \frac{e^x - 1}{2x}$$

e l'ultima l'espressione ammette limite per $x \rightarrow 0$. Concludiamo che

$$\lim_{x \rightarrow 0} \left(\frac{e^x - 1 - x}{x^2} \right) = \lim_{x \rightarrow 0} \frac{f'(x)}{g'(x)} = \lim_{x \rightarrow 0} \frac{e^x - 1}{2x} = \frac{1}{2}.$$

ESERCIZIO 5.4. Calcolare, se esistono, i seguenti limiti.

$$\lim_{x \rightarrow 0^+} \frac{\operatorname{arctg}(x^2)}{x \ln x}, \quad \lim_{x \rightarrow 0} \frac{\ln(1-x) - x}{-\cos(x)}, \quad \lim_{x \rightarrow 0^+} x e^{\frac{1}{\sqrt{x}}}$$

CONTROESEMPIO 5.5. Può succedere che esista il limite di f/g senza che esista il limite di f'/g' . Ad esempio, scegliamo

$$(5.6) \quad f(x) = x^2 \operatorname{sen} \left(\frac{1}{x} \right), \quad g(x) = x \quad \text{con } x \in (0, 1).$$

Ovviamente

$$\lim_{x \rightarrow 0^+} \frac{f(x)}{g(x)} = \lim_{x \rightarrow 0^+} \left[x \cdot \operatorname{sen} \left(\frac{1}{x} \right) \right] = 0,$$

ma

$$\frac{f'(x)}{g'(x)} = \frac{2x \operatorname{sen} \left(\frac{1}{x} \right) - \cos \left(\frac{1}{x} \right)}{1}$$

e tale rapporto non ammette limite per $x \rightarrow 0$, come si verifica facilmente.

6. La formula di Taylor

Abbiamo visto che se una funzione $f : (a, b) \rightarrow \mathbb{R}$ è derivabile in un punto $x_0 \in (a, b)$, allora si può scrivere

$$(6.1) \quad f(x_0 + h) = f(x_0) + f'(x_0)h + o(h)$$

dove $o(h)$ è un infinitesimo di ordine superiore a 1 per $h \rightarrow 0$:

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$$

Generalizziamo questa notazione a infinitesimi di ordine più alto. Se una funzione $g(h)$ ha la proprietà

$$\lim_{h \rightarrow 0} \frac{g(h)}{h^n} = 0$$

per un certo intero $n \geq 0$, diciamo che $g(h)$ è un *infinitesimo di ordine superiore ad n* . Anche la notazione degli o piccoli si generalizza a questo caso:

$$(6.2) \quad g(h) = o(h^n) \quad \text{vuol dire} \quad \lim_{h \rightarrow 0} \frac{g(h)}{h^n} = 0.$$

Analogamente scriviamo $o((x - x_0)^n)$ per indicare un infinitesimo di ordine superiore a n nel punto x_0 , cioè

$$R(x) = o((x - x_0)^n) \quad \iff \quad \lim_{x \rightarrow x_0} \frac{R(x)}{(x - x_0)^n} = 0.$$

Ricordiamo che una funzione f è derivabile in x_0 se e solo se

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + o(x - x_0)$$

Notare che il grafico della funzione $P_1(x) = f(x_0) + f'(x_0) \cdot (x - x_0)$ è proprio quello della retta tangente al grafico di f nel punto $(x_0, f(x_0))$. Quindi la precedente formula

ci dice che: se approssimiamo la funzione $f(x)$ con la sua retta tangente, l'errore che si commette è un infinitesimo di ordine superiore ad 1 per $x \rightarrow x_0$.

Il *polinomio* e la *formula di Taylor* sono una generalizzazione di queste considerazioni agli "ordini superiori". Ossia, non ci limitiamo ad approssimare la funzione con una retta (polinomio di grado 1) ma tentiamo di ottenere approssimazioni più precise utilizzando polinomi di ordine più alto. Questo è possibile solo quando la funzione è più regolare, ossia possiede derivate di ordine superiore.

Andiamo con ordine. Consideriamo un qualunque polinomio $P(x)$ di grado N . Di solito scriviamo $P(x)$ nella forma

$$P(x) = c_0 + c_1x + \cdots + c_Nx^n.$$

Se fissiamo un punto x_0 , possiamo scrivere P anche nella forma

$$(6.3) \quad P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_N(x - x_0)^N.$$

Infatti basta scrivere $x = (x - x_0) + x_0$ e sviluppare le potenze $((x - x_0) + x_0)^k$ per ottenere la seconda forma, "centrata" nel punto x_0 ; naturalmente i coefficienti a_j saranno diversi dai coefficienti c_j . In modo più sintetico possiamo scrivere

$$P(x) = \sum_{k=0}^N a_k(x - x_0)^k.$$

Notiamo una proprietà che ci servirà fra poco: l'unico polinomio di grado N con la proprietà $P(x) = o((x - x_0)^N)$ è il polinomio nullo!

TEOREMA 6.1. *Sia $P(x)$ un polinomio di grado N tale che*

$$\lim_{x \rightarrow x_0} \frac{P(x)}{(x - x_0)^N} = 0$$

Allora $P \equiv 0$.

Dimostrazione. Procediamo per induzione sul grado N .

Se $N = 0$, il grado di P è 0 ossia P è una costante $P = a_0$. In questo caso abbiamo $\lim_{x \rightarrow x_0} P(x)/(x - x_0)^0 = \lim_{x \rightarrow x_0} a_0 = a_0$. Quindi dall'ipotesi otteniamo subito $P \equiv a_0 \equiv 0$.

Supponiamo ora che il risultato sia vero per i polinomi di grado N , e dimostriamolo per un polinomio P di grado $N + 1$:

$$P(x) = a_0 + a_1(x - x_0) + \cdots + a_{N+1}(x - x_0)^{N+1}.$$

Dall'ipotesi

$$\lim_{x \rightarrow x_0} \frac{P(x)}{(x - x_0)^{N+1}} = 0$$

segue subito che $P(x_0) = 0$ (altrimenti il limite non potrebbe essere 0) cioè $a_0 = 0$. Allora possiamo scrivere

$$P(x) = a_1(x - x_0) + \cdots + a_{N+1}(x - x_0)^{N+1} = (x - x_0) \cdot Q(x)$$

dove $Q(x)$ è un polinomio di grado N , e si ha

$$\frac{P(x)}{(x - x_0)^{N+1}} = \frac{Q(x)}{(x - x_0)^N} \rightarrow 0 \quad \text{per } x \rightarrow x_0.$$

Usando l'ipotesi di induzione, otteniamo che $Q \equiv 0$ e in conclusione anche $P \equiv 0$.

Una conseguenza evidente è che se P, Q sono due polinomi di grado N e sappiamo che

$$\lim_{x \rightarrow x_0} \frac{P(x) - Q(x)}{(x - x_0)^N} = 0$$

allora $P - Q \equiv 0$ cioè i due polinomi coincidono. ■

Una seconda proprietà da notare è la seguente: i coefficienti a_k di P nella forma (6.3) si possono esprimere tramite i valori delle derivate di P nel punto x_0 . Infatti, è chiaro che

$$P(x_0) = a_0$$

perché tutti i termini si annullano tranne il primo. Inoltre la derivata è uguale a

$$P'(x) = a_1 + 2a_2(x - x_0) + \dots + Na_N(x - x_0)^{N-1}$$

e quindi

$$P'(x_0) = a_1.$$

Se continuiamo a derivare, ogni derivata abbassa il grado del polinomio e il primo termine è esattamente, per $k \leq N$,

$$P^{(k)}(x) = k!a_k + \dots$$

dove tutti gli altri termini (i puntini) si annullano quando poniamo $x = x_0$ (una dimostrazione completa si può fare facilmente per induzione). Quindi otteniamo, per $k = 0, 1, \dots, N$

$$P^{(k)}(x_0) = k!a_k$$

e ovviamente $P^{(k)} \equiv 0$ se $k > N$. Possiamo anche scrivere

$$a_k = \frac{P^{(k)}(x_0)}{k!}$$

e quindi vediamo che i coefficienti a_k si possono calcolare tutti usando solo i valori di P e delle sue derivate nel punto x_0 . In altri termini, *per qualunque polinomio di grado N vale la formula*

$$(6.4) \quad P(x) = \sum_{k=0}^N \frac{P^{(k)}(x_0)}{k!} (x - x_0)^k.$$

Questa formula non ci dice niente di troppo interessante su P , però suggerisce la strada da seguire per approssimare una funzione arbitraria con un polinomio.

DOMANDA: data una funzione $f : (a, b) \rightarrow \mathbb{R}$, possiamo trovare un polinomio $P(x)$ di grado N tale che la differenza $f - P$ soddisfi

$$(6.5) \quad \lim_{x \rightarrow x_0} \frac{f(x) - P(x)}{(x - x_0)^N} = 0?$$

Detto in un altro modo, possiamo *approssimare la funzione f con un polinomio di grado N* in modo che *l'errore* (cioè appunto la differenza $f - P$) sia un infinitesimo di ordine superiore ad N nel punto x_0 ? Notate che nel caso $N = 1$ sappiamo risolvere questo problema: se f è derivabile, il polinomio di grado 1 che verifica la proprietà precedente è la retta tangente $P_1(x) = f(x_0) + f'(x_0)(x - x_0)$.

La risposta è contenuta nella seguente definizione:

DEFINIZIONE 6.2. Sia $f : (a, b) \rightarrow \mathbb{R}$ una funzione derivabile N volte e sia $x_0 \in (a, b)$. Il *polinomio di Taylor* di ordine N di f nel punto x_0 è dato da

$$P_N(x) = \sum_{k=0}^N \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

e il *resto di Taylor* di ordine N è semplicemente la differenza $R_N(x) = f(x) - P_N(x)$. L'identità

$$f(x) = P_N + R_N = \sum_{k=0}^N \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + R_N(x)$$

si chiama *formula di Taylor* nel punto x_0 .

Notare che se f è un polinomio di grado N , abbiamo subito $P_N \equiv f$, $R_N \equiv 0$, e la formula di Taylor è proprio la (6.4). Osserviamo anche che le notazioni $P_N(x)$, $R_N(x)$ sono leggermente ambigue: infatti P_N e R_N dipendono non solo dal punto x ma anche dal punto x_0 in cui si calcola lo sviluppo; tuttavia, dato che nei ragionamenti seguenti il punto x_0 è fissato e non varia, preferiamo usare la notazione più comoda $P_N(x)$ invece di quella più precisa ma scomoda $P_N(x; x_0)$.

La seguente proprietà del polinomio di Taylor è immediata da verificare:

$$f(x_0) = P(x_0),$$

e derivando si ottiene subito, con dei calcoli simili a quelli precedenti,

$$f^{(k)}(x_0) = P_N^{(k)}(x_0), \quad k = 0, 1, \dots, N$$

cioè le derivate di f e P_N nel punto x_0 coincidono fino all'ordine N (provate a verificare). Però le derivate di P_N di ordine $\geq N+1$ sono tutte nulle mentre quelle di f , in generale, no.

Siamo pronti per dare una risposta affermativa alla DOMANDA: il polinomio cercato è esattamente il polinomio di Taylor.

TEOREMA 6.3. Sia $f : (a, b) \rightarrow \mathbb{R}$ una funzione derivabile N volte e sia $x_0 \in (a, b)$. Allora il resto di Taylor di ordine N ha la proprietà

$$\lim_{x \rightarrow x_0} \frac{R_N(x)}{(x - x_0)^N} = 0.$$

Dimostrazione. Dobbiamo dimostrare che

$$\lim_{x \rightarrow x_0} \frac{f(x) - P_N(x)}{(x - x_0)^N} = 0.$$

Procediamo per induzione su $N \geq 1$. Se $N = 1$ abbiamo subito

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{(x - x_0)^1} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) = f'(x_0) - f'(x_0) = 0.$$

Supponiamo adesso di aver dimostrato il risultato per un certo N e dimostriamo che se f è derivabile $N+1$ volte si ha

$$\lim_{x \rightarrow x_0} \frac{f(x) - P_{N+1}(x)}{(x - x_0)^{N+1}} = 0.$$

Dato che $P_{N+1}(x_0) = f(x_0)$, il limite precedente è una forma indeterminata del tipo $0/0$ e possiamo applicare il Teorema di de l'Hôpital. Quindi proviamo a calcolare il limite

$$\frac{1}{N+1} \cdot \lim_{x \rightarrow x_0} \frac{f'(x) - P'_{N+1}(x)}{(x - x_0)^N}.$$

Notiamo ora che la derivata $P'_{N+1}(x)$ del polinomio di Taylor di ordine $N + 1$ coincide con il polinomio di Taylor di ordine N della derivata f' , infatti

$$P'_{N+1}(x) = \sum_{k=0}^{N+1} \frac{f^{(k)}(x_0)}{k!} ((x-x_0)^k)' = \sum_{k=1}^{N+1} \frac{f^{(k)}(x_0)}{(k-1)!} (x-x_0)^{k-1} = \sum_{k=0}^N \frac{(f')^{(k)}(x_0)}{k!} (x-x_0)^k.$$

Dato che per ipotesi di induzione il risultato è vero per i polinomi di Taylor di ordine N , otteniamo che il limite è uguale a 0 e la dimostrazione è conclusa.

Resta un ultimo dubbio: la risposta alla DOMANDA è unica? Ossia, può esistere un altro polinomio Q di grado N che verifica (6.5)? La risposta è no: il polinomio di Taylor P_N è l'*unico* polinomio di grado N che soddisfa (6.5).

TEOREMA 6.4. *Sia f come nel Teorema 6.3. Se Q è un polinomio di grado N tale che*

$$\lim_{x \rightarrow x_0} \frac{f(x) - Q(x)}{(x - x_0)^N} = 0,$$

allora Q coincide con il polinomio di Taylor P_N .

Dimostrazione. Sappiamo che

$$\lim_{x \rightarrow x_0} \frac{f(x) - Q(x)}{(x - x_0)^N} = \lim_{x \rightarrow x_0} \frac{f(x) - P_N(x)}{(x - x_0)^N} = 0.$$

Quindi abbiamo anche

$$\lim_{x \rightarrow x_0} \frac{P_N(x) - Q(x)}{(x - x_0)^N} = 0$$

perché basta scrivere

$$\frac{P_N(x) - Q(x)}{(x - x_0)^N} = \frac{f(x) - Q(x)}{(x - x_0)^N} - \frac{f(x) - P_N(x)}{(x - x_0)^N}.$$

Come abbiamo visto all'inizio (Teorema 6.1) questo implica che $Q = P_N$.

ESEMPIO 6.5. Calcoliamo lo sviluppo di Taylor della funzione $f(x) = \sin x$ nel punto $x_0 = 0$. Il polinomio di Taylor è dato da

$$P_N(x) = \sum_{k=0}^N \frac{\sin^{(k)}(0)}{k!} x^k.$$

Come si calcolano le derivate successive di $\sin x$? È facile:

$$\sin' x = \cos x, \quad \sin'' x = -\sin x, \quad \sin''' x = -\cos x, \quad \sin^{(4)} x = \sin x.$$

E poi? Si ricomincia, dato che al quarto passo abbiamo riottenuto il punto di partenza $\sin x$. Ma allora otteniamo subito che

$$\sin(0) = 0, \quad \sin'(0) = 1, \quad \sin''(0) = 0, \quad \sin'''(0) = -1, \quad \sin^{(4)}(0) = 0$$

eccetera. Vediamo che la successione delle derivate in 0 vale $0, 1, 0, -1, 0, 1, 0, -1, \dots$ e così via, quindi otteniamo subito

$$\begin{aligned} P_0 &= 0 \\ P_1 &= P_2 = x \\ P_3 &= P_4 = x - \frac{x^3}{3!} \\ P_5 &= P_6 = x - \frac{x^3}{3!} + \frac{x^5}{5!}. \end{aligned}$$

Sinteticamente scriviamo

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

(interrompere al punto giusto per ottenere il polinomio di Taylor del grado richiesto).

Un procedimento quasi identico dà lo sviluppo di $\cos x$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

mentre quello dell'esponenziale è facilissimo dato che le derivate di qualunque ordine sono tutte uguali a e^x e quindi in 0 valgono tutte 1:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

Con un piccolo sforzo in più si può dimostrare che le somme infinite precedenti, interpretate come serie, convergono e quindi rappresentano degli effettivi sviluppi in serie delle funzioni seno, coseno ed esponenziale; torneremo su questo punto nei corsi successivi di Analisi.

ESERCIZIO 6.1. Trovare i polinomi di Taylor per $x_0 = 0$ e di ordine 1, 2 e 3 per le seguenti funzioni.

$$\begin{aligned} & 3x + x^2, \quad x^4, \quad 2x + \operatorname{sen}(x^4), \quad \operatorname{sen}(x) + x^4, \quad e^{x^2}(x^2 + x^4), \\ & \operatorname{sen}(x), \quad \cos(x), \quad \ln(1-x), \quad \operatorname{tg}(x), \quad \frac{1}{1-x}, \quad \frac{1}{1+x}, \\ & \operatorname{sen}(x^2), \quad \cos(x^4), \quad \ln(1-x^2), \quad \operatorname{tg}(x^3), \quad \frac{1}{1-x^2}, \quad \frac{x^4}{1-x^2} \end{aligned}$$

ESERCIZIO 6.2. Completare le affermazioni seguenti:

- i) il polinomio di Taylor di una somma di funzioni $f + g$ è uguale a...
- ii) per calcolare il polinomio di Taylor di ordine N di un prodotto fg di due funzioni è sufficiente calcolare separatamente il polinomio di ordine N di f , quello di g , e poi...

OSSERVAZIONE 6.6. Il "calcolo degli o piccoli" è un'utile alternativa al teorema di de l'Hospital per calcolare forme indeterminate. Supponiamo ad esempio di voler calcolare il limite

$$\lim_{x \rightarrow 0} \frac{\operatorname{sen} x - x}{e^x - 1 - x - \frac{x^2}{2}}.$$

Dallo sviluppo di Taylor di $\operatorname{sen} x$ intorno a 0 abbiamo

$$\operatorname{sen} x = x - \frac{x^3}{3!} + o(x^3).$$

Notare che avremmo potuto scrivere anche

$$\operatorname{sen} x = x - \frac{x^3}{3!} + o(x^4)$$

perchè $P_3 = P_4$, ma NON È VERO CHE

$$\operatorname{sen} x = x - \frac{x^3}{3!} + o(x^5)$$

(lo sviluppo all'ordine 5 prevede il termine $x^5/5!$). Allo stesso modo abbiamo

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + o(x^3).$$

Allora possiamo scrivere, sostituendo gli sviluppi precedenti al terzo ordine,

$$\frac{\operatorname{sen} x - x}{e^x - 1 - x - \frac{x^2}{2}} = \frac{-\frac{x^3}{3!} + o(x^3)}{\frac{x^3}{3!} + o(x^3)} = \frac{-\frac{1}{6} + \frac{o(x^3)}{x^3}}{\frac{1}{6} + \frac{o(x^3)}{x^3}}$$

e quindi, ricordando che $o(x^3)/x^3 \rightarrow 0$ quando $x \rightarrow 0$, otteniamo che il limite è uguale a -1 .

OSSERVAZIONE 6.7. Dato che la principale applicazione della formula di Taylor è l'approssimazione delle funzioni, un problema fondamentale è capire se l'approssimazione è buona o cattiva, ossia vorremmo stimare l'errore R_N . A questo scopo esistono numerosi metodi e varie rappresentazioni alternative del resto.

Una delle più utili è la *forma di Lagrange del resto*: se $f : (a, b) \rightarrow \mathbb{R}$ è derivabile $N + 1$ volte e $x, x_0 \in (a, b)$, allora esiste un punto ξ che sta tra x_0 e x (e dipende da tutti e due) tale che

$$(6.6) \quad R_N(x) = \frac{f^{(N+1)}(\xi)}{(N+1)!} (x - x_0)^{N+1}.$$

Notare che questa forma del resto "assomiglia" al termine successivo dello sviluppo, con una differenza fondamentale: la derivata non è calcolata in x_0 ma in un punto intermedio imprecisato, un po' come nel Teorema di Lagrange.

Per dimostrarla utilizziamo la seguente versione generalizzata del Teorema di Rolle: sia $F : (a, b) \rightarrow \mathbb{R}$ continua e derivabile $N + 1$ volte e siano $x, x_0 \in (a, b)$. Supponiamo che $F(x) = 0$ e che $F^{(k)}(x_0) = 0$ per $0 \leq k \leq N$. Allora esiste un punto ξ fra x_0 e x tale che $F^{(N+1)}(\xi) = 0$.

La dimostrazione è facilissima: se $N = 0$ la proprietà è esattamente il Teorema di Rolle applicato all'intervallo di estremi x_0 e x . Se $N = 1$ applichiamo Rolle a F , quindi otteniamo che esiste un punto c fra x_0 e x tale che $F'(c) = 0$; e poi applichiamo di nuovo Rolle alla funzione F' sull'intervallo tra x_0 e c , e quindi otteniamo un punto ξ tale che $F''(\xi) = 0$. Il caso generale si dimostra allo stesso modo per induzione.

Ora, per dimostrare la formula del resto di Lagrange, consideriamo la funzione seguente

$$F(t) = f(t) - P_N(t) - \frac{R_N(x)}{(x - x_0)^{N+1}} (t - x_0)^{N+1}$$

(ricordare che x e x_0 adesso sono due punti fissati, quello che varia è il numero t) e applichiamo il precedente Teorema di Rolle generalizzato. Abbiamo subito $F(x_0) = F(x) = 0$; la derivata prima di $F(t)$ è uguale a

$$F'(t) = f'(t) - P'_N(t) - \frac{R_N(x)}{(x - x_0)^{N+1}} (N + 1)(t - x_0)^N$$

e ovviamente si ha $F'(x_0) = 0$, e analogamente per tutte le derivate di F in x_0 fino all'ordine N compreso. Allora possiamo concludere che esiste un punto ξ fra x_0 e x tale che $F^{(N+1)}(\xi) = 0$; ma dato che

$$F^{(N+1)}(t) = f^{(N+1)}(t) - \frac{R_N(x)}{(x - x_0)^{N+1}} (N + 1)!$$

(notare che $P_N^{(N+1)}$ è identicamente nullo) abbiamo ottenuto che

$$f^{(N+1)}(\xi) = \frac{R_N(x)}{(x - x_0)^{N+1}} (N + 1)!$$

ossia la (6.6).

OSSERVAZIONE 6.8. La formula di Taylor fornisce una buona approssimazione della funzione *vicino* al punto x_0 , ma quando x è lontano da x_0 la funzione e il suo polinomio di Taylor possono essere molto diversi. Ad esempio, i polinomi di Taylor di grado $N \geq 1$ tendono sempre all'infinito per $x \rightarrow \pm\infty$, quindi per x grande sono pessime approssimazioni di f per una funzione limitata come $f(x) = \sin x$.

In genere se aumentiamo il grado N , l'approssimazione migliora; ma non è sempre così, come mostra l'esempio seguente:

$$f(x) = \begin{cases} e^{-\frac{1}{x}}, & x > 0, \\ 0 & x \leq 0. \end{cases}$$

È evidente che la funzione f è continua su tutto \mathbb{R} . Inoltre è derivabile su $x > 0$ e su $x < 0$ con derivata

$$f'(x) = e^{-\frac{1}{x}} \cdot \frac{1}{x^2} \text{ per } x > 0, \quad f'(x) = 0 \text{ per } x < 0.$$

Dato che il limite di $f'(x)$ per $x \rightarrow 0$ esiste e vale 0, otteniamo che f è derivabile anche in 0 e $f'(0) = 0$. Se deriviamo ancora una volta ci accorgiamo che il ragionamento precedente funziona ancora (in quanto la presenza del fattore $e^{-\frac{1}{x}}$ fa sempre convergere il limite destro a 0, anche se compaiono potenze più alte di x a denominatore). Per induzione, non è difficile dimostrare che f è derivabile di qualunque ordine su tutto \mathbb{R} e si ha sempre

$$f^{(k)}(0) = 0 \quad \forall k = 0, 1, 2, 3, \dots$$

Ma allora i polinomi di Taylor di f in 0 sono tutti nulli! In altri termini, l'approssimazione che riusciamo a ottenere non migliora al crescere di N .

SVILUPPI DI TAYLOR E o PICCOLI

1. DEFINIZIONE DI $o(g)$

Se g è una funzione definita su un intervallo aperto contenente x_0 , indichiamo con $o(g)$ per $x \rightarrow x_0$ l'insieme di tutte le funzioni $f(x)$, definite per x vicino x_0 , tali che

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0$$

In questo caso si dice che f è un o piccolo di g per $x \rightarrow x_0$, e si scrive $f \in o(g)$ o anche $f = o(g)$. Quindi le funzioni $o(g)$ sono quelle che tendono a zero¹ più velocemente di g per $x \rightarrow x_0$. Nel caso delle potenze, quando $f \in o(x^N)$ si dice anche che f è un *infinitesimo di ordine superiore a N* . Analoghe definizioni per $x \rightarrow x_0^\pm$, $x \rightarrow \pm\infty$. Ad esempio:

$$\begin{aligned} x^3 &\in o(x^2) \text{ per } x \rightarrow 0 \text{ ma } x^3 \notin o(x^3) \\ x &\in o(1) \text{ per } x \rightarrow 0 \\ 2o(x^2) &= o(x^2) = -o(x^2) \end{aligned}$$

In generale:

- (1) $f \in o(1)$ per $x \rightarrow x_0$ vuol dire semplicemente che $f \rightarrow 0$ per $x \rightarrow x_0$
- (2) $x^p \in o(x^q)$ per $x \rightarrow 0$ se $p, q \in \mathbb{N}$ con $p > q$, ma $x^p \notin o(x^p)$
- (3) $o(x^p) \subseteq o(x^q)$ per $x \rightarrow 0$ se $p, q \in \mathbb{N}$ con $p \geq q$
- (4) $o(g) = -o(g) = 2 \cdot o(g)$

Notare che nella pratica di solito si fa un piccolo abuso di notazione:

$$\text{invece di scrivere } x^3 \in o(x^2), \text{ si scrive } x^3 = o(x^2).$$

In questo modo le formule sono molto più leggibili, ma bisogna fare attenzione a non fare confusione. Per esempio,

$$o(x) - o(x) \text{ non è uguale a zero!}$$

Infatti i due $o(x)$ indicano due funzioni infinitesime di ordine superiore a 1, anche diverse fra loro (la notazione rigorosa sarebbe $o(x) - o(x) \subseteq o(x)$).

2. SVILUPPI DI TAYLOR

Esempio fondamentale: il resto di Taylor R_N di $f(x)$ nel punto x_0

$$R_N(x) := f(x) - P_N(x), \quad P_N(x) := \sum_{k=0}^N \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

è un infinitesimo di ordine superiore a N per $x \rightarrow x_0$. Infatti si ha

$$\lim_{x \rightarrow x_0} \frac{R_N(x)}{(x - x_0)^N} = 0 \quad \text{cioè} \quad R_N \in o((x - x_0)^N) \text{ per } x \rightarrow x_0,$$

anzi il polinomio di Taylor P_N in x_0 è definito proprio come l'unico polinomio P tale che $f - P \in o((x - x_0)^N)$ per $x \rightarrow x_0$. Ricordiamo i principali sviluppi in $x_0 = 0$:

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \dots + \frac{x^N}{N!} + o(x^N) \\ \sin x &= x - \frac{x^3}{3!} + \dots + (-1)^k \frac{x^{2k+1}}{(2k+1)!} + o(x^{2k+2}) \\ \cos x &= 1 - \frac{x^2}{2!} + \dots + (-1)^k \frac{x^{2k}}{(2k)!} + o(x^{2k+1}) \\ \ln(1+x) &= x - \frac{x^2}{2} + \dots + (-1)^{N+1} \frac{x^N}{N} + o(x^N) \\ \arctan x &= x - \frac{x^3}{3} + \dots + (-1)^k \frac{x^{2k+1}}{2k+1} + o(x^{2k+2}) \end{aligned}$$

¹In effetti la notazione spesso si usa anche se f, g non tendono a 0, ad esempio si dice: e^x è un $o(x)$ per $x \rightarrow +\infty$, dato che $\frac{e^x}{x} \rightarrow 0$ per $x \rightarrow +\infty$. Oppure: un infinitesimo è una qualunque funzione $o(1)$. Eccetera...

$$(1+x)^\alpha = 1 + \binom{\alpha}{1}x + \dots + \binom{\alpha}{N}x^N + o(x^N)$$

dove per $\alpha \in \mathbb{R}$ e $k \in \mathbb{N}$ si definisce $\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!}$. Notare che $\binom{\alpha}{k} = 0$ se $\alpha \in \mathbb{N}$ e $0 < \alpha < k$, quindi in questo caso riotteniamo la formula per la potenza di un binomio.

Vediamo già qui quanto sia comodo l'abuso di notazione degli o piccoli: ad esempio invece di scrivere la formula poco maneggevole

$$e^x - 1 - x - \frac{x^2}{2} \in o(x^2)$$

scriviamo la formula

$$e^x = 1 + x + \frac{x^2}{2} + o(x^2)$$

che è molto più comoda da usare nei calcoli (basta non sbagliare ad usarla...)

3. ARITMETICA DEGLI o PICCOLI

Le regole di calcolo per gli o piccoli sono semplicissime e forniscono uno strumento efficace per il calcolo di molti limiti. Dimostrare per esercizio le seguenti formule, nelle quali naturalmente tutti gli o piccoli si riferiscono allo stesso punto $x \rightarrow x_0$ e tutte le funzioni hanno limite per $x \rightarrow x_0$:

(1) $f \cdot o(g) \subseteq o(f \cdot g)$ (vuol dire: se $\phi \in o(g)$, allora $f \cdot \phi \in o(fg)$)

(2) $o(f) \cdot o(g) \subseteq o(f \cdot g)$ (vuol dire: se $\phi \in o(f)$ e $\psi \in o(g)$, allora $\phi \cdot \psi \in o(fg)$)

Quindi ad esempio se $p, q \in \mathbb{N}$

$$x^p \cdot o(x^q) \subseteq o(x^{p+q}), \quad o(x^p) \cdot o(x^q) \subseteq o(x^{p+q})$$

e inoltre

$$o(x)^q \subseteq o(x^q), \quad o(x^p)^q \subseteq o(x^{pq}).$$

Se invece sommiamo due o piccoli, il risultato sarà sempre l' o piccolo di ordine più basso. Dimostrare per esercizio le formule seguenti, per $x \rightarrow 0$:

(3) se $p, q \in \mathbb{N}$ e $p > q$ allora $x^p + o(x^q) \subseteq o(x^q)$

(4) se $p, q \in \mathbb{N}$ e $p \geq q$ allora $o(x^p) + o(x^q) \subseteq o(x^q)$

(la seconda vuol dire: se $f \in o(x^p)$ e $g \in o(x^q)$ allora $f + g \in o(x^q)$). Sono necessarie anche alcune regole di calcolo per funzioni composte. Dimostrare per esercizio le formule seguenti:

(5) $o(o(g)) \subseteq o(g)$ (vuol dire: se $f \in o(g)$ e $\phi \in o(f)$ allora $\phi \in o(g)$)

(6) $o(x^p + o(x^p)) \subseteq o(x^p)$ per $x \rightarrow 0$ se $p \geq 0$.

Ad esempio, l'ultima formula si dimostra così: se $\phi \in o(x^p)$ e $f \in o(x^p + \phi)$ allora

$$\frac{f}{x^p} = \frac{f}{x^p + \phi} \cdot \frac{x^p + \phi}{x^p} = \frac{f}{x^p + \phi} \cdot \left(1 + \frac{\phi}{x^p}\right) \rightarrow 0 \cdot (1 + 0) = 0$$

quindi $f \in o(x^p)$.

ESEMPI DI CALCOLO CON GLI o PICCOLI. Possiamo scrivere, per $x \rightarrow 0$,

$$(x^2 + o(x^2))^3 = x^6 + 3x^4 o(x^2) + 3x^2 o(x^2)^2 + o(x^2)^3 = x^6 + o(x^6) + o(x^6) + o(x^6) = x^6 + o(x^6)$$

oppure

$$(1 + x + o(x)) \cdot (x + o(x)) = x + o(x) + x^2 + xo(x) + xo(x) + o(x)^2 = x + o(x).$$

Nel secondo esempio, dato che compare un termine $o(x)$, tutti gli ordini superiori (che in questo caso sono x^2 e $o(x^2)$) vengono assorbiti da $o(x)$.

4. PRATICA DI CALCOLO

ESERCIZIO 1. Calcolare P_3 in $x_0 = 0$ per la funzione $f(x) = \ln(1 - \sin^2 x)$.

[Proviamo a utilizzare lo sviluppo $\sin x = x + o(x^2)$; se questo non è sufficiente, proveremo poi ad utilizzare un termine in più dello sviluppo di Taylor. Otteniamo

$$\sin^2 x = (x + o(x^2))^2 = x^2 + 2x \cdot o(x^2) + o(x^2)^2 = x^2 + o(x^3).$$

Fino a che ordine dobbiamo sviluppare \ln ? Proviamo con $\ln(1 + y) = y + o(y)$, sostituendo $y = -\sin^2 x = -x^2 + o(x^3)$ otteniamo

$$f(x) = -x^2 + o(x^3) + o(-x^2 + o(x^3)) = -x^2 + o(x^2).$$

Non basta! questo conto dimostra solo che $P_2 = -x^2$ (perché il resto che abbiamo ottenuto è solo $o(x^2)$). Proviamo con un termine in più ossia $\ln(1+y) = y - \frac{y^2}{2} + o(y^2)$; sostituendo $y = -\sin^2 x = -x^2 + o(x^3)$ otteniamo

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

I termini $(-x^2 + o(x^3))^2 = x^4 + o(x^5)$ e $o((x^2 + o(x^3))^2) = o(x^4)$ sono assorbiti da $o(x^3)$ e in conclusione

$$f(x) = -x^2 + o(x^3).$$

Questa formula implica $P_3(x) = -x^2$

ESERCIZIO 2. Calcolare P_3 in $x = 0$ per $f(x) = \frac{1}{1+2x+2x^3}$. [$P_3 = 1 - 2x + 4x^2 - 10x^3$]

ESERCIZIO 3. Calcolare P_3 in $x = 0$ per $f(x) = \ln(e^x - x)$. [$P_3 = \frac{x^2}{2} + \frac{x^3}{6}$]

ESERCIZIO 4. Calcolare P_3 in $x = 0$ per $f(x) = \sin x \cdot (e^x - \cos^2 x)$. [$P_3 = x^2 + \frac{3}{2}x^3$]

ESERCIZIO 5. Calcolare $f^{(4)}(0)$ per $f(x) = e^x \sin x$.

[Invece di calcolare la derivata quarta e poi porre $x = 0$, è più rapido scrivere

$$f(x) = (1 + x + \frac{x^2}{2} + \frac{x^3}{6} + o(x^2)) \cdot (x - \frac{x^3}{6} + o(x^4)) = x + x^2 + \frac{x^3}{2} + \frac{x^4}{6} - \frac{x^3}{6} - \frac{x^4}{6} + o(x^4)$$

da cui

$$f(x) = x + x^2 + \frac{x^3}{6} + o(x^4).$$

Quindi $P_4 = x + x^2 + \frac{x^3}{6}$ il che vuol dire $f(0) = 0$, $f'(0) = 1$, $f''(0) = 2$, $f'''(0) = 1$ e $f^{(4)}(0) = 0$]

ESERCIZIO 6. Calcolare $f^{(4)}(0)$ per $f(x) = \sqrt[5]{1 + 2x^2 + x^3}$. [$f^{(4)}(0) = -\frac{192}{25}$]

ESERCIZIO 7. Calcolare $f^{(4)}(0)$ per $f(x) = (\cos(\sin x))^{1/3}$. [$f^{(4)}(0) = 1$]

ESERCIZIO 8. Usando Taylor, calcolare $\lim_{x \rightarrow 0} \frac{e^x - \sin x - \cos x}{e^{x^2} - e^{x^3}}$.

[Si ha $e^x - \sin x - \cos x = 1 + x + \frac{x^2}{2} + o(x^2) - x + o(x^2) - 1 + \frac{x^2}{2} + o(x^3) = x^2 + o(x^2)$ e per il denominatore, sostituendo in $e^y = 1 + y + o(y)$ prima $y = x^2$ e poi $y = x^3$, $e^{x^2} - e^{x^3} = 1 + x^2 + o(x^2) - 1 - x^3 + o(x^3) = x^2 + o(x^2)$. Quindi il limite cercato diventa

$$\lim_{x \rightarrow 0} \frac{x^2 + o(x^2)}{x^2 + o(x^2)} = \lim_{x \rightarrow 0} \frac{x^2(1 + \frac{o(x^2)}{x^2})}{x^2(1 + \frac{o(x^2)}{x^2})} = 1.$$

Il limite si può calcolare anche usando il Teorema di de l'Hospital]

ESERCIZIO 9. Usando Taylor, calcolare $\lim_{x \rightarrow 0} \frac{\sin x - \ln(\cos x)}{x \sin x}$.

[Scriviamo $\ln(\cos x) = \ln(1 + \cos x - 1)$, e usiamo $\ln(1+y) = y - \frac{y^2}{2} + o(y^2)$ con $y = \cos x - 1 = -\frac{x^2}{2} + o(x^3)$, da cui

$$\ln(\cos x) = -\frac{x^2}{2} + o(x^3) - \frac{1}{2}(-\frac{x^2}{2} + o(x^3))^2 + o((-\frac{x^2}{2} + o(x^3))^2) = -\frac{x^2}{2} + o(x^3)$$

e quindi $\sin x - \log(\cos x) = x + o(x^2) + \frac{x^2}{2} + o(x^3) = x + o(x^2)$. Invece per il denominatore basta scrivere $x \sin x = x(x + o(x^2)) = x^2 + o(x^3)$. In conclusione il limite diventa

$$\lim_{x \rightarrow 0} \frac{x + o(x^2)}{x^2 + o(x^3)} = \lim_{x \rightarrow 0} \frac{1 + \frac{o(x^2)}{x}}{x(1 + \frac{o(x^3)}{x^2})}$$

e questo limite non esiste perché vale $+\infty$ per $x \rightarrow 0^+$ e $-\infty$ per $x \rightarrow 0^-$]

ESERCIZIO 10. Usando Taylor, calcolare $\lim_{x \rightarrow 0} \frac{\sin(x^2) - \ln(\cos x)}{x \sin x}$. [= $\frac{3}{2}$]

ESERCIZIO 11. Usando Taylor, calcolare $\lim_{x \rightarrow 0} \frac{\cos x - 1 + \frac{x}{2} \sin x}{(\ln(1+x))^4}$. [= $-\frac{1}{24}$]

ESERCIZIO 12. Usando Taylor, calcolare $\lim_{x \rightarrow 0} \frac{(\ln(1+x))^3}{\cos x \cdot \sin^2 x - x^2}$. [= 0]

ESERCIZIO 13. Usando Taylor, calcolare $\lim_{x \rightarrow +\infty} (1 + \frac{1}{x})^{x^2} \cdot e^{-x}$.

[Scriviamo

$$(1 + \frac{1}{x})^{x^2} \cdot e^{-x} = \exp[x^2 \ln(1 + \frac{1}{x}) - x]$$

e usando $\ln(1+y) = y - \frac{y^2}{2} + o(y^2)$

$$x^2 \ln(1 + \frac{1}{x}) - x = x^2(\frac{1}{x} - \frac{1}{2x^2} + o(\frac{1}{x^2})) - x = -\frac{1}{2} + \frac{o(1/x^2)}{1/x^2} \rightarrow -\frac{1}{2}$$

quindi il limite cercato è $e^{-1/2}$]

ESERCIZIO 14. Usando Taylor, calcolare $\lim_{x \rightarrow 0^+} (\cos x)^{\frac{1}{\sin x}}$.

$$[(\cos x)^{\frac{1}{\sin x}} = \exp\left[\frac{\log(\cos x)}{\sin x}\right] \rightarrow 1]$$

ESERCIZIO 15. Per quali valori di a, b, c esiste il $\lim_{x \rightarrow 0} \frac{e^x - a - bx}{\sin x - cx}$?

$$\left[\frac{e^x - a - bx}{\sin x - cx} = \frac{(1-a) + (1-b)x + \frac{x^2}{2} + o(x^2)}{(1-c)x - \frac{x^3}{6} + o(x^3)} \right]. \text{ Se } a \neq 1 \text{ il limite per } x \rightarrow 0^\pm \text{ è } \pm\infty \text{ o } \mp\infty, \text{ quindi in questo caso}$$

il limite per $x \rightarrow 0$ non esiste. Dunque deve essere $a = 1$. Se $c \neq 1$ il limite esiste ed è $\frac{1-b}{1-c}$. Se $c = 1$ il limite non esiste per qualunque valore di b . In conclusione il limite esiste se e solo se $a = 1, c \neq 1$]

Come calcolare alcuni polinomi di Taylor senza fare nessuna derivata (o quasi)

Calcoliamo i polinomi di Taylor di

$$f(x) = \ln(1+x).$$

Ricordiamo che si ha, per $-1 < x < 1$,

$$g(x) = \frac{1}{1+x} = \frac{1}{1-(-x)} = \sum_{k=0}^{+\infty} (-x)^k = \sum_{k=0}^{+\infty} (-1)^k x^k.$$

D'altra parte, se $S_n(x; 0)$ è il polinomio di Taylor di ordine n in $x_0 = 0$ di $g(x)$, dall'espressione di $g(x)$ come somma di una serie si ricava facilmente che

$$S_n(x; 0) = \sum_{k=0}^n (-1)^k x^k = 1 - x + x^2 - x^3 + \dots + (-1)^n x^n.$$

Confrontando questa espressione con l'espressione "standard" del polinomio di Taylor per $g(x)$,

$$S_n(x; 0) = \sum_{k=0}^n \frac{g^{(k)}(0)}{k!} x^k,$$

si vede facilmente che $g^{(k)}(0) = (-1)^k k!$ per ogni $k \geq 0$.

Osserviamo ora che si ha

$$f'(x) = [\ln(1+x)]' = \frac{1}{1+x} = g(x),$$

cosicché per ogni $k \geq 1$ si ha

$$f^{(k)}(x) = g^{(k-1)}(x),$$

da cui segue che

$$f^{(k)}(0) = g^{(k-1)}(0) = (-1)^{k-1} (k-1)!.$$

Dato poi che $f(0) = 0$, si ha, se $T_n(x; 0)$ è il polinomio di ordine n di $f(x)$ centrato in $x_0 = 0$,

$$T_n(x; 0) = \sum_{k=0}^n \frac{f^{(k)}(0)}{k!} x^k = \sum_{k=1}^n \frac{(-1)^{k-1} (k-1)!}{k!} x^k = \sum_{k=1}^n \frac{(-1)^{k-1}}{k} x^k.$$

Si ha dunque

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots + \frac{(-1)^{n-1}}{n} x^n + o(x^n).$$

Con lo stesso "trucco" possiamo calcolare il polinomio di Taylor dell'arcotangente $f(x) = \arctan(x)$. Infatti, detta $g(x) = \frac{1}{1+x^2}$ si ha (per $-1 < x < 1$)

$$g(x) = \frac{1}{1+x^2} = \frac{1}{1-(-x^2)} = \sum_{k=0}^{+\infty} (-x^2)^k = \sum_{k=0}^{+\infty} (-1)^k x^{2k}.$$

Confrontando questa espressione con quella del polinomio di Taylor per $g(x)$:

$$S_n(x; 0) = \sum_{k=0}^n \frac{g^{(k)}(0)}{k!} x^k,$$

1

si ricava che $g^{(h)}(0) = 0$ se h è dispari, mentre se $h = 2k$ è un numero pari, allora

$$g^{(2k)}(0) = (-1)^k (2k)!.$$

Dal momento che $f'(x) = g(x)$, otteniamo (come prima) che $f^{(k)}(0) = g^{(k-1)}(0)$ per ogni $k \geq 1$, e quindi che $f^{(h)}(0) = 0$ per ogni h dispari, mentre

$$f^{(2k+1)}(0) = g^{(2k)}(0) = (-1)^k (2k)!.$$

Si ha pertanto (usando anche che $f(0) = 0$) che

$$T_{2n+1}(x; 0) = \sum_{k=0}^n \frac{f^{(2k+1)}(0)}{(2k+1)!} x^{2k+1} = \sum_{k=0}^n \frac{(-1)^k (2k)!}{(2k+1)!} x^{2k+1} = \sum_{k=0}^n \frac{(-1)^k}{2k+1} x^{2k+1}.$$

In definitiva,

$$\arctan(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots + (-1)^n \frac{x^{2n+1}}{2n+1} + o(x^{2n+1}).$$

Calcoliamo ora il polinomio di Taylor, centrato in $x_0 = 0$, della funzione $\tan(x)$. Ricordiamo che

$$\tan(x) = \frac{\sin(x)}{\cos(x)} \iff \sin(x) = \tan(x) \cos(x).$$

Dal momento che la funzione tangente è una funzione dispari, nel suo polinomio di Taylor compaiono solo potenze dispari: in altre parole, se scriviamo $\tan(x)$ come somma di una serie, si ha

$$\tan(x) = \sum_{j=0}^{+\infty} a_j x^{2j+1},$$

e il nostro compito sarà quello di calcolare la successione dei coefficienti a_j al variare di j in \mathbb{N} . Ricordando lo sviluppo di seno e coseno, si ha

$$\sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = \tan(x) \sum_{h=0}^{+\infty} \frac{(-1)^h}{(2h)!} x^{2h}.$$

Scrivendo $\tan(x)$ come somma di una serie, abbiamo

$$\sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = \left(\sum_{j=0}^{+\infty} a_j x^{2j+1} \right) \left(\sum_{h=0}^{+\infty} \frac{(-1)^h}{(2h)!} x^{2h} \right),$$

che riscriviamo come

$$\sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = \sum_{j=0}^{+\infty} \sum_{h=0}^{+\infty} \frac{(-1)^h}{(2h)!} a_j x^{2j+2h+1}.$$

Osserviamo ora che l'esponente $2j + 2h + 1$ cui è elevata la x nell'espressione a destra varia da 1 (quando $j = h = 0$) fino a $+\infty$, e che fissato come esponente $2m + 1$, con $m \geq 0$, esistono esattamente $m + 1$ coppie (j, h) tali che $2j + 2h + 1 = 2m + 1$: ovvero le coppie $(m, 0)$, $(m-1, 1)$, \dots , $(1, m-1)$ e $(0, m)$. In altre parole, se chiamiamo $m = j + h$, abbiamo che m varia da 0 a $+\infty$, e che in corrispondenza di ogni m esistono $m + 1$ addendi nella serie a destra che hanno come monomio x^{2m+1} : tali addendi sono della forma $(j, h) = (\ell, m - \ell)$ con ℓ che varia da 0 a m . Abbiamo dunque

$$\sum_{j=0}^{+\infty} \sum_{h=0}^{+\infty} \frac{(-1)^h}{(2h)!} a_j x^{2j+2h+1} = \sum_{m=0}^{+\infty} \left(\sum_{\ell=0}^m a_\ell \frac{(-1)^{m-\ell}}{(2(m-\ell))!} \right) x^{2m+1}.$$

In definitiva, deve essere (cambiamo nome all'indice di sommatoria m nella seconda, tornando a k)

$$\sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = \sum_{k=0}^{+\infty} \left(\sum_{\ell=0}^k a_\ell \frac{(-1)^{k-\ell}}{(2(k-\ell))!} \right) x^{2k+1}.$$

Confrontando le due espressioni, deve quindi essere, per ogni $k \geq 0$,

$$\frac{(-1)^k}{(2k+1)!} = \sum_{\ell=0}^k a_\ell \frac{(-1)^{k-\ell}}{(2(k-\ell))!}.$$

Osserviamo ora che sia a sinistra che a destra compare un fattore $(-1)^k$ (che si può dunque semplificare) e che $(-1)^{-\ell} = (-1)^\ell$ qualsiasi sia ℓ . Deve dunque essere, per ogni $k \geq 0$,

$$\frac{1}{(2k+1)!} = \sum_{\ell=0}^k a_\ell \frac{(-1)^\ell}{(2(k-\ell))!}.$$

Se $k = 0$ la relazione diventa

$$\frac{1}{1!} = \sum_{\ell=0}^0 a_\ell \frac{(-1)^\ell}{(2(0-\ell))!} = a_0 \frac{1}{0!} = a_0,$$

da cui $a_0 = 1$. Se, invece $k \geq 1$, si ha

$$\frac{1}{(2k+1)!} = \sum_{\ell=0}^k a_\ell \frac{(-1)^\ell}{(2(k-\ell))!} = \sum_{\ell=0}^{k-1} a_\ell \frac{(-1)^\ell}{(2(k-\ell))!} + a_k \frac{(-1)^k}{(2(k-k))!},$$

da cui segue che

$$(-1)^k a_k = \frac{1}{(2k+1)!} - \sum_{\ell=0}^{k-1} a_\ell \frac{(-1)^\ell}{(2(k-\ell))!},$$

e quindi che (moltiplicando per $(-1)^k$ e usando che $(-1)^{k+\ell} = (-1)^{k+2\ell-\ell} = (-1)^{k-\ell}$)

$$a_k = \frac{(-1)^k}{(2k+1)!} - \sum_{\ell=0}^{k-1} a_\ell \frac{(-1)^{k-\ell}}{(2(k-\ell))!},$$

formula che permette di calcolare a_k una volta noti a_0, a_1, \dots, a_{k-1} .

Ad esempio, da $a_0 = 1$ si ricava

$$a_1 = \frac{(-1)^1}{(2+1)!} - \sum_{\ell=0}^0 a_\ell \frac{(-1)^{1-\ell}}{(2(1-\ell))!} = -\frac{1}{6} - a_0 \frac{(-1)^1}{2} = -\frac{1}{6} + \frac{1}{2} = \frac{1}{3}.$$

Successivamente

$$a_2 = \frac{1}{5!} - a_0 \frac{1}{4!} + a_1 \frac{1}{2!} = \frac{1}{120} - \frac{1}{24} + \frac{1}{6} = \frac{2}{15},$$

e

$$a_3 = -\frac{1}{7!} + a_0 \frac{1}{6!} - a_1 \frac{1}{4!} + a_2 \frac{1}{2!} = -\frac{1}{5040} + \frac{1}{720} - \frac{1}{72} + \frac{1}{15} = \frac{17}{315}.$$

Proseguendo, si trova ad esempio,

$$\tan(x) = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \frac{17}{315}x^7 + \frac{62}{2835}x^9 + \frac{1382}{155925}x^{11} + \frac{21844}{6081075}x^{13} + o(x^{13}).$$

Si può dimostrare (non senza fatica...) che

$$\lim_{j \rightarrow +\infty} \frac{a_{j+1}}{a_j} = \frac{4}{\pi^2}.$$

Da questa relazione segue che, data la serie

$$\tan(x) = \sum_{j=0}^{+\infty} a_j x^{2j+1} = x \sum_{j=0}^{+\infty} a_j x^{2j},$$

per ogni $x \geq 0$ fissato abbiamo una serie a termini positivi $b_j = a_j x^{2j}$. Applicando il criterio del rapporto, si ha

$$L = \lim_{j \rightarrow +\infty} \frac{b_{j+1}}{b_j} = \lim_{j \rightarrow +\infty} \frac{a_{j+1}}{a_j} \frac{x^{2j+2}}{x^{2j}} = \frac{4x^2}{\pi^2}.$$

Affinché la serie converga, deve essere $0 < L < 1$, ovvero

$$0 < \frac{4x^2}{\pi^2} < 1 \quad \iff \quad 0 < x^2 < \frac{\pi^2}{4} \quad \iff \quad |x| < \frac{\pi}{2},$$

cosa che, d'altra parte, ci si poteva aspettare!

Se, poi, uno si volesse veramente divertire, potrebbe provare a calcolare i coefficienti del polinomio di Taylor di $\tan(x)$ usando il fatto che

$$\tan(\arctan(x)) = x.$$

Scrivendo

$$\tan(x) = \sum_{j=0}^{+\infty} a_j x^{2j+1},$$

e ricordando lo sviluppo dell'arcotangente, deve dunque essere

$$x = \tan(\arctan(x)) = \sum_{j=0}^{+\infty} a_j \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{2k+1} x^{2k+1} \right)^{2j+1}.$$

Per calcolare gli a_j , è "sufficiente" richiedere che il coefficiente del termine di grado 1 nella somma a destra sia uguale a 1 (perché a sinistra c'è $1 \cdot x$), e che siano zero i coefficienti di tutti i termini di grado (dispari, ci sono solo quelli) maggiori o uguali a tre. Per capire meglio, osserviamo che i primi tre addendi della serie a destra sono:

$$a_0 \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{2k+1} x^{2k+1} \right) + a_1 \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{2k+1} x^{2k+1} \right)^3 + a_2 \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{2k+1} x^{2k+1} \right)^5,$$

che riscriviamo come:

$$a_0 x \left(1 - \frac{x^2}{3} + \frac{x^4}{5} + \dots \right) + a_1 x^3 \left(1 - \frac{x^2}{3} + \frac{x^4}{5} + \dots \right)^3 + a_2 x^5 \left(1 - \frac{x^2}{3} + \frac{x^4}{5} + \dots \right)^5.$$

Ora, nella seconda e terza somma gli esponenti minimi a cui compare x sono (rispettivamente) x^3 e x^5 : è chiaro quindi che a_1 e a_2 (così come tutti i successivi) non contribuiscono a "creare" il coefficiente del termine di grado 1. D'altra parte, a moltiplicare $a_0 x$ c'è

$$1 - \frac{x^2}{3} + \frac{x^4}{5} + \dots,$$

e quindi c'è un solo termine di grado 1. In definitiva, si ha

$$a_0 \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{2k+1} x^{2k+1} \right) + a_1 \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{2k+1} x^{2k+1} \right)^3 + a_2 \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{2k+1} x^{2k+1} \right)^5 = a_0 x,$$

più termini di grado superiore al primo. Dovendo essere la sommatoria uguale a x , si ha $a_0 = 1$ (e uno!). Cerchiamo ora i termini di grado 3: ce ne sono esattamente due, dato che la sommatoria moltiplicata per a_2 "parte" da x^5 . Abbiamo

$$-a_0 \frac{x^3}{3} + a_1 x^3,$$

dato che tutti gli altri termini sono di grado maggiore di 3. Deve quindi essere

$$-\frac{a_0}{3} + a_1 = 0,$$

da cui (ricordando che $a_0 = 1$) segue che $a_1 = \frac{1}{3}$ (e due!). Infine (e qui le cose si fanno complicate, se ve ne fosse ulteriore bisogno...) i termini di grado 5 sono dati da:

$$a_0 \frac{x^5}{5} + a_2 x^5 - a_1 x^5.$$

Da dove viene l'ultimo? A moltiplicare a_1 abbiamo (sviluppando il cubo):

$$x^3 \left(\sum_{k=0}^{+\infty} \frac{(-1)^k}{2k+1} x^{2k} \right)^3 = x^3 \left(1 - \frac{x^2}{3} + \dots \right)^3 = x^3 + x^3 \cdot 3 \cdot 1^2 \cdot \frac{-x^2}{3} + \dots = x^3 - x^5 + \dots,$$

Deve quindi essere

$$\frac{a_0}{5} - a_1 + a_2 = 0 \quad \iff \quad a_2 = \frac{1}{3} - \frac{1}{15} = \frac{2}{15},$$

ritrovando il valore già noto (e tre!). Per i successivi, le cose sono veramente pessime: è sufficiente osservare che di termini con x^7 dallo sviluppo del termine moltiplicato per a_1 ce ne sono 2: uno facendo il "quadrato" di $-\frac{x^2}{3}$ (moltiplicato per 1), e un altro facendo il quadrato di 1, e moltiplicandolo per $\frac{x^4}{5}$ (che è il termine successivo). Andando avanti, sono sempre di più le possibilità di ottenere x^{2k+1} dalla seconda sommatoria, e dalla terza e dalle successive...

CAPITOLO 1

Integrazione

1. L'area di un sottografico: strategia d'attacco

Lo sport di calcolare aree di figure geometriche è molto antico: i greci erano cultori accaniti, anche se avevano a disposizione solo strumenti geometrici. Questo rendeva anche il calcolo di un segmento di parabola un tour de force piuttosto impegnativo. Forse in origine questo interesse nasceva da problemi pratici, come la misurazione di appezzamenti di terreno per motivi economici e burocratici. Comunque, lo sport continuò ad essere praticato anche in secoli più recenti, e con rinnovato interesse dopo che Galileo mostrò che questo problema non era solo un passatempo ma portava alla risoluzione di problemi fisici importanti. La situazione fu rivoluzionata dall'invenzione del Calcolo, che rese la misurazione delle aree in molti casi un problema banale.

L'oggetto standard di cui vogliamo misurare l'area (o per essere più precisi, la cui area vogliamo definire) è il sottografico di una funzione positiva:

DEFINIZIONE 1.1. Sia $f : [a, b] \rightarrow [0, +\infty)$, cioè tale che $f(x) \geq 0$ per ogni $x \in [a, b]$. L'insieme

$$S_f = \{(x, y) \mid a \leq x \leq b, \quad 0 \leq y \leq f(x)\}$$

si chiama il *sottografico di f* nell'intervallo $[a, b]$.

Si noti che, date due funzioni f e g , il sottografico S_f di f è contenuto in S_g , sottografico di g , se e solo se $f \leq g$.

Vorremmo dare un senso all'area $\mathcal{A}(S_f)$ del sottografico di f . Per funzioni molto semplici potremmo ricorrere a risultati di geometria elementare. Ad esempio, se $f = C$ è una funzione costante positiva, il sottografico S_f in $[a, b]$ è semplicemente un rettangolo, e quindi potremmo definire $\mathcal{A}(S_f) = C(b - a)$ (ossia base per altezza). Se una funzione è costante a tratti, il suo sottografico è una unione di rettangoli, e anche in questo caso la definizione più naturale di area è la somma delle aree di questi rettangoli. E per funzioni più complicate?

DEFINIZIONE 1.2. Dato un intervallo $[a, b]$, una *partizione* di $[a, b]$ è un insieme $P = \{x_0, x_1, \dots, x_n\}$ tale che $a = x_0 < x_1 < \dots < x_n = b$. 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

$$f(x) = \begin{cases} \alpha_i & x \in [x_{i-1}, x_i) & i = 1, \dots, n - 1 \\ \alpha_n & x \in [x_{n-1}, x_n] \end{cases}$$

dove $\alpha_1, \dots, \alpha_n$ sono numeri reali.

Per una funzione f costante a tratti e non negativa il sottografico non è altro che un'unione finita di rettangoli, vedi Figura 1. Definiamo l'area di S_f come la 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 dell'intervallo $[x_{i-1}, x_i]$ (la scelta del punto è ininfluente dato che f è costante in $[x_{i-1}, x_i]$!).

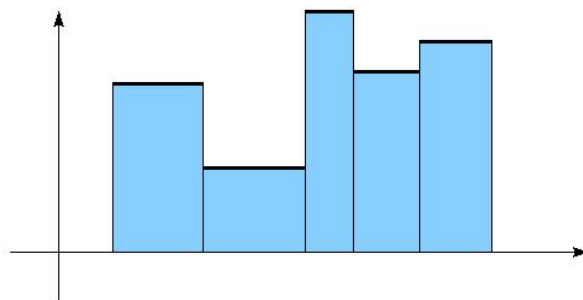


FIGURA 1. Una funzione costante a tratti non negativa ed il suo sottografo.

Per funzioni più generali, come potremmo definire $\mathcal{A}(S_f)$? Potremmo tentare di approssimare la funzione con funzioni costanti a tratti. Non è chiaro se sia possibile approssimare in modo efficiente ossia con errori piccoli a piacere. Per sicurezza, approssimiamo sia dal basso che dall'alto; ossia, data una funzione non negativa $f : [a, b] \rightarrow \mathbb{R}$, consideriamo due funzioni costanti a tratti g, h con $0 \leq g \leq f \leq h$; qualunque sia il numero $\mathcal{A}(S_f)$ che andiamo cercando, dovrà sicuramente valere $\mathcal{A}(S_g) \leq \mathcal{A}(S_f) \leq \mathcal{A}(S_h)$ (vedi Figura 2(a)), e se riusciamo ad avvicinare a piacere il primo e il terzo numero, vorrà dire che il nostro procedimento di approssimazione ha successo.

Notiamo subito che per poter trovare una $h \geq f$ occorre assumere che la funzione f sia superiormente limitata: se $\sup_{[a,b]} f(x) = +\infty$, non è possibile trovare una funzione costante a tratti h per cui $f \leq h$, vedi Figura 2(b). (L'ipotesi di limitatezza della funzione f è fondamentale qui, così come in tutto il Capitolo e in tutta la definizione dell'integrale definito. Invece il problema corrispondente per g non sussiste dato che stiamo considerando tutte funzioni positive).

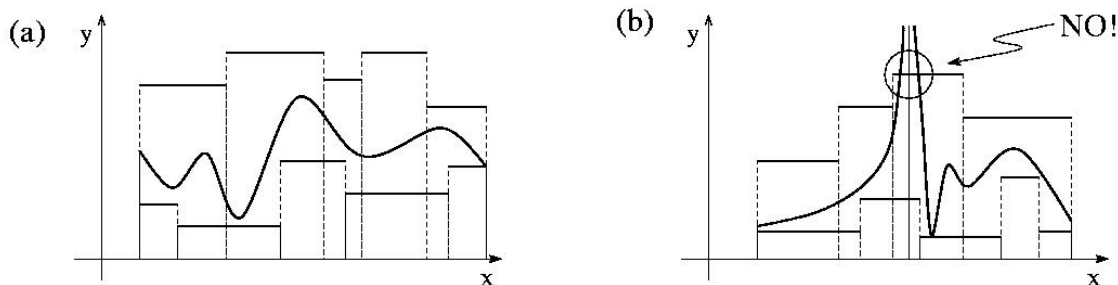


FIGURA 2. (a) Una funzione approssimata per eccesso e per difetto con due funzioni costanti a tratti. (b) Un tentativo (fallito) di approssimazione per eccesso di una funzione non limitata superiormente.

È chiaro che per semplificare i calcoli conviene scegliere funzioni g e h basate sulla stessa partizione P . Inoltre, approssimeremo f nel modo più preciso se scegliamo in

ciascun intervallo come valore della funzione h il sup di f sull'intervallo, e come valore della funzione g l'inf di f sull'intervallo (vedi Figura 3). Questo ci porta a considerare le seguenti funzioni costanti a tratti:

$$\underline{f}(x) = \alpha_i := \inf_{[x_{i-1}, x_i]} f(x) \quad \text{su } [x_{i-1}, x_i),$$

$$\bar{f}(x) = \beta_i := \sup_{[x_{i-1}, x_i]} f(x) \quad \text{su } [x_{i-1}, x_i).$$

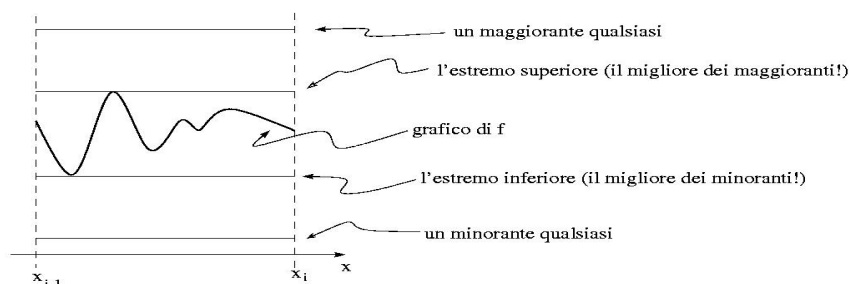


FIGURA 3. Stime qualsiasi e stime ottimali

In definitiva, data una partizione P dell'intervallo $[a, b]$, possiamo definire due numeri

$$\underline{\mathcal{A}}(S_f; P) := \mathcal{A}(S_{\underline{f}}) = \sum_{i=1}^n \alpha_i (x_i - x_{i-1}) \quad \bar{\mathcal{A}}(S_f; P) := \mathcal{A}(S_{\bar{f}}) = \sum_{i=1}^n \beta_i (x_i - x_{i-1}),$$

con $\alpha_i = \inf_{(x_{i-1}, x_i)} f(x)$, $\beta_i = \sup_{(x_{i-1}, x_i)} f(x)$ e ci aspettiamo che sia

$$\underline{\mathcal{A}}(S_f; P) \leq \mathcal{A}(S_f) \leq \bar{\mathcal{A}}(S_f; P)$$

Le due aree $\underline{\mathcal{A}}(S_f; P)$ e $\bar{\mathcal{A}}(S_f; P)$ sono colorate in azzurro nella Figura 4.

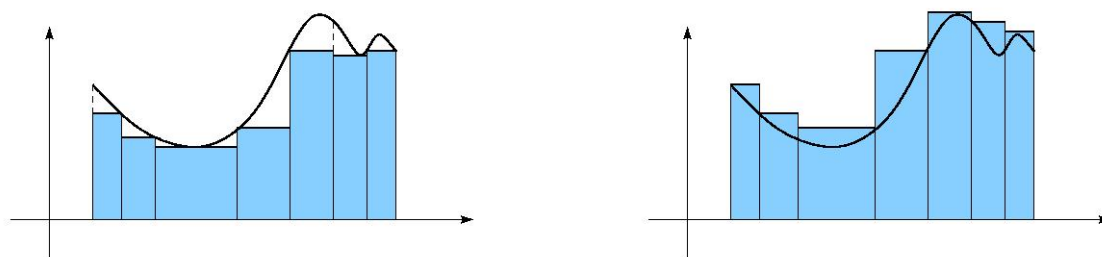


FIGURA 4. Due approssimazioni per S_f tramite le funzioni \underline{f} e \bar{f} .

Il passo seguente è migliorare l'approssimazione tramite la scelta di partizioni P_n con ampiezze sempre più piccole, ossia $|P_n| \rightarrow 0$ per $n \rightarrow \infty$. In questo modo, passando al limite nelle due successioni numeriche $\underline{\mathcal{A}}(S_f; P_n)$ e $\bar{\mathcal{A}}(S_f; P_n)$ si dovrebbe (incrociando le dita) ottenere proprio l'area richiesta. Proviamo a farlo per due esempi concreti.

Esempio I. Sia $f(x) = x$ per $x \in [a, b]$ con $0 \leq a < b$. Il sottografico della funzione f nell'intervallo $[a, b]$ è semplicemente un triangolo se $a = 0$, e se $a > 0$ è un trapezio. In entrambi i casi, la geometria elementare fornisce una formula per il calcolo dell'area, che è data da:

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

(“base maggiore più base minore per altezza diviso due”). Vediamo se il procedimento di approssimazione che abbiamo immaginato ci restituisce questo risultato o no.

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, si ha ovviamente

$$\alpha_k = \inf_{[x_{k-1}, x_k]} x = x_{k-1} = a + (k-1)h, \quad \beta_k = \sup_{[x_{k-1}, x_k]} x = x_k = a + kh.$$

Dato che $x_k - x_{k-1} = h$, otteniamo

$$\begin{aligned} \underline{\mathcal{A}}(S_f; P_n) &= ah + \dots + (a + (n-1)h)h = h \sum_{k=0}^{n-1} (a + kh) = hna + h^2 \sum_{k=0}^{n-1} k, \\ \overline{\mathcal{A}}(S_f; P_n) &= (a+h)h + \dots + (a+nh)h = h \sum_{k=1}^n (a + kh) = hna + h^2 \sum_{k=1}^n k. \end{aligned}$$

Ricordando la formula $\sum_{k=1}^n k = \frac{n(n+1)}{2}$ e osservando che $h = \frac{b-a}{n}$, concludiamo

$$\begin{aligned} \underline{\mathcal{A}}(S_f; P_n) &= hna + h^2 \frac{n(n-1)}{2} = (b-a)a + (b-a)^2 \frac{n(n-1)}{2n^2} \\ \overline{\mathcal{A}}(S_f; P_n) &= hna + h^2 \frac{n(n+1)}{2} = (b-a)a + (b-a)^2 \frac{n(n+1)}{2n^2}. \end{aligned}$$

Per $n \rightarrow +\infty$, le due quantità tendono allo stesso limite:

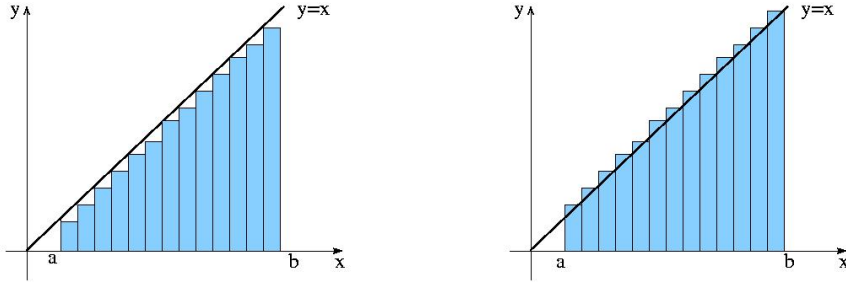


FIGURA 5. La funzione $f(x) = x$ e le approssimazioni determinate dai punti $a + \frac{k(b-a)}{n}$ con $k = 0, \dots, n$.

$$\lim_{n \rightarrow +\infty} \underline{\mathcal{A}}(E; P_n) = \lim_{n \rightarrow +\infty} \overline{\mathcal{A}}(E; P_n) = (b-a)a + \frac{(b-a)^2}{2} = \frac{b^2 - a^2}{2},$$

esattamente il valore previsto dalla geometria elementare.

Esempio II. Consideriamo $f(x) = x^2$ in $[a, b]$ con $0 \leq a < b$; ora la geometria elementare non ci aiuta (Archimede era riuscito a calcolare quest'area spingendo al limite i suoi metodi geometrici, va detto con un certo sforzo). Scegliamo la stessa partizione P_n di prima. Dato che x^2 è crescente su $[a, b]$ abbiamo

$$\alpha_i = \inf_{[x_{k-1}, x_k]} x^2 = (a + (k-1)h)^2, \quad \beta_i = \sup_{[x_{k-1}, x_k]} x^2 = (a + kh)^2,$$

da cui segue

$$\begin{aligned} \underline{\mathcal{A}}(S_f; P_n) &= a^2h + (a+h)^2h + (a+2h)^2h + \dots + (a+(n-1)h)^2h \\ \overline{\mathcal{A}}(S_f; P_n) &= (a+h)^2h + (a+2h)^2h + \dots + (a+nh)^2h \end{aligned}$$

Svolti i quadrati e tenuto presente che $h = \frac{b-a}{n}$, si ottiene

$$\begin{aligned}\underline{\mathcal{A}}(S_f; P_n) &= h \left\{ na^2 + 2ah \sum_{k=0}^{n-1} k + h^2 \sum_{k=0}^{n-1} k^2 \right\} \\ &= (b-a) \left\{ a^2 + \frac{2a(b-a)}{n^2} \sum_{k=0}^{n-1} k + \frac{(b-a)^2}{n^3} \sum_{k=0}^{n-1} k^2 \right\} \\ \overline{\mathcal{A}}(S_f; P_n) &= h \left\{ na^2 + 2ah \sum_{k=1}^n k + h^2 \sum_{k=1}^n k^2 \right\} \\ &= (b-a) \left\{ a^2 + \frac{2a(b-a)}{n^2} \sum_{k=1}^n k + \frac{(b-a)^2}{n^3} \sum_{k=1}^n k^2 \right\}\end{aligned}$$

Utilizzando le formule (si veda il primo Capitolo)

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

e sostituendo nelle precedenti,

$$\begin{aligned}\underline{\mathcal{A}}(S_f; P_n) &= (b-a) \left\{ a^2 + a(b-a) \frac{n-1}{n} + (b-a)^2 \frac{(n-1)(2n-1)}{6n^2} \right\} \\ \overline{\mathcal{A}}(S_f; P_n) &= (b-a) \left\{ a^2 + a(b-a) \frac{n+1}{n} + (b-a)^2 \frac{(n+1)(2n+1)}{6n^2} \right\}\end{aligned}$$

Quindi, passando al limite per $n \rightarrow \infty$,

$$\lim_{n \rightarrow +\infty} \underline{\mathcal{A}}(S_f; P_n) = \lim_{n \rightarrow +\infty} \overline{\mathcal{A}}(S_f; P_n) = (b-a)a^2 + a(b-a)^2 + \frac{(b-a)^3}{3} = \frac{b^3 - a^3}{3},$$

cioè

$$\mathcal{A}(S_f) = \frac{b^3}{3} - \frac{a^3}{3}.$$

Esempio III. Proviamo ora con una funzione molto più irregolare, la *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 qualunque partizione P si ha $\underline{\mathcal{A}}(S_D; P) = 0$ 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. Vediamo che in questo caso il nostro procedimento fallisce e non riusciamo ad assegnare un valore all'area del sottografico di $D(x)$.

Esempio IV. Come ultimo esempio, calcoliamo l'area del sottografico di $f(x) = \cos x$ nell'intervallo $[0, \pi/2]$. Prima di tutto dimostriamo il seguente Lemma.

LEMMA 1.3. Vale

$$(1.1) \quad \sum_{k=0}^n \cos(ky) = \frac{1}{2} \left[\frac{\operatorname{sen}(n + \frac{1}{2})y}{\operatorname{sen}(\frac{y}{2})} + 1 \right] \quad \forall y \in \mathcal{R}, \quad \forall n \in \mathcal{N}.$$

Dimostrazione. (Per induzione) Per $n = 0$, l'identità è banalmente verificata. Supponiamola vera per $n - 1$ e dimostriamola per n :

$$\begin{aligned} \sum_{k=0}^n \cos(ky) &= \frac{1}{2} \left[\frac{\operatorname{sen}(n - \frac{1}{2})y}{\operatorname{sen}(\frac{y}{2})} + 1 \right] + \cos(ny) \\ &= \frac{1}{2} \left[\frac{\operatorname{sen}(n - \frac{1}{2})y + 2 \cos(ny) \operatorname{sen}(\frac{y}{2})}{\operatorname{sen}(\frac{y}{2})} + 1 \right]. \end{aligned}$$

Dato che

$$\operatorname{sen} \left(n \pm \frac{1}{2} \right) y = \operatorname{sen}(ny) \cos \left(\frac{y}{2} \right) \mp \cos(ny) \operatorname{sen} \left(\frac{y}{2} \right),$$

ne segue la conclusione. ■

Per calcolare l'area di

$$S_f = \left\{ (x, y) \mid 0 \leq x \leq \frac{\pi}{2}, \quad 0 \leq y \leq \cos x \right\},$$

fissato $n \in \mathcal{N}$, scegliamo la partizione P_n determinata dai punti $x_k := \frac{k\pi}{2n}$ per $k = 0, \dots, n$. La funzione $\cos x$ è decrescente in $[0, \pi/2]$, quindi

$$\alpha_i = \inf_{[x_{k-1}, x_k]} \cos x = \cos \left(\frac{k\pi}{2n} \right), \quad \beta_i = \sup_{[x_{k-1}, x_k]} \cos x = \cos \left(\frac{(k-1)\pi}{2n} \right).$$

Perciò, utilizzando la formula (1.1) con $y = \frac{\pi}{2n}$ otteniamo

$$\begin{aligned} \underline{\mathcal{A}}(S_f; P_n) &= \frac{\pi}{2n} \sum_{k=1}^n \cos \left(\frac{k\pi}{2n} \right) = \frac{\pi}{4n} \left[\frac{\operatorname{sen}(n + \frac{1}{2})\frac{\pi}{2n}}{\operatorname{sen}(\frac{\pi}{4n})} - 1 \right] = \frac{\pi}{4n} \left[\frac{\operatorname{sen}(\frac{\pi}{2} + \frac{\pi}{4n})}{\operatorname{sen}(\frac{\pi}{4n})} - 1 \right] \\ \overline{\mathcal{A}}(S_f; P_n) &= \frac{\pi}{2n} \sum_{k=0}^{n-1} \cos \left(\frac{k\pi}{2n} \right) = \frac{\pi}{4n} \left[\frac{\operatorname{sen}(n + \frac{1}{2})\frac{\pi}{2n}}{\operatorname{sen}(\frac{\pi}{4n})} + 1 \right] = \frac{\pi}{4n} \left[\frac{\operatorname{sen}(\frac{\pi}{2} + \frac{\pi}{4n})}{\operatorname{sen}(\frac{\pi}{4n})} + 1 \right] \end{aligned}$$

Dato che $\operatorname{sen}(\frac{\pi}{2} + \alpha) = \cos \alpha$,

$$\underline{\mathcal{A}}(S_f; P_n) = \frac{\pi/4n}{\operatorname{tg}(\pi/4n)} - \frac{\pi}{4n}, \quad \overline{\mathcal{A}}(S_f; P_n) = \frac{\pi/4n}{\operatorname{tg}(\pi/4n)} + \frac{\pi}{4n}$$

Per $n \rightarrow +\infty$, $\lim_{n \rightarrow +\infty} \underline{\mathcal{A}}(S_f; P_n) = \lim_{n \rightarrow +\infty} \overline{\mathcal{A}}(S_f; P_n) = 1$ e quindi $\mathcal{A}(S_f) = 1$. Faticoso, eh?

ESERCIZIO 1.4. Dato $b > 0$, dimostrare che il sottografico di $\cos x$ in $[0, b]$ è misurabile e calcolarne l'area.

2. L'integrale di Riemann

Gli esempi del paragrafo precedente suggeriscono che il procedimento immaginato sia quello giusto per definire il numero $\mathcal{A}(S_f)$, almeno per molte funzioni interessanti, anche se sarebbe veramente opportuno trovare metodi più pratici per calcolarlo.

Dedichiamoci allora alla definizione rigorosa di $\mathcal{A}(S_f)$, il cui nome ufficiale è *l'integrale definito di Riemann*.

DEFINIZIONE 2.1. Una *partizione* dell'intervallo $[a, b]$ è un insieme finito di punti $P = \{x_0, x_1, \dots, x_n\}$ tali che $a = x_0 < x_1 < \dots < x_n = b$. L'*ampiezza* della partizione P è il numero $|P| := \max\{x_i - x_{i-1} : i = 1, \dots, n\}$. Diciamo che una partizione P_1 è *più fine* di un'altra partizione P_2 quando $P_2 \subset P_1$.

DEFINIZIONE 2.2. Sia $f : [a, b] \rightarrow \mathbb{R}$ limitata e sia $P = \{x_0, x_1, \dots, x_n\}$ una partizione dell'intervallo $[a, b]$. Poniamo

$$\alpha_i = \inf_{[x_{i-1}, x_i]} f(x) \quad \text{e} \quad \beta_i = \sup_{[x_{i-1}, x_i]} f(x).$$

Si chiamano *somma integrale inferiore e superiore di f* relative alla partizione P le somme

$$\underline{S}(f; P) = \sum_{i=1}^n \alpha_i (x_i - x_{i-1}) \quad \text{e} \quad \overline{S}(f; P) = \sum_{i=1}^n \beta_i (x_i - x_{i-1}).$$

Naturalmente, la somma parziale inferiore è semplicemente l'area $\mathcal{A}(S_{\underline{f}})$ della funzione costante a tratti $\underline{f} \leq f$ "ottimale" costruita nel Paragrafo 1:

$$\underline{S}(f; P) = \mathcal{A}(S_{\underline{f}}), \quad \underline{f}(x) = \sum_{i=1}^n \alpha_i (x_i - x_{i-1})$$

mentre la somma superiore è l'area associata alla funzione \overline{f}

$$\overline{S}(f; P) = \mathcal{A}(S_{\overline{f}}), \quad \overline{f}(x) = \sum_{i=1}^n \beta_i (x_i - x_{i-1}).$$

Vediamo quindi che stiamo seguendo proprio il procedimento descritto nel Paragrafo 1 per approssimare l'area del sottografico di f .

LEMMA 2.3. Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione limitata e siano P_1 e P_2 due partizioni di $[a, b]$. Supponiamo che $P_1 \subseteq P_2$; allora si ha

$$(2.1) \quad \underline{S}(f; P_1) \leq \underline{S}(f; P_2) \leq \overline{S}(f; P_2) \leq \overline{S}(f; P_1)$$

e quindi $\overline{S}(f; P_2) - \underline{S}(f; P_2) \leq \overline{S}(f; P_1) - \underline{S}(f; P_1)$.

Dimostrazione. Consideriamo prima il caso molto semplice in cui P_2 ha soltanto un punto in più rispetto a P_1 :

$$P_1 = \{x_0, \dots, x_n\}, \quad P_2 = P_1 \cup \{z\}, \quad x_{k-1} < z < x_k$$

per un certo k . Se chiamiamo come al solito α_k l'inf di f sull'intervallo $[x_{k-1}, x_k]$ e invece chiamiamo α'_k, α''_k rispettivamente gli inf di f sui due intervalli più piccoli $[x_{k-1}, z]$ e $[z, x_k]$, chiaramente abbiamo $\alpha_k \leq \alpha'_k$ e $\alpha_k \leq \alpha''_k$, quindi

$$\alpha_k (x_k - x_{k-1}) \leq \alpha'_k (z - x_{k-1}) + \alpha''_k (x_k - z).$$

Aggiungendo gli altri termini della somma parziale (che sono gli stessi per P_1 e P_2) otteniamo $\underline{S}(f; P_1) \leq \underline{S}(f; P_2)$. Le altre disuguaglianze si dimostrano nello stesso modo.

Nel caso generale P_2 si ottiene aggiungendo a P_1 un numero finito di punti, quindi il precedente ragionamento applicato un numero finito di volte ci dà la tesi (con il solito argomento per induzione).

OSSERVAZIONE 2.4. Se P_1, P_2 sono due partizioni qualunque di $[a, b]$, la loro unione $P = P_1 \cup P_2$ è più fine di entrambe, quindi dal lemma precedente otteniamo

$$\underline{S}(f, P_1) \leq \underline{S}(f, P) \leq \overline{S}(f, P) \leq \overline{S}(f, P_2).$$

In particolare si ha sempre

$$\underline{S}(f, P_1) \leq \overline{S}(f, P_2),$$

ossia *le somme inferiori sono sempre minori o uguali alle somme superiori*.

DEFINIZIONE 2.5. Sia $f : [a, b] \rightarrow \mathbb{R}$ limitata. Si chiamano *integrale inferiore e superiore secondo Riemann* della funzione f su $[a, b]$ i due numeri

$$\int_a^b f(x)dx = \sup_P \underline{S}(f, P), \quad \int_a^b f(x)dx = \inf_P \overline{S}(f, P)$$

dove inf e sup sono calcolati al variare di tutte le partizioni P dell'intervallo $[a, b]$.

Dall'osservazione precedente vediamo che l'integrale inferiore è sempre minore o uguale all'integrale superiore. Se i due integrali coincidono, diciamo che la funzione f è *integrabile secondo Riemann* sull'intervallo $[a, b]$, e scriviamo semplicemente

$$\int_a^b f(x)dx = \int_a^b f(x)dx = \int_a^b f(x)dx.$$

OSSERVAZIONE 2.6. Se la funzione f assume anche valori negativi, nelle somme parziali a questi valori corrispondono dei termini $\alpha_i(x_i - x_{i-1})$ negativi:

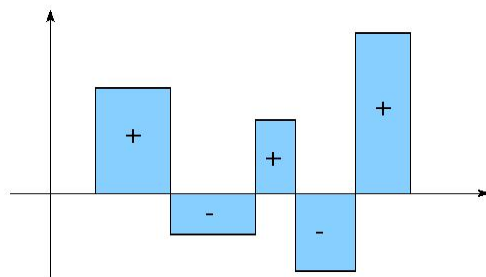


FIGURA 6. L'integrale è l'area *con segno*.

In generale, faremo la convenzione che l'area di una regione che si trova al di sotto dell'asse delle ascisse abbia un valore *negativo*. Per esempio, l'integrale della funzione

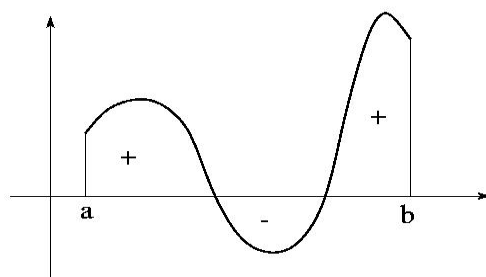


FIGURA 7. L'integrale è l'area *con segno*.

$\sin(x)$ da 0 a 2π fa esattamente 0: l'area da 0 a π è positiva, quella da π a 2π negativa, e le due aree si cancellano dando zero come totale. Il motivo di questa convenzione

è che in questo modo l'integrale definito ha proprietà molto più comode e utili: per esempio, l'integrale di una somma di funzioni è la somma degli integrali, l'integrale di $-f$ è l'opposto dell'integrale di f , eccetera.

Esempio. Attenzione: non tutte le funzioni limitate sono integrabili secondo Riemann. Un esempio è la funzione di Dirichlet $D(x)$ già studiata nel primo Paragrafo. Infatti per qualunque partizione P di $[0, 1]$, ciascun intervallo I_j della partizione contiene sia numeri razionali, sia irrazionali, e quindi i numeri $\alpha_j = \inf_{I_j} D(x)$ sono tutti uguali a 0 mentre i numeri $\beta_j = \sup_{I_j} D(x)$ sono tutti uguali a uno. Pertanto si ha

$$\underline{S}(D, P) = 0, \quad \overline{S}(D, P) = 1 \quad \text{per qualunque partizione } P$$

e quindi

$$\int_0^1 D(x)dx = 0, \quad \overline{\int_0^1 D(x)dx} = 1.$$

Vediamo che la funzione di Dirichlet $D(x)$ non è integrabile secondo Riemann su $[0, 1]$.

Come facciamo a riconoscere se una funzione è integrabile secondo Riemann?

PROPOSIZIONE 2.7 (Criterio di integrabilità). *Una funzione limitata $f : [a, b] \rightarrow \mathbb{R}$ è integrabile secondo Riemann su $[a, b]$ se e solo se per ogni $\epsilon > 0$ esiste una partizione P_ϵ di $[a, b]$ tale che*

$$(2.2) \quad \overline{S}(f, P_\epsilon) - \underline{S}(f, P_\epsilon) < \epsilon.$$

Dimostrazione. Se f è integrabile, fissato $\epsilon > 0$ possiamo trovare due partizioni P_1, P_2 tali che

$$\underline{S}(f; P_1) > \int_a^b f(x)dx - \frac{\epsilon}{2} \quad \text{e} \quad \overline{S}(f; P_2) < \int_a^b f(x)dx + \frac{\epsilon}{2}$$

perché l'integrale di f è uguale sia al sup delle somme inferiori sia all'inf delle somme superiori. Se ora scegliamo $P_\epsilon = P_1 \cup P_2$ abbiamo

$$\underline{S}(f; P_1) \leq \underline{S}(f; P_\epsilon) \leq \overline{S}(f; P_\epsilon) \leq \overline{S}(f; P_2)$$

e quindi

$$\overline{S}(f; P_\epsilon) - \underline{S}(f; P_\epsilon) \leq \overline{S}(f; P_2) - \underline{S}(f; P_1) < \int_a^b f(x)dx + \frac{\epsilon}{2} - \int_a^b f(x)dx + \frac{\epsilon}{2} = \epsilon.$$

Viceversa, supponiamo che valga la proprietà (2.2). Dato che

$$\inf_P \overline{S}(f; P) - \sup_P \underline{S}(f; P) \leq \overline{S}(f; P_\epsilon) - \underline{S}(f; P_\epsilon)$$

vediamo che per ogni $\epsilon > 0$ si ha

$$\inf_P \overline{S}(f; P) - \sup_P \underline{S}(f; P) < \epsilon$$

e quindi la differenza a primo membro deve essere uguale a zero, cioè f è integrabile.

Studiamo ora le proprietà principali dell'integrale di Riemann: la *linearità*

$$\int_a^b (cf(x) + g(x))dx = c \int_a^b f(x)dx + \int_a^b g(x)dx,$$

dove c è una costante; la *monotonia*

$$f \leq g \implies \int_a^b f(x)dx \leq \int_a^b g(x)dx,$$

l'*additività*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx, \quad a < c < b$$

e la proprietà

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx.$$

PROPOSIZIONE 2.8 (Linearità I). *Siano $f, g : [a, b] \rightarrow \mathbb{R}$ due funzioni integrabili su $[a, b]$. Allora anche $f + g$ è integrabile su $[a, b]$ e vale la formula*

$$\int_a^b (f(x) + g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx.$$

Dimostrazione. Notiamo che se due funzioni sono definite su un insieme A , vale sempre

$$\sup_A (f + g) \leq \sup_A f + \sup_A g.$$

Infatti per ogni $x \in A$ si ha $f(x) + g(x) \leq \sup_A f + \sup_A g$, e prendendo il sup in x otteniamo la proprietà.

Dimostriamo il risultato. Sia $P = \{x_0, \dots, x_N\}$ una partizione di $[a, b]$, e sia $I_j = [x_{j-1}, x_j]$. Allora abbiamo

$$(2.3) \quad \sup_{I_j} (f + g) \leq \sup_{I_j} f + \sup_{I_j} g.$$

Moltiplicando la (2.3) per $(x_j - x_{j-1})$ e sommando su j otteniamo

$$(2.4) \quad \overline{S}(f + g, P) \leq \overline{S}(f, P) + \overline{S}(g, P)$$

per qualunque partizione P .

Siano ora P_1, P_2 due partizioni, e sia $P = P_1 \cup P_2$. Dal Lemma 2.3 abbiamo

$$\overline{S}(f, P_1) + \overline{S}(g, P_2) \geq \overline{S}(f, P) + \overline{S}(g, P) \geq \overline{S}(f + g, P) \geq \int_a^{\overline{b}} (f + g)dx.$$

Prendendo l'inf al variare di tutte le partizioni P_1 e P_2 , otteniamo

$$\int_a^{\overline{b}} f(x)dx + \int_a^{\overline{b}} g(x)dx \geq \int_a^{\overline{b}} (f + g)dx.$$

Esattamente allo stesso modo si dimostra la disuguaglianza

$$\int_a^{\underline{a}} f(x)dx + \int_a^{\underline{a}} g(x)dx \leq \int_a^{\underline{a}} (f + g)dx.$$

Mettendo tutto insieme abbiamo

$$\int_a^{\underline{a}} f(x)dx + \int_a^{\underline{a}} g(x)dx \leq \int_a^{\underline{a}} (f + g)dx \leq \int_a^{\overline{b}} (f + g)dx \leq \int_a^{\overline{b}} f(x)dx + \int_a^{\overline{b}} g(x)dx$$

Dato che f, g per ipotesi sono integrabili, il primo e l'ultimo termine coincidono e quindi queste disuguaglianze sono tutte uguaglianze. Ne segue che anche $f + g$ è integrabile, e l'integrale coincide con la somma degli integrali.

PROPOSIZIONE 2.9 (Linearità II). *Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione integrabile su $[a, b]$ e sia $c \in \mathbb{R}$. Allora anche la funzione cf è integrabile su $[a, b]$ e $\int_a^b cf(x)dx = c \int_a^b f(x)dx$.*

Dimostrazione. Se $c = 0$ non c'è nulla da dimostrare.

Supponiamo $c > 0$. Presa una qualunque partizione $P = \{x_0, \dots, x_N\}$, abbiamo

$$\sup_{[x_{j-1}, x_j]} cf = c \cdot \sup_{[x_{j-1}, x_j]} f$$

e moltiplicando per $(x_j - x_{j-1})$ e sommando su j otteniamo

$$\overline{S}(cf, P) = c \cdot \overline{S}(f, P).$$

Prendendo l'inf al variare di tutte le partizioni P , otteniamo quindi

$$\overline{\int_a^b} (cf)dx = c \overline{\int_a^b} f dx.$$

Esattamente allo stesso modo si dimostra che

$$\underline{\int_a^b} (cf)dx = c \underline{\int_a^b} f dx$$

ed essendo f integrabile otteniamo subito la tesi.

Se invece $c < 0$ abbiamo

$$\sup_{[x_{j-1}, x_j]} cf = c \cdot \inf_{[x_{j-1}, x_j]} f$$

e quindi come prima

$$\overline{S}(cf, P) = c \cdot \underline{S}(f, P).$$

Prendendo l'inf su tutte le partizioni P otteniamo

$$\overline{\int_a^b} (cf)dx = \inf_P [c \cdot \underline{S}(f, P)] = c \cdot \sup_P \underline{S}(f, P) = c \underline{\int_a^b} f dx.$$

Allo stesso modo si dimostra che

$$\underline{\int_a^b} (cf)dx = c \overline{\int_a^b} f dx$$

ed essendo f integrabile otteniamo subito la tesi.

PROPOSIZIONE 2.10 (Monotonia). *Siano $f, g : [a, b] \rightarrow \mathbb{R}$ integrabili secondo Riemann su $[a, b]$, e supponiamo che $f(x) \leq g(x)$ per tutti gli x . Allora si ha $\int_a^b f(x)dx \leq \int_a^b g(x)dx$.*

Dimostrazione. Basta osservare che, per qualunque partizione $P = \{x_0, \dots, x_N\}$,

$$\inf_{[x_{j-1}, x_j]} f \leq \inf_{[x_{j-1}, x_j]} g$$

e quindi

$$\underline{S}(f, P) \leq \underline{S}(g, P).$$

Prendendo il sup in P otteniamo la tesi.

PROPOSIZIONE 2.11 (Additività). *Sia $f : [a, b] \rightarrow \mathbb{R}$ limitata e sia c un punto di (a, b) . f è integrabile secondo Riemann su $[a, b]$ se e solo se è integrabile sia su $[a, c]$ sia su $[c, b]$. In tal caso vale la formula*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Dimostrazione. Supponiamo che f sia integrabile su $[a, b]$. Per il criterio di integrabilità, dato $\epsilon > 0$ esiste una partizione P_ϵ di $[a, b]$ tale che

$$\overline{S}(f, P_\epsilon) - \underline{S}(f, P_\epsilon) < \epsilon.$$

Consideriamo la partizione $P = P_\epsilon \cup \{c\}$. Dal Lemma 2.3 abbiamo

$$\overline{S}(f, P) - \underline{S}(f, P) \leq \overline{S}(f, P_\epsilon) - \underline{S}(f, P_\epsilon) < \epsilon.$$

Notiamo che si ha

$$\overline{S}(f, P) - \underline{S}(f, P) = \sum_{j=1}^N (\beta_j - \alpha_j)(x_j - x_{j-1})$$

e tutti i termini della somma sono positivi; quindi se scriviamo $P = P_1 \cup P_2$ dove P_1 è una partizione di $[a, c]$ e P_2 una partizione di $[c, b]$, e scartiamo dalla somma tutti i termini relativi all'intervallo $[c, b]$, otteniamo

$$\overline{S}(f, P_1) - \underline{S}(f, P_1) \leq \overline{S}(f, P) - \underline{S}(f, P) < \epsilon.$$

Quindi abbiamo verificato il criterio di integrabilità sull'intervallo $[a, c]$, ossia la funzione f è integrabile secondo Riemann sull'intervallo $[a, c]$. Analogamente si dimostra che f è integrabile su $[c, b]$.

Viceversa, supponiamo che f sia integrabile su $[a, c]$ e su $[c, b]$. Allora per il solito criterio, fissato $\epsilon > 0$, possiamo trovare una partizione P_1 di $[a, c]$ e una partizione P_2 di $[c, b]$ tali che

$$\overline{S}(f, P_1) - \underline{S}(f, P_1) < \epsilon, \quad \overline{S}(f, P_2) - \underline{S}(f, P_2) < \epsilon.$$

Se uniamo le due partizioni otteniamo una partizione $P = P_1 \cup P_2$ di $[a, b]$ e si ha subito

$$\overline{S}(f, P) - \underline{S}(f, P) = \overline{S}(f, P_1) - \underline{S}(f, P_1) + \overline{S}(f, P_2) - \underline{S}(f, P_2) < 2\epsilon.$$

Pertanto il criterio è verificato anche su $[a, b]$ e la funzione f è integrabile su $[a, b]$.

Resta da dimostrare la formula. Sia P una qualunque partizione di $[a, b]$, sia $P' = P \cup \{c\}$, e scriviamo $P' = P_1 \cup P_2$ come unione di una partizione P_1 di $[a, c]$ e una partizione P_2 di $[c, b]$. Abbiamo allora

$$\overline{\int_a^c f dx} + \overline{\int_c^b f dx} \leq \overline{S}(f, P_1) + \overline{S}(f, P_2) = \overline{S}(f, P') \leq \overline{S}(f, P)$$

e prendendo l'inf in P otteniamo

$$\overline{\int_a^c f dx} + \overline{\int_c^b f dx} \leq \overline{\int_a^b f dx} \quad \implies \quad \int_a^c f dx + \int_c^b f dx \leq \int_a^b f dx.$$

La disuguaglianza opposta si dimostra allo stesso modo usando le somme inferiori.

È prassi comune estendere la notazione dell'integrale definito ponendo

$$\int_b^a f(x)dx := - \int_a^b f(x)dx, \quad \int_a^a f(x)dx = 0.$$

Con queste definizioni, il simbolo \int_a^b ha senso sia se $a < b$, sia se $a = b$, sia se $a > b$. Si vede facilmente che allora la formula

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

è valida per tre punti a, b, c in qualunque posizione, non necessariamente nell'ordine $a < c < b$.

PROPOSIZIONE 2.12. *Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione integrabile secondo Riemann su $[a, b]$. Allora anche la funzione $|f|$ è integrabile su $[a, b]$, e vale la disuguaglianza*

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx.$$

Dimostrazione. Notiamo che se $f : A \rightarrow \mathbb{R}$, si ha sempre

$$\sup_A |f| - \inf_A |f| \leq \sup_A f - \inf_A f.$$

(anche se f assume valori negativi!). Infatti per tutti i punti $x, y \in A$ si ha

$$f(x) - f(y) \leq \sup_A f - \inf_A f \quad \text{e} \quad f(y) - f(x) \leq \sup_A f - \inf_A f$$

e questo vuol dire $|f(x) - f(y)| \leq \sup_A f - \inf_A f$. Ma allora dalla disuguaglianza triangolare abbiamo

$$|f(x)| - |f(y)| \leq |f(x) - f(y)| \leq \sup_A f - \inf_A f$$

e quindi

$$\sup_A |f| - \inf_A |f| = \sup_{x, y \in A} [|f(x)| - |f(y)|] \leq \sup_A f - \inf_A f$$

come volevasi dimostrare.

Sia $P = \{x_0, \dots, x_N\}$ una qualunque partizione di $[a, b]$ e sia I_j l'intervallo $[x_{j-1}, x_j]$. Come dimostrato sopra, possiamo scrivere

$$\sup_{I_j} |f| - \inf_{I_j} |f| \leq \sup_{I_j} f - \inf_{I_j} f.$$

Ora moltiplichiamo la disuguaglianza per $(x_j - x_{j-1})$ e sommiamo su j . Otteniamo che per qualunque partizione P vale la disuguaglianza

$$\overline{S}(|f|, P) - \underline{S}(|f|, P) \leq \overline{S}(f, P) - \underline{S}(f, P).$$

f è integrabile per ipotesi, quindi, grazie al criterio di integrabilità, per ogni $\epsilon > 0$ esiste una partizione P_ϵ tale che

$$\overline{S}(|f|, P_\epsilon) - \underline{S}(|f|, P_\epsilon) \leq \overline{S}(f, P_\epsilon) - \underline{S}(f, P_\epsilon) < \epsilon$$

e quindi, sempre per il criterio, anche $|f|$ è integrabile.

Infine notiamo che dalla disuguaglianza

$$-|f(x)| \leq f(x) \leq |f(x)|$$

e dalla monotonia dell'integrale si ha subito

$$-\int_a^b |f(x)|dx \leq \int_a^b f(x)dx \leq \int_a^b |f(x)|dx$$

e questa è esattamente la disuguaglianza da dimostrare.

ESERCIZIO 2.1. Sapreste trovare una funzione *non* integrabile secondo Riemann ma tale che $|f|$ invece è integrabile? (Suggerimento: è una variazione della funzione di Dirichlet $D(x)$, che assume valori $+1$ e -1 invece che $+1$ e 0 . Quando se ne prende il modulo la funzione ottenuta vale sempre $+1$...)

3. Classi di funzioni integrabili

Una volta definito l'integrale e studiate le sue proprietà principali, il secondo problema è determinare classi sufficientemente ampie di funzioni integrabili, in modo da potere applicare la definizione. Per dimostrare che una certa funzione f è integrabile useremo sempre il Criterio di Integrabilità: mostreremo che per qualunque $\epsilon > 0$ si può trovare una partizione P_ϵ di $[a, b]$ tale che

$$\overline{S}(f; P_\epsilon) - \underline{S}(f; P_\epsilon) \leq C\epsilon$$

dove C è una costante fissata (ma che non dipende da ϵ o dalla partizione!).

Integrabilità delle funzioni monotone. La prima classe che consideriamo è quella delle funzioni monotone.

TEOREMA 3.1. *Se $f : [a, b] \rightarrow \mathbb{R}$ monotona, allora essa è integrabile in $[a, b]$.*

Dimostrazione. Consideriamo il caso di una funzione crescente (il caso decrescente è analogo). Anzitutto, la limitatezza di f è automaticamente verificata, dato che

$$f(a) \leq f(x) \leq f(b) \quad \forall x \in [a, b].$$

Inoltre, data una partizione $P = \{x_0, \dots, x_n\}$, sup e inf di f sugli intervalli $I_j = [x_{j-1}, x_j]$ sono raggiunti agli estremi:

$$\alpha_j = \inf_{I_j} f(x) = f(x_{j-1}) \quad \text{e} \quad \beta_j = \sup_{I_j} f(x) = f(x_j).$$

Quindi le somme integrali di f si scrivono come

$$\underline{S}(f; P) = \sum_{i=1}^n f(x_{i-1})(x_i - x_{i-1}) \quad \text{e} \quad \overline{S}(f; P) = \sum_{i=1}^n f(x_i)(x_i - x_{i-1}).$$

e la loro differenza si può stimare nel modo seguente:

$$\overline{S}(f; P) - \underline{S}(f; P) = \sum_{i=1}^n [f(x_i) - f(x_{i-1})] (x_i - x_{i-1})$$

Indicando con $|P|$ l'ampiezza di P , cioè il massimo delle lunghezze $x_i - x_{i-1}$, si ha

$$\overline{S}(f; P) - \underline{S}(f; P) \leq |P| \sum_{i=1}^n [f(x_i) - f(x_{i-1})] = |P| [f(b) - f(a)].$$

A questo punto, per ogni $\epsilon > 0$ basta prendere una partizione P_ϵ con ampiezza $|P_\epsilon| < \epsilon$ per ottenere

$$\overline{S}(f; P_\epsilon) - \underline{S}(f; P_\epsilon) < (f(b) - f(a))\epsilon$$

e per il criterio di integrabilità possiamo concludere che f è integrabile. ■

A partire dalle funzioni monotone è possibile costruire una classe molto ampia di funzioni integrabili. Consideriamo ad esempio una funzione f definita in $[a, b]$ tale che, per qualche $c \in [a, b]$, la funzione f è crescente in $[a, c]$ e decrescente in $[c, b]$. Dal Teorema precedente sappiamo che f è integrabile su $[a, c]$ ed è integrabile su $[c, b]$; e quindi dalla Proposizione 2.11 concludiamo che f è integrabile su tutto $[a, b]$.

Con una costruzione analoga, si mostra che *tutte le funzioni con un numero finito di cambi di monotonia sono integrabili*. Si noti che quasi tutte le funzioni considerate in questo corso rientrano in questa classe. Ad esempio, tutti i polinomi e, più in generale, tutte le funzioni razionali sono integrabili secondo Riemann su un intervallo $[a, b]$ (se il denominatore non si annulla su $[a, b]$).

Integrabilità delle funzioni continue. La seconda classe che consideriamo è quella delle funzioni continue. Introduciamo anzitutto una nozione leggermente più forte della continuità.

Ricordiamo che f è continua in un punto y_0 se

$$\lim_{x \rightarrow y_0} f(x) = f(y_0).$$

Dalla definizione di limite, questo equivale a dire: per ogni $\epsilon > 0$ esiste $\delta > 0$ tale che

$$(3.1) \quad |f(x) - f(y_0)| < \epsilon \quad \text{per tutti gli } x \in I \text{ tali che } |x - y_0| < \delta.$$

La nuova definizione è la seguente:

DEFINIZIONE 3.2. Sia $f : I \rightarrow \mathbb{R}$ una funzione definita su un intervallo $I \subseteq \mathbb{R}$. La funzione si dice *uniformemente continua su I* se per ogni $\epsilon > 0$ esiste $\delta > 0$ tale che

$$(3.2) \quad |f(x) - f(y)| < \epsilon \quad \text{per tutti gli } x, y \in I \text{ tali che } |x - y| < \delta.$$

Confrontando la (3.1) con la (3.2), vediamo subito che se una funzione è uniformemente continua in un intervallo allora essa è anche continua in ogni punto y_0 dell'intervallo: basta “tenere fermo” il punto $y = y_0$ e applicare la definizione.

Ma la proprietà (3.2) è *più forte* della (3.1). Infatti il numero δ nella (3.1) dipende dal punto y_0 ; cambiando punto y_0 , potrebbe essere necessario rimpicciolire δ perché la (3.1) sia soddisfatta. Al contrario, nella (3.2) il parametro δ è *lo stesso* per tutti i punti y dell'intervallo: la funzione è “ugualmente continua” in tutti i punti.

C'è un caso particolarmente importante in cui le due definizioni *coincidono*: quando l'intervallo I è *chiuso e limitato*, ossia $I = [a, b]$. Vediamo i dettagli.

TEOREMA 3.3 (Heine–Borel). *Una funzione continua su un intervallo $[a, b]$ chiuso e limitato è anche uniformemente continua*

Dimostrazione. Procediamo per assurdo. Supponiamo che f non sia uniformemente continua. Negando la definizione, otteniamo: esiste $\epsilon_0 > 0$ tale che per ogni $\delta > 0$ possiamo trovare $x, y \in [a, b]$ che verificano

$$|x - y| < \delta \quad \text{ma} \quad |f(x) - f(y)| \geq \epsilon_0.$$

Applichiamo questa proprietà scegliendo successivamente $\delta = \frac{1}{n}$ con $n \geq 1$ intero; quindi per ogni n otteniamo due punti $x_n, y_n \in [a, b]$ tali che

$$|x_n - y_n| < \frac{1}{n} \quad \text{ma} \quad |f(x_n) - f(y_n)| \geq \epsilon_0.$$

Per il Teorema di Bolzano–Weierstrass, possiamo estrarre da x_n una sottosuccessione x_{k_n} convergente

$$x_{k_n} \rightarrow \bar{x} \in [a, b].$$

Analogamente, possiamo estrarre da y_{k_n} una sotto-sottosuccessione convergente

$$y_{h_{k_n}} \rightarrow \bar{y} \in [a, b].$$

Naturalmente abbiamo anche $x_{h_{k_n}} \rightarrow \bar{x}$. Osserviamo adesso che per costruzione

$$|x_{h_{k_n}} - y_{h_{k_n}}| < \frac{1}{h_{k_n}} \rightarrow 0$$

e quindi passando al limite otteniamo

$$|\bar{x} - \bar{y}| = 0 \quad \text{ossia} \quad \bar{x} = \bar{y}.$$

In altri termini, le due successioni estratte convergono allo stesso punto $\bar{x} = \bar{y}$. D'altra parte abbiamo anche

$$|f(x_{h_{k_n}}) - f(y_{h_{k_n}})| \geq \epsilon_0$$

e dato che la f è continua, passando al limite otteniamo

$$|f(\bar{x}) - f(\bar{x})| \geq \epsilon_0 > 0$$

il che evidentemente è assurdo. ■

Armati del Teorema di Heine–Borel, siamo in grado di dimostrare il risultato seguente:

TEOREMA 3.4. *Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua. Allora f è integrabile secondo Riemann su $[a, b]$.*

Dimostrazione. Se P è una partizione di $[a, b]$, di punti $a = x_0 < x_1 < \dots < x_N = b$, possiamo scrivere come al solito

$$\bar{S}(f, P) - \underline{S}(f, P) = \sum_{i=1}^N (\beta_i - \alpha_i)(x_i - x_{i-1})$$

dove

$$\alpha_i = \inf_{I_i} f, \quad \beta_i = \sup_{I_i} f, \quad I_i = [x_{i-1}, x_i].$$

Dato che f è continua, per il Teorema di Weierstrass essa ha massimo e minimo su ciascun intervallo I_i , quindi possiamo trovare $p_i, q_i \in I_i$ per cui

$$\alpha_i = f(p_i), \quad \beta_i = f(q_i).$$

Notiamo che p_i e q_i appartengono allo stesso intervallo I_i e quindi abbiamo

$$|p_i - q_i| \leq |P|$$

dove $|P|$ è l'ampiezza della partizione P .

Ora, fissiamo $\epsilon > 0$ e applichiamo ora il Teorema di Heine–Borel: la funzione f è in realtà uniformemente continua, quindi esiste $\delta > 0$ tale che se $|x - y| < \delta$ allora $|f(x) - f(y)| < \epsilon$. Se scegliamo una qualunque partizione di ampiezza $|P| < \delta$, mettendo tutto insieme otteniamo

$$|p_i - q_i| \leq |P| < \delta \quad \text{e quindi} \quad \beta_i - \alpha_i = |f(q_i) - f(p_i)| < \epsilon.$$

In conclusione

$$\bar{S}(f, P) - \underline{S}(f, P) \leq \sum_{i=1}^N (\beta_i - \alpha_i)(x_i - x_{i-1}) \leq \epsilon \sum_{i=1}^N (x_i - x_{i-1}) = \epsilon(b - a)$$

e quindi f è integrabile su $[a, b]$. ■

4. Teorema della media integrale

Avviciniamoci al risultato più importante della teoria (il Teorema Fondamentale del Calcolo). Un passaggio importante è il risultato seguente.

TEOREMA 4.1 (Teorema della Media Integrale). *Sia $f : [a, b] \rightarrow \mathbb{R}$ continua. Allora esiste $\xi \in [a, b]$ tale che*

$$(4.1) \quad f(\xi) = \frac{1}{b-a} \int_a^b f(x) dx.$$

Dimostrazione. Per il teorema di Weierstrass, esistono due punti $p, q \in [a, b]$ tali che

$$f(p) = m = \min_{[a,b]} f, \quad f(q) = M = \max_{[a,b]} f.$$

La funzione f soddisfa

$$m \leq f(x) \leq M \quad \text{per tutti gli } x \in [a, b].$$

Quindi, per la monotonia dell'integrale, possiamo scrivere

$$m(b-a) = \int_a^b m dx \leq \int_a^b f(x) dx \leq \int_a^b M dx = M(b-a)$$

e dividendo per $b-a$ otteniamo

$$m \leq \frac{1}{b-a} \int_a^b f(x) dx \leq M.$$

In altri termini, il numero

$$\mu = \frac{1}{b-a} \int_a^b f(x) dx$$

è intermedio fra il minimo m e il massimo M di f . Ma allora, per il Teorema dei Valori Intermedi, esiste un punto $\xi \in [a, b]$ nel quale la funzione assume il valore μ e questo conclude la dimostrazione. ■

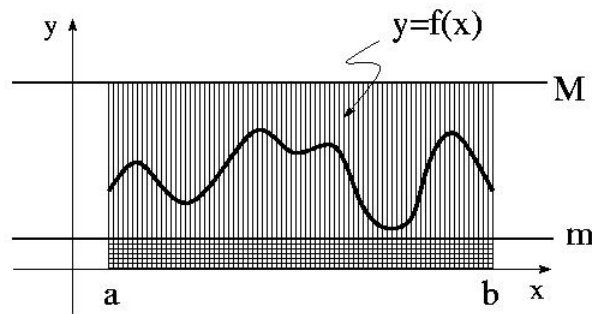


FIGURA 8. Significato geometrico della stima $m(b-a) \leq \int_a^b f(x) dx \leq M(b-a)$.

Il numero

$$\mu = \frac{1}{b-a} \int_a^b f(x) dx$$

si chiama la *media integrale* di f su $[a, b]$. Il Teorema precedente si può riassumere dicendo: la media integrale di una funzione continua f è un valore intermedio fra il minimo e il massimo di f , e quindi tale valore viene assunto dalla funzione almeno in

un punto dell'intervallo. Nel caso di una funzione non negativa f , il teorema equivale ad affermare che esiste un rettangolo di base $[a, b]$ ed altezza $f(\xi)$ che ha la stessa area del sottografico di f in $[a, b]$.

Si noti che la formula (4.1) è valida anche scambiando b con a , semplicemente perché

$$\frac{1}{b-a} \int_a^b f(x) dx = \frac{1}{a-b} \int_b^a f(x) dx.$$

Controesempio. Se la funzione f non è continua su tutto $[a, b]$, il Teorema della media potrebbe essere falso. Ad esempio, si consideri la funzione

$$f(x) = \operatorname{sgn} x = \begin{cases} -1 & x < 0, \\ 0 & x = 0, \\ +1 & x > 0, \end{cases}$$

definita sull'intervallo $[-1, 2]$. Allora f è integrabile (è monotona), e si ha

$$\mu = \frac{1}{2 - (-1)} \int_{-1}^2 \operatorname{sgn}(x) dx = \frac{1}{3} (-1 + 2) = \frac{1}{3}.$$

Tuttavia, $\frac{1}{3}$ non appartiene all'immagine della funzione $\operatorname{sgn} x$.

Il nome “media integrale” discende dal fatto che il valore μ è una naturale generalizzazione della media aritmetica. La *media aritmetica* di n quantità f_1, f_2, \dots, f_n si definisce come

$$\frac{f_1 + f_2 + \dots + f_n}{n}.$$

Nella costruzione dell'integrale di una funzione in un intervallo, se usiamo una partizione in n intervalli della stessa lunghezza $(b-a)/n$ e scegliamo dei punti $\xi_i \in (x_{i-1}, x_i)$, otteniamo come valore approssimato dell'integrale la quantità

$$\sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) = (b-a) \frac{f(\xi_1) + f(\xi_2) + \dots + f(\xi_n)}{n}$$

La media aritmetica di $f(\xi_1), \dots, f(\xi_n)$ è pari a

$$\frac{f(\xi_1) + \dots + f(\xi_n)}{n} = \frac{1}{b-a} \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}),$$

che, per $x_i - x_{i-1} \rightarrow 0$ converge proprio alla media integrale.

5. Teorema fondamentale del calcolo integrale

Arriviamo ora al risultato più importante del calcolo (non a caso si chiama il Teorema Fondamentale). Esso rivela che, in un senso opportuno, il calcolo di un integrale è l'operazione inversa della derivata. Dato che siamo in grado di derivare tutte le funzioni elementari, questo risultato ci mette a disposizione una quantità enorme di metodi di integrazione: basta leggere al contrario una tabella di derivate per avere una tabella di integrali.

L'idea chiave è la seguente: data una funzione integrabile $f : [a, b] \rightarrow \mathbb{R}$, invece di considerare l'integrale definito su $[a, b]$ cioè il numero $\int_a^b f(x) dx$, definiamo una nuova funzione facendo variare uno degli estremi di integrazione:

$$(5.1) \quad \phi(x) = \int_a^x f(t) dt, \quad x \in [a, b].$$

La nuova funzione $\phi : [a, b] \rightarrow \mathbb{R}$ si chiama la *funzione integrale* di f .

Alcune proprietà della funzione integrale si verificano facilmente. Ad esempio, dato che f è integrabile e quindi limitata, ossia $|f(x)| \leq M$ per una opportuna costante M , possiamo scrivere per tutti i punti $x < y$ nell'intervallo $[a, b]$

$$|\phi(y) - \phi(x)| = \left| \int_x^y \phi(t) dt \right| \leq \int_x^y |\phi(t)| dt \leq \int_x^y M dt = M|x - y|.$$

Scopriamo così che la funzione integrale di una f integrabile è Lipschitziana, con una costante di Lipschitz uguale alla costante M che maggiora $|f|$. In particolare, ϕ è sempre una funzione *continua*.

Ma si può dire molto di più. Mostriamo adesso che se f è una funzione continua allora la sua funzione integrale è addirittura *derivabile*, e la derivata di ϕ è proprio f . Introduciamo un nuovo termine: diciamo che

$$F \text{ è una } \textit{primitiva} \text{ di } f \text{ se } F' = f.$$

Ad esempio, $\sin x$ è una primitiva di $\cos x$, e $\ln x$ è una primitiva di $\frac{1}{x}$.

TEOREMA 5.1 (Teorema Fondamentale del Calcolo Integrale). *Data una funzione continua $f : [a, b] \rightarrow \mathbb{R}$ e un punto $\alpha \in [a, b]$, sia*

$$\phi(x) = \int_{\alpha}^x f(t) dt \quad x \in [a, b].$$

Allora la funzione ϕ è derivabile in (a, b) e $\phi'(x) = f(x)$, ossia ϕ è una primitiva di f . Tutte le altre primitive F di f si scrivono nella forma

$$F(x) = \phi(x) + C = \int_{\alpha}^x f(t) dt + C$$

dove C è una costante.

Dimostrazione. Il rapporto incrementale di ϕ nel punto x è uguale a

$$\frac{\phi(x+h) - \phi(x)}{h} = \frac{1}{h} \left[\int_{\alpha}^{x+h} f(t) dt - \int_{\alpha}^x f(t) dt \right] = \frac{1}{h} \int_x^{x+h} f(t) dt$$

(l'ultima uguaglianza è conseguenza delle proprietà di additività dell'integrale). Dato che la funzione f è continua, è possibile applicare il Teorema della Media Integrale, sull'intervallo $[x, x+h]$ quando $h > 0$ oppure sull'intervallo $[x+h, x]$ quando $h < 0$: otteniamo che esiste un punto ξ compreso tra x e $x+h$ tale che

$$\frac{\phi(x+h) - \phi(x)}{h} = f(\xi) \quad (\xi \text{ compreso tra } x \text{ e } x+h).$$

Per h che tende a 0, ξ tende ad x e, dato che f è continua, $\lim_{h \rightarrow 0} f(\xi) = f(x)$. In conclusione

$$\phi'(x) = \lim_{h \rightarrow 0} \frac{\phi(x+h) - \phi(x)}{h} = \lim_{h \rightarrow 0} f(\xi) = f(x)$$

e quindi ϕ è effettivamente una primitiva di f .

Supponiamo ora di avere una seconda primitiva F di f . Questo vuol dire che

$$F' = f \text{ e } \phi' = f, \text{ e quindi } (F - \phi)' = f - f = 0$$

ossia la differenza $F - \phi$ ha derivata identicamente nulla e quindi è una funzione costante C , e questo dimostra anche l'ultima affermazione. ■

ESEMPIO 5.2. Naturalmente se f è integrabile ma non continua, la sua funzione integrale potrebbe essere non derivabile (anche se è sicuramente continua e anzi Lipschitziana, come dimostrato sopra). Consideriamo ad esempio la funzione $f(x) = \operatorname{sgn} x$, che è continua a tratti con un unico punto di salto in $x = 0$. Calcoliamo la sua funzione integrale: si ha subito

$$\begin{aligned} x > 0 &\Rightarrow \phi(x) = \int_0^x \operatorname{sgn}(t) dt = \int_0^x 1 dt = x \\ x < 0 &\Rightarrow \phi(x) = \int_0^x \operatorname{sgn}(t) dt = \int_0^x (-1) dt = -x. \end{aligned}$$

Quindi la funzione $\phi(x)$ è la funzione modulo: $\phi(x) = |x|$, che non è derivabile in $x = 0$.

Il Teorema fondamentale del calcolo afferma che se f è una funzione continua la sua funzione integrale ϕ è una primitiva di f cioè soddisfa $\phi' = f$. In pratica, il Teorema afferma che la derivazione è l'operazione inversa dell'integrazione (in un senso opportuno).

L'insieme di tutte le primitive della funzione f si indica con

$$\int f(x) dx \quad \text{e spesso si scrive esplicitamente} \quad \int f(x) dx + C,$$

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, però tutte le funzioni della classe differiscono tra loro solo per una costante additiva.

Attenzione: conosciamo già moltissime primitive! (È questo il bello del Teorema Fondamentale del Calcolo). Ogni regola di derivazione che conosciamo ci fornisce una regola per calcolare una primitiva. Esempi:

$$\begin{aligned} (\operatorname{sen} x)' = \cos x &\quad \text{equivale a} \quad \int \cos x dx = \operatorname{sen} x + C \\ (\ln x)' = \frac{1}{x} &\quad \text{equivale a} \quad \int \frac{1}{x} dx = \ln x + C \\ (\arctan x)' = \frac{1}{1+x^2} &\quad \text{equivale a} \quad \int \frac{1}{1+x^2} dx = \arctan x + C \end{aligned}$$

eccetera. Studieremo come calcolare integrali indefiniti e definiti nei prossimi paragrafi.

Primitive e calcolo degli integrali definiti. Esaminiamo in dettaglio come si applica il Teorema Fondamentale al calcolo degli integrali. Ricordiamo l'ultima formula del Teorema: qualunque primitiva di f si scrive nella forma

$$F(x) = \int_a^x f(t) dt + C$$

dove abbiamo scelto come primo estremo di integrazione $\alpha = a$. Se poniamo $x = a$, l'integrale si annulla, e otteniamo $F(a) = C$. Quindi la formula si scrive in modo equivalente

$$F(x) - F(a) = \int_a^x f(t) dt.$$

Se infine scegliamo proprio $x = b$, otteniamo una formula per l'integrale definito di f :

$$(5.2) \quad \boxed{\int_a^b f(t) dt = F(b) - F(a).}$$

Molto spesso si usano le notazioni abbreviate

$$F|_a^b \quad \text{o} \quad [F]_a^b \quad \text{per indicare la differenza} \quad F(b) - F(a).$$

Allora la ricetta per calcolare un integrale definito $\int_a^b f(x)dx$ è molto semplice:

- (1) **trovare una primitiva** F di f , ossia calcolare $\int f(x)dx$
- (2) **calcolare la differenza** $F|_a^b = F(b) - F(a)$, che per la formula (5.2) è proprio l'integrale definito di f .

Naturalmente, tra il dire e il fare... Il passo difficile è (1): non sempre è facile trovare una primitiva di f ossia una funzione F la cui derivata è proprio f . Il passo (2) invece è un calcolo meccanico e non presenta problemi.

Vediamo un esempio. Nel primo paragrafo abbiamo calcolato “a mano”, con una certa fatica, l'integrale

$$\int_a^b x^2 dx = \frac{1}{3}(b^3 - a^3).$$

Utilizzando il Teorema Fondamentale il calcolo diventa banale:

- (1) una primitiva di $f(x) = x^2$ è evidentemente $F(x) = \frac{1}{3}x^3$, dato che $(\frac{1}{3}x^3)' = x^2$
- (2) quindi $\int_a^b x^2 dx = \frac{1}{3}x^3|_a^b = \frac{1}{3}b^3 - \frac{1}{3}a^3$.

Quasi indolore. Anzi, possiamo calcolare l'integrale di qualunque potenza x^α , una cosa che probabilmente non saremmo riusciti a fare “a mano”: basta osservare che per $\alpha \neq -1$

$$\left(\frac{1}{\alpha + 1} x^{\alpha+1} \right)' = x^\alpha,$$

ossia $F(x) = \frac{x^{\alpha+1}}{\alpha+1}$ è una primitiva di $f(x) = x^\alpha$, e quindi

$$\int_a^b x^\alpha dx = \frac{1}{\alpha + 1} x^{\alpha+1} \Big|_a^b = \frac{1}{\alpha + 1} (b^{\alpha+1} - a^{\alpha+1}) \quad \alpha \neq -1.$$

Nello stesso modo possiamo ottenere formule di integrazione per altre funzioni. Ad esempio, $(\text{sen } x)' = \cos x$, e perciò

$$\int_0^{\pi/2} \cos x dx = \text{sen } x \Big|_0^{\pi/2} = \text{sen}(\pi/2) - \text{sen } 0 = 1.$$

Si suggerisce di confrontare l'ultimo calcolo con quello molto più complesso fatto “a mano” dello stesso integrale. Il Teorema Fondamentale rende banali molti calcoli di aree che con metodi geometrici sono impervi o addirittura impossibili.

6. Metodi di calcolo per gli integrali

Dedichiamoci ora al calcolo di integrali. Abbiamo visto che il problema principale da risolvere è il calcolo di una primitiva. Anzitutto, raccogliamo in una tabella quelle che sappiamo già. Riscrivendo le regole di derivazione elementari sotto forma di primitive, abbiamo:

$$\int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha + 1} + \text{costante} \quad \forall \alpha \neq -1, \quad \int \frac{dx}{x} = \ln|x| + C$$

$$\begin{aligned}
\int \operatorname{sen} x \, dx &= -\cos x + C, & \int \cos x \, dx &= \operatorname{sen} x + C, \\
\int e^x \, dx &= e^x + C, & \int a^x \, dx &= \frac{a^x}{\ln a} + C, \\
\int \frac{dx}{1+x^2} &= \arctan x + C, & \int c \, dx &= cx + C, \quad c \text{ costante} \\
\int \frac{dx}{\cos^2 x} &= \operatorname{tg} x + C = \frac{\operatorname{sen} x}{\cos x} + C, & \int \frac{dx}{\operatorname{sen}^2 x} &= -\cot x + C = -\frac{\cos x}{\operatorname{sen} x} + C, \\
\int \frac{dx}{\sqrt{1-x^2}} &= \operatorname{arc} \operatorname{sen} x + C, & \int \frac{dx}{\sqrt{1-x^2}} &= -\operatorname{arccos} x + C
\end{aligned}$$

(notare che spesso invece di scrivere $\int \frac{1}{f(x)} dx$ si scrive $\int \frac{dx}{f(x)}$ trattando il simbolo dx come se fosse un fattore: è una notazione tradizionale, come ce ne sono parecchie nel Calcolo).

Inoltre, grazie alla linearità dell'integrale, anche combinazioni lineari di funzioni di cui si conosce la primitiva, possono essere integrate esplicitamente, dato che

$$\int (cf(x) + g(x)) dx = c \int f(x) dx + \int g(x) dx.$$

Ad esempio,

$$\int (1 + 2x + 3e^x) dx = \int 1 dx + 2 \int x dx + 3 \int e^x dx = x + x^2 + 3e^x + C.$$

Si può essere soddisfatti: per un buon numero di funzioni è possibile conoscere la famiglia delle primitive. Ma questo non vale per *tutte* le funzioni: si può dimostrare che esistono combinazioni di funzioni elementari la cui primitiva (che esiste per i Teoremi che abbiamo dimostrato) *non si può rappresentare* come combinazione di funzioni elementari. Si tratta di un fenomeno molto simile alla non risolubilità delle equazioni di grado maggiore o uguale a 5 (e infatti la teoria che se ne occupa si chiama Teoria Differenziale di Galois). L'esempio più semplice è dato dalla funzione $f(x) = e^{-x^2}$ la cui primitiva $\operatorname{Erf}(x) = \int_0^x e^{-t^2} dt$ non si può rappresentare come combinazione di funzioni elementari. Chiaramente questo è un modo per “generare nuove funzioni elementari” (e se ci fate caso, dalla tabella precedente si vede che anche \ln , sen , \cos , tg si possono generare a partire da integrali di funzioni ancora più elementari).

Potremmo aggiungere alla lista precedente alcune altre regole di uso frequente. Dalla regola $(e^{f(x)})' = e^{f(x)} f'(x)$ otteniamo

$$\int f'(x) e^{f(x)} dx = e^{f(x)} + C$$

mentre dalla regola $(\ln |f(x)|)' = \frac{f'(x)}{f(x)}$ otteniamo

$$\int \frac{f'(x)}{f(x)} dx = \ln |f(x)| + C.$$

Attenzione: l'ultima formula, come anche la formula $\int \frac{dx}{x} = \ln |x| + C$, va applicata con cautela, nel senso seguente: posso scrivere senz'altro

$$\int_1^2 \frac{d}{x} = \ln |x| \Big|_1^2 = \ln 2 - \ln 1 = \ln 2$$

oppure

$$\int_{-5}^{-3} \frac{d}{x} = \ln |x| \Big|_{-5}^{-3} = \ln 3 - \ln 5 = \ln \frac{3}{5}.$$

Me non posso scrivere

$$\int_{-1}^2 \frac{d}{x} = \ln |x| \Big|_{-1}^2 = \ln 2 - \ln 1 = \ln 2 \quad \text{NO!}$$

perché la funzione $f(x) = \frac{1}{x}$ non è *limitata* sull'intervallo $[a, b] = [-1, 2]$ e quindi non è integrabile secondo Riemann su tale intervallo (non è neanche definita in 0). La formula per l'integrale di $1/x$ si può applicare solo su un intervallo $[a, b]$ che non contiene 0. Analogamente, la formula per l'integrale di f'/f si può utilizzare solo se f non si annulla nell'intervallo $[a, b]$.

Studiamo ora in dettaglio i principali metodi di integrazione. Ricordare sempre che ogni metodo di integrazione non è nient'altro che un metodo di derivazione letto al contrario.

6.1. Metodo di sostituzione. La formula di derivazione di funzioni composte si scrive

$$\left(F(\phi(x)) \right)' = F'(\phi(x))\phi'(x).$$

Quindi $F(\phi(x))$ è una primitiva di $F'(\phi(x))\phi'(x)$, cioè

$$(6.1) \quad \int F'(\phi(x))\phi'(x) dx = F(\phi(x)) + C.$$

Con questa semplice osservazione, possiamo già risolvere molti problemi nuovi. Ad esempio, qual è una primitiva di $\sin x \cos x$? Se $\phi(x) = \sin x$ e $F(t) = \frac{1}{2}t^2$, allora $F'(t) = t$,

$$F'(\phi(x)) = \sin x, \quad \phi'(x) = \cos x$$

e quindi

$$\int \sin x \cos x dx = \int F'(\phi(x))\phi'(x) dx = F(\phi(x)) + C = \frac{1}{2}\sin^2 x + C.$$

Adesso l'integrale definito è calcolato con facilità:

$$\int_0^1 \sin x \cos x dx = \frac{1}{2}\sin^2 x \Big|_0^1 = \frac{1}{2}\sin^2 1.$$

Scrivendo $f(x) := F'(x)$, la formula (6.1) si può riformulare come

$$(6.2) \quad \boxed{\int f(\phi(x))\phi'(x) dx = \int f(t) dt \quad (\text{formula di sostituzione, I versione})}$$

Questa formula esprime come si trasforma l'espressione dell'integrale passando dalla variabile di integrazione x alla nuova variabile $t = \phi(x)$. L'operazione si può ricordare mnemonicamente così: se chiamo $\phi(x) = t$, allora devo sostituire formalmente $dt = \phi'(x) dx$:

$$t = \phi(x) \quad \Rightarrow \quad "dt = \phi'(x) dx".$$

(L'uso delle virgolette “ ” sta a sottolineare che non è stato dato senso ai simboli dt e dx fuori dall'integrale; la regola sopra scritta è solo formale). Se $F' = f$, cioè se si conosce una primitiva di f , l'integrale a destra in (6.2) è pari a $F(t) + C$ con $C \in \mathcal{R}$

e, per trovare la primitiva della funzione di partenza $f(\phi(x))\phi'(x)$ basta sostituire al posto di t la funzione $\phi(x)$ del cambio di variabile:

$$\int f(\phi(x))\phi'(x) dx = \int f(t) dt = F(t) + C = F(\phi(x)) + C.$$

Ad esempio, ponendo $t = x^2$,

$$\int 2x\text{sen}(x^2) dx = \int \text{sen}(t) dt = -\cos t + C = -\cos(x^2) + C.$$

La domanda naturale è: *come individuare una decomposizione della funzione integranda del tipo $f(\phi(x))\phi'(x)$?* Un metodo generale non c'è, anche se con un po' di pratica si impara a fare questo passaggio senza problemi.

Anche per gli integrali definiti esiste una formula di sostituzione del genere della (6.2). Integrando in $[\alpha, \beta]$ la relazione $F'(\phi(x))\phi'(x)$,

$$\int_{\alpha}^{\beta} F'(\phi(x))\phi'(x) dx = \int_{\alpha}^{\beta} (F(\phi(x)))' dx = F(\phi(\beta)) - F(\phi(\alpha)).$$

Ponendo $a = \phi(\alpha)$ e $b = \phi(\beta)$,

$$F(\phi(\beta)) - F(\phi(\alpha)) = F(b) - F(a) = \int_a^b F'(t) dt.$$

Quindi otteniamo la formula

$$\int_a^b F'(t) dt = \int_{\alpha}^{\beta} F'(\phi(x))\phi'(x) dx,$$

che, chiamando come prima $f := F'$, può essere riscritta come

$$(6.3) \quad \boxed{\int_{\phi(\alpha)}^{\phi(\beta)} f(t) dt = \int_{\alpha}^{\beta} f(\phi(x))\phi'(x) dx \quad (\text{formula di sostituzione, II versione})}$$

Detto altrimenti: se si vuole applicare il metodo di sostituzione direttamente ad un integrale definito, bisogna ricordarsi di cambiare *anche* gli estremi di integrazione, sempre secondo la sostituzione $t = \phi(x)$.

ESEMPIO 6.1. Calcoliamo

$$\int_0^1 \frac{e^t}{1+e^{2t}} dt,$$

utilizzando la sostituzione $t = \ln x$. Dato che $\phi(x) = \ln x$, ne segue $\phi'(x) = 1/x$ e quindi

$$\int_0^1 \frac{e^t}{1+e^{2t}} dx = \int_{\alpha}^{\beta} \frac{x}{1+x^2} \frac{1}{x} dx = \int_{\alpha}^{\beta} \frac{dx}{1+x^2}.$$

Rimangono da calcolare α e β :

$$\begin{cases} 0 = \phi(\alpha) = \ln \alpha \\ 1 = \phi(\beta) = \ln \beta \end{cases} \iff \begin{cases} \alpha = \phi^{-1}(0) = e^0 = 1 \\ \beta = \phi^{-1}(1) = e^1 = e. \end{cases}$$

In definitiva

$$\int_0^1 \frac{e^t}{1+e^{2t}} dt = \int_1^e \frac{dx}{1+x^2} = \arctan e - \arctan 1 = \arctan e - \frac{\pi}{4}.$$

Esempi. 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,

$$\int \frac{dx}{x \ln x} = \ln |\ln x| + C, \quad \int \operatorname{tg} x \, dx = - \int \frac{-\operatorname{sen} x}{\cos x} dx = - \ln |\cos x| + C.$$

Allo stesso modo, ponendo $x = \phi(u)$,

$$\int [\phi(u)]^\alpha \phi'(u) du = \int x^\alpha dx = \frac{x^{\alpha+1}}{\alpha+1} + C = \frac{1}{\alpha+1} [\phi(u)]^{\alpha+1} + C \quad \alpha \neq -1.$$

Ad esempio,

$$\int \operatorname{sen}^k x \cos x \, dx = \frac{1}{k+1} \operatorname{sen}^{k+1} x + C.$$

La formula di sostituzione è sempre conveniente nel caso di funzioni composte di cui l'ultima sia lineare: ponendo $x = au + b$

$$\int f(au + b) du = \frac{1}{a} \int f(x) dx.$$

Non è detto che l'integrale di destra sia risolvibile, ma l'espressione è comunque più semplice. Vediamo un esempio di questo genere

$$\int \frac{1}{\cos^2(2u+3)} du = \frac{1}{2} \int \frac{1}{\cos^2 x} dx = \frac{1}{2} \operatorname{tg} x + C = \frac{1}{2} \operatorname{tg}(2u+3) + C.$$

Ancora un'espressione per la formula di sostituzione. 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 $x = \phi(u)$? E, in caso affermativo, in che modo? Se la funzione ϕ è invertibile tutto fila liscio, occorre solo un po' di pazienza. Sia ψ la funzione inversa di ϕ ,

$$\psi := \phi^{-1}$$

e chiamiamo $f(u) := h(\phi(u))$. Allora, con il cambio di variabile $u = \phi(x)$,

$$\int h(\phi(u)) du = \int f(u) du = \int f(\psi(x)) \psi'(x) dx = \int h(x) \psi'(x) dx.$$

dato che $f(\psi(x)) = h(\phi(\psi(x))) = h(x)$. Guardando il primo e l'ultimo termine:

$$\int h(\phi(u)) du = \int h(x) \psi'(x) dx. \quad (\text{formula di sostituzione, III versione}).$$

Nel caso di integrali definiti occorre cambiare gli estremi di integrazione coerentemente con la nuova variabile introdotta:

$$\int_a^b h(\phi(u)) du = \int_{\phi(a)}^{\phi(b)} h(x) \psi'(x) dx. \quad (\text{formula di sostituzione, IV versione}).$$

ESERCIZIO 6.2. Calcolare l'integrale indefinito

$$\int (1 + e^x)^2 dx.$$

Soluzione. Poniamo $t = 1 + e^x$, allora $x = \ln(t - 1)$ e $dx = \frac{1}{t-1} dt$. Quindi

$$\int (1 + e^x)^2 dx = \int \frac{t^2}{t-1} dt.$$

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 scrivendo la funzione integranda nella forma $(1 + e^x)^2 = 1 + 2e^x + e^{2x}$.

ESERCIZIO 6.3. Calcolare l'integrale indefinito

$$\int \cos^2 x dx.$$

Soluzione. Dato che $\cos(2x) = 2 \cos^2 x - 1$,

$$\int \cos^2 x dx = \frac{1}{2} \int (\cos(2x) + 1) dx = \frac{1}{2} \int \cos(2x) dx + \frac{1}{2}x.$$

Ponendo $t = 2x$,

$$\int \cos(2x) dx = \frac{1}{2} \int \cos t dt = \frac{1}{2} \text{sen} t + C = \frac{1}{2} \text{sen}(2x) + C = \text{sen} x \cos x + C.$$

Quindi

$$\int \cos^2 x dx = \frac{1}{2} (\text{sen} x \cos x + x) + C.$$

ESERCIZIO 6.4. Calcolare l'integrale definito

$$\int_0^{\pi/2} \cos^2 x dx.$$

Soluzione. Procedendo come sopra l'integrale può essere riscritto nella forma

$$\int_0^{\pi/2} \cos^2 x dx = \frac{1}{2} \int_0^{\pi/2} (\cos(2x) + 1) dx = \frac{1}{2} \int_0^{\pi/2} \cos(2x) dx + \frac{\pi}{4}.$$

Poniamo nell'integrale $t = 2x$:

$$\int_0^{\pi/2} \cos(2x) dx = \frac{1}{2} \int_0^{\pi} \cos t dt = \frac{1}{2} \text{sen} t \Big|_0^{\pi} = 0.$$

Quindi il valore dell'integrale è $\pi/4$.

ESERCIZIO 6.5. Fissato $a > 0$, calcolare gli integrali (indefinito e definito)

$$\int x\sqrt{x^2 - a^2} dx \quad \text{e} \quad \int_0^a x\sqrt{x^2 - a^2} dx.$$

Soluzione. Moltiplichiamo e dividiamo per 2 e poniamo $t = x^2 - a^2$:

$$\begin{aligned} \int x\sqrt{x^2 - a^2} dx &= \frac{1}{2} \int 2x(x^2 - a^2)^{1/2} dx = \frac{1}{2} \int t^{1/2} dt \\ &= \frac{t^{3/2}}{3} + C = \frac{(x^2 - a^2)^{3/2}}{3} + C. \end{aligned}$$

Per il calcolo dell'integrale definito basta calcolare la differenza di una primitiva tra i due estremi di integrazione

$$\int_0^a x\sqrt{x^2 - a^2} dx = \frac{(x^2 - a^2)^{3/2}}{3} \Big|_0^a = \frac{a^3}{3}.$$

6.2. Metodo di integrazione per parti. Il secondo metodo generale di integrazione è una riformulazione della regola per la derivata di un prodotto

$$(fg)' = f'g + fg'.$$

Integrando ed usando il Teorema Fondamentale del Calcolo Integrale:

$$f(x)g(x) = \int g(x) f'(x) dx + \int g'(x) f(x) dx$$

o, equivalentemente,

$$(6.4) \quad \boxed{\int g'(x) f(x) dx = f(x)g(x) - \int g(x) f'(x) dx \quad (\text{integrazione per parti})}$$

Ecco spiegato perché si parla di *integrazione per parti*: l'integrale della funzione $g'f$ viene trasformato in una parte integrata, il termine fg , sommato all'integrale della funzione gf' . Il metodo è vantaggioso se per il termine gf' si conosce un metodo di integrazione.

Per gli integrali definiti, la formula (6.4) diviene

$$(6.5) \quad \boxed{\begin{aligned} \int_a^b g'(x) f(x) dx &= f(x)g(x) \Big|_a^b - \int_a^b g(x) f'(x) dx \\ &= f(b)g(b) - f(a)g(a) - \int_a^b g(x) f'(x) dx \end{aligned}}$$

Vediamo una classe di integrali risolvibili tramite la formula (6.4). Partiamo da un caso semplice

$$\int x e^x dx = \int x(e^x)' dx = x e^x - \int (x)' e^x dx = x e^x - \int e^x dx = (x - 1)e^x + C.$$

Anche nel caso della funzione $x^2 e^x$ si può procedere in modo analogo, iterando due volte l'applicazione della formula di integrazione per parti,

$$\begin{aligned} \int x^2 e^x dx &= \int x^2 (e^x)' dx = x^2 e^x - 2 \int x e^x dx = x^2 e^x - 2 \left[x e^x - \int e^x dx \right] \\ &= x^2 e^x - 2 [(x - 1)e^x + C] = (x^2 - 2x + 2)e^x + C. \end{aligned}$$

E' chiaro a questo punto che è possibile, iterando n volte l'uso della formula, risolvere integrali del tipo

$$\int p(x) e^x dx \quad p \text{ polinomio di grado } n,$$

In modo analogo è possibile calcolare anche integrali della forma

$$\int p(x)e^{\alpha x} dx \quad \alpha \in \mathcal{R},$$

(basta osservare che $e^{\alpha x} = \frac{1}{\alpha}(e^{\alpha x})'$). Ad esempio,

$$\begin{aligned} \int x^2 e^{3x} dx &= \frac{1}{3}x^2 e^{3x} - \frac{2}{3} \int x e^{3x} dx = \frac{1}{3}x^2 e^{3x} \\ &\quad - \frac{2}{9} \left(x e^{3x} - \int e^{3x} dx \right) = \left(\frac{1}{3}x^2 - \frac{2}{9}x + \frac{2}{27} \right) e^{3x} + C \end{aligned}$$

Allo stesso modo si calcolano

$$\int x \operatorname{sen} x dx = \int x(-\cos x)' dx = -x \cos x + \int \cos x dx = \operatorname{sen} x - x \cos x + C;$$

$$\int x \cos x dx = \int x(\operatorname{sen} x)' dx = x \operatorname{sen} x - \int \operatorname{sen} x dx = \cos x + x \operatorname{sen} x + C.$$

In generale iterando il procedimento un numero opportuno di volte si calcolano gli integrali

$$\int p(x) \operatorname{sen}(ax) dx, \quad \int p(x) \cos(ax) dx \quad (a \in \mathcal{R}).$$

dove p è un polinomio.

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 $k \in \mathcal{N}$,

$$\int x^k \ln x dx = \frac{x^{k+1}}{k+1} \ln x - \frac{1}{k+1} \int x^k dx = \frac{x^{k+1}}{k+1} \left(\ln x - \frac{1}{k+1} \right) + C.$$

ESERCIZIO 6.6. Calcolare gli integrali indefiniti

$$\int \arctan x dx, \quad \int x \arctan x dx.$$

Soluzione. Procediamo come visto in precedenza per $\ln x$

$$\int \arctan x dx = \int (x)' \arctan x dx = x \arctan x - \int \frac{x}{1+x^2} dx = \dots$$

Per calcolare l'ultimo integrale, moltiplichiamo e dividiamo per due, in modo da ricondurci alla forma $\int \phi'(x)/\phi(x) dx$

$$\dots = x \arctan x - \frac{1}{2} \int \frac{(1+x^2)'}{1+x^2} dx = x \arctan x - \frac{1}{2} \ln(1+x^2) + C.$$

Analogamente, per il secondo integrale

$$\int x \arctan x dx = \frac{1}{2} x^2 \arctan x - \frac{1}{2} \int \frac{x^2}{1+x^2} dx.$$

Consideriamo separatamente l'integrale a secondo membro:

$$\int \frac{x^2}{1+x^2} dx = \int \frac{1+x^2-1}{1+x^2} dx = \int dx - \int \frac{dx}{1+x^2} = x - \arctan x + C.$$

Quindi

$$\int x \arctan x dx = \frac{1}{2} [(x^2+1) \arctan x - x] + C.$$

Adesso usiamo l'integrazione per parti in un modo leggermente diverso: iterando l'applicazione di (6.4) torniamo all'integrale originale, ottenendo in questo modo un'equazione per la primitiva. In questo modo risolviamo integrali della forma

$$\int e^{ax} \operatorname{sen}(bx) dx, \quad \int e^{ax} \operatorname{cos}(bx) dx.$$

Ad esempio,

$$\begin{aligned} \int e^{2x} \operatorname{sen}(3x) dx &= \frac{1}{3} \int e^{2x} (-\operatorname{cos}(3x))' dx = -\frac{1}{3} \operatorname{cos}(3x) e^{2x} + \frac{2}{3} \int e^{2x} \operatorname{cos}(3x) dx \\ &= -\frac{1}{3} \operatorname{cos}(3x) e^{2x} + \frac{2}{9} \int e^{2x} (\operatorname{sen}(3x))' dx \\ &= \frac{1}{9} (2\operatorname{sen}(3x) - 3\operatorname{cos}(3x)) e^{2x} - \frac{4}{9} \int e^{2x} \operatorname{sen}(3x) dx. \end{aligned}$$

Guardando il primo e l'ultimo termine

$$\int e^{2x} \operatorname{sen}(3x) dx = \frac{1}{9} (2\operatorname{sen}(3x) - 3\operatorname{cos}(3x)) e^{2x} - \frac{4}{9} \int e^{2x} \operatorname{sen}(3x) dx.$$

da cui, esplicitando rispetto all'integrale richiesto,

$$\int e^{2x} \operatorname{sen}(3x) dx = \frac{1}{13} (2\operatorname{sen}(3x) - 3\operatorname{cos}(3x)) e^{2x} + C.$$

In generale si ottengono le formule

$$\begin{aligned} \int e^{ax} \operatorname{sen}(bx) dx &= \frac{1}{a^2 + b^2} (a \operatorname{sen}(bx) - b \operatorname{cos}(bx)) e^{ax} + C, \\ \int e^{ax} \operatorname{cos}(bx) dx &= \frac{1}{a^2 + b^2} (a \operatorname{cos}(bx) + b \operatorname{sen}(bx)) e^{ax} + C. \end{aligned}$$

Formula ricorsiva. 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

$$I_n = \int \operatorname{sen}^{2n} x dx, \quad n \in \mathbb{N}$$

(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

$$I_{n+1} = \frac{1}{2n} \{ (2n+1)I_n - \sin^{2n+1} x \cos x \} + C.$$

Ad esempio, per $n = 1$,

$$\begin{aligned} \int \sin^4 x dx &= I_2 = \frac{1}{2} (3I_1 - \sin^3 x \cos x) + C \\ &= \frac{1}{4} (3x - 3\sin x \cos x - 2\sin^3 x \cos x) + C. \end{aligned}$$

6.3. Integrazione di funzioni razionali. Affrontiamo ora il problema di integrare funzioni razionali:

$$\int \frac{P(x)}{Q(x)} dx \quad P, Q \text{ polinomi.}$$

Vedremo che è sempre possibile esprimere le primitive di una qualsiasi funzione razionale in termini di funzioni elementari.

Più precisamente, questo integrale è sempre una combinazione di funzioni elementari, ma se si vuole scrivere esplicitamente il risultato è necessario conoscere le radici del polinomio a denominatore $Q(x)$. Se sappiamo fattorizzare Q nel prodotto di termini irriducibili, cioè polinomi di primo grado (con molteplicità opportuna) e polinomi di secondo grado irriducibili (con molteplicità opportuna), allora possiamo completare il calcolo.

In questo Paragrafo ci limiteremo a mostrare come si integrano le funzioni razionali P/Q quando il polinomio Q sia di grado al più due.

Denominatore Q di grado 1. Sia $Q(x) = a(x - x_0)$ con $a, x_0 \in \mathbb{R}$. 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 risolubile 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

$$Q(x) = a(x^2 + 2bx + c) \quad a, b, c \in \mathbb{R}.$$

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

$$\frac{P(x)}{Q(x)} = P_1(x) + \frac{R(x)}{Q(x)},$$

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 risolubile 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 $\int \phi'(x)/\phi(x) dx$; 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

$$(6.6) \quad \int \frac{dx}{x^2 + 2bx + c}.$$

Nel caso in cui R sia di grado 0 ci si riconduce direttamente a questa situazione. La risoluzione dell'integrale (6.6) 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

$$\frac{(A_1 + A_2)x - (A_1x_2 + A_2x_1)}{x^2 + 2bx + c},$$

quindi le costanti A_1, A_2 sono le soluzioni del sistema lineare⁽¹⁾

$$A_1 + A_2 = 0, \quad A_1x_2 + A_2x_1 = -1.$$

⁽¹⁾Il determinante di questo sistema è $x_1 - x_2$ che, nel caso $b^2 > c$ è diverso da zero.

Individuati i valori di A_1 e A_2 , l'integrale è risolto, infatti

$$\begin{aligned}\int \frac{dx}{x^2 + 2bx + c} &= A_1 \int \frac{dx}{x - x_1} + A_2 \int \frac{dx}{x - x_2} \\ &= A_1 \ln |x - x_1| + A_2 \ln |x - x_2| + C.\end{aligned}$$

ESERCIZIO 6.7. Calcolare

$$\int \frac{x^3}{x^2 - x - 2} dx.$$

Soluzione. Tramite la divisione di polinomi $\frac{x^3}{x^2 - x - 2} = x + 1 + \frac{3x+2}{x^2 - x - 2}$.

Quindi

$$\int \frac{x^3}{x^2 - x - 2} dx = \int \left(x + 1 + \frac{3x+2}{x^2 - x - 2} \right) dx = \frac{x^2}{2} + x + 3 \int \frac{x + \frac{2}{3}}{x^2 - x - 2} dx.$$

Per risolvere l'integrale a secondo membro, moltiplichiamo e dividiamo per 2 e, successivamente, sommiamo e sottraiamo -1

$$\begin{aligned}3 \int \frac{x + \frac{2}{3}}{x^2 - x - 2} dx &= \frac{3}{2} \int \frac{(2x - 1) + 1 + \frac{4}{3}}{x^2 - x - 2} dx \\ &= \frac{3}{2} \int \frac{2x - 1}{x^2 - x - 2} dx + \frac{7}{2} \int \frac{dx}{x^2 - x - 2} = \frac{3}{2} \ln |x^2 - x - 2| + \frac{7}{2} \int \frac{dx}{x^2 - x - 2}.\end{aligned}$$

Dato che le radici del polinomio $x^2 - x - 2$ sono -1 e 2 , esistono A e B tali che

$$\frac{1}{x^2 - x - 2} = \frac{A}{x - 2} + \frac{B}{x + 1}.$$

Il sistema lineare soddisfatto da A e B è: $A + B = 0$, $A - 2B = 1$, quindi $A = \frac{1}{3}$ e $B = -\frac{1}{3}$. Quindi

$$\int \frac{dx}{x^2 - x - 2} = \frac{1}{3} \int \frac{dx}{x - 2} - \frac{1}{3} \int \frac{dx}{x + 1} = \frac{1}{3} \ln |x - 2| - \frac{1}{3} \ln |x + 1| + C.$$

In conclusione, l'integrale richiesto è

$$\begin{aligned}\int \frac{x^3}{x^2 - x - 2} dx &= \frac{x^2}{2} + x + \frac{3}{2} \ln |x - 2| - \frac{3}{2} \ln |x + 1| + \frac{7}{6} \ln |x - 2| + \frac{7}{6} \ln |x + 1| + C \\ &= \frac{x^2}{2} + x + \frac{8}{3} \ln |x - 2| + \frac{1}{3} \ln |x + 1| + C.\end{aligned}$$

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 6.8. Calcolare

$$\int \frac{x(x+3)}{(x-1)^2} dx.$$

Soluzione. Tramite la divisione di polinomi $\frac{x(x+3)}{(x-1)^2} = 1 + \frac{5x-1}{x^2-2x+1}$. Quindi

$$\begin{aligned}\int \frac{x(x+3)}{(x-1)^2} dx &= x + 5 \int \frac{x - \frac{1}{5}}{x^2 - 2x + 1} dx = x + \frac{5}{2} \int \frac{(2x - 2) + 2 + \frac{2}{5}}{x^2 - 2x + 1} dx \\ &= x + \frac{5}{2} \ln |(x - 1)^2| + 6 \int \frac{dx}{(x - 1)^2} = x + 5 \ln |x - 1| - \frac{6}{x - 1} + C\end{aligned}$$

Caso III. $b^2 < c$. In questo caso il polinomio 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 $\nu := \sqrt{c-b^2} > 0$ e riscriviamo in maniera opportuna il denominatore

$$x^2 + 2bx + c = x^2 + 2bx + b^2 + (c-b^2) = (x+b)^2 + \nu^2 = \nu^2 \left\{ \left(\frac{x+b}{\nu} \right)^2 + 1 \right\}.$$

Ponendo $t = (x+b)/\nu$

$$\begin{aligned} \int \frac{dx}{x^2 + 2bx + c} &= \frac{1}{\nu^2} \int \frac{dx}{1 + \left(\frac{x+b}{\nu}\right)^2} = \frac{1}{\nu} \int \frac{dt}{1+t^2} \\ &= \frac{1}{\nu} \arctan t + C = \frac{1}{\nu} \arctan \left(\frac{x+b}{\nu} \right) + C. \end{aligned}$$

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 6.9. Calcolare

$$\int \frac{3x-2}{x^2-2x+2} dx.$$

Soluzione. Come al solito, ricostruiamo a numeratore la derivata del denominatore:

$$\int \frac{3x-2}{x^2-2x+2} dx = \frac{3}{2} \int \frac{2x-2+\frac{2}{3}}{x^2-2x+2} dx = \frac{3}{2} \ln|x^2-2x+2| + \int \frac{dx}{x^2-2x+2}.$$

L'ultimo integrale può essere risolto come sopra

$$\int \frac{dx}{x^2-2x+2} = \int \frac{dx}{1+(x-1)^2} = \int \frac{dt}{1+t^2} = \arctan t + C = \arctan(x-1) + C.$$

Quindi

$$\int \frac{3x-2}{x^2-2x+2} dx = \frac{3}{2} \ln|x^2-2x+2| + \arctan(x-1) + C.$$

7. Breve campionario incompleto

È facile capire che esistono molti altri metodi di integrazione: basta calcolare una derivata non compresa nella lista precedente e leggerla come una primitiva. Ci limitiamo qui a elencare qualche altro caso utile.

Esempio 1. Supponiamo di voler calcolare

$$\int R(\sin x, \cos x) dx$$

dove R è una funzione razionale dei suoi argomenti. Dalle relazioni

$$\sin x = \frac{2t}{1+t^2} \quad \text{e} \quad \cos x = \frac{1-t^2}{1+t^2} \quad \text{dove} \quad t = \operatorname{tg} \left(\frac{x}{2} \right),$$

ponendo $t = \operatorname{tg}(x/2)$ o, equivalentemente, $x = 2 \arctan t$, dato che $dx = 2/(1+t^2)dt$, l'integrale si trasforma in

$$\int R\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right) \frac{2}{1+t^2} dt,$$

Ad esempio,

$$\int \frac{dx}{\operatorname{sen} x} = \int \frac{1+t^2}{2t} \frac{2}{1+t^2} dt = \int \frac{dt}{t} = \ln |t| + C = \ln |\operatorname{tg}(x/2)| + C.$$

$$\int \frac{dx}{\cos x} = \int \frac{2 dt}{1-t^2} = \int \left(\frac{1}{1+t} + \frac{1}{1-t} \right) dt = \ln \left| \frac{1+\operatorname{tg}(x/2)}{1-\operatorname{tg}(x/2)} \right| + C.$$

Esempio 2. Proviamo a calcolare l'area dell'ellisse

$$\Omega = \left\{ (x, y) \in \mathbb{R} : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \right\} \quad a, b > 0.$$

Per evidenti ragioni di simmetria, l'area $\mathcal{A}(\Omega)$ di Ω è pari al valore dell'integrale definito

$$\mathcal{A}(\Omega) = 4b \int_0^a \sqrt{1 - (x^2/a^2)} dx.$$

Introduciamo la variabile t definita da $x = a \cos t$, da cui $dx = -a \operatorname{sen} t dt$:

$$\mathcal{A}(\Omega) = 4ab \int_0^{\pi/2} \sqrt{1 - \cos^2 t} \operatorname{sen} t dt = 4ab \int_0^{\pi/2} \operatorname{sen}^2 t dt = 2ab [t - \operatorname{sen} t \cos t]_0^{\pi/2} = \pi ab.$$

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

$$\int R(x, \sqrt{1 - (x^2/a^2)}) dx = -a \int R(a \cos t, \operatorname{sen} t) \operatorname{sen} t dt.$$

dove $x = a \cos t$, e il secondo membro è razionale in $\operatorname{sen} t$ e $\cos t$.

Esempio 3. Torniamo al caso dell'integrazione di funzioni razionali $\frac{P(x)}{Q(x)}$, e supponiamo che Q abbia solo radici reali *distinte*, cioè

$$Q(x) = a(x - x_1) \cdots (x - x_n) \quad x_1, \dots, x_n \in \mathbb{R} \quad \text{con } x_i \neq x_j \quad \text{se } i \neq j.$$

Supponiamo che il grado di P sia minore del grado di Q (altrimenti basta utilizzare il solito algoritmo della divisione di polinomi), e sfruttiamo la fattorizzazione di Q per riscrivere la funzione razionale come somma di funzioni razionali più semplici. Cerchiamo n costanti A_1, \dots, A_n tali che

$$\frac{P(x)}{a(x - x_1) \cdots (x - x_n)} = \frac{1}{a} \left(\frac{A_1}{x - x_1} + \cdots + \frac{A_n}{x - x_n} \right).$$

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

$$\frac{P(x)}{a(x - x_2) \cdots (x - x_n)} = \frac{1}{a} \left\{ A_1 + \frac{A_2(x - x_1)}{x - x_2} + \cdots + \frac{A_n(x - x_1)}{x - x_n} \right\}.$$

e successivamente porre $x = x_1$, ottenendo il valore di A_1

$$A_1 = \frac{P(x_1)}{(x_1 - x_2) \cdots (x_1 - x_n)}$$

Analogamente per A_2, \dots, A_n . Determinate le costanti A_i ,

$$\begin{aligned} \int \frac{P(x)}{Q(x)} dx &= \frac{1}{a} \int \left(\frac{A_1}{x - x_1} + \cdots + \frac{A_n}{x - x_n} \right) dx \\ &= \frac{1}{a} (A_1 \ln |x - x_1| + \cdots + A_n \ln |x - x_n|) + C. \end{aligned}$$

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

$$\begin{aligned} \int \frac{dx}{(x+1)(x+2)(x+3)} &= \frac{1}{2} \int \left(\frac{1}{x+1} - \frac{2}{x+2} + \frac{1}{x+3} \right) dx \\ &= \frac{1}{2} \ln \left| \frac{(x+1)(x+3)}{(x+2)^2} \right| + C. \end{aligned}$$

Esempio 4. Vediamo come si procede per decomporre una funzione razionale

$$\frac{P(x)}{Q(x)}$$

nella somma di frazioni parziali, sotto l'ipotesi che $Q(x)$ abbia coefficienti reali.

PASSO 1. Se il grado di P è maggiore o uguale al grado di Q , si fa la divisione di polinomi P/Q che permette di scrivere $P(x) = P_1(x)Q(x) + R(x)$ dove $P_1(x)$ è il risultato della divisione e R è il resto. Quindi abbiamo

$$\frac{P}{Q} = N + \frac{P_1}{Q}.$$

Notare che adesso il grado di P_1 è strettamente minore del grado di Q , e il termine $N(x)$ è molto facile da integrare. Da adesso in poi possiamo supporre il grado di P strettamente minore del grado di Q .

PASSO 2. Si fattorizza il denominatore $Q(x)$. Questo passo naturalmente non si può portare a termine sempre (perché non sempre è possibile calcolare esplicitamente le radici di un polinomio!). Supponiamo di essere riusciti a fattorizzare Q nella forma seguente:

$$Q(x) = (x - r_1)^{n_1} \dots (x - r_k)^{n_k} \cdot (x^2 + p_1x + q_1)^{m_1} \dots (x^2 + p_\ell x + q_\ell)^{m_\ell}$$

dove tutti i p_j, q_j, r_j sono reali, i fattori sono tutti distinti, e i termini $x^2 + p_jx + q_j$ sono irriducibili sui reali cioè hanno discriminante < 0 . (In realtà non è difficile dimostrare che tutti i polinomi a coefficienti reali si possono fattorizzare in questo modo! però non sempre è possibile determinare i valori di q, a, b, c esplicitamente).

A questo punto possiamo già scrivere la forma della decomposizione in frazioni parziali: per ogni fattore del tipo $(x - r)^n$ dobbiamo aggiungere i termini

$$\frac{A_1}{x - q} + \frac{A_2}{(x - q)^2} + \dots + \frac{A_n}{(x - q)^n};$$

mentre per ogni fattore del tipo $(x^2 + px + q)^m$ dobbiamo aggiungere i termini

$$\frac{A_1x + B_1}{x^2 + px + q} + \frac{A_2x + B_2}{(x^2 + px + q)^2} + \dots + \frac{A_mx + B_m}{(x^2 + px + q)^m}.$$

Esempi:

$$\begin{aligned} \frac{2 - x - x^2}{(x - 2)^2(x + 3)} &= \frac{A_1}{x - 2} + \frac{A_2}{(x - 2)^2} + \frac{B_1}{x + 3} \\ \frac{x^3 - 4}{(x - 1)^2(x^2 + 2x + 3)} &= \frac{A_1}{x - 1} + \frac{A_2}{(x - 1)^2} + \frac{Bx + C}{x^2 + 2x + 3} \end{aligned}$$

PASSO 3. Si calcolano i coefficienti della decomposizione in frazioni parziali raccogliendo e uguagliando i numeratori. Per il primo esempio del passo 2 abbiamo

$$\frac{2 - x - x^2}{(x - 2)^2(x + 3)} = \frac{A_1(x - 2)(x + 3) + A_2(x + 3) + B_1(x - 2)^2}{(x - 2)^2(x + 3)}$$

e quindi vogliamo

$$2 - x - x^2 = A_1(x - 2)(x + 3) + A_2(x + 3) + B_1(x - 2)^2.$$

Ponendo $x = 2$ si ottiene $-4 = 5A_2$ da cui $A_2 = -\frac{4}{5}$. Ponendo $x = -3$ si ottiene $-4 = 25B_1$ da cui $B_1 = -\frac{4}{25}$. Infine uguagliando i coefficienti di x^2 si ottiene $-1 = A_1 + B_1$ da cui $A_1 = -B_1 - 1 = -\frac{21}{25}$:

$$\frac{2 - x - x^2}{(x - 2)^2(x + 3)} = -\frac{21}{25(x - 2)} - \frac{4}{5(x - 2)^2} - \frac{4}{25(x + 3)}.$$

Invece per il secondo esempio del passo 2 abbiamo

$$\begin{aligned} \frac{x^3 - 4}{(x - 1)^2(x^2 + 2x + 3)} &= \\ &= \frac{A_1(x - 1)(x^2 + 2x + 3) + A_2(x^2 + 2x + 3) + (Bx + C)(x - 1)^2}{(x - 1)^2(x^2 + 2x + 3)} \end{aligned}$$

e anche qui uguagliamo i numeratori. Ponendo $x = 1$ abbiamo $-3 = 6A_2$ da cui $A_2 = -\frac{1}{2}$. Ponendo $x = 0$ si ottiene $-4 = -3A_1 + 3A_2 + C$ da cui $C - 3A_1 = -\frac{5}{2}$. Ponendo $x = 2$ si ottiene $4 = 11A_1 + 11A_2 + 2B + C$ da cui $2B + C + 11A_1 = \frac{19}{2}$.

Uguagliando i coefficienti di x^3 si ottiene $1 = A_1 + B$. Da queste tre relazioni si possono ricavare A_1, B, C e si ottiene $A_1 = \frac{5}{6}$, $B = \frac{1}{6}$, $C = 0$ ossia

$$\frac{x^3 - 4}{(x-1)^2(x^2 + 2x + 3)} = \frac{5}{6(x-1)} - \frac{1}{2(x-1)^2} + \frac{x}{6(x^2 + 2x + 3)}.$$

NATURALMENTE, dopo aver decomposto in frazioni parziali, dobbiamo integrare i vari termini. Per le frazioni del tipo $\frac{A}{(x-r)^n}$ questo è facilissimo; un po' meno facile integrare i termini del tipo $\frac{Ax+B}{x^2+px+q}$, che però abbiamo studiato in dettaglio nelle sezioni precedenti; infine, non è difficile integrare anche i termini del tipo $\frac{Ax+B}{(x^2+px+q)^m}$ con $m > 1$, ma il procedimento è più lungo e qui preferiamo omettere i dettagli.

Vediamo un esempio concreto: calcoliamo la primitiva

$$\int \frac{x^2}{(x+1)^2(x-1)} dx.$$

Decomponiamo: si ha

$$\frac{x^2}{(x+1)^2(x-1)} = \frac{A}{x+1} + \frac{B}{(x+1)^2} + \frac{C}{x-1}$$

da cui

$$x^2 = A(x+1)(x-1) + B(x-1) + C(x+1)^2;$$

ponendo $x = \pm 1$ si ottiene $C = \frac{1}{4}$, $B = -\frac{1}{2}$, e uguagliando i coefficienti di x^2 si ottiene $A + C = 1$ da cui $A = \frac{3}{4}$:

$$\frac{x^2}{(x+1)^2(x-1)} = \frac{3}{4(x+1)} - \frac{1}{2(x+1)^2} + \frac{1}{4(x-1)}.$$

A questo punto si ha subito

$$\int \frac{x^2}{(x+1)^2(x-1)} dx = \frac{3}{4} \ln(x+1) + \frac{1}{2(x+1)} + \frac{1}{4} \ln(x-1) + C.$$

Esempio 5. Se calcoliamo la derivata di $F(x) = \ln(x + \sqrt{x^2 + 1})$ otteniamo

$$F'(x) = \frac{1 + \frac{x}{\sqrt{x^2+1}}}{x + \sqrt{x^2+1}} = \frac{1}{\sqrt{x^2+1}}$$

e questo vuol dire che

$$\int \frac{1}{\sqrt{x^2+1}} dx = \ln(x + \sqrt{x^2+1}) + C.$$

Come nel caso della formula $\int \frac{1}{x} dx = \ln|x| + C$, il calcolo precedente dà lo stesso risultato anche se si parte da $F(x) = \ln(-x - \sqrt{x^2 + 1})$, quindi la formula precedente si può anche scrivere nella forma più generale

$$\int \frac{1}{\sqrt{x^2+1}} dx = \ln| + \sqrt{x^2+1} | + C.$$

Si noti che un conto simile applicato a $F(x) = \ln(x + \sqrt{x^2 - 1})$ (oppure applicato a $\ln(-x - \sqrt{x^2 - 1})$) dà

$$F'(x) = \frac{1}{\sqrt{x^2-1}}$$

e quindi otteniamo anche

$$\int \frac{1}{\sqrt{x^2 - 1}} dx = \ln |x + \sqrt{x^2 - 1}| + C.$$

Tabella degli integrali elementari

(C indica una costante arbitraria)

funzione f	primitiva $\int f dx$	funzione f	primitiva $\int f dx$
0	C	$x^\alpha \quad (\alpha \neq -1)$	$\frac{x^{\alpha+1}}{\alpha+1} + C$
$\text{sen} x$	$-\cos x + C$	$\cos x$	$\text{sen} x + C$
e^x	$e^x + C$	$\frac{1}{x}$	$\ln x + C$
$\frac{1}{1+x^2}$	$\arctan x + C$	a^x	$\frac{a^x}{\ln a} + C$
$\frac{1}{\cos^2 x}$	$\text{tg} x + C$	$\frac{1}{\text{sen}^2 x}$	$-\cot x + C$
$\frac{1}{\sqrt{1-x^2}}$	$\text{arc sen} x + C$	$\frac{1}{\sqrt{1-x^2}}$	$-\arccos x + C$
$\frac{1}{\sqrt{x^2+1}}$	$\ln x + \sqrt{x^2+1} + C$	$\frac{1}{\sqrt{x^2-1}}$	$\ln x + \sqrt{x^2-1} + C$

Qualche altro integrale

funzione f	primitiva $\int f dx$	funzione f	primitiva $\int f dx$
$\ln x$	$x(\ln x - 1) + C$	$\arctan x$	$x \arctan x - \frac{1}{2} \ln(1+x^2) + C$
$\text{sen}^2 x$	$\frac{1}{2}(x - \text{sen} x \cos x) + C$	$\cos^2 x$	$\frac{1}{2}(x + \text{sen} x \cos x) + C$
$\text{arc sen} x$	$x \text{arc sen} x + \sqrt{1-x^2} + C$	$\text{arccos} x$	$x \text{arccos} x - \sqrt{1-x^2} + C$

Gli integrali impropri

Nello sviluppare la teoria dell'integrazione secondo Riemann, abbiamo sempre considerato funzioni **limitate** definite su un intervallo chiuso e **limitato** $[a, b]$. La limitatezza dell'intervallo e la limitatezza della funzione sono necessarie affinché siano ben definite le somme superiori e le somme inferiori relative a una partizione:

$$S^-(f; P) = \sum_{k=0}^{n-1} \alpha_k (x_{k+1} - x_k), \quad S^+(f; P) = \sum_{k=0}^{n-1} \beta_k (x_{k+1} - x_k),$$

dove

$$\alpha_k = \inf\{f(x), x \in [x_k, x_{k+1}]\}, \quad \beta_k = \sup\{f(x), x \in [x_k, x_{k+1}]\}.$$

La condizione di limitatezza dell'intervallo fa sì che la quantità $(x_{k+1} - x_k)$ sia sempre finita, mentre la condizione di limitatezza su f rende finite per ogni k le quantità α_k e β_k . In altre parole, “inscatolando” il grafico della f in un grande rettangolo di area finita, di base l'intervallo $[a, b]$ e di altezza l'intervallo $[-M, M]$ (dove $M \geq 0$ è tale che $|f(x)| \leq M$), ci siamo garantiti il fatto che tutte le quantità con cui abbiamo a che fare sono finite.

Che succede, però, se l'intervallo su cui è definita la funzione f non è limitato, o se la funzione f non è limitata? Chiaramente, la definizione di integrabilità e di integrale che abbiamo dato non funziona più, perché almeno uno degli addendi che compaiono in $S^+(f; P)$ o $S^-(f; P)$ non è finito (o perché uno degli α_k o β_k non è finito, o perché una delle differenze $(x_{k+1} - x_k)$ non lo è).

Possiamo però ovviare a questa “impossibilità” nel seguente modo. Supponiamo, ad esempio, che $f : (a, b) \rightarrow \mathbb{R}$ sia una funzione illimitata in un intorno destro di $x = a$, e integrabile su $[c, b]$ per ogni $c > a$. In altre parole, appena ci spostiamo, sia pure di poco, da $x = a$, otteniamo una funzione integrabile secondo Riemann (osserviamo che questo fatto implica che la funzione f è illimitata “solo” in $x = a$ e non in altri punti dell'intervallo (a, b)). Essendo la funzione f integrabile su $[c, b]$, è ben definito il numero reale (dipendente da c)

$$\int_c^b f(x) dx, \quad c > a.$$

Possiamo allora chiederci se esista o meno (e, qualora esista, se sia finito o meno) il limite di tale numero per c tendente ad a (da destra):

$$\lim_{c \rightarrow a^+} \int_c^b f(x) dx.$$

Ebbene: diciamo che **la funzione f è integrabile in senso improprio su (a, b)** se il limite è un numero reale (ovvero: esiste ed è finito). In tal caso scriviamo

$$\int_a^b f(x) dx = \lim_{c \rightarrow a^+} \int_c^b f(x) dx,$$

ovvero, ponendo $c = a + \varepsilon$, con $\varepsilon > 0$,

$$\int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dx.$$

Ricapitolando, abbiamo la seguente definizione.

DEFINIZIONE 1. Sia $f : (a, b) \rightarrow \mathbb{R}$ una funzione tale che:

- f è illimitata in un intorno destro di $x = a$;
- f è integrabile su $[a + \varepsilon, b]$ per ogni $\varepsilon > 0$;
- esiste ed è finito il limite

$$\lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dx.$$

Diciamo allora che f è integrabile in senso improprio su $(a, b]$ e definiamo

$$(1) \quad \int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dx.$$

OSSERVAZIONE 2. Se f è integrabile su $[a, b]$, è sempre vero che

$$(2) \quad \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dx = \int_a^b f(x) dx.$$

Infatti si ha

$$\left| \int_a^b f(x) dx - \int_{a+\varepsilon}^b f(x) dx \right| = \left| \int_a^{a+\varepsilon} f(x) dx \right| \leq \int_a^{a+\varepsilon} |f(x)| dx \leq M \varepsilon,$$

dove $M \geq 0$ è tale che $|f(x)| \leq M$ per ogni x in $[a, b]$ (si noti che tale costante esiste perché f , essendo integrabile su $[a, b]$, è limitata su $[a, b]$). Pertanto,

$$0 \leq \lim_{\varepsilon \rightarrow 0^+} \left| \int_a^b f(x) dx - \int_{a+\varepsilon}^b f(x) dx \right| \leq \lim_{\varepsilon \rightarrow 0^+} M \varepsilon = 0,$$

da cui la (2). Pertanto, la definizione di integrale improprio “estende” la nozione di integrale secondo Riemann.

OSSERVAZIONE 3. Se la funzione f ha segno costante su $(a, b]$ (ad esempio, è tutta positiva, o tutta negativa), allora il limite dell’integrale esiste sempre. Per vederlo, definiamo, per $t > 0$,

$$G(t) = \int_{a+t}^b f(x) dx,$$

e supponiamo che $f(x) \geq 0$ per ogni x in $(a, b]$. Se $0 < s < t$, allora

$$G(s) - G(t) = \int_{a+s}^b f(x) dx - \int_{a+t}^b f(x) dx = \int_{a+s}^{a+t} f(x) dx \geq 0,$$

dato che f è positiva e $a + s < a + t$ per ipotesi. Dunque si ha

$$s < t \implies G(s) > G(t),$$

e quindi $G(t)$ è una funzione decrescente (se $f(x) \leq 0$, $G(t)$ è crescente). Essendo monotona, esiste

$$\lim_{\varepsilon \rightarrow 0^+} G(\varepsilon) = \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dx,$$

che è il limite che ci dice se $f(x)$ è integrabile o meno in senso improprio su $(a, b]$.

Chiaramente, l’illimitatezza “a destra di $x = a$ ” non ha niente di speciale: possiamo considerare funzioni che siano illimitate “a sinistra di $x = b$ ”, e dare la stessa definizione.

DEFINIZIONE 4. Sia $f : [a, b) \rightarrow \mathbb{R}$ una funzione tale che:

- f è illimitata in un intorno sinistro di $x = b$;
- f è integrabile su $[a, b - \varepsilon]$ per ogni $\varepsilon > 0$;

- esiste ed è finito il limite

$$\lim_{\varepsilon \rightarrow 0^+} \int_a^{b-\varepsilon} f(x) dx.$$

Diciamo allora che f è integrabile in senso improprio su $[a, b)$ e definiamo

$$(3) \quad \int_a^b f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \int_a^{b-\varepsilon} f(x) dx.$$

ESEMPIO 5. Consideriamo la funzione $f(x)$ definita su $(0, 1]$ da

$$f(x) = \frac{1}{x^\alpha}, \quad \alpha > 0,$$

e ci chiediamo per quali valori di α la funzione sia integrabile in senso improprio su $(0, 1]$. Osserviamo che la funzione è illimitata in un intorno destro di $x = 0$ (perché $\alpha > 0$) ed è integrabile su $[\varepsilon, 1]$ per ogni $\varepsilon > 0$ (dato che è continua — o monotona decrescente — su tale intervallo). Abbiamo poi:

$$\int_\varepsilon^1 f(x) dx = \int_\varepsilon^1 \frac{dx}{x^\alpha} = \begin{cases} \ln(|x|) \Big|_\varepsilon^1 & \text{se } \alpha = 1, \\ \frac{x^{1-\alpha}}{1-\alpha} \Big|_\varepsilon^1 & \text{se } \alpha \neq 1. \end{cases} = \begin{cases} -\ln(\varepsilon) & \text{se } \alpha = 1, \\ \frac{1 - \varepsilon^{1-\alpha}}{1-\alpha} & \text{se } \alpha \neq 1. \end{cases}$$

Si ha dunque

$$\lim_{\varepsilon \rightarrow 0^+} \int_\varepsilon^1 \frac{dx}{x^\alpha} = \begin{cases} +\infty & \text{se } \alpha \geq 1, \\ \frac{1}{1-\alpha} & \text{se } 0 < \alpha < 1. \end{cases}$$

Se ne deduce, pertanto, che la funzione $f(x) = \frac{1}{x^\alpha}$ è integrabile in senso improprio su $(0, 1]$ se e solo se $0 < \alpha < 1$; in tal caso si ha

$$\int_0^1 \frac{dx}{x^\alpha} = \frac{1}{1-\alpha}.$$

E se invece la funzione f è definita su $[a, c) \cup (c, b]$ ed è illimitata “intorno a $x = c$ ”? In questo caso diciamo che f è integrabile in senso improprio su $[a, b]$ se:

- f è integrabile in senso improprio su $[a, c)$;
- f è integrabile in senso improprio su $(c, b]$.

In tal caso, scriviamo

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx,$$

cosicché la formula “standard” per lo spezzamento degli integrali continua a valere.

OSSERVAZIONE 6. Un’osservazione importante: se la funzione f è illimitata sia a sinistra che a destra di un punto $x = c$, diciamo che la funzione è integrabile in senso improprio se lo è a sinistra di $x = c$ e a destra di $x = c$. In altre parole, chiediamo che siano finiti i limiti

$$(4) \quad \lim_{\varepsilon \rightarrow 0^+} \int_a^{c-\varepsilon} f(x) dx \quad \text{e} \quad \lim_{\varepsilon \rightarrow 0^+} \int_{c+\varepsilon}^b f(x) dx.$$

Si noti che questa condizione è più forte rispetto a chiedere che sia finito il limite

$$\lim_{\varepsilon \rightarrow 0^+} \left[\int_a^{c-\varepsilon} f(x) dx + \int_{c+\varepsilon}^b f(x) dx \right],$$

dato che quest'ultimo limite può essere finito anche se entrambi i limiti di (4) sono infiniti.

ESEMPIO 7. Consideriamo la funzione $f(x)$ definita su $[-1, 1] \setminus \{0\}$ da $f(x) = \frac{1}{x}$. Si ha

$$\lim_{\varepsilon \rightarrow 0^+} \int_{\varepsilon}^1 f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \ln(|x|) \Big|_{\varepsilon}^1 = \lim_{\varepsilon \rightarrow 0^+} -\ln(\varepsilon) = +\infty,$$

e

$$\lim_{\varepsilon \rightarrow 0^+} \int_{-1}^{-\varepsilon} f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \ln(|x|) \Big|_{-1}^{-\varepsilon} = \lim_{\varepsilon \rightarrow 0^+} \ln(\varepsilon) = -\infty,$$

e quindi $f(x)$ non è integrabile in senso improprio né a sinistra di $x = 0$ né a destra di $x = 0$, cosicché f non è integrabile in senso improprio su $[-1, 1]$. Però, a causa della simmetria ($f(x)$ è dispari, e l'intervallo è simmetrico)

$$\lim_{\varepsilon \rightarrow 0^+} \left[\int_{-1}^{-\varepsilon} f(x) dx + \int_{\varepsilon}^1 f(x) dx \right] = \lim_{\varepsilon \rightarrow 0^+} [\ln(\varepsilon) - \ln(\varepsilon)] = 0.$$

Fin qui abbiamo considerato funzioni illimitate su un intervallo limitato. Che succede se invece ad essere illimitato è l'intervallo? Se, ad esempio, è una semiretta $[a, +\infty)$ o $(-\infty, b]$? L'idea è la stessa.

DEFINIZIONE 8. Sia $f : [a, +\infty) \rightarrow \mathbb{R}$ una funzione tale che:

- f è integrabile su $[a, M]$ per ogni $M > a$;
- esiste ed è finito il limite

$$\lim_{M \rightarrow +\infty} \int_a^M f(x) dx.$$

Diciamo allora che f è integrabile in senso improprio su $[a, +\infty)$ e definiamo

$$(5) \quad \int_a^{\infty} f(x) dx = \lim_{M \rightarrow +\infty} \int_a^M f(x) dx.$$

Analogamente:

DEFINIZIONE 9. Sia $f : (-\infty, b] \rightarrow \mathbb{R}$ una funzione tale che:

- f è integrabile su $[M, b]$ per ogni $M < b$;
- esiste ed è finito il limite

$$\lim_{M \rightarrow -\infty} \int_M^b f(x) dx.$$

Diciamo allora che f è integrabile in senso improprio su $(-\infty, b]$ e definiamo

$$(6) \quad \int_{-\infty}^b f(x) dx = \lim_{M \rightarrow -\infty} \int_M^b f(x) dx.$$

OSSERVAZIONE 10. Osserviamo che, come nel caso degli integrali impropri su intervalli limitati, se $f(x)$ ha segno costante su $[a, +\infty)$ o su $(-\infty, b]$, allora esistono sempre i limiti

$$\lim_{M \rightarrow +\infty} \int_a^M f(x) dx \quad \text{e} \quad \lim_{M \rightarrow -\infty} \int_M^b f(x) dx,$$

dato che si tratta di limiti di funzioni monotone.

ESEMPIO 11. Consideriamo la funzione $f(x)$ definita su $[1, +\infty)$ da

$$f(x) = \frac{1}{x^\alpha}, \quad \alpha > 0,$$

e ci chiediamo per quali valori di α la funzione sia integrabile in senso improprio su $[1, +\infty)$. Osserviamo che la funzione è integrabile su $[1, M]$ per ogni $M > 1$ (dato che è continua — o monotona decrescente — su tale intervallo). Abbiamo poi:

$$\int_1^M f(x) dx = \int_1^M \frac{dx}{x^\alpha} = \begin{cases} \ln(|x|) \Big|_1^M & \text{se } \alpha = 1, \\ \frac{x^{1-\alpha}}{1-\alpha} \Big|_1^M & \text{se } \alpha \neq 1. \end{cases} = \begin{cases} \ln(M) & \text{se } \alpha = 1, \\ \frac{M^{1-\alpha} - 1}{1-\alpha} & \text{se } \alpha \neq 1. \end{cases}$$

Si ha dunque

$$\lim_{M \rightarrow +\infty} \int_\varepsilon^1 \frac{dx}{x^\alpha} = \begin{cases} +\infty & \text{se } \alpha \leq 1, \\ \frac{1}{\alpha - 1} & \text{se } \alpha > 1. \end{cases}$$

Se ne deduce, pertanto, che la funzione $f(x) = \frac{1}{x^\alpha}$ è integrabile in senso improprio su $[1, +\infty)$ se e solo se $\alpha > 1$; in tal caso si ha

$$\int_1^{+\infty} \frac{dx}{x^\alpha} = \frac{1}{\alpha - 1}.$$

Esattamente come prima diciamo che una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ è integrabile in senso improprio su \mathbb{R} se è integrabile in senso improprio su $(-\infty, 0]$ e su $[0, +\infty)$; vediamo con il prossimo esempio perché è necessario “spezzare gli integrali”.

ESEMPIO 12. Sia $f(x) = \frac{2x}{1+x^2}$. Si ha

$$\int_0^M f(x) dx = \int_0^M \frac{2x}{1+x^2} dx = \ln(1+x^2) \Big|_0^M = \ln(1+M^2),$$

cosicché

$$\lim_{M \rightarrow +\infty} \int_0^M f(x) dx = \lim_{M \rightarrow +\infty} \ln(1+M^2) = +\infty,$$

e f non è integrabile in senso improprio su $[0, +\infty)$. Analogamente,

$$\int_M^0 f(x) dx = \int_M^0 \frac{2x}{1+x^2} dx = \ln(1+x^2) \Big|_M^0 = -\ln(1+M^2),$$

e quindi

$$\lim_{M \rightarrow -\infty} \int_M^0 f(x) dx = \lim_{M \rightarrow -\infty} \ln(1+M^2) = +\infty.$$

Però, dato che f è dispari e l'intervallo è simmetrico rispetto all'origine, si ha

$$\int_{-M}^M f(x) dx = 0,$$

e quindi

$$\lim_{M \rightarrow +\infty} \int_{-M}^M f(x) dx = 0.$$

Chiaramente, per una funzione generica definita su \mathbb{R} è possibile studiare l'integrazione impropria sia “al finito” che all'infinito. Ad esempio, per vedere se è integrabile in senso improprio su \mathbb{R} la funzione

$$f(x) = \frac{1}{\sqrt{|x|} \sqrt{|x-1|} \sqrt{|x+1|}},$$

dobbiamo verificare se lo è in un intorno di $x = 0$, di $x = 1$, di $x = -1$, a $+\infty$ e a $-\infty$ (sono, in tutto, otto limiti da calcolare: lasciamo al lettore l'arduo compito di verificare che la funzione è integrabile su \mathbb{R}).

Già: come si risolve l'esercizio precedente? Perché il calcolo esplicito dell'integrale (ovvero, di una primitiva di $f(x)$) è assai complicato! Se solo si potesse dire se una funzione è integrabile (o meno) soltanto dal suo comportamento intorno ad un punto (o all'infinito) senza calcolare l'integrale... Per fortuna, è possibile dirlo, grazie ai seguenti criteri (qualche campanello d'allarme dovrebbe iniziare a suonare nella testa di ognuno di voi).

TEOREMA 13. *Siano $f : (a, b] \rightarrow \mathbb{R}$ e $g : (a, b] \rightarrow \mathbb{R}$ due funzioni **positive** tali che*

$$0 \leq f(x) \leq g(x), \quad \forall x \in (a, b].$$

Se g è integrabile in senso improprio su $(a, b]$ allora anche f lo è, mentre se f non è integrabile in senso improprio su $(a, b]$, neanche g lo è.

PROOF. È sufficiente osservare che si ha, per la monotonia dell'integrale, e per ogni $\varepsilon > 0$,

$$0 \leq \int_{a+\varepsilon}^b f(x) dx \leq \int_{a+\varepsilon}^b g(x) dx.$$

Se g è integrabile in senso improprio su $(a, b]$, allora

$$\lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b g(x) dx = \int_a^b g(x) dx < +\infty.$$

D'altra parte, dato che

$$\varepsilon \mapsto \int_{a+\varepsilon}^b f(x) dx$$

è una funzione decrescente rispetto a ε ¹²³, esiste

$$\lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dx,$$

e si ha

$$0 \leq \lim_{\varepsilon \rightarrow 0^+} \int_{a+\varepsilon}^b f(x) dx \leq \int_a^b g(x) dx < +\infty,$$

e quindi f è integrabile in senso improprio su $(a, b]$. La dimostrazione nel caso in cui f non sia integrabile è analoga. \square

TEOREMA 14. *Siano $f : (a, b] \rightarrow \mathbb{R}$ e $g : (a, b] \rightarrow \mathbb{R}$ due funzioni **positive** tali che*

$$\lim_{x \rightarrow a^-} \frac{f(x)}{g(x)} = L \in (0, +\infty).$$

Allora f è integrabile in senso improprio su $(a, b]$ se e solo se g è integrabile in senso improprio su $(a, b]$.

DIMOSTRAZIONE. Dato che

$$\lim_{x \rightarrow a^+} \frac{f(x)}{g(x)} = L > 0,$$

esiste $\delta > 0$ tale che

$$\frac{L}{2} \leq \frac{f(x)}{g(x)} \leq \frac{3L}{2}, \quad \forall x \in (a, a + \delta).$$

Si ha pertanto

$$\frac{L}{2} g(x) \leq f(x) \leq \frac{3L}{2} g(x), \quad \forall x \in (a, a + \delta).$$

Applicando il teorema precedente, si vede che f è integrabile in senso improprio su $(a, a + \delta)$ (e quindi su $(a, b]$) se e solo se lo è g . \square

¹Si veda quanto detto nell'Osservazione 3.

²È una nota a piè di pagina, non un esponente.

³Anche questa!

Un criterio analogo (che non enunciamo per brevità) vale su intervalli del tipo $[a, b)$. In modo analogo si può enunciare un criterio di integrabilità impropria a $+\infty$ (o a $-\infty$):

TEOREMA 15. *Siano $f : [a, +\infty) \rightarrow \mathbb{R}$ e $g : [a, +\infty) \rightarrow \mathbb{R}$ due funzioni **positive** tali che*

$$0 \leq f(x) \leq g(x), \quad \forall x \geq a.$$

Se g è integrabile in senso improprio su $[a, +\infty)$ anche f lo è, mentre se f non è integrabile in senso improprio su $[a, +\infty)$ neanche g lo è. Se invece si ha

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = L \in (0, +\infty),$$

allora f è integrabile in senso improprio su $[a, +\infty)$ se e solo se g è integrabile in senso improprio su $[a, +\infty)$.

ESEMPIO 16. Sia

$$f(x) = \frac{1}{\sqrt{|x|} \sqrt{|x-1|} \sqrt{|x+1|}}.$$

Diciamo che $f(x) \approx g(x)$ per x tendente a x_0 se

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = L \in (0, +\infty),$$

e analogamente che $f(x) \approx g(x)$ per x tendente a $+\infty$ o a $-\infty$. Allora

$$f(x) \approx g_1(x) = \frac{1}{\sqrt{|x|}} \quad \text{per } x \text{ tendente a } x_0 = 0 \quad (L = 1),$$

$$f(x) \approx g_2(x) = \frac{1}{\sqrt{|x-1|}} \quad \text{per } x \text{ tendente a } x_0 = 1 \quad (L = \sqrt{2}/2),$$

$$f(x) \approx g_3(x) = \frac{1}{\sqrt{|x+1|}} \quad \text{per } x \text{ tendente a } x_0 = -1 \quad (L = \sqrt{2}/2),$$

$$f(x) \approx g_4(x) = \frac{1}{|x|^{\frac{3}{2}}} \quad \text{per } x \text{ tendente a } +\infty \quad (L = 1),$$

e

$$f(x) \approx g_4(x) = \frac{1}{|x|^{\frac{3}{2}}} \quad \text{per } x \text{ tendente a } -\infty \quad (L = 1),$$

Siccome g_1, g_2, g_3 e g_4 sono integrabili in senso improprio (rispettivamente intorno a $x_0 = 0, x_0 = 1, x_0 = -1$ e $\pm\infty$, si vedano gli esempi 5 e 11), la funzione $f(x)$ è integrabile in senso improprio su \mathbb{R} .

Per il lettore interessato, si ha

$$\int_{-\infty}^{+\infty} \frac{dx}{\sqrt{|x|} \sqrt{|x-1|} \sqrt{|x+1|}} = \frac{8\sqrt{\pi} \Gamma(5/4)}{\Gamma(3/4)},$$

dove $\Gamma(x)$ è la funzione Gamma di Eulero (vedere oltre).

Il prossimo esempio chiarisce perché i criteri di integrabilità impropria visti in precedenza sembrano noti (e perché la funzione $\frac{1}{x^\alpha}$ è integrabile “a $+\infty$ ” se e solo se $\alpha > 1$).

ESEMPIO 17. Sia ora $f : [1 + \infty) \rightarrow \mathbb{R}$ una funzione **decescente** e **positiva**. Si ha allora

$$f(k+1) \leq f(x) \leq f(k), \quad \forall x \in [k, k+1],$$

e quindi, integrando tra k e $k+1$,

$$f(k+1) = \int_k^{k+1} f(k+1) dx \leq \int_k^{k+1} f(x) dx \leq \int_k^{k+1} f(k) dx = f(k).$$

Pertanto

$$\sum_{k=1}^{n-1} f(k+1) \leq \sum_{k=1}^{n-1} \int_k^{k+1} f(x) dx = \int_1^n f(x) dx.$$

Analogamente

$$\int_1^n f(x) dx = \sum_{k=1}^{n-1} \int_k^{k+1} f(x) dx \leq \sum_{k=1}^{n-1} f(k).$$

In definitiva (cambiando indici nella prima sommatoria),

$$\sum_{k=2}^n f(k) \leq \int_1^n f(x) dx \leq \sum_{k=1}^{n-1} f(k),$$

da cui segue, facendo tendere n all'infinito (si noti che tutti i limiti esistono perché abbiamo a che fare con successioni crescenti),

$$\sum_{k=2}^{+\infty} f(k) \leq \int_1^{+\infty} f(x) dx \leq \sum_{k=1}^{+\infty} f(k).$$

Da questa formula traiamo la seguente conclusione: la funzione $f(x)$ è integrabile in senso improprio su $[1, +\infty)$ se e solo se è convergente la serie di termine generico $f(k)$.

ESERCIZIO 18. Dire se sono finiti i seguenti integrali impropri:

$$\begin{aligned} & \int_0^1 \frac{dx}{x^2}, \quad \int_0^1 \frac{dx}{\sqrt{x}(2-x)}, \quad \int_0^1 \frac{dx}{\sqrt[3]{x}(1-x)}, \quad \int_0^1 \frac{dx}{\ln(x)}, \quad \int_0^1 \frac{dx}{\sqrt[4]{\sin(x)}}, \\ & \int_0^1 \frac{\sin(x)}{x^2} dx, \quad \int_0^1 \frac{e^x - 1 - x}{\sin(x^3)} dx, \quad \int_{-1}^1 e^{-\frac{1}{x}} dx, \quad \int_{-\pi/2}^{\pi/2} \tan(x) dx, \quad \int_{-1}^1 \frac{dx}{1-x^2}, \\ & \int_1^{+\infty} \frac{dx}{x^2}, \quad \int_1^{+\infty} x^2 e^{-x} dx, \quad \int_1^{+\infty} \frac{\arctan(x)}{1+x^2} dx, \quad \int_0^{+\infty} \frac{dx}{\sqrt{x}(x+1)}. \end{aligned}$$

ESEMPIO 19. Sia $x > 0$ e

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

Dimostriamo che $\Gamma(x)$ è ben definita per ogni $x > 0$. Per farlo, dobbiamo far vedere che $f(t) = t^{x-1} e^{-t}$ è integrabile in senso improprio sia a destra di $t = 0$ (per ogni x tra 0 e 1, altrimenti la funzione è continua e l'integrale è quello di Riemann) che a $+\infty$. Si ha, se $0 < x < 1$,

$$\lim_{t \rightarrow 0^+} \frac{t^{x-1} e^{-t}}{t^{x-1}} = 1,$$

e quindi $f(t)$ è integrabile in senso improprio vicino a $t = 0$ se e solo se $g(t) = t^{x-1} = \frac{1}{t^{1-x}}$ lo è. Dal momento che $0 < 1 - x < 1$, la funzione $t \mapsto t^{x-1}$ è integrabile in senso improprio (si veda l'Esempio 5), e quindi anche $f(t)$ lo è. All'infinito, invece, si ha, qualsiasi sia x ,

$$\lim_{t \rightarrow +\infty} \frac{t^{x-1} e^{-t}}{e^{-t/2}} = \lim_{t \rightarrow +\infty} t^{x-1} e^{-t/2} = 0,$$

dato che $e^{-t/2} \asymp t^{x-1}$ per ogni x . Dunque, esiste $t_0 > 0$ tale che

$$0 \leq t^{x-1} e^{-t} \leq e^{-t/2}, \quad \forall t \geq t_0.$$

Ma allora

$$0 \leq \int_{t_0}^M t^{x-1} e^{-t} dt \leq \int_{t_0}^M e^{-t/2} dt = -2e^{-t/2} \Big|_{t_0}^M = 2e^{-t_0/2} - 2e^{-M/2},$$

e quindi

$$0 \leq \lim_{M \rightarrow +\infty} \int_{t_0}^M t^{x-1} e^{-t} dt \leq \lim_{M \rightarrow +\infty} \int_{t_0}^M e^{-t/2} dt = 2e^{-t_0/2} < +\infty,$$

da cui segue che $f(t)$ è integrabile in senso improprio su $[t_0, +\infty)$, e quindi “all’infinito”.

Dunque $\Gamma(x)$ è ben definita. Dimostriamo (per induzione) che si ha

$$\Gamma(n) = (n-1)! \quad \forall n \geq 1,$$

ovvero che la funzione Γ “interpola” il fattoriale. Intanto,

$$\Gamma(1) = \int_0^{+\infty} e^{-t} dt = \lim_{M \rightarrow +\infty} \int_0^M e^{-t} dt = \lim_{M \rightarrow +\infty} -e^{-t} \Big|_0^M = \lim_{M \rightarrow +\infty} [1 - e^{-M}] = 1,$$

e il primo passo è fatto. Supponiamo ora che $\Gamma(n) = (n-1)!$ e dimostriamo che $\Gamma(n+1) = n!$. Osserviamo innanzitutto che, integrando per parti (derivando t^n e integrando e^{-t}) si ha

$$\int t^n e^{-t} dt = -t^n e^{-t} + \int n t^{n-1} e^{-t} dt.$$

Pertanto,

$$\int_0^M t^n e^{-t} dt = -t^n e^{-t} \Big|_0^M + n \int_0^M t^{n-1} e^{-t} dt = -M^n e^{-M} + n \int_0^M t^{n-1} e^{-t} dt.$$

Facendo tendere M a $+\infty$ si ha, dato che $M^n e^{-M}$ tende a zero (ricordiamo che $e^n \gg M^n$), si ha

$$\Gamma(n+1) = \int_0^{+\infty} t^n e^{-t} dt = n \int_0^{+\infty} t^{n-1} e^{-t} dt = n \Gamma(n) = n(n-1)! = n!,$$

come volevasi dimostrare.

Si noti che la relazione che abbiamo dimostrato non vale solo per n intero: si ha $\Gamma(x+1) = x\Gamma(x)$ per ogni $x > 0$ (come si vede facilmente integrando per parti). In altre parole,

$$\Gamma(x) = \frac{\Gamma(x+1)}{x}, \quad \forall x > 0.$$

Volendo, si può usare questa formula per **definire** $\Gamma(x)$ per ogni $x < 0$ (diverso da un intero negativo). Infatti: se $-1 < x < 0$, allora $0 < x+1 < 1$ e quindi

$$\Gamma(x) = \frac{\Gamma(x+1)}{x} = \frac{1}{x} \int_0^{+\infty} t^x e^{-t} dt.$$

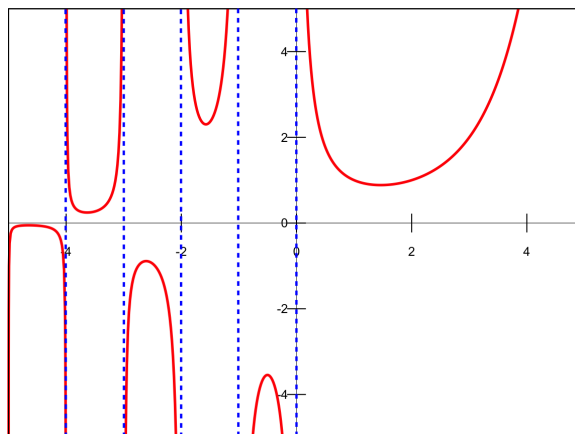
Se $-2 < x < -1$, allora $-1 < x+1 < 0$ e quindi

$$\Gamma(x) = \frac{\Gamma(x+1)}{x} = \frac{1}{x} \frac{1}{x+1} \int_0^{+\infty} t^{x+1} e^{-t} dt,$$

e così via. Si noti che

$$\Gamma(0) = \int_0^{+\infty} t^{-1} e^{-t} dt = +\infty,$$

dato che $f(t) = \frac{e^{-t}}{t}$ si comporta vicino a zero come $\frac{1}{t}$, che non è integrabile in senso improprio (si veda l’Esempio 5). Ne segue che $\Gamma(-n) = \pm\infty$ (a seconda se ci si avvicina da destra o da sinistra, e a seconda della parità / disparità di n).



Il grafico della funzione Γ

ESEMPIO 20. Sia $f(x) = \frac{\sin(x)}{x}$. Vogliamo sapere se f sia integrabile in senso improprio su $[\pi, +\infty)$ o no. Dato che $f(x)$ cambia segno, non possiamo applicare nessuno dei criteri visti in precedenza; d'altra parte, il limite di $\sin(x)$ a $+\infty$ non esiste, e quindi non è chiaro quale sia il comportamento della funzione all'infinito.

Per verificare l'integrabilità, sia n un intero maggiore di 1, e consideriamo

$$S_n = \int_{\pi}^{(n+1)\pi} \frac{\sin(x)}{x} dx = \sum_{k=1}^n \int_{k\pi}^{(k+1)\pi} \frac{\sin(x)}{x} dx = \sum_{k=1}^n a_k,$$

dove

$$a_k = \int_{k\pi}^{(k+1)\pi} \frac{\sin(x)}{x} dx.$$

Osserviamo che $\sin(x)$ è positivo su $[k\pi, (k+1)\pi]$ se k è pari, mentre è negativo se k è dispari. Pertanto, si ha

$$\sin(x) = (-1)^k |\sin(x)| \quad \forall x \in [k\pi, (k+1)\pi],$$

cosicché

$$a_k = (-1)^k \int_{k\pi}^{(k+1)\pi} \frac{|\sin(x)|}{x} dx = (-1)^k b_k,$$

dove

$$b_k = \int_{k\pi}^{(k+1)\pi} \frac{|\sin(x)|}{x} dx.$$

Chiaramente $b_k \geq 0$ per ogni k , essendo l'integrale di una funzione positiva. Dimostriamo che b_k è decrescente. Infatti, con la sostituzione $y = x - \pi$, si ha

$$b_{k+1} = \int_{(k+1)\pi}^{(k+2)\pi} \frac{|\sin(x)|}{x} dx = \int_{k\pi}^{(k+1)\pi} \frac{|\sin(y+\pi)|}{y+\pi} dy.$$

Osserviamo ora che $x \mapsto |\sin(x)|$ è una funzione periodica di periodo π (dato che $\sin(x+\pi) = -\sin(x)$); pertanto, nell'ultimo integrale si ha $|\sin(y+\pi)| = |\sin(y)|$, e dunque

$$b_{k+1} = \int_{k\pi}^{(k+1)\pi} \frac{|\sin(y+\pi)|}{y+\pi} dy = \int_{k\pi}^{(k+1)\pi} \frac{|\sin(y)|}{y+\pi} dy \leq \int_{k\pi}^{(k+1)\pi} \frac{|\sin(y)|}{y} dy = b_k,$$

dove nell'ultimo passaggio si è usata la monotonia dell'integrale e il fatto che

$$\frac{|\sin(y)|}{y+\pi} \leq \frac{|\sin(y)|}{y}, \quad \forall y \in [k\pi, (k+1)\pi].$$

Infine, b_k tende a zero quando k tende ad infinito; per dimostrarlo, osserviamo che si ha, essendo $|\sin(x)| \leq 1$,

$$0 \leq b_k = \int_{k\pi}^{(k+1)\pi} \frac{|\sin(x)|}{x} dx \leq \int_{k\pi}^{(k+1)\pi} \frac{1}{x} dx = \ln(x) \Big|_{k\pi}^{(k+1)\pi} = \ln\left(1 + \frac{1}{k}\right).$$

Dato che $\ln(1 + 1/k)$ tende a zero, anche b_k tende a zero per il teorema dei carabinieri.

Ricapitolando, abbiamo dimostrato che

$$S_n = \int_{\pi}^{(n+1)\pi} \frac{\sin(x)}{x} dx = \sum_{k=1}^n (-1)^k b_k,$$

dove b_k è una successione di numeri reali positivi, decrescente e infinitesima. Per il criterio di Leibnitz, si ha

$$\sum_{k=1}^{+\infty} (-1)^k b_k = \lim_{n \rightarrow +\infty} \sum_{k=1}^n (-1)^k b_k < +\infty,$$

e quindi

$$\int_{\pi}^{+\infty} \frac{\sin(x)}{x} dx = \lim_{n \rightarrow +\infty} S_n = \sum_{k=1}^{+\infty} (-1)^k b_k < +\infty.$$

Qualora, colti da improvvisa follia, ci chiedessimo se fosse integrabile in senso improprio su $[\pi, +\infty)$ la funzione $\frac{|\sin(x)|}{x}$, dovremmo capire se è convergente o meno la serie

$$\sum_{k=1}^{+\infty} b_k,$$

ovvero se è convergente assolutamente la serie la cui somma è l'integrale improprio di $\frac{\sin(x)}{x}$. Orbene, dimostriamo che si ha

$$b_k \geq \frac{1}{2} \ln\left(1 + \frac{4}{1+6k}\right).$$

Si ha infatti, dato che la funzione integranda è positiva,

$$b_k = \int_{k\pi}^{(k+1)\pi} \frac{|\sin(x)|}{x} dx \geq \int_{k\pi + \frac{\pi}{6}}^{(k+1)\pi - \frac{\pi}{6}} \frac{|\sin(x)|}{x} dx.$$

Osserviamo che si ha

$$|\sin(x)| \geq \frac{1}{2}, \quad \forall x \in \left[k\pi + \frac{\pi}{6}, (k+1)\pi - \frac{\pi}{6}\right],$$

e quindi

$$b_k \geq \int_k^{(k+1)\pi - \frac{\pi}{6}} \frac{|\sin(x)|}{x} dx \geq \frac{1}{2} \int_k^{(k+1)\pi - \frac{\pi}{6}} \frac{1}{x} dx = \frac{1}{2} \ln \left(\frac{k + 1 - \frac{1}{6}}{k + \frac{1}{6}} \right).$$

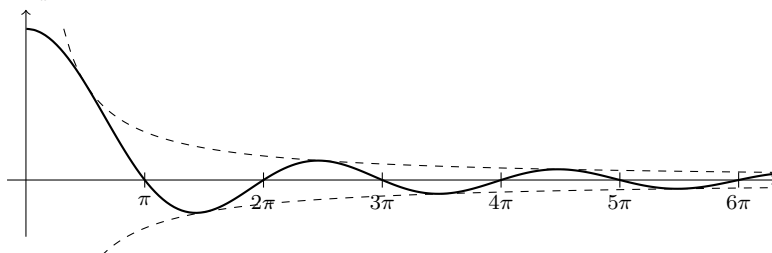
Semplificando, si ha

$$b_k \geq \frac{1}{2} \ln \left(1 + \frac{4}{1+6k} \right),$$

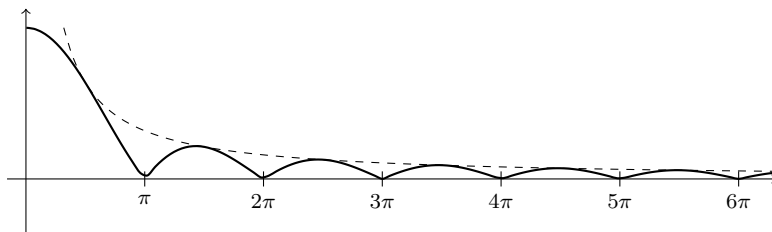
come volevasi dimostrare. Dato che la serie di termine generico $\ln \left(1 + \frac{4}{1+6k} \right)$ si comporta come la serie di termine generico $\frac{4}{1+6k}$, che diverge, si ha

$$\int_{\pi}^{+\infty} \frac{|\sin(x)|}{x} dx = \sum_{k=1}^{+\infty} b_k \geq \frac{1}{2} \sum_{k=1}^{+\infty} \ln \left(1 + \frac{4}{1+6k} \right) = +\infty,$$

e quindi la funzione $\frac{|\sin(x)|}{x}$ non è integrabile in senso improprio su $[\pi, +\infty)$.



I grafici di $\frac{\sin(x)}{x}$ e di $\pm \frac{1}{x}$ (tratteggiati)



I grafici di $\frac{\sin(x)}{x}$ e di $\frac{1}{x}$ (tratteggiato)