

Fioravante PATRONE

APPUNTI DI
ANALISI MATEMATICA I

per

SCIENZE DELL'INFORMAZIONE

A.A. 1993/94

CAPITOLO I

I NUMERI REALI

1. I numeri razionali: proprietà formali delle operazioni e della relazione d'ordine

Cominceremo col richiamare un po' di proprietà e di terminologia relativa ai numeri.

Ci sono i numeri "naturali": 1, 2, 3, ... Poi i numeri interi (vale a dire, i naturali cui aggiungiamo lo 0 e i numeri negativi). I numeri interi vengono anche detti numeri relativi, o numeri interi col segno. Ricordiamo che tra numeri interi si può sempre fare la sottrazione, ma la divisione no.

L'idea che si usa per poter fare "comunque" la divisione, è quella di introdurre le frazioni. Un modo curioso di procedere: visto che uno non sa fare la divisione tra 3 e 5, la lascia indicata (cambia solo simbolo, cioè scrive $3/5$ anziché $3:5$). Eppure, funziona. Non è comunque questa l'unica ragione per introdurre i numeri razionali (cioè le frazioni). E' chiaro che servono i numeri decimali se si vuole affrontare il problema della misura (non è altro che un confrontare multipli e sottomultipli...). Ricordiamo che i numeri decimali sono "l'altra faccia" delle frazioni. In effetti i numeri decimali non sono altro che un modo per rappresentare una frazione che ha a denominatore una potenza di 10. Ma questo va bene solo per i decimali finiti. Invece se facciamo la divisione di 1 per 3, non ci fermiamo mai: otteniamo un numero decimale illimitato. Con una caratteristica, però: è periodico. Anzi, dovrebbe essere noto che i numeri razionali (le frazioni, cioè) coincidono proprio con i numeri decimali finiti o periodici. Anche se pochi se la ricordano con esattezza, più o meno tutti sanno che c'è una regola per passare da un numero decimale periodico alla sua frazione generatrice. Di meno sanno come si faccia a garantire che da una divisione tra numeri interi viene fuori necessariamente un numero decimale finito o periodico, anche se la dimostrazione non è difficile. Ancor meno sanno che l'essere decimale finito o infinito dipende dal sistema di numerazione usato (per esempio: $1/5$ con la numerazione

binaria viene $0.001100110011\dots$ cioè $\overline{0.0011}$).

Comunque, ricapitolando, abbiamo a disposizione i cosiddetti numeri razionali (ovverossia le frazioni, ovverossia i numeri decimali finiti o periodici). Con questi possiamo fare tutte e quattro le operazioni. Vi è un solo problema: non possiamo dividere per 0 .

Quello che è importante richiamare, a questo punto, per quanto verrà in seguito, sono le cosiddette "proprietà formali" delle operazioni. Di queste, la più famosa è la proprietà commutativa: è $x+y=y+x$, qualunque siano i numeri razionali x ed y . Essa vale non solo per l'addizione ma anche per la moltiplicazione (la formuletta recita: "invertendo l'ordine dei fattori il prodotto non cambia"). Ma c'è un'altra proprietà importante, anche se meno famosa: la proprietà associativa, valida anch'essa sia per l'addizione che per la moltiplicazione. E' utile per potere scrivere con poche parentesi: per l'addizione ci dice che $x+(y+z) = (x+y)+z$, qualunque siano x , y e z . Analogamente per la moltiplicazione.

Le altre proprietà formali delle operazioni da mettere in evidenza sono forse meno conosciute, ma non meno importanti. Anzi, ad esse è legato il punto di vista più "moderno", che riduce il numero delle operazioni da 4 a 2. Basta riflettere su cosa significhi fare $7-5$: significa sommare a 7 l'opposto di 5 . Idem, $7/5$ significa moltiplicare 7 per il reciproco di 5 .

Vediamo le proprietà formali connesse a questo approccio, cominciando con l'addizione. Esiste un elemento speciale con la curiosa proprietà che sommato a qualsiasi numero lo lascia invariato. Si tratta evidentemente dello 0 , detto anche "elemento neutro della somma". Cioè: per ogni numero x razionale si ha: $x+0=x$. Possiamo sfruttare la presenza dello zero per introdurre l'idea di "opposto": infatti, per ogni numero razionale ce n'è un altro (il suo opposto) t.c. sommato a lui dà come risultato 0 . Ad esempio, $1/3$ ha come opposto $-1/3$; $-5/7$ ha come opposto $5/7$, etc. A questo punto, come già detto, non abbiamo più bisogno dell'operazione di sottrazione: interpreteremo $x-y$ come $x+(-y)$, dove $-y$ indica l'opposto di y .

Cose del tutto analoghe valgono per l'operazione di moltiplicazione. C'è l'elemento neutro, che stavolta è 1 , e poi ogni numero razionale (con l'eccezione di 0) ha il reciproco. E la divisione, $x:y$, viene semplicemente letta come $x \cdot (1/y)$, dove $1/y$ indica il reciproco di y (ovviamente con $y \neq 0$) .

A questo punto, spero di avere messo in evidenza quanto mi interessava, e cioè che sui razionali sono definite due operazioni: addizione e moltiplicazione. Esse godono delle proprietà formali sopra evidenziate. Vi è però una ulteriore proprietà che va ricordata, il cui ruolo è altrettanto essenziale delle precedenti, ma che ha una caratteristica diversa. Essa infatti coinvolge direttamente entrambe le operazioni. Si tratta della proprietà distributiva: $x \cdot (y+z) = x \cdot y + x \cdot z$ (che vale qualunque siano i numeri razionali x , y e z).

Vedremo in seguito che queste proprietà sono centrali anche per i numeri reali.

Passiamo ora ad esaminare un altro aspetto dei numeri razionali: l'ordinamento. Cioè l'argomento "disequazioni". Viene spesso sottovalutata o non adeguatamente compresa l'importanza della relazione d'ordine sui razionali: vale allora la pena di ricordare che essa è alla base del processo stesso di misurazione. Invito a riflettere su cosa significhi dire che una sbarra è lunga 1.2 metri oppure 1.28 metri.

Come è noto, vi sono quattro simboli a disposizione per l'ordinamento: " \leq " (minore o uguale), " \geq " (maggiore o uguale), " $<$ " (minore) e " $>$ " (maggiore). Per ragioni di chiarezza e per facilità di esposizione nel seguito, userò solo un simbolo, cioè quello di " \leq ", per discutere le proprietà dell'ordinamento.

Intanto per cominciare, la relazione di minore o uguale gode delle tre proprietà chiave che identificano una relazione d'ordine. E cioè la proprietà riflessiva, antisimmetrica e transitiva.

Presi infatti tre numeri razionali x , y e z , si ha che:

- i) $x \leq x$ (proprietà riflessiva)
- ii) se $x \leq y$ e $y \leq x$, allora $x = y$ (proprietà antisimmetrica)
- iii) se $x \leq y$ e $y \leq z$, allora $x \leq z$ (proprietà transitiva)

Queste tre proprietà formali sono quelle che identificano una relazione come relazione d'ordine. Non è certo il \leq in \mathbb{Q} l'unica relazione che gode di quelle proprietà: esse sono soddisfatte anche dalla relazione "essere contenuto in" (\subseteq) tra insiemi. Però essa non gode di una ulteriore proprietà, cioè quella di essere "totale". Presi comunque due insiemi A e B , non è sempre necessariamente vero che sia $A \subseteq B$ o $B \subseteq A$: possono cioè essere false entrambe le relazioni. Ad esempio, se $A = \{\text{Piero, Mario}\}$ e $B = \{\text{Piero, Giovanna}\}$. I numeri razionali invece sono sempre confrontabili tra loro. Cioè, presi due numeri razionali x e

y si ha che

iv) $x \leq y$ oppure $y \leq x$ (proprietà totale)

Si noti l'uso di "oppure", che non è da intendersi nella sua accezione "esclusiva" (corrisponde al latino "vel" , nonché ad "or" nella cosiddetta "logica circuitale"). In altre parole, può essere benissimo vero sia $x \leq y$ che $y \leq x$ (ad esempio, se $x = y = 4$).

Le quattro proprietà precedenti, identificate con i numeri i)+iv) , ci dicono che il \leq è una relazione d'ordine totale sull'insieme dei numeri razionali.

Anche qui, però, non ci si può soffermare solo sulle proprietà del \leq a sé stante. Così come prima avevamo messo in evidenza la proprietà distributiva, che "lega" tra di loro le due operazioni di addizione e di moltiplicazione, così ora ricordiamo due importanti proprietà che "legano" tra loro la relazione d'ordine e le due operazioni.

Si sa che, presi comunque tre numeri razionali x , y e z :

se $x \leq y$, allora $x+z \leq y+z$

se $x \leq y$ e $0 \leq z$, allora $x \cdot z \leq y \cdot z$.

Tutto qui. Abbiamo finito il ripasso delle proprietà chiave dei razionali che ci interessano. Ci interessano perchè le analoghe proprietà per i numeri reali, con l'aggiunta di una sola ulteriore proprietà, saranno tutto quello di cui avremo bisogno per costruire l'edificio dell'analisi.

2. I numeri reali: gli assiomi

Da quanto abbiamo detto in precedenza, dovrebbero essere evidenti due caratteristiche di questo paragrafo. Una caratteristica è la sua importanza cruciale, chiave per tutto quello che vedremo in seguito. Un'altra caratteristica è che questo paragrafo ed i successivi sono astratti, di comprensione probabilmente faticosa. Per cui potrebbe essere un'utile strategia di studio ritornare su questo primo capitolo più volte: con questo non invito a leggerlo in modo superficiale, ma dico solo che una sua rilettura potrebbe essere proficua dopo aver visto altre cose (per esempio, dopo aver fatto i limiti di successioni, o dopo aver visto il teorema sulle successioni monotone, o dopo aver acquisito un po' di esperienza dal corso parallelo di algebra).

Quello che normalmente viene indicato col nome di "numeri reali" è un costruito matematico costituito da un insieme sul quale vengono definite due operazioni e una relazione che godono di certe proprietà. Più precisamente, abbiamo la seguente:

Definizione 1 I "numeri reali" sono un insieme, che usualmente viene indicato con \mathbb{R} , sul quale sono definite due operazioni, indicate con $+$ e \cdot rispettivamente, ed una relazione indicata con \leq . Esse godono delle proprietà 1÷16 elencate nella tabella che si trova alla fine di questo capitolo. \square

Useremo solo questo, assieme alle regole del corretto argomentare, per tutta l'analisi che faremo. In tal modo, le proprietà scelte avranno ovviamente un'importanza cruciale.

Conformemente a quanto abbiamo appena detto, noi ricaveremo tutte le proprietà dei numeri reali a partire da questa tabella. Prima di intraprendere questa fatica, voglio far notare la presenza dell'assioma n. 10. Tale assioma è una novità, rispetto alla discussione fatta prima a proposito di \mathbb{Q} : ne abbiamo però bisogno per escludere che \mathbb{R} possa essere costituito da un solo elemento.

I primi 15 assiomi di \mathbb{R} sono le proprietà che ho messo in evidenza nel paragrafo precedente, relativamente alle operazioni e alla relazione d'ordine in \mathbb{Q} . Quindi \mathbb{Q} soddisfa tutti gli assiomi, tranne che per l'ultimo. E' proprio tale assioma la differenza chiave tra \mathbb{Q} e \mathbb{R} . Soffermiamoci dunque su tale proprietà. La prima cosa da osservare è che essa davvero non vale in \mathbb{Q} . Per provare questo, basta esibire una coppia di classi separate che non hanno elemento separa-

tore. Per farci venire un'idea su come si possa fare, possiamo sfruttare il fatto che, bene o male, una idea di chi siano i numeri reali ce l'abbiamo. Per esempio, è piuttosto diffusa la conoscenza del fatto che $\sqrt{2}$ non è un numero razionale (cioè: non esiste alcun numero razionale il cui quadrato sia uguale a 2). Mettiamoci allora per un momento a lavorare, ovviamente informalmente, nei numeri reali. Potremmo costruirci una coppia di classi separate in questo modo:

$$\begin{array}{c} A \qquad \sqrt{2} \qquad B \\ \text{---}//////////) (//////////\text{---} \end{array} .$$

Se il disegno non è chiaro, l'idea comunque è di prendere tutti i numeri, tranne $\sqrt{2}$. Più precisamente, mettiamo in A tutti i numeri più piccoli di $\sqrt{2}$ e in B tutti quelli più grandi. Allora l'elemento separatore dovrebbe essere proprio $\sqrt{2}$ (e solo lui). Se quindi ci mettiamo a lavorare in \mathbb{Q} , cioè se in A mettiamo tutti i razionali minori di $\sqrt{2}$ e in B i razionali maggiori di $\sqrt{2}$, tra A e B dovrebbe esserci un "buco", poiché l'unico candidato ad essere l'elemento separatore dovrebbe essere $\sqrt{2}$ che però non va bene perché non sta in \mathbb{Q} .

Questa sopra delineata, convincente o no che sia, è l'idea che effettivamente funziona. Si può infatti dimostrare che $A = \{ x \in \mathbb{Q} : x^2 < 2 \}$ e $B = \{ x \in \mathbb{Q} : x > 0 \text{ e } x^2 > 2 \}$ sono classi separate in \mathbb{Q} , senza però elemento separatore in \mathbb{Q} . Non vedremo i dettagli di questa dimostrazione. Ma l'idea è semplice. Supponiamo che sia $z \in \mathbb{Q}$ elemento di separazione: si considera z^2 e si fa vedere che non può essere né $z^2 > 2$, né $z^2 < 2$. Visto che non può neanche essere $z^2 = 2$, dato che in \mathbb{Q} non c'è alcun numero il cui quadrato faccia 2, ne deduciamo che non ci sono elementi di separazione.

Non ci si può però fermare alla sola constatazione che la proprietà di completezza vale in \mathbb{R} ma non in \mathbb{Q} . Perché questo non rende minimamente l'idea del perché ci interessi così tanto quella proprietà. Non è facile da dirsi. Si vedrà a posteriori la portata di questa proprietà. Per ora mi limito a dire che essa appare come conclusione (almeno, per ora) di tutto il processo di sistemazione dei fondamenti dell'analisi, una fatica che è durata secoli. Tale fatica ha trovato il suo compimento nel secolo scorso, portando però anche nel contempo alla cosiddetta "crisi dei fondamenti" della matematica. Penso che queste brevi annotazioni servano almeno a dare l'idea di quanto peso abbia avuto tale proprietà nello sviluppo della matematica, e di come non si tratti di una questione liquidabile in poche parole.

3. Conseguenze elementari degli assiomi

Non vedremo certo tutte le conseguenze: quel che più mi interessa mettere in evidenza è, per così dire, l'idea guida, e cioè che se volessimo potremmo davvero ricavarle tutte. Meglio però cominciare subito, di modo che si capisca dove voglio andare a parare, e rinviare alla fine del paragrafo i commenti.

Una proprietà interessante che si ottiene è l'unicità dell'elemento neutro. Siamo oltretutto moralmente obbligati a provarlo, in quanto abbiamo utilizzato questo fatto nell'elenco stesso degli assiomi. Vediamo subito "come si fa", notando che useremo solo gli assiomi n. 1, 2 e 3.

Siano a, a' due elementi neutri. Per definizione di elemento neutro, ciò significa che:

$$\begin{cases} a+x = x & \text{per ogni } x \in \mathbb{R} \\ a'+y = y & \text{per ogni } y \in \mathbb{R} \end{cases}$$

Allora, in particolare (scegliendo a' come x ed a come y): $a+a' = a'$ e $a'+a = a$. La proprietà commutativa ci dice che $a+a' = a'+a$, da cui si ha che $a' = a$. Pertanto l'elemento neutro per l'addizione è unico.

Si dimostra facilmente anche che l'opposto di un numero reale è unico. Siano y, z due opposti di x , cioè t.c. $x+y=0$ e $x+z=0$. Allora:

$$\begin{aligned} z &= && \text{aggiungendo a } z \text{ lo } 0, \text{ elemento neutro per il } + \\ = z+0 &= && \text{per la proprietà commutativa del } + \\ = 0+z &= && \text{aggiungendo } z \text{ a entrambi i membri della prima relazione} \\ = (x+y)+z &= && \text{per la proprietà commutativa del } + \\ = (y+x)+z &= && \text{per la proprietà associativa del } + \\ = y+(x+z) &= && \text{perché } z \text{ è opposto di } x \\ = y+0 &= && \text{per definizione di elemento neutro} \\ = y & . \end{aligned}$$

Abbiamo ottenuto che $y=z$. Quindi l'opposto di un numero reale è unico.

Da qui segue anche che $-(-x)=x$. Infatti, essendo $-x$ l'opposto di x , per definizione si ha $x+(-x)=0$. Ma la proprietà commutativa ci dà $(-x)+x = 0$. E quindi x è opposto di $-x$. Per l'unicità dell'opposto di un numero, possiamo concludere che $x = -(-x)$.

Trucchi simili permettono di provare l'unicità dell'elemento neutro per la moltiplicazione.

Un'altra proprietà interessante da notare e da provare è che $a \cdot 0 = 0$,

qualunque sia $a \in \mathbb{R}$. Come si dimostra? Notiamo, per cominciare, che:

$$a \cdot 0 = a \cdot (0+0) = a \cdot 0 + a \cdot 0$$

($0 = 0+0$ perché 0 è l'elemento neutro per l'addizione; la seconda uguaglianza è conseguenza della proprietà distributiva).

Aggiungendo ad entrambi i membri $-(a \cdot 0)$, cioè l'opposto di $a \cdot 0$, si ha:

$$\begin{aligned} 0 &= && \text{per definizione di opposto} \\ &= a \cdot 0 + (-(a \cdot 0)) = && \text{per quanto visto sopra} \\ &= (a \cdot 0 + a \cdot 0) + (-(a \cdot 0)) = && \text{per la proprietà associativa del } + \\ &= a \cdot 0 + (a \cdot 0 + (-(a \cdot 0))) = && \text{per definizione di opposto} \\ &= a \cdot 0 . \end{aligned}$$

Sembrano giochi di prestigio. In realtà, per capire davvero quello che si sta facendo, non bisogna scordarsi quale è l'obbiettivo: con i mattoni a nostra disposizione (gli assiomi) stiamo costruendo dei pezzi di casa che poi useremo così già belli fatti per proseguire nella costruzione. E' chiaro che all'inizio il lavoro e' noioso e richiede molta attenzione (sennò poi tutto verrà storto). Una fonte di disagio può essere che uno non capisce perché si fa tutto questo: sembra che stiamo dimostrando cose ovvie. E' vero se si pensa a tutto quanto uno ha già imparato di matematica: ma il punto di vista qui è che vogliamo far vedere che possiamo "rifare tutto" a partire solo dagli assiomi 1÷16. Quindi non possiamo dare per scontato nulla. Un'altra fonte di disagio può essere che uno non capisce come possono venire in mente certi trucchi. Detto altrimenti, se uno avesse dovuto dimostrare da solo che $a \cdot 0 = 0$, magari costui non avrebbe saputo da che parte voltarsi. Credo che sia sensata questa sensazione di disagio: ad un muratore apprendista può sembrare un'impresa immane riuscire a tirar su un muretto di mattoni in costa. La situazione di chi legge queste righe può essere simile: con l'esperienza, queste difficoltà iniziali appariranno inspiegabili. Si può comunque dare qualche idea per far capire che non bisogna essere marziani per affrontare questi problemi. Tanto per cominciare, la proprietà in questione, e cioè $a \cdot 0 = 0$, coinvolge le due operazioni (l'addizione è coinvolta per la presenza di 0 che è l'elemento neutro per l'addizione). Era quindi praticamente scontato che si dovesse usare la proprietà distributiva, in quanto proprietà chiave che "lega" le due operazioni. Essendo poi coinvolto 0 e non avendo per giunta molto di più su cui lavorare, era plausibile che bisognasse sfruttare la sua caratteristica, cioè quella di essere elemento neutro per l'addizione: queste sono esattamente le idee

che sono servite per ottenere il risultato preliminare che $a \cdot 0 = a \cdot 0 + a \cdot 0$.

La proprietà che $a \cdot 0 = 0$ ha varie conseguenze interessanti. La più importante è che preclude la possibilità di trovare il reciproco di 0 per la moltiplicazione. Infatti, l'esistenza del reciproco di 0, cioè di un numero $x \in \mathbb{R}$ t.c. $0 \cdot x = 1$, è in contraddizione con quanto appena trovato, che ci garantisce che $0 \cdot x = 0$ (ricordare che $0 \neq 1$ per l'assioma 10!).

Si noti che, "fortunatamente", nell'assioma 8 abbiamo posto la condizione $x \neq 0$: se avessimo voluto "pretendere di più", cioè garantire l'esistenza del reciproco di ogni numero reale, avremmo ottenuto il risultato di trovarci con un sistema assiomatico contraddittorio e quindi del tutto inutile.

Già che parliamo dell'assioma 8, voglio far notare che in esso non c'è scritto esplicitamente che 0 non ha reciproco! E' detto che i numeri diversi da zero hanno reciproco, il che è ben altra cosa. Che 0 non abbia reciproco l'abbiamo appena finito di dimostrare ora.

Un'altra conseguenza di $a \cdot 0 = 0$ sono le "regole dei segni". Cioè che $+++ = +$, $+- = -$, etc. La validità di queste regole è molto spesso accettata più per ragioni di sopravvivenza scolastica che per intima convinzione: non è mia intenzione dare qui delle motivazioni applicative a supporto di queste regole (la più difficile da accettare è che $-- = +$), ma almeno far vedere che esse sono conseguenza degli assiomi sì.

Prima però di vedere questo, ho bisogno di un momento di riflessione, e cioè: che cosa significa dire che "più per meno uguale meno"? Ci sono due interpretazioni possibili. Una è che se moltiplico un numero positivo¹ per un numero negativo, ho come risultato un numero negativo. Ed è forse questa la interpretazione usuale. Ma ce n'è un'altra. E cioè che se moltiplico x per l'opposto di y , cioè faccio $x \cdot (-y)$, quello che ottengo è l'opposto di $x \cdot y$. Insomma, in formule: $x \cdot (-y) = -(x \cdot y)$.

Ebbene, le regole dei segni possono essere dimostrate a partire dagli assiomi in entrambe le loro accezioni. Io dimostrerò quest'ultima versione.

Cominciamo a vedere che per ogni numero reale z si ha $-z = (-1) \cdot z$. Infatti: $z + ((-1) \cdot z) = 1 \cdot z + ((-1) \cdot z) = (1 + (-1)) \cdot z = 0 \cdot z = 0$. Quindi $(-1) \cdot z$ è

¹ Cioè maggiore di zero. Vale a dire che, se x è questo numero, si ha $0 \leq x$ e $x \neq 0$. Non è pedanteria, questa: ho detto che avremmo ricavato tutto dagli assiomi, quindi devo stare attento a non introdurre una terminologia che non sia definibile a partire dagli assiomi.

l'opposto di z (ricordiamo che l'opposto è unico!), come si era affermato. A questo punto:

$$\begin{aligned}
 x \cdot (-y) &= && \text{per quel che abbiamo appena dimostrato} \\
 = x \cdot ((-1) \cdot y) &= && \text{per la proprietà associativa della moltiplicazione} \\
 = (x \cdot (-1)) \cdot y &= && \text{per la proprietà commutativa della moltiplicazione} \\
 = ((-1) \cdot x) \cdot y &= && \text{per la proprietà associativa della moltiplicazione} \\
 = (-1) \cdot (x \cdot y) &= && \text{per quel che abbiamo appena dimostrato} \\
 = -(x \cdot y)
 \end{aligned}$$

Un'altra conseguenza importante degli assiomi, ed un legame importante tra operazioni e relazione d'ordine, è la possibilità di definire il valore assoluto di un numero, con le sue proprietà (in particolare, la cosiddetta disuguaglianza triangolare).

Dato $x \in \mathbb{R}$, definiamo $|x|$ nel modo seguente:

$$|x| = \begin{cases} x & \text{se } 0 \leq x \\ -x & \text{se } x \leq 0 \end{cases}$$

La prima osservazione da fare è che si ha $0 \leq |x|$ per ogni $x \in \mathbb{R}$. Questo è ovvio se $0 \leq x$ (perché allora $|x| = x$). Se invece $x \leq 0$, aggiungendo membro a membro $-x$, otteniamo per l'assioma 15 che $x + (-x) \leq 0 + (-x)$ e quindi $0 \leq -x = |x|$.

Un'altra proprietà utile, per quanto riguarda il valore assoluto è l'equivalenza tra le relazioni $|x| \leq a$ e $-a \leq x \leq a$, dove a è un numero reale dato soddisfacente la condizione $0 \leq a$. Suggestivo di provare a dimostrare questa osservazione.

Come detto, la proprietà importante è la disuguaglianza triangolare: per ogni x ed y numeri reali si ha: $|x+y| \leq |x| + |y|$. Ma si ha anche $|x \cdot y| = |x| \cdot |y|$. Vedremo come si fa a dimostrare la disuguaglianza triangolare. Il modo in cui è definito il valore assoluto suggerisce una strada per la dimostrazione: distinguere i casi da analizzare a seconda che x o y siano minori o uguali di zero o viceversa. I casi possibili sono dunque quattro:

PRIMO CASO: $0 \leq x$ e $0 \leq y$. Dall'assioma 15 si ha che $0+y \leq x+y$, ma $0+y = y$ perché 0 è l'elemento neutro per l'addizione: quindi abbiamo $y \leq x+y$. Da $0 \leq y$ e $y \leq x+y$ otteniamo, per la proprietà transitiva (assioma 13) $0 \leq x+y$. Ma allora $|x+y| = x+y$, per definizione di valore assoluto. Sempre per definizione di valore assoluto, $|x| = x$ e $|y| = y$. Quindi in questo caso

$|x+y| = x+y = |x|+|y|$. Cioè $|x+y| = |x|+|y|$ e quindi per la proprietà riflessiva (assioma 11) si ottiene $|x+y| \leq |x|+|y|$.

SECONDO CASO: $0 \leq x$ e $y \leq 0$. Consideriamo $x+y$. Per l'assioma 14 sono possibili due casi: o $x+y \leq 0$ o $0 \leq x+y$. Esaminiamo il primo sottocaso. Il lettore dovrebbe controllare la validità di tutti i passaggi, sulla base degli assiomi o di quel che abbiamo dimostrato prima. E' $y+x \leq 0+x = x \leq x+|y| = |x|+|y|$. E inoltre $-(|x|+|y|) = (-1) \cdot (|x|+|y|) = (-1)|x|+(-1)|y| = -|x|+(-|y|) = -|x|+(-(-y)) = -|x|+y \leq y \leq y+x$. Avendo ottenuto $y+x \leq |x|+|y|$ e $-(|x|+|y|) \leq y+x$, per quanto notato a suo tempo possiamo affermare che $|x+y| \leq |x|+|y|$.

La dimostrazione dell'altro sottocaso e dei due casi rimanenti è lasciata al lettore (che naturalmente può anche consultare il C-S o altri manuali).

Esercizio 1 Giustificare tutti i passaggi fatti nella dimostrazione del secondo caso. □

Come promesso all'inizio di questo paragrafo, ecco i commenti.

Immagino che le conseguenze degli assiomi che ho dimostrato in questo paragrafo non abbiano riservato nessuna sorpresa sensazionale. Erano cose presumibilmente già note al lettore. Ma è proprio da questo fatto che di solito sorge il maggior imbarazzo negli studenti, i quali si chiedono smarriti se (in particolare all'esame) gli sarà chiesto anche di dimostrare che $2+2=4$. Per fortuna la risposta è no² . La risposta può essere no perché le proprietà elementari dei numeri reali ottenibili dagli assiomi sono sostanzialmente quelle già conosciute: se uno va a vedere la lista delle R.1÷R.20 sul C-S, vedrà che esse esprimono proprietà ben note.

"Ma allora?" dirà, giustamente, il lettore impaziente. Per rispondere, mi limito a ribadire quanto ho affermato all'inizio del paragrafo: lo scopo è far vedere come si fa. Per rendere comprensibile come tutte quelle regolette relative all'algebra, alle equazioni e alle disequazioni che lo studente medio ha in testa (al di fuori, però, di una chiara trama), si possano ottenere a partire da quei soli 16 assiomi di \mathbb{R} . Quindi lo scopo principale è di offrire uno schema unitario per la comprensione di ciò che si è fatto nelle scuole precedenti in "algebra". Non intendo tuttavia tacere un'altra ragione per questo paragrafo: e

² Anche se nel paragrafo 6 faremo ben di peggio. Addirittura cercheremo di capire cosa voglia dire che $1+1=2$!!

cioè quella di abituare lo studente a un modo rigoroso di ragionare. Chi fosse recalcitrante, deve rassegnarsi: lo "standard" di rigore richiesto è sensibilmente più alto³ di quello richiesto nelle scuole secondarie. E allora questo paragrafo, essendo fondato su una breve lista di assiomi, ben si presta come "allenamento" per la "ginnastica" mentale che è l'analisi matematica⁴.

³ Non è il massimo livello di rigore possibile (non esiste, un massimo livello).

⁴ A costo di essere noioso con queste note a piè pagina, ci tengo a dire che l'analisi matematica non è solo una ginnastica mentale. Vederla così sarebbe sciocco, anche se è fra i principali aspetti "formativi" dell'analisi vista come disciplina di insegnamento.

4. Relazione d'ordine e intervalli

Mentre il paragrafo precedente è stato dedicato in prevalenza alle operazioni in \mathbb{R} , in questo paragrafo concentreremo l'attenzione sulla relazione d'ordine.

Abbiamo a disposizione \leq , relazione d'ordine totale e completa su \mathbb{R} . Vogliamo vedere come si possano definire semplicemente a partire da questa le relazioni $<$, \geq , $>$. Casomai uno non se ne fosse accorto, faccio notare che nel precedente paragrafo sono stato attento ad usare sempre e solo \leq (neanche \geq), in quanto gli assiomi di \mathbb{R} fanno riferimento esclusivamente a tale relazione.

Definizione 1 Dati $x, y \in \mathbb{R}$, diciamo che:

$x < y$ se $x \leq y$ e $x \neq y$

$x \geq y$ se $y \leq x$

$x > y$ se $y < x$. \square

Queste relazioni "nuove" godono delle proprietà consuete. Tanto per fare un esempio, la relazione $<$ è transitiva. Siano infatti $x, y, z \in \mathbb{R}$ t.c. $x < y$ e $y < z$. Allora, per la definizione 1, abbiamo $x \leq y$ e $y \leq z$, da cui per la transitività di \leq (assioma 13), si ha che $x \leq z$. Resta da provare che $x \neq z$, per poter dire che $x < z$. Se per assurdo fosse $x = z$, allora sarebbe anche $z \leq x$ (il \leq è riflessivo, assioma 11) e quindi $z \leq y$ (perchè? perchè \leq è transitiva e abbiamo $z \leq x$ e $x \leq y$). Ma $z \leq y$ è incompatibile con l'assunto $y < z$. In effetti, $y < z$ implica $y \leq z$, che assieme alla relazione $z \leq y$ e all'antisimmetria (assioma 12) ci dà $y = z$, il che però è incompatibile con $y < z$.

Un'altra proprietà che dimostriamo in quanto ci servirà nel prossimo paragrafo, è la seguente.

Proposizione 1 Dati $x, y, z \in \mathbb{R}$, se $x < y$ allora $x + z < y + z$. \square

Dimostrazione Poiché $x < y$, allora $x \leq y$ e quindi per l'assioma 15 si ha $x + z \leq y + z$. Dobbiamo pertanto solo garantire che non può essere $x + z = y + z$. Ma se così fosse, aggiungendo ad entrambi i membri l'opposto di z , cioè $-z$, avremmo: $(x+z)+(-z) = (y+z)+(-z)$. Ovverossia: $x+(z+(-z)) = y+(z+(-z))$. Ovvero $x = y$, il che è in contrasto con la ipotesi che fosse $x < y$. \square

Avendo a disposizione sia \leq che $<$, possiamo introdurre una notazione utilizzata per descrivere particolari sottoinsiemi di \mathbb{R} , detti intervalli.

Definizione 2 Dati $a, b \in \mathbb{R}$, definiamo:

$$[a, b] = \{ x \in \mathbb{R} : a \leq x \leq b \}$$

$$[a, b[= \{ x \in \mathbb{R} : a \leq x < b \}$$

$$[a, +\infty[= \{ x \in \mathbb{R} : a \leq x \}$$

$$]a, +\infty[= \{ x \in \mathbb{R} : a < x \}$$

$$]-\infty, +\infty[= \mathbb{R} . \square$$

Esercizio 1 Definire $]a, b]$, $]a, b[$, $]-\infty, b]$, $]-\infty, b[$. \square

Osservazione 1 Si noti che gli intervalli di estremi a, b sono vuoti se $b < a$. Non solo, ma se $a = b$, abbiamo $[a, b] = \{a\}$ e $[a, b[=]a, b] =]a, b[= \emptyset$. Quindi anche \emptyset e $\{a\}$ sono intervalli. Questi si dicono però intervalli degeneri. Pertanto, d'ora in poi quando farò riferimento agli intervalli assumerò sempre implicitamente che sia $a < b$, salvo esplicito avviso contrario. \square

5. Estremo superiore

In questo paragrafo, dall'assioma di completezza dedurremo una proprietà molto importante, relativa al cosiddetto estremo superiore, che useremo molto spesso nel seguito. Prima però abbiamo bisogno di un po' di definizioni.

Definizione 1 Siano dati $A \subseteq \mathbb{R}$ e $k \in \mathbb{R}$. Diremo che k è un maggiorante per A se $a \leq k$ per ogni $a \in A$. \square

Notare che a k non si richiede di appartenere ad A . Se $k \in A$, allora diciamo che k è massimo per A e lo indichiamo con $\max A$. Se si ha $k \leq a$ per ogni $a \in A$, allora diremo che k è un minorante (e se $k \in A$ sarà detto minimo e lo indicheremo con $\min A$).

Definizione 2 $A \subseteq \mathbb{R}$ si dice superiormente limitato se esiste $k \in \mathbb{R}$ t.c. k è un maggiorante per A . \square

Analoga è la definizione di inferiormente limitato. Se infine esistono $h, k \in \mathbb{R}$, rispettivamente minorante e maggiorante per A , allora si dice che A è limitato.

Fatte queste premesse, osserviamo che non è garantito che un insieme superiormente limitato abbia massimo.

Esempio 1 $[0,1[$ non ha massimo. Infatti 1 è un maggiorante, ma $1 \notin [0,1[$. Neanche $k > 1$ è un massimo per $[0,1[$ (perchè $k \notin [0,1[$). Se prendiamo $k \in [0,1[$, abbiamo $(1+k)/2 \in [0,1[$ e $k < (1+k)/2$, pertanto k non è un maggiorante. Infine, se $k < 0$, di nuovo k non può essere massimo perchè $k \notin [0,1[$. \square

Per molte applicazioni è importante però avere a disposizione un "surrogato" del massimo. L'idea è semplice. Nell'esempio, 1 non è massimo, però è il più piccolo di tutti i maggioranti. Possiamo allora sperare che possa svolgere certe funzioni di solito assegnate al massimo, pur senza esserlo. Per questa ragione, introduciamo la seguente definizione.

Definizione 3 Sia $A \subseteq \mathbb{R}$. E sia $M_A = \{ k \in \mathbb{R} : k \text{ è maggiorante per } A \}$. Se esiste $\min M_A$, diciamo che tale minimo è l'estremo superiore di A e lo indichiamo con $\sup A$. \square

L'interesse della definizione 3 sta nel fatto (che dimostreremo) che ogni insieme superiormente limitato di \mathbb{R} ha estremo superiore, con una sola ecce-

zione. L'eccezione è l'insieme vuoto, \emptyset : poiché \emptyset non ha elementi, ogni elemento di \mathbb{R} è un suo maggiorante⁵. Abbiamo così che $M_{\emptyset} = \mathbb{R}$, ed \mathbb{R} non ha minimo.

Esercizio 1 Dimostrare che, per ogni $x \in \mathbb{R}$, $x-1 < x$ (provare dapprima che $1 > 0$, e notare che $x-1$ è una scrittura abbreviata per $x+(-1)$). Dedurre da ciò che \mathbb{R} non ha minimo. \square

Proviamo quindi il risultato sopra indicato.

Teorema 1 (esistenza dell'estremo superiore) Sia $A \subseteq \mathbb{R}$, $A \neq \emptyset$ e superiormente limitato. Allora esiste $\lambda \in \mathbb{R}$ t.c. $\lambda = \sup A$. \square

Dimostrazione Sia $A \subseteq \mathbb{R}$, $A \neq \emptyset$ e superiormente limitato. Consideriamo $M_A = \{ m \in \mathbb{R} : a \leq m \text{ per ogni } a \in A \}$, cioè l'insieme dei maggioranti di A . Si verifica facilmente che $\{ A, M_A \}$ è una coppia di classi separate. Pertanto esiste $z \in \mathbb{R}$ t.c. per ogni $a \in A$ e $b \in M_A$, si ha $a \leq z \leq b$. Essendo $a \leq z$ per ogni $a \in A$, si ha che z è un maggiorante. Quindi $z \in M_A$. Poiché $z \leq b$ per ogni $b \in M_A$, abbiamo che z è proprio il minimo dei maggioranti, cioè $z = \sup A$. \square

E' difficile sopravvalutare l'importanza di questo teorema, tanto è vero che abbiamo riportato nella tabella la proprietà espressa dal precedente teorema : in effetti addirittura potremmo sostituire l'assioma 16 con la proprietà 17 (vedi tabella). Cioè: il fatto che un sottoinsieme di \mathbb{R} non vuoto e superiormente limitato abbia estremo superiore, è addirittura una formulazione della proprietà di completezza equivalente a quella che abbiamo scelto noi, basata sulle classi separate. Proveremo ora questo fatto. Cioè che se agli assiomi 1+15 aggiungiamo come ulteriore assioma la proprietà 17, otteniamo la proprietà 16 come teorema.

Ed eccone la formalizzazione e la dimostrazione.

Teorema 2 Le proprietà 1+15 e 17 implicano la proprietà 16. \square

Dimostrazione Sia quindi (A, B) una coppia di classi separate. Allora $A \neq \emptyset$ per definizione di coppia di classi separate; poiché $B \neq \emptyset$ per la stessa ragione, abbiamo che esiste $b \in B$ e quindi che A ha almeno un maggiorante, cioè è superiormente limitato. In altre parole, il fatto che (A, B) sia una coppia di classi separate garantisce che $A \neq \emptyset$ e che A è superiormente limitato. Quindi

⁵ Invito chi non ci crede, o ha dei dubbi, a pazientare ed attendere la lettura del § II.3.

esiste $z = \sup A$. Proviamo che per ogni $a \in A$, per ogni $b \in B$, $a \leq z \leq b$. Che sia $a \leq z \quad \forall a \in A$ è ovvio in quanto il \sup è un maggiorante. Per l'altra disuguaglianza, notiamo che gli elementi di B sono tutti dei maggioranti per A (conseguenza della definizione di coppia di classi separate) e pertanto $z \leq b$ per ogni $b \in B$ perché z , essendo il \sup di A , è il più piccolo dei maggioranti di A . \square

Osservazione 1 Se A non è superiormente limitato, ovviamente non può esservi il minimo dei maggioranti, in quanto di maggioranti non ce ne sono. In tal caso, se $A \neq \emptyset$, si usa dire che $\sup A = +\infty$. Se $A = \emptyset$, si usa dire che $\sup A = -\infty$. \square

Deve essere ben chiaro che la scrittura $\sup A = +\infty$ è solo un modo per dire sbrigativamente che A non è superiormente limitato ed è non vuoto: non si intende dire che $+\infty$ è un numero reale! Ribadito che $+\infty$ e $-\infty$ non sono numeri reali, va anche detto che queste convenzioni non sono "a caso": volendo, si potrebbe "estendere" l'insieme \mathbb{R} con due nuovi elementi ($-\infty$ e $+\infty$ appunto). A tale insieme si può estendere anche la relazione d'ordine \leq , senza alcuna difficoltà. Qualche difficoltà presenta invece l'estensione delle due operazioni a questo insieme: esse non godono più delle buone proprietà che hanno in \mathbb{R} . Io non userò questa estensione di \mathbb{R} . Chi fosse interessato può consultare C-S a pag. 36 (definizione 9.6). Faccio solo notare che in questa estensione effettivamente $\sup A = +\infty$, quando A è non vuoto e superiormente illimitato: da qui la ragione per l'uso di questa notazione.

Osservazione 2 Naturalmente quel che è stato appena detto è una sciocchezza colossale, messa lì come esca per attirare i formalisti più sfrenati. Io la penso ben diversamente, come cercherò di spiegare qui appresso.

Vi sono molte situazioni in cui è interessante l'idea di "enormemente grande", o di "allontanarsi indefinitamente", o di "crescere senza limiti", e simili. Ad esempio, se abbiamo due cariche elettriche dello stesso segno, la loro "tendenza spontanea" è quella di allontanarsi sempre di più. E' allora utile avere a disposizione, in vari contesti, una formalizzazione matematica di queste idee, le quali presentano, si badi bene, delle diverse sfumature di significato. Dato che le idee in gioco hanno comunque qualcosa in comune, non è strano che si usi lo stesso simbolo per denotare il concetto matematico che le traduce a livello for-

male. Ecco quindi perché lo stesso simbolo " ∞ " è usato in casi distinti:

- nei limiti
- per indicare il "sup" di un insieme che non ha estremo superiore in \mathbb{R} , non essendo superiormente limitato.

La raffinatezza maggiore (per i formalisti incalliti) è poi quella di riuscire a escogitare un modo per avere a disposizione un numero "enormemente grande": ovviamente il simbolo che verrà usato per indicare tale "numero" sarà ∞ .

Ed ecco allora spuntare fuori il sistema dei "numeri reali estesi", dentro al quale $+\infty$ e $-\infty$ hanno lo stesso diritto di cittadinanza che 257. All'interno di questo sistema dei numeri reali estesi, è vero che ogni sottoinsieme ha estremo superiore, senza eccezione alcuna: quindi ad esempio il sup dei numeri positivi è $+\infty$. Ma da qui a dire che questa è la "vera ragione" per cui uno dice che $\sup A = +\infty$ quando A non è superiormente limitato (e non vuoto), ce ne corre! \square

Concluderò questo paragrafo parlando brevemente di "classi contigue". Lo faccio perché esse saranno utili nel seguito (nel capitolo V), ma anche per offrire a chi già avesse sentito parlare di classi contigue un'occasione di rilettura ed anche di inquadramento rispetto all'approccio che ho seguito. Ma soprattutto lo faccio perché mi offre un'occasione magnifica per introdurre il paragrafo successivo.

Definizione 4 Una coppia (A, B) si dice coppia di classi contigue se è una coppia di classi separate ed inoltre gode della seguente proprietà:

per ogni numero reale $\varepsilon > 0$, esistono due numeri $a \in A$ e $b \in B$
t.c. $b - a < \varepsilon$. \square

L'idea è ovvia: le due classi sono "appiccicate", ovverossia non c'è spazio vuoto tra di loro. Se questa è l'idea, dovrebbe essere ovvio anche il seguente teorema (almeno nella prima parte in cui afferma che vi è un unico elemento separatore).

Teorema 3 Sia (A, B) una coppia di classi contigue. Allora esiste uno ed un solo $z \in \mathbb{R}$, elemento separatore per le due classi; inoltre $z = \sup A$ e $z = \inf B$. \square

Dimostrazione Che un elemento separatore ci sia, è garantito dall'assioma 16, essendo (A, B) una coppia di classi separate. Supponiamo allora per assurdo che

non sia unico, e siano quindi $z_1, z_2 \in \mathbb{R}$ con $z_1 \neq z_2$ entrambi elementi separatori. Supponiamo $z_1 < z_2$ (altrimenti scambiamo gli indici). Basta prendere $\varepsilon = \frac{z_2 - z_1}{1+1}$ per avere una contraddizione. Infatti, per definizione di coppia di classi contigue, dovrebbero esistere $a \in A$ e $b \in B$ t.c. $b - a < \varepsilon$. Ma $a \leq z_1 < z_2 \leq b$. E quindi $z_2 - z_1 \leq b - a < \varepsilon = \frac{z_2 - z_1}{1+1}$, il che è impossibile.

Quindi l'elemento separatore è unico. Ma allora deve essere per forza $z = \sup A$ (indicando, come di consueto, con z l'elemento separatore), poiché abbiamo già visto nella dimostrazione del teorema 2 che $\sup A$ è un elemento separatore: essendovi un solo elemento separatore, esso non può essere altri che $\sup A$. Per l'affermazione che $z = \inf B$, la dimostrazione del teorema 2, opportunamente adattata, garantisce che $\inf B$ è un elemento separatore: al resto ci pensa l'unicità. \square

Esercizio 1 Provare che $a \leq z_1 < z_2 \leq b$ implica che $z_2 - z_1 \leq b - a$. \square

Esercizio 2 Provare che, se $\alpha \geq 0$, allora $\frac{\alpha}{1+1} \leq \alpha$. Dedurre che $z_2 - z_1 < \frac{z_2 - z_1}{1+1}$ è impossibile quando $z_1 < z_2$. Se non ce la fate, leggete l'osservazione seguente. \square

Osservazione 3 Spero proprio che ogni lettore si sia chiesto perché ho usato $1+1$ anziché 2 . La ragione è ovvia, per chi abbia dimestichezza col metodo assiomatico. Infatti, non c'è scritto da nessuna parte che cosa voglia dire 2 . Risolveremo questo problema nel prossimo paragrafo. \square

Osservazione 4 Chissà se qualcuno si è accorto che ho usato 2 nella dimostrazione del teorema 3! E allora, perché mi sono permesso di usare il simbolo 2 come indice ma non l'ho usato al posto di $1+1$? Anche questo problema, un po' più "sottile" del precedente, lo risolveremo nel prossimo paragrafo. \square

Esercizio 3 In questo capitolo ho già usato 2 in modo illegale. Dove? \square

6. I naturali, interi e razionali come sottoinsiemi di \mathbb{R}

L'approccio assiomatico ad \mathbb{R} che ho adottato ha alcuni pregi, ma anche difetti. Non mi riferisco solo al fatto che è un approccio astratto e difficile da assimilare, all'inizio. Ci sono anche delle difficoltà intrinseche di questo approccio. Una è che molte cose sono da ricostruire di sana pianta. Ad esempio, proprietà ben note e largamente sfruttate già nelle scuole precedenti, come il fatto che $x^2 \geq 0$ per ogni numero reale x . Noi abbiamo visto la dimostrazione solo di un campione estremamente ristretto di tali regole, nello spirito di: "vedete, si fa così e così: se volessimo, potremmo perdere il nostro tempo a provare tutto, ma non ne vale la pena".

In questo paragrafo affrontiamo (anche se lo spirito rimane sempre quello) un problema importante, che può essere esemplificato dalla domanda: "chi è 2 ?". Tutti sanno che 2 è il simbolo usato per indicare un particolare numero naturale, nel consueto sistema di numerazione decimale (ed usando le cosiddette cifre arabe, anche se gli arabi usano un simbolo diverso...). Ma 2 è un elemento di \mathbb{R} ? L'idea è che la risposta dovrebbe essere sì: dopotutto, con i numeri reali si cerca di allargare il più possibile il campo dei numeri. Ma dove sta scritto negli assiomi che 2 sta in \mathbb{R} ? O, almeno, come si fa a ricavarlo? Generalizzando, abbiamo il seguente problema: possiamo "ricostruire" dentro ad \mathbb{R} dei sottoinsiemi che si comportano in tutto e per tutto come i naturali, o gli interi, o i razionali?

Come facciamo, tanto per cominciare, ad identificare i naturali? La chiave di tutto sta proprio nella risposta alla domanda da cui siamo partiti: chi è 2 ? Esso è il successore di 1 , si potrebbe dire: ma questo ci aiuta poco, in \mathbb{R} . Più utile è osservare che $2 = 1 + 1$, visto che abbiamo a disposizione 1 (attenzione: quello che conta è che c'è un numero reale che si comporta sotto molti aspetti come il numero naturale 1 , non certo il fatto che sia stato usato lo stesso simbolo per indicarlo! Se avessi usato \cdot per indicare l'elemento neutro per la moltiplicazione, definirei 2 come $\cdot + \cdot$). Vedremo che questo approccio funziona, ma occorre un po' di cautela. Chi ci garantisce che 2 sia un numero reale "nuovo", cioè diverso da quelli che conosciamo, vale a dire 0 e 1 ?

Proposizione 1 $1+1 \neq 0$ e $1+1 \neq 1$. \square

Dimostrazione Osserviamo che $0 < 1$. Infatti, dati 0 e 1 , essendo diversi

per l'assioma 10 ed essendo l'ordine totale, deve essere vero o che $0 < 1$ o che $1 < 0$. Se supponiamo $1 < 0$, abbiamo (per la Proposizione 1, § 4) $1 + (-1) < 0 + (-1)$, cioè $0 < -1$. Moltiplicando $1 < 0$ membro a membro per -1 otteniamo (per l'assioma 15) che $1 \cdot (-1) < 0 \cdot (-1)$. Cioè $-1 < 0$. Da cui $(-1) + 1 < 0 + 1$, cioè $0 < 1$, contraddizione. Pertanto non può essere altro che $0 < 1$. Ma allora $0 + 1 < 1 + 1$ (per la proposizione 1. § 4), cioè $1 < 1 + 1$ e quindi in particolare $1 \neq 1 + 1$.

Poiché $0 < 1$ e $1 < 1 + 1$, per la proprietà transitiva abbiamo $0 < 1 + 1$ e quindi $0 \neq 1 + 1$. \square

Abbiamo quindi provato che \mathbb{R} contiene almeno tre elementi. Indicheremo con 2 il numero $1 + 1$. Naturalmente si può proseguire e mostrare che $2 + 1$ (indicato, guarda caso, con 3) è diverso da ognuno dei numeri $0, 1$, e 2 , e così via. Si può dimostrare che il sottoinsieme costituito da $1, 2, 3, \dots$ "si comporta come i numeri naturali" e lo denoteremo pertanto con \mathbb{N} . Poi, se consideriamo il sottoinsieme di \mathbb{R} costituito dagli elementi di \mathbb{N} , da tutti i loro opposti e da 0 , abbiamo che esso si comporta come gli interi e lo indicheremo con \mathbb{Z} . Ancora, l'insieme di tutti i numeri del tipo p/q con $p \in \mathbb{Z}$ e $q \in \mathbb{N}$, può essere identificato con l'insieme dei numeri razionali, e cioè \mathbb{Q} . Abbiamo così ricostruito dentro \mathbb{R} , come sottoinsiemi di \mathbb{R} , gli insiemi numerici \mathbb{N} , \mathbb{Z} e \mathbb{Q} . Spero sia ovvio che ho ommesso un numero enorme di dimostrazioni.

Vorrei essere molto chiaro su un punto. E cioè: questo capitolo è dedicato ai numeri reali, basandosi sul presupposto che i numeri naturali, interi e razionali siano noti con le loro proprietà. Per esempio, nel primo paragrafo, non ho fatto altro che richiamare proprietà dei numeri razionali note al lettore⁶: non avevo lo scopo di dimostrarle o di "stabilirle" in alcun modo, ma volevo solo, per l'appunto, richiamarle. Analogamente, proprietà riguardanti i numeri interi come quella che fa riferimento alla distinzione tra pari e dispari o alla scomposizione in fattori primi, l'idea stessa di numero primo, sono state date per scontate.

Quindi l'obbiettivo di questo paragrafo non poteva essere quello di "inventare" $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ a partire da \mathbb{R} . Ma solo quello di scoprire dentro \mathbb{R} dei sottoinsiemi che possono essere agevolmente identificati con gli insiemi numerici $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ che ho assunti come noti, anche per quanto riguarda le loro proprietà. Ad

⁶ O che avrebbero dovuto essere tali. In effetti il § 1 aveva anche la funzione, per così dire, di rinfrescare la memoria.

esempio, quando faccio riferimento ai numeri 3 e 5 come elementi di \mathbb{R} , voglio che sia vero che $3 < 5$, esattamente come succedeva per i corrispondenti numeri naturali. Oppure che 17 sia ancora un numero primo, cioè che non si riescano a trovare dentro il sottoinsieme di \mathbb{R} che rappresenta i naturali due numeri (a parte 1 e 17) il cui prodotto mi dia 17. Etc, etc.

Quanto ho appena finito di dire dovrebbe rendere possibile la soluzione del quesito posto dall'osservazione 5.3. Infatti, un conto è usare il ben noto numero naturale 2 come indice per una variabile, un conto è invece fare riferimento al numero 2 all'interno dei numeri reali!

7. Alcune domande di fondo sull'approccio assiomatico seguito

Abbiamo detto che per noi \mathbb{R} è un insieme su cui sono definite due operazioni e una relazione, soddisfacenti gli assiomi indicati nella tabella.

Ma c'è un tale insieme? E' unico?

La risposta ad entrambe le domande è affermativa. Per la prima, la proveremo nel paragrafo 9 costruendo esplicitamente un modello di \mathbb{R} , a partire dai numeri razionali. Vi sono molte strade percorribili, tutte un po' complicate: quella che ho scelto ha un vantaggio: permette di giungere facilmente ad ottenere le proprietà chiave, cioè la completezza di \mathbb{R} .

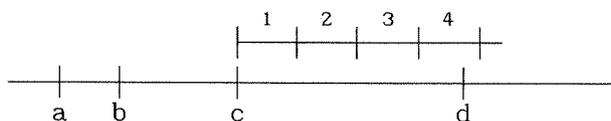
Per quanto riguarda l'unicità di \mathbb{R} , la situazione è la seguente: si può dimostrare che, presi due insiemi \mathbb{R}_1 ed \mathbb{R}_2 , su entrambi i quali sono definite due operazioni e una relazione soddisfacenti gli assiomi 1÷16, essi risultano essere tra loro isomorfi. Isomorfi significa che c'è una corrispondenza biunivoca $\phi: \mathbb{R}_1 \longrightarrow \mathbb{R}_2$ t.c. per ogni $x, y \in \mathbb{R}_1$ si ha: $\phi(x+y) = \phi(x) + \phi(y)$, $\phi(x \cdot y) = \phi(x) \cdot \phi(y)$ e $(x \leq y \text{ se e solo se } \phi(x) \leq \phi(y))$. Il fatto che siano isomorfi significa che sono a tutti gli effetti indistinguibili tra loro, nella misura in cui sono coinvolti gli assiomi e le loro conseguenze. Concludo dicendo che non proverò l'esistenza di questo isomorfismo, anche se non è particolarmente difficile ottenerlo.

Osservazione 1 Ho molto insistito sul fatto che i razionali godono di tutte le proprietà dei numeri reali, eccettuata la completezza. Forse c'è chi si è convinto che se io sostituissi \mathbb{Q} ad \mathbb{R} negli assiomi 1÷15 della tabella, io otterrei gli assiomi per i numeri razionali. Così non è. Ovviamente i razionali godono delle proprietà espresse dagli assiomi 1÷15 (con \mathbb{Q} al posto di \mathbb{R}), come abbiamo già visto nel primo paragrafo; tuttavia questi assiomi non bastano a caratterizzare \mathbb{Q} (a meno di isomorfismi), contrariamente a quel che abbiamo appena visto succedere per \mathbb{R} con gli assiomi 1÷16. Esistono delle caratterizzazioni assiomatiche di \mathbb{Q} , ma non è mia intenzione presentarle. □

8. La proprietà archimede

Questo paragrafo presenta una conseguenza degli assiomi di \mathbb{R} particolarmente utile per i limiti. Si tratta della proprietà archimede. Questa proprietà dovrebbe essere stata già vista nelle scuole secondarie, nel contesto della geometria euclidea (più precisamente, la geometria della retta euclidea, nel capitolo relativo alla misurazione dei segmenti). L'idea geometrica è che se abbiamo due segmenti, pur di "riportare" un numero sufficientemente grande di volte il più piccolo dei due si otterrà alla fine un segmento più grande dell'altro.

In figura:



Tale idea può essere sfruttata anche nel contesto dei numeri reali. Siano dati due numeri reali x ed y . Supponiamo che sia $0 < x < y$: è allora possibile trovare un multiplo di x più grande di y ? Ricordiamo che multiplo vuol dire $n \cdot x$, con $n \in \mathbb{N}$ (cioè $n = \underbrace{1+1+\dots+1}_{n \text{ volte}}$, ricordando come abbiamo ricostruito

i naturali all'interno dei numeri reali). La risposta a questa domanda è sì. Ma come facciamo a dimostrarlo? Un po' inaspettatamente, forse, abbiamo bisogno dell'assioma di completezza. Consideriamo $A = \{ n \cdot x : n \in \mathbb{N} \}$. Se per assurdo non riusciamo a trovare n t.c. $n \cdot x > y$, allora ciò vuol dire che y è un maggiorante per A : vale a dire che A è superiormente limitato. Sia dunque $\lambda = \sup A$ (A è non vuoto): è $n \cdot x \leq \lambda$ per ogni $n \in \mathbb{N}$. Ma allora si ha anche $(k+1) \cdot x \leq \lambda$ per ogni $k \in \mathbb{N}$. Cioè $k \cdot x \leq \lambda - x$ per ogni $k \in \mathbb{N}$. Cioè $\lambda - x$ è anch'esso un maggiorante per A . Assurdo, perché λ è $\sup A$ e quindi dovrebbe essere il minimo dei maggioranti.

Una importante conseguenza della proprietà di Archimede è che in \mathbb{R} non ci sono elementi "infinitesimi". Dire che un elemento è un infinitesimo vorrebbe dire che tale elemento è "radicalmente" più piccolo degli altri elementi "ordinari". Ciò si potrebbe esprimere dicendo che, se x è un elemento infinitesimo ed y è un elemento ordinario, x è più piccolo di ogni sottomultiplo di y . In formule, che $x < y/n$ per ogni $n \in \mathbb{N}$. E' evidente che questa affermazione è in contrasto proprio con la proprietà di Archimede. Si noti che l'approccio di Leibniz all'analisi era invece proprio basato sull'uso di elementi infinitesimi.

Un'altra conseguenza della proprietà di Archimede è la "densità" dei

razionali nei numeri reali. Cioè il fatto che tra due numeri reali qualsiasi, purché distinti, c'è sempre un numero razionale. Per questa proprietà rinvio direttamente al C-S.

9. Un modello per \mathbb{R}

Un modello del sistema dei numeri reali è ottenibile a partire da \mathbb{Q} in vari modi: sezioni, classi contigue, allineamenti decimali illimitati, classi di equivalenza di successioni di Cauchy. Noi utilizzeremo le "semirette razionali senza massimo proprie". Grazie a loro costruiremo un insieme \mathcal{R} , sul quale definiremo due operazioni \oplus e \odot , nonché una relazione d'ordine \otimes in modo che siano soddisfatti tutti gli assiomi di \mathbb{R} .

Definizione 1 $\rho \subseteq \mathbb{Q}$ si dice semiretta razionale senza massimo propria se:

- a) per ogni $q \in \rho$, e per ogni $r \in \mathbb{Q}$, (se $r \leq q$, allora $r \in \rho$)
 ["semiretta razionale"]
 b) per ogni $q \in \rho$, esiste $r \in \rho$ t.c. $q < r$ ["senza massimo"]
 c) $\rho \neq \emptyset$, $\rho \neq \mathbb{Q}$ ["propria"] \square

Indicheremo con $\bar{\mathcal{R}}$ l'insieme delle semirette razionali senza massimo, mentre l'insieme che più ci interessa, cioè \mathcal{R} , sarà l'insieme delle semirette razionali senza massimo proprie: ovverossia, $\mathcal{R} = \bar{\mathcal{R}} \setminus \{\emptyset, \mathbb{Q}\}$.

Si vede facilmente che una semiretta razionale senza massimo propria è ad esempio $\rho = \{ r \in \mathbb{Q} : r < 3 \}$; più in generale, dato $\bar{r} \in \mathbb{Q}$, $\{ r \in \mathbb{Q} : r < \bar{r} \}$ è un elemento di \mathcal{R} . Mi sembra abbastanza evidente che c'è una corrispondenza biunivoca tra questo genere di semirette e i numeri razionali. Se, per così dire, tutto finisse lì, non avremmo ottenuto niente di nuovo. La cosa invece interessante è che ci sono altri tipi di semirette razionali. Cioè, data $\rho \in \mathcal{R}$, non è detto che $\exists \bar{r} \in \mathbb{Q}$ t.c. $\rho = \{ r \in \mathbb{Q} : r < \bar{r} \}$. Ragionando in termini non formali, abbiamo per esempio la semiretta costituita da tutti i razionali "minori di $\sqrt{2}$ ": ovverossia $\{ x \in \mathbb{Q} : x^2 < 2 \} \cup \{ x \in \mathbb{Q} : x < 0 \}$. Osservi il lettore che l'insieme appena descritto è effettivamente una semiretta razionale senza massimo propria. Ebbene, questa semiretta non individua alcun numero razionale \bar{r} ! Cioè, non esiste alcun numero razionale \bar{r} t.c. essa si possa scrivere come $\{ r \in \mathbb{Q} : r < \bar{r} \}$. Volendo immaginare quale mai sia il numero che essa individua, dobbiamo "uscire" da \mathbb{Q} , in quanto pesantissimi indizi portano ad indicare come sospettato numero uno $\sqrt{2}$. E così è, infatti.

Il trucco è tutto qui. L'insieme \mathcal{R} delle semirette razionali senza massimo proprie contiene anche altre semirette oltre a quelle del tipo $\{ r \in \mathbb{Q} :$

$r < \bar{r}$ } ! Quindi l'insieme \mathcal{R} in un certo senso contiene \mathbb{Q} , ma anche dell'altro: possiamo quindi sperare che esso ci dia quella particolarissima "estensione" di \mathbb{Q} che sono i numeri reali (anticipo che sarà proprio così, anche se questa certezza l'avremo solo alla fine del paragrafo).

Naturalmente tutti avranno ben chiaro che per poter parlare di "numeri reali" non basta solo avere un insieme, ma dobbiamo anche indicare quali sono le operazioni e la relazione d'ordine. Dobbiamo pertanto affrontare un nuovo problema: come definire la "somma" di due elementi di \mathcal{R} , cioè: come si fa a sommare due semirette razionali senza massimo proprie?

Se cerchiamo di farci venire delle buone idee a partire dalle semirette "speciali", cioè quelle del tipo $\{ r \in \mathbb{Q} : r < \bar{r} \}$, rischiamo di farci portare fuori strada perché sono, per così dire, troppo semplici. In effetti, uno potrebbe pensare, date $\rho = \{ r \in \mathbb{Q} : r < \bar{r} \}$ e $\sigma = \{ r \in \mathbb{Q} : r < \bar{s} \}$, di definire $\rho + \sigma = \{ r \in \mathbb{Q} : r < \bar{r} + \bar{s} \}$. Senza dubbio è un'idea sensata, ma sfortunatamente si applica solo a queste semirette "speciali". Se prendo una generica $\rho \in \mathcal{R}$, non ho nessuna garanzia che ci sia $\bar{r} \in \mathbb{Q}$ t.c. $\rho = \{ r \in \mathbb{Q} : r < \bar{r} \}$: ma allora come facciamo?

L'idea è abbastanza semplice: possiamo sfruttare abilmente le stesse semirette "speciali" per farcela venire in mente o, quantomeno, per capirla. Date $\rho = \{ r \in \mathbb{Q} : r < \bar{r} \}$ e $\sigma = \{ r \in \mathbb{Q} : r < \bar{s} \}$, si vede facilmente che $\{ r \in \mathbb{Q} : r < \bar{r} + \bar{s} \}$ è anche uguale a

$$(\star) \quad \{ r \in \mathbb{Q} : \text{esiste } p \in \rho \text{ ed esiste } q \in \sigma \text{ t.c. } r = p + q \}$$

a parole, è la semiretta ottenuta nel modo seguente: prendiamo un elemento di ρ ed un elemento di σ e li sommiamo, e facciamo così per ogni possibile scelta di un elemento in ρ e di un elemento in σ .

Ebbene, l'espressione (\star) non solo ci ridà $\rho + \sigma$, ma ha anche un pregio fondamentale: non fa intervenire affatto \bar{r} ed \bar{s} ! Quindi possiamo usare la stessa idea per definire in generale la somma di due semirette, anche se non sono di quelle "speciali". Ed è proprio quello che faremo.

Passando alla definizione formale, definiremo la somma $\oplus : \mathcal{R} \times \mathcal{R} \longrightarrow \mathcal{R}$ nel modo seguente: date $\rho, \sigma \in \mathcal{R}$, poniamo:

$$\rho \oplus \sigma = \{ r \in \mathbb{Q} : \text{esiste } p \in \rho \text{ ed esiste } q \in \sigma \text{ t.c. } r = p + q \}$$

E' lasciato al lettore vedere che effettivamente \oplus è definita su tutto $\mathcal{R} \times \mathcal{R}$

ed è a valori in \mathcal{R} (quindi, in particolare, che $\rho \otimes \sigma$ è un elemento di \mathcal{R}). Ugualmente il lettore proverà che \otimes soddisfa gli assiomi 1+4: si noti che " 0 " = $\{ q \in \mathbb{Q} : q < 0 \}$ è l'elemento neutro di \mathcal{R} rispetto a \otimes .

La definizione di $\circ : \mathcal{R} \times \mathcal{R} \longrightarrow \mathcal{R}$ è un po' più elaborata. Premettiamo che, dato A sottoinsieme di \mathbb{Q} che soddisfa b), indicheremo con τ_A il seguente insieme: $\{ q \in \mathbb{Q} : \text{esiste } r \in A \text{ t.c. } q \leq r \}$. Intuitivamente, τ_A è la semi-retta "generata" da A . Ovviamente è lasciato al lettore provare che $\tau_A \in \bar{\mathcal{R}}$. Indichiamo con $\mathcal{R}_\otimes = \{ \rho \in \mathcal{R} : \text{esiste } q > 0 \text{ t.c. } q \in \rho \}$. Data $\rho \in \mathcal{R}_\otimes$, indichiamo con $\rho^+ = \{ q \in \rho : q > 0 \}$: si noti che ρ^+ soddisfa la condizione b). Dati $\rho, \sigma \in \mathcal{R}_\otimes$, definiamo $\rho \circ \sigma = \tau_G$, dove $G = \{ r \in \mathbb{Q} : \text{esiste } p \in \rho^+ \text{ ed esiste } q \in \sigma^+ \text{ t.c. } r = p \cdot q \}$. Si noti che G soddisfa b) e quindi $\tau_G \in \bar{\mathcal{R}}$. Inoltre, poiché $G \neq \emptyset$ ed esiste $s \in \mathbb{Q}$ t.c. $s \notin G$, possiamo affermare che $\tau_G \in \mathcal{R}$.

Per definire \circ su tutto $\mathcal{R} \times \mathcal{R}$, ci si ispira alle "regole dei segni". Intanto, occorre provare che, data $\rho \in \mathcal{R}$, si possono verificare solo tre possibilità mutuamente escludentesi: $\rho \in \mathcal{R}_\otimes$, $\ominus \rho \in \mathcal{R}_\otimes$ (con $\ominus \rho$ indichiamo l'opposto di ρ rispetto a \otimes), $\rho = "0" = \{ q \in \mathbb{Q} : q < 0 \}$. Provato ciò, la definizione di \circ si può dare caso per caso. Ad esempio, se $\rho \in \mathcal{R}_\otimes$ e σ è t.c. $\ominus \sigma \in \mathcal{R}_\otimes$, definiamo $\rho \circ \sigma = \ominus(\rho \circ (\ominus \sigma))$. E così via... E' lasciata al lettore molto paziente la verifica che \otimes e \circ soddisfano tutti gli assiomi 1+10.

Resta la relazione d'ordine su \mathcal{R} .

Date $\rho, \sigma \in \mathcal{R}$, diremo che $\rho \leq \sigma$ se $\rho \circ \sigma$. E' immediato verificare che \leq gode delle proprietà 11, 12 e 13 (è conseguenza del fatto che \leq è una relazione d'ordine). Meno scontato è che \leq sia totale, cosa che ora proveremo. Siano dunque $\rho, \sigma \in \mathcal{R}$ t.c. $\rho \circ \sigma$ e $\sigma \circ \rho$ siano entrambe false. Ciò significa che $\rho \leq \sigma$ e $\sigma \leq \rho$ sono entrambe false: pertanto esiste $x \in \rho$ t.c. $x \notin \sigma$ ed esiste $y \in \sigma$ t.c. $y \notin \rho$. Ovviamente $x \neq y$. Supponiamo che sia $x < y$ (altrimenti basterà scambiare i ruoli di x e y): si ha subito una contraddizione, perché F è una semi-retta ed essendo $x < y$ deve essere $x \in \sigma$.

Verifichiamo infine che \leq soddisfa la proprietà 17, ovvero "l'esistenza del sup". Sia dunque $A \subseteq \mathcal{R}$, A non vuoto e superiormente limitato. Sia $\zeta = \bigcup_{\alpha \in A} \alpha$: proveremo che $\zeta \in \mathcal{R}$ e che $\zeta = \sup A$.

Poiché gli elementi di \mathcal{A} sono semirette razionali, è $\alpha \leq \mathbb{Q}$ per ogni $\alpha \in \mathcal{A}$ e quindi $\zeta = \bigcup_{\alpha \in \mathcal{A}} \alpha \subseteq \mathbb{Q}$. Verifichiamo che ζ soddisfa a), b) e c).

Cominciamo con c). Poiché \mathcal{A} è superiormente limitato, esiste $\sigma \in \mathcal{R}$ t.c. $\alpha \leq \sigma \forall \alpha \in \mathcal{A}$. Vale a dire, esiste $\sigma \in \mathcal{R}$ t.c. $\alpha \leq \sigma$ per ogni $\alpha \in \mathcal{A}$. E quindi $\zeta = \bigcup_{\alpha \in \mathcal{A}} \alpha \subseteq \sigma$: poiché $\sigma \in \mathcal{R}$, $\sigma \neq \mathbb{Q}$ e quindi $\zeta \neq \mathbb{Q}$. Poiché $\alpha \neq \emptyset$ per ogni $\alpha \in \mathcal{A}$, e poiché $\mathcal{A} \neq \emptyset$, segue che $\bigcup_{\alpha \in \mathcal{A}} \alpha \neq \emptyset$ [la unione di una famiglia non vuota di insiemi non vuoti è non vuota. Si noti che se fosse $\mathcal{A} = \emptyset$ avremmo $\bigcup_{\alpha \in \mathcal{A}} \alpha = \emptyset$, che è una semiretta razionale senza massimo ma non è propria, quindi sta in $\bar{\mathcal{R}}$ ma non sta in \mathcal{R}].

Per quanto riguarda a), sia $q \in \zeta$ e $r \leq q$. Poiché $q \in \zeta$, esiste $\bar{\alpha} \in \mathcal{A}$ t.c. $q \in \bar{\alpha}$. Dato che $\bar{\alpha}$ soddisfa a), $r \in \bar{\alpha}$. E allora si ha $r \in \bigcup_{\alpha \in \mathcal{A}} \alpha$.

Per il punto b), infine, supponiamo per assurdo che esista $\bar{r} \in \zeta$ t.c. $q \leq \bar{r}$ per ogni $q \in \zeta$. Se $\bar{r} \in \zeta$, esiste $\bar{\alpha} \in \mathcal{A}$ t.c. $\bar{r} \in \bar{\alpha}$. Non può esistere $q \in \bar{\alpha}$ t.c. $q > \bar{r}$ poiché se così fosse avremmo che $q \in \zeta$ e $q > \bar{r}$, quindi \bar{r} non sarebbe massimo per ζ . Ma allora \bar{r} è massimo per $\bar{\alpha}$, il che è assurdo perché $\bar{\alpha}$ soddisfa b). Abbiamo quindi provato che $\zeta \in \mathcal{R}$.

Resta da provare che $\zeta = \sup \mathcal{A}$. Che ζ sia maggiorante per \mathcal{A} , cioè che $\beta \leq \zeta$ per ogni $\beta \in \mathcal{A}$, è conseguenza del fatto che se $\beta \in \mathcal{A}$, allora $\beta \subseteq \bigcup_{\alpha \in \mathcal{A}} \alpha = \zeta$. Infine, ζ è il minimo dei maggioranti. Infatti, sia $\mu \in \mathcal{R}$ un maggiorante per \mathcal{A} . Ciò vuol dire che $\alpha \leq \mu$ per ogni $\alpha \in \mathcal{A}$. Cioè $\alpha \subseteq \mu$ per ogni $\alpha \in \mathcal{A}$. E quindi $\zeta = \bigcup_{\alpha \in \mathcal{A}} \alpha \subseteq \mu$. Cioè, $\zeta \leq \mu$.

Due considerazioni finali

*) si invita a riflettere sul problema 1 a pag 37 del C-S

**) Si può provare che $\bar{\mathcal{R}}$, cioè l'insieme delle semirette razionali senza massimo, fornisce un modello per $\bar{\mathcal{R}}$ [vedasi C-S a pag 36]. In particolare, $+\infty$ sarà rappresentato dalla semiretta \mathbb{Q} , mentre $-\infty$ sarà rappresentato da \emptyset .

10. Confronto con gli assiomi di C-S

In questo breve paragrafo proverò a grandi linee l'equivalenza tra gli assiomi per \mathbb{R} che ho utilizzato e quelli usati nel C-S. Tenuto conto dell'approccio seguito da questo testo che privilegia l'assioma del \sup , proverò l'equivalenza tra gli assiomi del C-S e gli assiomi 1+15 e 17 della tabella. Abbiamo già dimostrato l'equivalenza tra gli assiomi 1+16 e il gruppo di assiomi 1+15 e 17, pertanto per questa via otteniamo il risultato desiderato.

Abbiamo quindi da dimostrare che gli assiomi 1+15 e 17 congiuntamente sono equivalenti al complesso delle proprietà 8.1+8.7 del C-S.

Cominciamo a dimostrare che gli assiomi 1+15 e 17 implicano gli assiomi del C-S.

Che 1+10 implichino le proprietà 8.1+8.5 è ovvio (si vede anzi immediatamente che si tratta di due gruppi di assiomi tra di loro equivalenti). Poiché 17 non è altro che la proprietà 8.7, resta solo da ottenere la proprietà 8.6. La prima cosa da fare è individuare \mathbb{R}_+ . Definiamo $\mathbb{R}_+ = \{ x \in \mathbb{R} : 0 \leq x \text{ e } 0 \neq x \}$.

Proviamo che vale 8.6, parte i), per quanto riguarda \cdot . Se $a, b \in \mathbb{R}_+$, grazie a 15 ottengo che $0 \leq a \cdot b$ (si scelga $x=0$, $y=a$, $z=b$ e si tenga conto del fatto che $0 \cdot b = 0$, come a suo tempo dimostrato). Per poter affermare che $a \cdot b \in \mathbb{R}_+$, devo ancora dimostrare che $a \cdot b \neq 0$. Ma non può essere $a \cdot b = 0$, essendo $a, b \neq 0$ (che $a, b \neq 0$ è dovuto al fatto che essi appartengono ad \mathbb{R}_+ ; che da ciò si possa dedurre $a \cdot b \neq 0$ è dovuto alla R.9 a pag. 29 del C-S).

Con considerazioni similari si prova anche 8.6, parte i), per il $+$.

Per provare 8.6, parte ii), sia $a \neq 0$. Dobbiamo provare che $a \in \mathbb{R}_+$ oppure (or esclusivo, overossia aut) $-a \in \mathbb{R}_+$. Intanto proviamo che almeno una delle due circostanze si verifica. Sia dunque $a \in \mathbb{R}$, $a \neq 0$. Se $a \notin \mathbb{R}_+$, vuol dire che $a \leq 0$ (grazie all'assioma 14). Usando 15, otteniamo che $a + (-a) \leq -a$. Cioè $0 \leq -a$. E quindi che $-a \in \mathbb{R}_+$ (essendo $a \neq 0$, e pertanto anche $-a \neq 0$, come si prova facilmente: vedi R.2 e R.5). Resta ormai solo da ottenere che non si verificano contemporaneamente $a \in \mathbb{R}_+$ e $-a \in \mathbb{R}_+$. Ma se così fosse, si avrebbe $0 \leq a$ e $0 \leq -a$. Da cui, per l'assioma 15, $a = 0 + a \leq (-a) + a = 0$. Abbiamo quindi $0 \leq a$ e $a \leq 0$: grazie all'assioma 12 si ottiene che $a = 0$. Impossibile perché avevamo supposto $a \neq 0$.

Proviamo che 8.1+8.7 implicano 1+15 e 17. Per quanto già osservato, il problema è riuscire a provare che valgono 11+15. A tale scopo, occorre per prima

cosa definire \leq . Lo faremo, ovviamente, usando il sottoinsieme di \mathbb{R} fornito dalla 8.6, cioè \mathbb{R}_+ . Diremo, dati $x, y \in \mathbb{R}$, che $x \leq y$ se $y - x \in \mathbb{R}_+$ oppure (ve) $y - x = 0$. Che valgano le proprietà 11÷14, lo si ottiene con ragionamenti simili a quelli esposti nel C-S a pag. 29. Per la proprietà 15, si rinvia alla dimostrazione delle R.11 e R.12, sempre su C-S a pag. 30.

11. I numeri complessi

Come si può desumere dai termini usati in questo contesto (unità immaginaria, numeri complessi), c'è una certa aura di mistero attorno a tali numeri. In realtà, non presentano nulla di misterioso, certo non più dei numeri "reali": se uno non ha obiezioni da fare a proposito dei numeri reali, è alquanto strano che abbia da ridire sulla presentazione usuale, "moderna" dei numeri complessi.

I numeri complessi non sono altro che $\mathbb{R} \times \mathbb{R}$, cioè \mathbb{R}^2 , sul quale sono definite in modo opportuno due operazioni.

Definizione 1

$$(a,b) + (c,d) = (a+c, b+d)$$

$$(a,b) \cdot (c,d) = (a \cdot c - b \cdot d, a \cdot d + b \cdot c) \quad \square$$

Per noi i numeri complessi non saranno altro che \mathbb{R}^2 con queste due operazioni⁷, ed useremo il simbolo \mathbb{C} per riferirci appunto a questa struttura.

E' più noioso che difficile verificare che queste operazioni godono delle proprietà 1÷10, cioè che \mathbb{R}^2 con queste proprietà risulta essere un "corpo commutativo" o "campo" (si noti che non solo i numeri reali, ma anche i razionali sono un corpo commutativo).

E' per esempio immediato vedere che $(0,0)$ è l'elemento neutro per l'operazione $+$ e che $(1,0)$ è l'elemento neutro per l'operazione \cdot .

La validità delle proprietà associativa e commutativa è una conseguenza immediata del fatto che tali proprietà valgono per le operazioni in \mathbb{R} . Ad esempio:

$$\begin{aligned} (a,b) \cdot (c,d) &= && \text{per definizione di } \cdot \text{ in } \mathbb{C} \\ = (a \cdot c - b \cdot d, a \cdot d + b \cdot c) &= && \text{per la proprietà commutativa in } \mathbb{R} \\ = (c \cdot a - d \cdot b, d \cdot a + c \cdot b) &= && \text{per definizione di } \cdot \text{ in } \mathbb{C} \\ = (c,d) \cdot (a,b) \end{aligned}$$

L'opposto di (a,b) è ovviamente $(-a,-b)$. Un po' meno ovvio è chi sia il reciproco di (a,b) , con $(a,b) \neq (0,0)$: si tratta di trovare (x,y) t.c.

$$(a,b) \cdot (x,y) = (1,0) \quad \text{. Ovverossia} \quad \begin{cases} a \cdot x - b \cdot y = 1 \\ a \cdot y + b \cdot x = 0 \end{cases} \quad \text{. Da cui} \quad (x,y) = \left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2} \right) \quad \text{.}$$

⁷ Sarebbe forse stato più opportuno usare due simboli diversi dai soliti $+$ e \cdot , per non fare confusione con le operazioni di \mathbb{R} , ma spero che questa annotazione sia sufficiente.

Lascio al lettore il compito di provare che vale la proprietà distributiva.

Questi sono i numeri complessi.

Invito però a questo punto lo studente troppo disposto ad accettare qualunque cosa gli venga propinata, a chiedersi: ma che fine ha fatto i ? In fondo, i numeri complessi vengono fuori proprio dal problema di risolvere equazioni come $x^2 = -1$ che non hanno soluzioni reali (anche se questa non è l'origine storica dei numeri complessi). E poi si è abituati a vedere un numero complesso come un qualcosa del tipo $a+ib$, non come un elemento di \mathbb{R}^2 (anche se di solito si sa che i numeri complessi si possono rappresentare sul cosiddetto piano di Argand-Gauss⁸).

Proveremo pertanto che esiste in \mathbb{C} un numero il cui quadrato fa -1 , e pertanto lo possiamo chiamare i secondo la consuetudine. Non solo, faremo anche vedere come (a,b) non sia altro che il solito $a+ib$.

Proposizione 1 $(0,1) \cdot (0,1) = (-1,0)$. \square

Dimostrazione Basta usare la definizione di \cdot in \mathbb{C} . \square

Abbiamo quindi che $(0,1)^2 = (-1,0)$. Il mio problema è di far capire come mai $(-1,0)$ sia sostanzialmente lo stesso che $-1 \in \mathbb{R}$.

L'idea chiave è che i numeri complessi del tipo $(a,0)$ si comportano in tutto e per tutto come i numeri reali. In effetti, $(x,0)+(y,0) = (x+y,0)$ e $(x,0) \cdot (y,0) = (x \cdot y,0)$: come si vede, le operazioni su \mathbb{C} , ristrette al sottoinsieme $E = \{ (x,y) \in \mathbb{C} : y=0 \}$, si comportano esattamente come le operazioni in \mathbb{R} . Possiamo anche vedere le cose nel modo seguente: è come se usassimo delle notazioni diverse per indicare i numeri reali, scrivendo $(x,0)$ anziché x (come se scrivessimo α o \tilde{x} , anziché x). Da un altro punto di vista ancora, possiamo stabilire una corrispondenza biunivoca $\psi: \mathbb{R} \longrightarrow E$, definendo $\psi(x) = (x,0)$. Tale ψ non solo è una corrispondenza biunivoca, ma rispetta anche le operazioni: è cioè un isomorfismo tra \mathbb{R} ed E .

Pertanto, possiamo identificare E con \mathbb{R} : quindi è tollerabile scrivere a anziché $(a,0)$ per brevità. La presenza dell'isomorfismo sopra individuato permette di fare i calcoli indifferentemente usando \mathbb{R} oppure E .

Proposizione 2 Per ogni numero complesso (a,b) si ha:

⁸ A dire la verità, spesso non è molto chiaro cosa ci sia di diverso tra il piano di Argand-Gauss e il solito "piano cartesiano": semplicemente ci si disinteressa del problema (ci si accontenta di dire che l'asse delle y è l'asse immaginario e tutto finisce lì), alimentando l'aura di mistero attorno ai numeri complessi.

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

Dimostrazione Basta usare la definizione delle operazioni in \mathbb{C} . \square

Siamo a posto. Se decidiamo di usare il simbolo i per indicare il numero complesso $(0,1)$, possiamo ottenere la cosiddetta forma algebrica dei numeri complessi. Cioè, dato (a,b) , abbiamo che $(a,b) = (a,0) + (0,1) \cdot (b,0)$ e quindi pos-

$$\begin{array}{ccc} \downarrow & \downarrow & \downarrow \\ a & i & b \end{array}$$

siamo scrivere (a,b) come $a+ib$.

Non solo: $(0,1)^2 = (-1,0)$, cioè possiamo scrivere $(0,1)^2 = -1$, ovvero sia $i^2 = -1$.

Abbiamo quindi ritrovato la cosiddetta "forma algebrica" dei numeri complessi. Abbiamo cioè che (a,b) può essere scritto come $a+ib$. Non solo, ma se con questa scrittura usiamo le solite regole dell'algebra applicate ai polinomi (cioè trattiamo i come una incognita, con la curiosa proprietà che $i^2 = -1$), otteniamo esattamente le operazioni come le avevamo definite in \mathbb{C} . Infatti:

$$(a+ib) + (c+id) = (a+c) + i(b+d) \quad \text{e}$$

$$(a+ib) \cdot (c+id) = ac + i(ad+bc) + i^2bd = (ac-bd) + i(ad+bc) .$$

Questo mostra che i numeri complessi, come li abbiamo visti in questo paragrafo, sono esattamente la stessa cosa che uno eventualmente già conosceva. Ho scelto quel tipo di approccio ai complessi perché ha il pregio di essere il meno "misterioso": avrei anche potuto usare direttamente la simbologia $a+ib$, intendendola come scrittura formale, ma seguire questa strada richiede una maggiore astrazione.

Concludo accennando ad un aspetto interessante di \mathbb{C} .

Probabilmente a qualcuno sarà capitato di osservare che i numeri complessi non si confrontano tra loro (nelle scuole precedenti non avete mai imparato a risolvere disequazioni in \mathbb{C} , né io ho introdotto una relazione d'ordine su \mathbb{C}). Come mai? La risposta è molto drastica: non è possibile definire su \mathbb{C} una relazione d'ordine totale che "rispetti le operazioni". Cioè non si può definire una relazione \preceq che sia una relazione d'ordine totale e per cui valgano le seguenti proprietà, (analoghe alla proprietà 15 dei numeri reali, vedi tabella):

$$\left\{ \begin{array}{l} \text{per ogni terna } x,y,z \text{ di numeri complessi, si ha:} \\ x \preceq y \text{ implica } x+z \preceq y+z \quad \text{e} \\ (x \preceq y \text{ e } 0 \preceq z) \text{ implica } x \cdot z \preceq y \cdot z \end{array} \right.$$

Potremmo anche provare questo fatto (non è neanche troppo difficile) ma a me

interessava solo far notare questo aspetto curioso dei numeri complessi.

12. Sommario e conclusioni

In questo primo capitolo ci siamo solo occupati di numeri. Sono partito col richiamare succintamente le proprietà delle operazioni e della relazione d'ordine in \mathbb{Q} , principalmente a scopo introduttivo per gli assiomi di \mathbb{R} . Ripeto che a partire da quegli assiomi costruiremo tutto l'edificio dell'analisi, come se fossero dei mattoni.

A partire dagli assiomi ho prima di tutto dimostrato alcune delle proprietà dei numeri reali, più che altro a scopo esemplificativo.

Ho poi affrontato le conseguenze più rilevanti, cioè quelle che mettono in gioco l'assioma di completezza, ottenendo così la proprietà archimedeica e la proprietà del \sup . Il ruolo di queste due proprietà sarà evidente sin dal prossimo capitolo, dedicato alle successioni e ai limiti di successioni.

Ho provato l'equivalenza tra il sistema assiomatico per \mathbb{R} presentato dal C-S e quello adottato in queste dispense. Uno potrebbe ragionevolmente chiedersi perché, visto che sono equivalenti, non ho adottato direttamente la lista degli assiomi usata sul C-S. La ragione è la seguente: ho preferito, a scapito della rapidità, presentare un sistema di assiomi che esaltasse le connessioni tra le diverse strutture presenti su \mathbb{R} . Si tratta degli assiomi 9 e 15: il primo traccia un legame evidente tra le due operazioni, il secondo connette le operazioni alla struttura d'ordine presente su \mathbb{R} . Questa è la ragione fondamentale per cui ho adottato gli assiomi della tabella anziché direttamente quelli di C-S.

L'approccio assiomatico che ho adottato pone due ordini di problemi.

Uno è di carattere didattico: è un approccio che può risultare eccessivamente astratto, difficile da capire veramente. Questo può essere lo scotto da pagare per la chiarezza, rapidità ed efficienza che invece offre. Non va comunque dimenticato, d'altronde, che i numeri reali sono "intrinsecamente difficili": non c'è nessuna strada facile per arrivarci.

L'altro ordine di problemi è di carattere sostanziale: descrivendo i numeri reali solo mediante una lista di assiomi, non si ha la garanzia che un tale ente ci sia. Ecco la ragione per la quale mi sono sentito in dovere di offrire un modello di \mathbb{R} , elaborato a partire dai numeri razionali. Ma ci sono anche altri problemi, come quello della sostanziale unicità o meno dell'ente individuato dagli assiomi dati, problema al quale ho poco più che accennato. Altre problematiche, generalmente caratteristiche dell'approccio assiomatico, non le ho neppure sfio-

rate.

Il capitolo poi si chiude con una rapida descrizione dei numeri complessi, introdotti come coppie ordinate di numeri reali, ma per i quali ho anche indicato la connessione con l'usuale scrittura algebrica.

Appendice. Il principio di induzione

Vi sono varie formulazioni.

Ne richiamo due: una "logica" e una "insiemistica".

Ma prima una precisazione sulle notazioni: ricordo che per me $\mathbb{N} = \{1, 2, 3, \dots\}$. C'è chi mette anche lo zero in \mathbb{N} (tipicamente gli algebristi, ma non solo): io ho invece l'abitudine di non mettercelo. Per quello che vedremo non c'è alcuna differenza significativa.

Sia P una proprietà che dipende da $n \in \mathbb{N}$. Essa può dipendere da altre variabili, ma questo non ci interessa. Per sottolinearne la dipendenza da n indicheremo quindi con $P(n)$ tale proprietà.

1° FORMULAZIONE:

$$[P(1) \text{ e } (\forall m \in \mathbb{N} (P(m) \Rightarrow P(m+1)))] \Rightarrow [\forall k \in \mathbb{N} P(k)] . \square$$

Tradotto in parole, vuol dire che se abbiamo una proprietà tale che:

vale per 1

e tale che

ogni qual volta vale per n , allora vale per $n+1$,

allora tale proprietà vale per ogni n .

Un po' di terminologia: la proposizione $P(1)$ si dice "base", la proposizione $(\forall m \in \mathbb{N} (P(m) \Rightarrow P(m+1)))$ si dice "passo induttivo", la proposizione $P(m)$ si dice "ipotesi induttiva".

2° FORMULAZIONE: Sia $E \subseteq \mathbb{N}$.

$$[1 \in E \text{ e } (\forall m \in \mathbb{N} (m \in E \Rightarrow m+1 \in E))] \Rightarrow [E = \mathbb{N}] . \square$$

E' facile provare l'equivalenza di queste due formulazioni: è una sorta di gioco di traduzione da un linguaggio a un altro.

Esercizio 1 Dimostrarlo. Suggerimento (ovvio...): data P , si definisca E come l'insieme dei naturali per cui P è vera. Viceversa... \square

Di solito il principio di induzione viene presentato come un principio valido, senza preoccuparsi di vedere se derivi da altri principi "precedenti". Volendo, lo si può dedurre dagli assiomi della teoria degli insiemi, nel contesto degli insiemi "bene ordinati": chi fosse interessato può consultare per esempio Kelley (John L. Kelley: "General Topology", Van Nostrand, Princeton, 1955) a pagg. 250 e segg. (in particolare il teorema 137 a pag. 272). Il principio di induzione è anche al cuore della definizione assiomatica di \mathbb{N} data da Peano: è il 5° postulato. Ricordo, infine, che il principio di induzione matematica può essere visto come una conseguenza degli assiomi di \mathbb{R} (campo ordinato e completo):

vedasi, per esempio, ancora Kelley oppure Apostol (Tom M. Apostol: "Calculus", vol. 1, Blaisdell, Waltham (MA), 1967; 1° edizione: 1961; c'è traduzione italiana edita da Boringhieri)¹.

E' interessante notare come dal principio di induzione si ottengano le seguenti proprietà:

- principio di "buon ordinamento". Ovverossia, ogni sottoinsieme di \mathbb{N} ha minimo (in realtà questo principio è equivalente al principio di induzione)

- principio di induzione "esteso"

- ogni sottoinsieme finito di \mathbb{N} ha massimo

- possibilità di dare definizioni per ricorrenza (vedi sotto)

Proviamo, tanto per fare un po' di esercizio, che si può ricavare il principio di "buon ordinamento" dal principio di induzione.

Teorema 1 Ogni sottoinsieme di \mathbb{N} ha elemento minimo.

Dimostrazione (per induzione, ovviamente!).

Lo dimostro per induzione sul numero di elementi di questo insieme. Più che altro stiamo attenti a darne una formulazione precisa. Seguiamo l'approccio "logico".

Indichiamo con $P(n)$ la seguente proposizione:

"ogni s.i. di \mathbb{N} che ha n elementi ha minimo"

Ovviamente $P(1)$ è vera.

Sia quindi $k \in \mathbb{N}$ e supponendo che sia vera $P(k)$ dimostriamo che vale $P(k+1)$. Consideriamo allora un s.i. A di \mathbb{N} con $k+1$ elementi: $A = \{n_1, \dots, n_{k+1}\}$. Consideriamo $\{n_1\}$ ed $\{n_2, \dots, n_{k+1}\}$. Poiché quest'ultimo ha k elementi, esso ha minimo (per l'ipotesi induttiva): sia n_j questo elemento minimo.

Consideriamo ora $\{n_1, n_j\}$. Esso ha ovviamente minimo per la ipotesi induttiva e quindi non resta che fare l'ovvia verifica che tale minimo è anche il minimo di A . \square

¹ Brevemente, si fa così. Abbiamo a disposizione \mathbb{R} che è un campo ordinato e completo. Diciamo che un sottoinsieme A è un "insieme induttivo" se:

$$1 \in A \text{ e } \forall x \in \mathbb{R} (x \in A \Rightarrow x+1 \in A);$$

dopo di che si definiscono "naturali" gli elementi di \mathbb{R} che stanno in ogni insieme induttivo. Detto ω l'insieme dei "naturali", è evidente che ω è un insieme induttivo e che ω è contenuto in ogni insieme induttivo. A questo punto il principio di induzione si può esprimere come teorema: ogni s.i. induttivo di ω coincide con ω .

Esercizio 2 La dimostrazione precedente è sbagliata. Dove è l'errore?□

Per quanto concerne il principio di induzione esteso, esso consiste nel sostituire all'ipotesi induttiva $P(m)$ la seguente:

$$\forall j \in \{1, \dots, m\} P(j)$$

ovverossia, meno formalmente, si tratta di supporre la validità della proposizione non solo per m ma anche per tutti i naturali minori o uguali di m . E' evidente che questa versione è più forte del principio di induzione tradizionale, d'altronde è facile dimostrare che le due versioni sono in realtà equivalenti.

Esercizio 3 Provare l'equivalenza di queste due versioni.□

Esercizio 4 Provare che ogni sottoinsieme finito di \mathbb{N} ha massimo.□

Riguardo alle definizioni per ricorrenza, esse godono un po' dello stesso "status" del principio di induzione. Ovvero, tutti sanno che ci sono, però potrebbero avere imbarazzo a dover dire precisamente come diavolo sia fatta una definizione per ricorrenza. Si tratta però di un imbarazzo molto minore se ci si pone il problema di capire come dal principio di induzione si possa introdurre una formulazione sensata di "come è fatta" una definizione per induzione. Un approccio molto semplice lo si trova, per esempio, su C-S.

Teorema 2 (definizione per ricorrenza) Sia dato un insieme A ed una applicazione $f: \mathbb{N} \times A \longrightarrow A$, nonché $a \in A$. Allora esiste una ed una sola applicazione $g: \mathbb{N} \longrightarrow A$ t.c.:

1 $g(1) = a$

2 $\forall n \in \mathbb{N} \quad g(n+1) = f(g(n), n)$.□

Dimostrazione Ovviamente per induzione.

Riporto la dimostrazione di C-S, che usa la "2° formulazione", o "insiemistica".

Dimostriamo dapprima che g è unica.

Siano $g, g': \mathbb{N} \longrightarrow A$ due applicazioni soddisfacenti 1 e 2.

Sia $S = \{ n \in \mathbb{N} : g(n) = g'(n) \}$.

Allora, per 1, $1 \in S$. Se poi $n \in S$, dall'ipotesi induttiva $g(n) = g'(n)$ segue che

$$g(n+1) = f(g(n), n) = f(g'(n), n) = g'(n+1).$$

Vediamo ora che davvero esiste una funzione g come affermato. Sia $E \subseteq \mathbb{N}$ l'insieme dove è definita una funzione g soddisfacente 1 e 2. Proviamo che $E = \mathbb{N}$, ovviamente per induzione e seguendo la strada insiemistica. Basta osservare che, grazie ad 1, $1 \in E$ e che, grazie a 2, se $n \in E$, allora $n+1 \in E$.□

Esercizio 5 Sia $a = 1$ e $f: \mathbb{N} \times \mathbb{R} \longrightarrow \mathbb{R}$ definita come $f(n, x) = x + 2n + 1$. E' vero che $g(n) = n^2$? \square

Chi ama invece le cose complicate, può deliziarsi con la definizione per ricorrenza (e teoremi relativi) che si trova su Kelley.

Esercizio 6 Dimostrare che $\sum_{k=1}^n k = \frac{n(n+1)}{2}$ e che $\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$. \square

Esercizio 7 (da Apostol, che lo accredita a G. Pólya) Trovare dove è l'errore...

"Dimostriamo" per induzione la seguente affermazione:

"sia dato un insieme di n ragazzi biondi: se uno di loro ha gli occhi azzurri, allora hanno tutti gli occhi azzurri".

L'affermazione è ovvia per $n = 1$. Per passare da k a $k+1$, si può procedere nel modo descritto qui appresso per $k = 3$. Supponiamo quindi che l'affermazione sia vera per $k = 3$ e siano B_1, B_2, B_3, B_4 quattro ragazzi biondi di cui almeno uno (assumiamo sia B_1) ha gli occhi azzurri. Prendendo B_1, B_2, B_3 insieme e usando il fatto che l'affermazione è vera per $n = 3$, otteniamo che anche B_2 e B_3 hanno gli occhi azzurri. Ripetiamo la procedura con B_1, B_2, B_4 e troviamo che anche B_4 ha gli occhi azzurri. E' ovvio come queste considerazioni si possano adattare per passare in generale da k a $k+1$. \square

TABELLA DEGLI ASSIOMI DI \mathbb{R}

nome della proprietà	proprietà (descrizione formale)	note
1) COMMUTATIVA (+)	$\forall x, y \in \mathbb{R} \quad x+y=y+x$	} si dimostra che l'elemento neutro per + è unico lo si indica con 0
2) ASSOCIATIVA (+)	$\forall x, y, z \in \mathbb{R} \quad (x+y)+z=x+(y+z)$	
3) ESISTENZA ELEMENTO NEUTRO (+)	$\exists a \in \mathbb{R} \text{ t.c. } \forall x \in \mathbb{R} \quad a+x=x$	
4) ESISTENZA OPPOSTO (+)	$\forall x \in \mathbb{R} \exists y \in \mathbb{R} \text{ t.c. } x+y=0$	
5) COMMUTATIVA (·)	$\forall x, y \in \mathbb{R} \quad x \cdot y = y \cdot x$	} si dimostra che l'elemento neutro per · è unico lo si indica con 1
6) ASSOCIATIVA (·)	$\forall x, y, z \in \mathbb{R} \quad (x \cdot y) \cdot z = x \cdot (y \cdot z)$	
7) ESISTENZA ELEMENTO NEUTRO (·)	$\exists a \in \mathbb{R} \text{ t.c. } \forall x \in \mathbb{R} \quad a \cdot x = x$	} si dimostra che il reciproco di x è unico e lo si indica con x^{-1} o con $1/x$
8) ESISTENZA OPPOSTO (+, ·)	$\forall x \in \mathbb{R} \setminus \{0\} \exists y \in \mathbb{R} \text{ t.c. } x \cdot y = 1$	
9) DISTRIBUTIVA (+, ·)	$\forall x, y, z \in \mathbb{R} \quad (x+y) \cdot z = x \cdot z + y \cdot z$	
10) ESCLUSIONE CASI BANALI (+, ·)	$0 \neq 1$	
11) RIFLESSIVA (\leq)	$\forall x \in \mathbb{R} \quad x \leq x$	} le proprietà 11, 12 e 13 dicono che \leq è una relazione d'ordine
12) ANTISIMMETRICA (\leq)	$\forall x, y \in \mathbb{R} \quad ((x \leq y \text{ e } y \leq x) \Rightarrow x=y)$	
13) TRANSITIVA (\leq)	$\forall x, y, z \in \mathbb{R} \quad ((x \leq y \text{ e } y \leq z) \Rightarrow x \leq z)$	} la 14 ci dice che è anche totale
14) TOTALE (\leq)	$\forall x, y \in \mathbb{R} \quad x \leq y \text{ o } y \leq x$	
15) COMPATIBILITA' TRA ORDINE E OPERAZ. (+, ·, \leq)	$\forall x, y, z \in \mathbb{R} \quad (x \leq y \Rightarrow x+z \leq y+z)$ $\forall x, y, z \in \mathbb{R} \quad ((x \leq y \text{ e } 0 \leq z) \Rightarrow x \cdot z \leq y \cdot z)$	
16) COMPLETEZZA (\leq)	$\forall (A, B)$, coppia di classi separate, $\exists z \in \mathbb{R} \text{ t.c.}:$ $\forall a \in A, \forall b \in B, a \leq z \leq b$	(A, B) è una coppia di classi separate se $A, B \subseteq \mathbb{R}; A, B \neq \emptyset;$ $\forall a \in A, b \in B$ si ha $a \leq b$

è importante anche la seguente proprietà

nome della proprietà	proprietà (descrizione formale)	note
17) ESISTENZA DELL'ESTREMO SUPERIORE	$\forall A \subseteq \mathbb{R}$, non vuoto e limitato superiormente, $\exists z \in \mathbb{R}$ t.c. $z = \sup A$	$\mathbb{A}\mathbb{S}\mathbb{R}$ è superiormente limitato se ha un maggiorante, cioè se $\exists m \in \mathbb{R}$ t.c. $\forall x \in A, x \leq m$ $\sup A$ è il minimo dei maggioranti di A

CAPITOLO II

LIMITI DI SUCCESSIONI

1. Definizione di limite per successioni

Definizione 1 Una successione di numeri reali è una applicazione avente come dominio \mathbb{N} e come codominio \mathbb{R} . \square

Quindi, se chiamiamo a una successione, abbiamo che è $a: \mathbb{N} \longrightarrow \mathbb{R}$. Quello che abbiamo definito è una successione di numeri reali: se il codominio dell'applicazione è un insieme di fagioli, avremo una successione di fagioli. Noi ci occuperemo quasi esclusivamente di successioni di numeri reali.

Dato un numero naturale n , è consuetudine indicare il valore che la successione a assume in n con a_n , anziché con $a(n)$, come di solito si fa con le applicazioni. Dato $n \in \mathbb{N}$, a_n viene detto "termine n -esimo della successione a ". Vale la pena di notare che l'insieme dei termini della successione a non è altro che l'immagine di a , cioè $a(\mathbb{N})$. Tale insieme è importante, tanto è vero che diremo che una successione a è limitata (o che ha massimo, o che il suo estremo superiore è λ , etc.) se $a(\mathbb{N})$ è limitato (o ha massimo, o il suo estremo superiore è λ , etc.).

Una nozione che non possiamo invece definire usando solo $a(\mathbb{N})$ è l'idea di successione crescente e decrescente. Guardando solo ad $a(\mathbb{N})$ non possiamo capire come stanno le cose a questo riguardo. Ad esempio, $a_n = 1/n$ è decrescente, mentre $b_n = \begin{cases} 1/(2k-1) & \text{per } n=2k, k \in \mathbb{N} \\ 1/2k & \text{per } n=2k-1, k \in \mathbb{N} \end{cases}$, cioè $b_1 = 1/2$, $b_2 = 1$, $b_3 = 1/4$, $b_4 = 1/3$, $b_5 = 1/6$, $b_6 = 1/5$, ... , non è decrescente. Eppure, $a(\mathbb{N}) = b(\mathbb{N})$. Abbiamo pertanto bisogno della seguente:

Definizione 2 Una successione $a: \mathbb{N} \longrightarrow \mathbb{R}$ si dice debolmente crescente (rispettivamente: debolmente decrescente, strettamente crescente, strettamente decrescente) se per ogni $n \in \mathbb{N}$ $a_n \leq a_{n+1}$ (rispettivamente: $a_n \geq a_{n+1}$, $a_n <$

$\langle a_{n+1}, a_n \rangle a_{n+1}$). Una successione si dice monotona se è debolmente crescente o debolmente decrescente; si dice strettamente monotona se è strettamente crescente o strettamente decrescente. \square

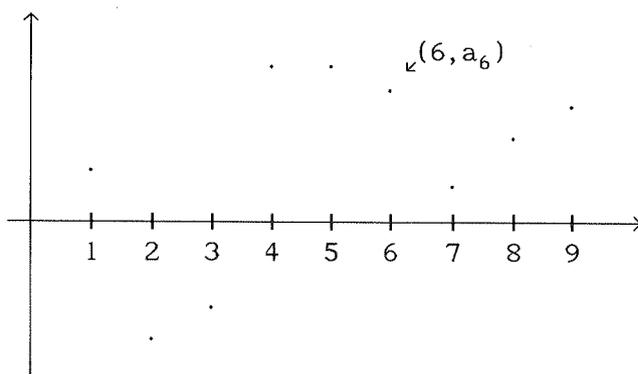
Esercizio 1 Fare un esempio di successione non monotona. \square

Un altro concetto importante è quello di grafico di una successione¹. Una successione a , essendo una applicazione, ha come tutte le applicazioni il suo grafico che è:

$$\text{gph}(a) = \{ (n, y) \in \mathbb{N} \times \mathbb{R} : n \in \mathbb{N} \text{ e } y = a(n) \}.$$

Naturalmente, essendo $\mathbb{N} \subseteq \mathbb{R}$, $\text{gph}(a) \subseteq \mathbb{R}^2$.

Come per tutte le applicazioni, la conoscenza del grafico di una successione è equivalente alla conoscenza della successione stessa. Pertanto, un modo sensato di "rappresentare" una successione è quello di "tracciarne il grafico" nel piano cartesiano.



Molto spesso, tuttavia, una successione viene individuata tramite l'insieme dei suoi termini. Tale rappresentazione può essere utile, anche più comoda di quella mediante il grafico, però può dare luogo a gravi malintesi. Occorre tenere presente che l'insieme dei termini di una successione non è altro che l'immagine della applicazione a : è ben noto che applicazioni tra di loro diverse possono avere la stessa immagine. Per esempio, la successione $a_n = (-1)^n$ e la successione b così definita: $b_n = \begin{cases} -1 & \text{per } n=1 \\ 1 & \text{per } n>1 \end{cases}$ hanno la stessa immagine. Uno si potrebbe anzi chiedere perché venga così tanto utilizzata questa rappresentazione

¹ Dedicherò il paragrafo III.2 ad un approfondimento di alcune questioni concernenti la rappresentazione dei grafici.

per le successioni: la ragione è che tale rappresentazione sottintende spesso una "dinamica" che non si può riprodurre su un mezzo di comunicazione statico come un foglio di carta. Per essere più esplicito, di solito succede che chi spiega (per esempio) indichi sulla lavagna via via i termini della successione. In tal modo si riesce a notare se i termini della successione si avvicinano (con lo scorrere del tempo) verso un numero reale ℓ . Oppure si nota se si vanno a sovrapporre l'uno all'altro, ed in quale modo (per esempio con la successione $a_n = (-1)^n$ il gesso saltella sempre tra -1 e 1 , mentre con la successione b_n sopra definita il gesso, a parte la prima volta, va poi sempre a cadere nel punto 1). Quindi, riflettendo un po', abbiamo che il modo in cui concretamente si usa la rappresentazione di una successione mediante l'insieme dei suoi termini, usa in effetti uno schema bidimensionale: oltre alla retta reale sulla quale si posizionano i termini della successione, c'è anche l'altra dimensione che è il tempo.

Fatte queste premesse, possiamo passare senza indugi al "cuore" di questo capitolo: la definizione di limite per successioni.

Prima serve però una convenzione "stenografica": useremo il simbolo \forall come sinonimo di "per ogni" e il simbolo \exists come sinonimo di "esiste". Questa convenzione, anche se probabilmente nota a tutti, non è stata utilizzata nel primo capitolo, tranne che per la tabella degli assiomi di \mathbb{R} . E' evidentemente servita in quella tabella per ragioni di spazio e di sinteticità: sono queste le stesse ragioni che mi inducono ad introdurla a questo punto. Userò anche il simbolo \Rightarrow come sinonimo di "se ... allora".

Definizione 3 Sia data una successione $a: \mathbb{N} \longrightarrow \mathbb{R}$ ed un numero reale ℓ . Diremo che ℓ è il limite della successione per "n tendente a infinito" se è soddisfatta la seguente condizione:

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

Terribile, vero? A parte la crudeltà di avere introdotto i simboli \forall ed \exists proprio allo scopo di far apparire più criptica di quanto non sia questa definizione, non è che dopo averla tradotta in un linguaggio un po' più simile all'italiano uno è molto più soddisfatto. E allora ricominciamo daccapo.

Proviamo a chiederci cosa voglia dire limite di una successione, facciamo un po' di esempi, cerchiamo di precisare la nostra intuizione: ci accorgeremo che è ragionevole giungere alla definizione che ho dato così "ex abrupto", anche se non

è banale.

In prima approssimazione, l'intuizione che vogliamo rendere precisa con una definizione è che i termini a_n si avvicinano sempre più ad un certo numero reale l , via via che n diventa grande.

Ciò può essere espresso dicendo che la "distanza" tra a_n ed l tende a diventare sempre più piccola al crescere di n . La "distanza" tra a_n ed l la misuriamo con $|a_n - l|$ e quindi potremmo dire che

$$(*) \quad |a_{n+1} - l| \leq |a_n - l| \quad \text{per ogni } n$$

il che esprime l'idea che tale distanza tende a diminuire. Ma questo chiaramente non basta: noi vogliamo che la suddetta distanza diventi sempre più piccola. Fac-

ciamo un esempio: consideriamo $a_n = 1 + \frac{1}{n}$ ed $l = 0$.

Allora $|a_{n+1} - l| = 1 + \frac{1}{n+1} < 1 + \frac{1}{n} = |a_n - l|$, quindi la condizione sopra indicata è soddisfatta. Ma $|a_n - l| = 1 + \frac{1}{n}$ non diventa "piccola" al crescere di n : resta sempre maggiore di 1.

Quindi occorre qualcos'altro: dobbiamo riuscire ad esprimere il fatto che $|a_n - l|$ diventa "piccola". Occorre trovare un modo per esprimere questa idea dinamica nel contesto dei numeri reali, che di per sé non hanno alcun aspetto dinamico. Ci vuole poco, però: basta fare una specie di gioco a due mosse. Con la prima mossa si fissa una "soglia" ε e con la seconda si trova un indice n tale che sia $|a_n - l| < \varepsilon$.

La chiave di tutto sta qui: prima si sceglie un ε e poi si cerca l'indice n . La dinamica dell'idea di limite si è spostata fuori dai numeri reali, è stata incorporata nella peculiare struttura della frase che usiamo per cercare di tradurre tale idea. Voglio sottolineare che capire questo punto significa capire il nocciolo della definizione di limite.

A questo punto, superato lo scoglio più arduo, possiamo occuparci di precisare meglio le cose.

Tanto per cominciare, abbiamo fatto una richiesta eccessiva. Perché mai $|a_n - l|$, per diventare sempre più piccola, deve soddisfare la condizione (*)? Facciamo un esempio. Consideriamo una pallina che rimbalza sul pavimento. Per attrito ed altri motivi, l'energia meccanica si trasforma in calore e quindi la pallina saltellerà sempre meno. Se immaginiamo di fotografare la pallina ad istanti suc-

cessivi, e su ogni foto misuriamo la distanza tra pallina e pavimento, otteniamo una successione la quale si avvicina sì a zero, ma non ci aspettiamo che in ogni fotogramma la distanza tra pallina e pavimento sia minore della distanza nel fotogramma precedente.

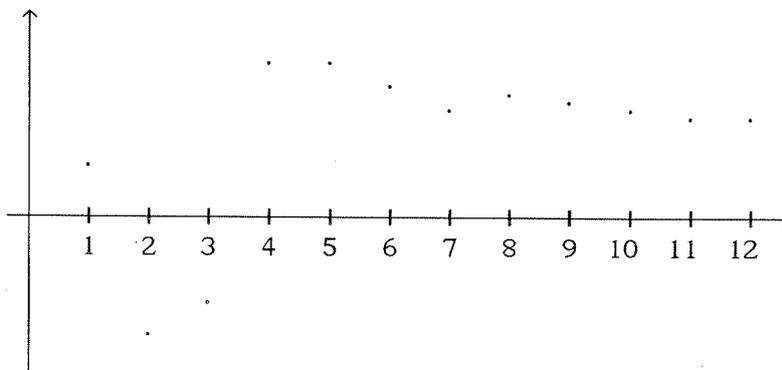
Quindi la condizione (*) si presenta come una condizione accessoria, non indispensabile per rendere l'idea di limite. Sarà quindi opportuno scartarla. Perché? Perché quando si vuole precisare un concetto, è opportuno depurarlo di tutti gli elementi estranei, di modo che si abbia una maggiore chiarezza.

Avendo tolto di mezzo la condizione (*), ci si presenta però una difficoltà. E cioè: la condizione che ci sembrava di avere enucleato per rendere l'idea di limite non va più bene. Non è sufficiente. Per capirlo, possiamo fare il seguente esempio. Supponiamo di avere una successione che quando n è pari vale 1 e quando n è dispari vale $1/n$. E' evidente che, fissato $\varepsilon > 0$, ci sarà n t.c. $1/n$ è più piccolo di ε . Quindi la condizione che avevamo trovato è soddisfatta con $l=0$. Ma non penso che si possa dire che tale successione tende a zero (solo i termini dispari tendono a zero). E allora? Come possiamo rimediare?

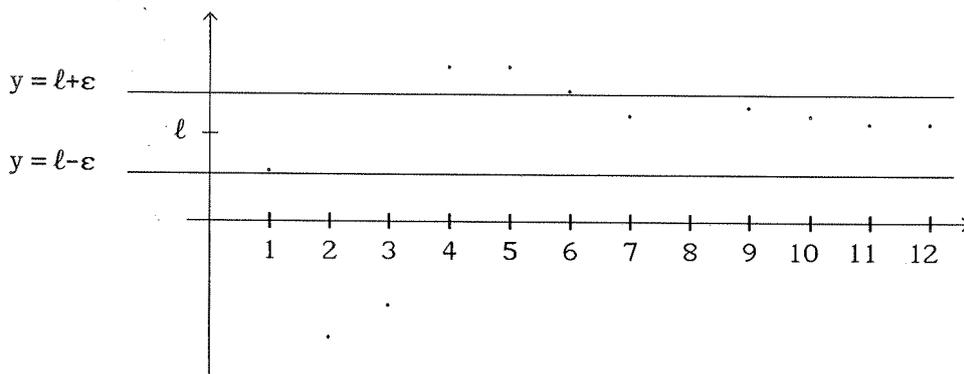
Credo che di idee ce ne possano essere molte, ma non è mia intenzione fare una disamina di tutto quel che può passare per la testa a chi voglia risolvere questo problema. E allora presenterò subito l'idea "buona". Tutto sommato, non è poi tanto strana. Prima avevamo detto che, dato un numero reale ε positivo, bisognava riuscire a trovare un indice n t.c. $|a_n - l| < \varepsilon$. Bene, si tratta di richiedere che non solo si abbia $|a_n - l| < \varepsilon$, ma anche $|a_m - l| < \varepsilon$ per tutti gli indici m successivi ad n .

Spero che sia chiara la differenza tra le due formulazioni. Prima ci accontentavamo di trovare un indice n t.c. $|a_n - l| < \varepsilon$. Ora pretendiamo ben di più: vogliamo trovare un indice n t.c. da quell'indice in poi, cioè per ogni indice m più grande di n , sia soddisfatta la condizione $|a_m - l| < \varepsilon$.

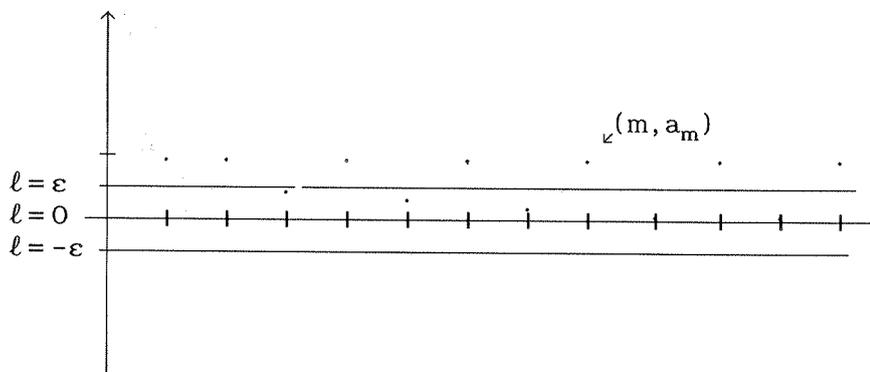
Graficamente, l'idea è questa. Esaminiamo il grafico di una successione.



Per poter affermare che l è il limite, noi fissiamo $\varepsilon > 0$ e consideriamo la striscia di piano compresa tra le due rette di equazione $y = l - \varepsilon$ ed $y = l + \varepsilon$. Avere che il punto (n, a_n) del grafico della successione a sta nella striscia, è equivalente ad affermare che $|a_n - l| < \varepsilon$.



La condizione precedente ci chiedeva di riuscire a trovare, per ogni prefissata striscia di ampiezza 2ε , un punto del grafico nella striscia. Cosa soddisfatta evidentemente dalla successione $a_n = 1$ per n pari ed $a_n = 1/n$ per n dispari:



La richiesta che facciamo è ben più impegnativa. Vogliamo trovare un indice n t.c. ogni punto (m, a_m) del grafico di a stia nella striscia, purché

$m > n$. Dall'ultimo disegno dovrebbe essere chiaro che, per quanto proviamo a prendere grande n , ci sono sempre punti (m, a_m) che "scappano" fuori dalla striscia: non diremo pertanto che questa successione tende a 1.

Ebbene, la condizione che abbiamo individuato e che abbiamo anche visto rappresentata graficamente è proprio la definizione di limite.

Ricapitolando, la condizione si può esprimere così. Data una successione a ed un numero reale l , possiamo affermare che a_n tende ad l quando n diventa grande se è vero quanto segue:

"per ogni numero reale $\varepsilon > 0$ fissato, si può trovare un indice n t.c., per ogni indice m , la condizione $m > n$ garantisce che $|a_m - l| < \varepsilon$ "

Togliendo un po' di fronzoli:

"per ogni $\varepsilon > 0$, esiste $n \in \mathbb{N}$ t.c. per ogni $m \in \mathbb{N}$, se $m > n$ allora $|a_m - l| < \varepsilon$ ".

E questa è proprio la definizione 3, che si differenzia da quanto scritto sopra solo per l'uso delle abbreviazioni stenografiche. Anzi, vale la pena di ritrascriverla qui:

Definizione 3 Sia data una successione $a: \mathbb{N} \rightarrow \mathbb{R}$ ed un numero reale l . Diremo che l è il "limite della successione per n tendente a infinito" se è soddisfatta la seguente condizione:

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon) . \square$$

Se una successione a e un numero reale l soddisfano la definizione 3, diremo che l è il "limite di a_n per n che tende all'infinito", cosa che verrà riassunta nella notazione $\lim_{n \rightarrow \infty} a_n = l$. Un'altra notazione, anche più comoda, è $a_n \xrightarrow{n \rightarrow \infty} l$. Anzi, addirittura a volte si scrive solo $a_n \rightarrow l$, senza stare a ricordare che $n \rightarrow \infty$ (tanto, dove può tendere, se non a ∞ ?).

Avendo a disposizione la definizione di limite, vediamo subito un paio di esempi.

Esempio 1 E' $\lim_{n \rightarrow \infty} 1/n = 0$. \square

Dettaglio Questo fatto era già stato notato, ma in un contesto di discussione informale. In realtà, là era stato dato per scontato un fatto per nulla banale. E cioè che, dato $\varepsilon > 0$, ci sia $n \in \mathbb{N}$ t.c. $1/n < \varepsilon$. Chi ci autorizza a fare questa affermazione? In quale assioma di \mathbb{R} sta scritto? Direttamente in nessuno.

In realtà tale affermazione è corretta grazie al principio di Archimede: in effetti, dire che $1/n < \varepsilon$ è equivalente a dire che $1/\varepsilon < n$. E' proprio il principio di Archimede che ci garantisce che, dato il numero reale $x = 1/\varepsilon$, c'è un $n \in \mathbb{N}$ t.c. $n > 1/\varepsilon$. Qui sta il nocciolo del discorso. Il resto è facile. Vediamo quindi che $1/n \rightarrow 0$. Sia $\varepsilon > 0$. Dobbiamo trovare $\nu \in \mathbb{N}$ t.c. per ogni $n > \nu$ si abbia $|a_n - 0| < \varepsilon$, cioè $|(1/n) - 0| < \varepsilon$, cioè $1/n < \varepsilon$. Per Archimede, $\exists \nu \in \mathbb{N}$ t.c. $1/\nu < \varepsilon$. Se poi prendiamo $n > \nu$, è $1/n < 1/\nu$ e quindi $1/n < 1/\nu < \varepsilon$, cioè $1/n < \varepsilon$, come volevamo. \square

Esempio 2 Sia $a_n = 1$ per n pari ed $a_n = 1/n$ per n dispari. E' falso che $\lim_{n \rightarrow \infty} a_n = 0$. \square

Dettaglio Basta fissare $\varepsilon = 1/2$. Se il limite fosse 0, dovremmo essere in grado di trovare $\nu \in \mathbb{N}$ t.c. $\forall n > \nu$ si abbia $a_n < 1/2$. Ma questo ovviamente non è possibile, perché $\forall \nu \in \mathbb{N}$ ci sarà sempre un n pari maggiore di ν^2 e quindi, visto che $a_n = 1$ se n è pari, non è possibile soddisfare la richiesta che sia $a_n < 1/2$ per ogni indice $n > \nu$. \square

² Questa affermazione mi permette di ribadire una cosa. Ho detto che assumerò per note le proprietà dei numeri naturali (e razionali): questo vale anche per il sottoinsieme di \mathbb{R} che abbiamo visto poter identificare con l'insieme dei naturali.

2. Unicità del limite e successioni convergenti

E' giunta l'ora di dimostrare uno dei risultati più importanti: il limite di una successione, se esiste, è unico. Penso che possa essere di per sé evidente l'utilità di un tale risultato, per cui non mi dilungo in chiacchiere e passo subito al teorema.

Teorema 1 Una successione non può avere più di un limite. \square

Dimostrazione L'affermazione del teorema può essere tradotta così: se l è il limite di una successione a , ed abbiamo $l \neq l'$, allora non è vero che $\lim_{n \rightarrow \infty} a_n = l'$. Detto altrimenti, dati l ed l' con $l \neq l'$, non può essere vero sia $\lim_{n \rightarrow \infty} a_n = l$ che $\lim_{n \rightarrow \infty} a_n = l'$. Ancora, equivalentemente: data una successione a e dati $l, l' \in \mathbb{R}$, se $\lim_{n \rightarrow \infty} a_n = l$ e $\lim_{n \rightarrow \infty} a_n = l'$, allora $l = l'$.

Si tratta di tre affermazioni non identiche ma tra loro equivalenti. Sta a noi scegliere quella che più ci ispira per la dimostrazione. Userò la seconda affermazione, che suggerisce implicitamente di tentare una dimostrazione per assurdo.

Siano quindi dati $a: \mathbb{N} \rightarrow \mathbb{R}$ ed $l, l' \in \mathbb{R}$ con $l \neq l'$. Supponiamo $\lim_{n \rightarrow \infty} a_n = l$ e $\lim_{n \rightarrow \infty} a_n = l'$. Vediamo di ricavare da ciò una contraddizione. Se $l \neq l'$, allora $|l - l'| > 0$. Dalla definizione di limite abbiamo, scelto $\varepsilon = \frac{|l - l'|}{2}$:

$$\exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon)$$

$$\exists \nu' \in \mathbb{N} \text{ t.c. } \forall n' \in \mathbb{N} (n' > \nu' \Rightarrow |a_{n'} - l'| < \varepsilon)$$

Osserviamo il fatto seguente. Chiamato $\nu'' = \max \{ \nu, \nu' \}$ ³, se $p > \nu''$, allora $p > \nu$ e $p > \nu'$.

Ma allora abbiamo che per $p > \nu''$ si ha $|a_p - l| < \varepsilon$ sia $|a_p - l'| < \varepsilon$. Sommando membro a membro:

$$|a_p - l| + |a_p - l'| < 2\varepsilon$$

Ma allora:

$$|l - l'| = \quad \quad \quad \text{per come abbiamo scelto } \varepsilon$$

³ La notazione dovrebbe essere chiara di per sé. Non essendo però consueta, può provocare qualche dubbio. Bene: ν'' indica l'elemento massimo nell'insieme i cui elementi sono ν e ν' . Quindi se $\nu > \nu'$ allora $\nu'' = \nu$, se $\nu' > \nu$ allora $\nu'' = \nu'$, se $\nu = \nu'$ allora $\nu'' = \nu = \nu'$.

$$\begin{aligned}
&= 2\varepsilon > && \text{per la disuguaglianza appena provata} \\
&> |a_p - l| + |a_p - l'| = && \text{perché } |x| = |-x| \quad \forall x \in \mathbb{R} \\
&= |l - a_p| + |a_p - l'| \geq && \text{per la disuguaglianza triangolare} \\
&\geq |(l - a_p) + (a_p - l')| = && \text{semplificando} \\
&= |l - l'|
\end{aligned}$$

Abbiamo così ottenuto $|l - l'| > |l - l'|$, una contraddizione. \square

Una volta provata l'unicità del limite, possiamo vedere un punto che di solito risulta essere un po' delicato, e cioè che cosa significa dire che una successione è convergente. La definizione è semplice.

Definizione 1 Data una successione a , diremo che essa è convergente se $\exists l \in \mathbb{R}$ t.c. $\lim_{n \rightarrow \infty} a_n = l$. \square

Apparentemente, niente di particolare. Per così dire, una successione è convergente se converge da qualche parte. In realtà, le difficoltà nascono quando si vuole dimostrare che una successione non è convergente. Per farlo, cominciamo a scrivere cosa vuol dire che una successione è convergente:

$$(\bullet) \quad \exists l \in \mathbb{R} \text{ t.c. } \forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon)$$

Come si fa a negare questa proposizione? Facile. Questa proposizione ci dice che $\exists l \in \mathbb{R}$ che gode della proprietà seguente:

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

Negare questo fatto, cioè dire che non c'è alcun $l \in \mathbb{R}$ che gode della proprietà (*), è come dire che qualunque sia $l \in \mathbb{R}$ che noi scegliamo, esso non gode della proprietà (*). Cioè, $\forall l \in \mathbb{R}$ la proprietà (*) è falsa. Si tratta di una cosa ovvia: negare che a Genova vivano marziani è equivalente a dire che ogni persona vivente a Genova non è marziana.

Quindi abbiamo la equivalenza tra:

$$\text{non } (\exists l \in \mathbb{R} \text{ t.c. } (\forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon)))$$

e

$$\forall l \in \mathbb{R} (\text{non } (\forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon)))$$

Se si guarda all'aspetto formale della faccenda, abbiamo utilizzato il fatto che "non $\exists l \in \mathbb{R}$ t.c." è equivalente a " $\forall l \in \mathbb{R}$, non".

Similmente, dire che "non $\forall \varepsilon > 0$ " è equivalente a dire " $\exists \varepsilon > 0$ t.c. non", e possiamo applicare questo fatto per riscrivere (\bullet) così:

$$\forall l \in \mathbb{R} \quad \exists \varepsilon > 0 \text{ t.c. non } (\exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon))$$

E possiamo proseguire, ottenendo:

$$\forall l \in \mathbb{R} \exists \varepsilon > 0 \text{ t.c. } \forall \nu \in \mathbb{N} \text{ non } (\forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon))$$

E ancora, sempre equivalentemente:

$$\forall l \in \mathbb{R} \exists \varepsilon > 0 \text{ t.c. } \forall \nu \in \mathbb{N} \exists n \in \mathbb{N} \text{ t.c. non } (n > \nu \Rightarrow |a_n - l| < \varepsilon)$$

E qui viene il bello. Si tratta di trovare una proposizione che sia equivalente a:

$$(\blacktriangleright) \quad \text{non } (n > \nu \Rightarrow |a_n - l| < \varepsilon)$$

A questo punto chi legge è invitato a riflettere su come si possa riscrivere. Per invogliare il lettore pigro a farlo, il testo prosegue nella pagina seguente, dove sono elencate una serie di proposizioni che non sono equivalenti a quella data. Anticipo che la risposta

$$(n > \nu \neq |a_n - l| < \varepsilon)$$

non è accettabile, perchè è solo un modo diverso di riscrivere la stessa cosa di (\blacktriangleright) .

Le seguenti proposizioni sono tutte sbagliate come negazione della proposizione $(n > \nu \Rightarrow |a_n - l| < \varepsilon)$

$$n > \nu \Rightarrow |a_n - l| \geq \varepsilon$$

$$|a_n - l| < \varepsilon \Rightarrow n > \nu$$

$$n > \nu \text{ e } |a_n - l| > \varepsilon$$

$$|a_n - l| \geq \varepsilon \Rightarrow n \leq \nu$$

La corretta negazione di (\blacktriangleright) è:

$$(n > \nu \text{ e } (\text{non } |a_n - l| < \varepsilon))$$

ovverossia

$$(n > \nu \text{ e } |a_n - l| \geq \varepsilon)$$

Perché? Perché la corretta negazione di " $p \Rightarrow q$ " è " p e non q ". Di solito si usa il simbolo " \neg " per indicare "non". Pertanto la negazione di $p \Rightarrow q$ la scriveremo così: p e $\neg q$. Un'altra convenzione usata è quella di indicare con " \wedge " la congiunzione "e" e con " \vee " la "o" nel suo significato non esclusivo ("or", ovvero "vel"). E quindi potremmo scrivere $p \wedge \neg q$ come negazione di $p \Rightarrow q$. Non userò molto spesso, comunque, i simboli " \wedge " e " \vee ", perché dal punto di vista "stenografico" e da quello della chiarezza di scrittura non ci si guadagna molto rispetto all'uso di "e" ed "o".

Riassumendo, ricordo che il problema era quello di esprimere cosa voglia dire che una successione non è convergente. Bene: dire che una successione non è convergente equivale ad affermare che:

$$\forall l \in \mathbb{R} \exists \varepsilon > 0 \text{ t.c. } \forall \nu \in \mathbb{N} \exists n \in \mathbb{N} \text{ t.c. } (n > \nu \text{ e } |a_n - l| \geq \varepsilon)$$

Esempio 1 Provare che $(-1)^n$ non è convergente. \square

Dettaglio Dobbiamo dimostrare che $\forall l \in \mathbb{R}$ vale una certa cosa. Ci conviene verificarlo facendo una distinzione di casi. E cioè per $l = -1$, $l = 1$ ed $l \in \mathbb{R} \setminus \{-1, 1\}$. Naturalmente i casi esaminati devono essere esaustivi, il che è ovvio, essendo $\{1\} \cup \{-1\} \cup (\mathbb{R} \setminus \{-1, 1\}) = \mathbb{R}$.

Se $l = -1$, dobbiamo provare (come per ogni $l \in \mathbb{R}$) che

$$\exists \varepsilon > 0 \text{ t.c. } \forall \nu \in \mathbb{N} \exists n \in \mathbb{N} \text{ t.c. non } (n > \nu \Rightarrow |a_n - l| < \varepsilon)$$

Prendiamo $\varepsilon = 1$. Osserviamo che $\forall \nu \in \mathbb{N}$, $2\nu > \nu$ e perciò possiamo prendere $n = 2\nu$, ottenendo $|a_n - l| = |a_{2\nu} - l| = |1 - (-1)| = 2 \geq \varepsilon$.

Se $l = 1$, la strada da seguire è simile (useremo $2\nu + 1$ anziché 2ν).

Se $l \neq -1, 1$, allora prendiamo $\varepsilon = \min \{ |l - (-1)|, |l - 1| \}$. Allora, $\forall \nu \in \mathbb{N}$ scegliamo $n = \nu + 1$ e abbiamo $|a_n - l| \geq \varepsilon$ proprio per come abbiamo definito ε . \square

Esercizio 1 Dimostrare che la successione $a_n = \begin{cases} 1 & \text{per } n \text{ pari} \\ 1/n & \text{per } n \text{ dispari} \end{cases}$ non è convergente. \square

3. Microcorso di logica

L'aspetto più semplice della logica sta nel cosiddetto "calcolo proposizionale" in cui non si va a vedere il contenuto delle proposizioni, ma ci si limita all'esame delle relazioni "logiche" tra loro.

Ad esempio: la proposizione composta $p \Rightarrow q$ è equivalente alla proposizione $\neg q \Rightarrow \neg p$ indipendentemente da quanto viene affermato nelle due proposizioni p e q .

Le proposizioni composte si ottengono, a partire da quelle "di base" o "atomiche" (nel senso di atomi, ovvero sia non più ulteriormente "divisibili"⁴) legando le proposizioni atomiche tra loro mediante dei "connettivi". Essi sono: " \neg ", " \vee ", " \wedge ", " \Rightarrow " e " \Leftrightarrow " (ovverossia, rispettivamente: "non", "o", "e", "implica" e "se e solo se").

Il significato dei connettivi è completamente racchiuso dalle "tavole di verità" qui appresso descritte.

p	$\neg p$
V	F
F	V

p	q	$p \vee q$
V	V	V
V	F	V
F	V	V
F	F	F

p	q	$p \wedge q$
V	V	V
V	F	F
F	V	F
F	F	F

p	q	$p \Rightarrow q$
V	V	V
V	F	F
F	V	V
F	F	V

p	q	$p \Leftrightarrow q$
V	V	V
V	F	F
F	V	F
F	F	V

Si tratta di tabelle dal significato evidente: dalla seconda riga della terza tabella si evince, ad esempio, che la proposizione composta $p \wedge q$ è falsa quando la proposizione p è falsa e la proposizione q è vera. Più in generale, la terza tabella ci dice che la proposizione $p \wedge q$ è vera solo quando sono vere sia p che q .

La seconda tabella mostra che stiamo considerando "o" nel suo significato non esclusivo: infatti consideriamo vera la proposizione $p \vee q$ anche nel caso in cui sia p che q siano entrambe vere. Se fossimo interessati al significato esclusivo, ovvero sia ad "aut", avremmo la tabella seguente:

⁴ La terminologia richiederebbe un certo aggiornamento...

p	q	p aut q
V	V	F
V	F	V
F	V	V
F	F	F

come si può notare agevolmente, la tabella di "aut" è la tabella di "↔" rovesciata. Possiamo quindi sostituire la proposizione $p \text{ aut } q$ con la equivalente $\neg(p \leftrightarrow q)$.

La tabella più interessante è quella di \Rightarrow per la presenza della discutibile e inquietante terza riga. Perché mai deve essere vero che $p \Rightarrow q$ quando p è falsa e q è vera? Forse il modo più drastico per esporre queste perplessità è di effettuare il seguente ragionamento: "se da una premessa p falsa sono riuscito ad ottenere una conseguenza q vera, devo aver per forza sbagliato qualcosa nel mio ragionamento, per cui la deduzione è di sicuro falsa". In realtà l'argomentazione tra virgolette è completamente fuori tema: non stiamo parlando affatto di deduzione (o di dimostrazione, per usare un termine più consueto). A noi non interessa minimamente la presenza o meno di una qualche connessione tra le proposizioni p e q . Tanto per fare un esempio, p può essere "oggi piove" e q "questa rosa è rossa".

Il modo in cui il simbolo di implicazione è usato costantemente nella matematica è proprio quello corrispondente alla tabella di verità sopra riportata.

Per quanto elementare, la "logica delle proposizioni" permette già di fare tutta una serie di interessanti considerazioni.

Tanto per cominciare, vi sono una serie di "tautologie", ovverossia proposizioni sempre vere, indipendentemente dal fatto che le proposizioni componenti siano vere o false (e indipendentemente dai loro "contenuti!"). Ad esempio:

$p \vee (\neg p)$	"terzo escluso"
$\neg(p \wedge (\neg p))$	"principio di non contraddizione"
$\neg(p \vee q) \Leftrightarrow ((\neg p) \wedge (\neg q))$	"leggi di De Morgan"
$\neg(p \wedge q) \Leftrightarrow ((\neg p) \vee (\neg q))$	
$(p \Rightarrow q) \Leftrightarrow ((p \Rightarrow q) \wedge (q \Rightarrow p))$	l'equivalenza logica è logicamente equivalente alla doppia implicazione ("se e solo se")

Altre tautologie interessanti sono le seguenti: su di esse sono fondati dei metodi di dimostrazione frequentemente usati.

$(p \Rightarrow q) \Leftrightarrow ((\neg q) \Rightarrow (\neg p))$ è alla base della dimostrazione per

"contrapposta" o "contronominale", talvolta confusa con la dimostrazione per assurdo.

$(p \Rightarrow q) \Leftrightarrow ((p \wedge (\neg q)) \Rightarrow ((r \wedge (\neg r))))$ è la dimostrazione per assurdo: si nega la tesi e da questa si ricava una contraddizione.

Esempio 1 Verificare che $(p \Rightarrow q) \Leftrightarrow ((p \wedge (\neg q)) \Rightarrow ((r \wedge (\neg r))))$ è una tautologia. \square

Dettaglio Basta costruire la "tavola di verità" di questa proposizione. Dato che sono coinvolte tre proposizioni "atomiche", p, q, r , vi saranno otto diverse possibili combinazioni dei loro valori di verità. Abbiamo quindi:

p	\Rightarrow	q	\Leftrightarrow	p	\wedge	\neg	q	\Rightarrow	r	\wedge	\neg	r
V	V	V	V	V	F	F	V	V	V	F	F	V
V	V	V	V	V	F	F	V	V	F	F	V	F
V	F	F	V	V	V	V	F	F	V	F	F	V
V	F	F	V	V	V	V	F	F	F	F	V	F
F	V	V	V	F	F	F	V	V	V	F	F	V
F	V	V	V	F	F	F	V	V	F	F	V	F
F	V	F	V	F	F	V	F	V	V	F	F	V
F	V	F	V	F	F	V	F	V	F	F	V	F
1	2	3	4	5	6	7	8	9	10	11	12	13

Le colonne sono state riempite nell'ordine seguente. Dapprima le colonne 1,3,5,8,10,13, corrispondenti alle proposizioni atomiche. Poi le colonne 7 e 12; poi le colonne 2,6,11. Infine la colonna 9 e, per ultima, la colonna 4. Come si vede la colonna 4 contiene solo V, il che significa che la proposizione composta risulta essere sempre vera, indipendentemente dal valore di verità delle proposizioni atomiche; è quindi una tautologia. \square

$(p \Leftrightarrow (q \vee r)) \Rightarrow ((p \Rightarrow s) \Leftrightarrow ((q \Rightarrow s) \wedge (r \Rightarrow s)))$ è la dimostrazione per distinzione dei casi.

Esercizio 1 (obbligatorio) Provare che $(p \Rightarrow q) \Leftrightarrow (\neg(p \wedge (\neg q)))$. \square

Esercizio 2 C'è qualche legame tra il precedente esercizio e il fatto che in Turbo Pascal sono presenti solo i connettivi NOT, OR e AND per variabili booleane? \square

A un livello meno elevato di astrazione, vi è la logica dei predicati. Qui si entra un pò più nel merito delle proposizioni. Tipico esempio è la frase: "ogni uomo è mortale", equivalente a "nessun uomo è immortale". Che le due frasi siano tra di loro equivalenti spero sia accolto da tutti; non solo: le due frasi sono

"logicamente" equivalenti tra di loro. Infatti, il nocciolo sta nel fatto che asserire che ogni tizio goda di una certa proprietà, è equivalente ad asserire che nessun tizio non gode di tale proprietà. Ritroviamo insomma l'equivalenza tra "per ogni x è ..." e "non esiste x t.c. non ...". Tale equivalenza non è esprimibile a livello di logica delle proposizioni: anzi, le due proposizioni stesse sono ricavate l'una dall'altra attraverso delle manipolazioni per così dire interne.

Non posso dilungarmi troppo sulla logica dei predicati. Essa tipicamente si occupa di proposizioni del tipo " $\forall x Px$ " e " $\exists x Px$ ", dove con Px intendo esprimere il fatto che x goda della proprietà P .

A volte viene specificato l'"universo" da dove si prendono le x : " $\forall x \in U Px$ ". Esempi: $\forall n \in \mathbb{N} \quad 2n \geq n$. Qui l'universo è \mathbb{N} e la proprietà di cui si parla a proposito di n è che $2n \geq n$. Poiché tale proprietà è vera per ogni x nell'universo considerato, la proposizione $\forall n \in \mathbb{N} \quad 2n \geq n$ è vera. Sono poi anche diffuse notazioni sintetiche, quali " $\forall \varepsilon > 0 \quad P\varepsilon$ ". Questo può essere inteso così: " $\forall \varepsilon \in \mathbb{R}_+ \quad P\varepsilon$ ". Cioè, l'universo nell'ambito del quale può variare ε viene specificato in modo non ultraortodosso (ovverossia, invece di specificare direttamente l'universo, esso viene individuato mediante una proprietà che lo individua).

Nell'ambito della logica dei predicati, mi limito ad indicare che sono valide le seguenti tautologie, che abbiamo già usato nel precedente paragrafo:

$$\neg(\forall x Px) \Leftrightarrow \exists x (\neg Px)$$

$$\neg(\exists x Px) \Leftrightarrow \forall x (\neg Px).$$

Concludo segnalando un errore che è facile commettere: la frase $(\forall x Px \Rightarrow \exists x Px)$ non è una tautologia! Può infatti essere falsa; si pensi ad esempio alla seguente affermazione: "Se ogni marziano ha la pelle verde, allora c'è un marziano con la pelle verde". E' evidente dove sta l'inghippo: l'universo dove prendono valore le x può essere vuoto.

4. Bastano gli epsilon piccoli

In questo paragrafo darò spazio ad una osservazione interessante, in quanto esplicita un'idea che si ha in testa quando si introduce la definizione di limite. E cioè: bastano gli ε piccoli. Ma come rendere in modo preciso tale idea? Nella definizione di limite non c'è, anzi è coinvolto ogni numero reale positivo, anche se "molto grande" (infatti, si dice solo " $\forall \varepsilon > 0$ "). D'altronde ho messo tra virgolette l'espressione "molto grande", in quanto non è chiaro cosa significhi dire che un numero reale è piccolo o grande e tanto meno cosa significhi essere molto grande. In effetti, 10^{20} , che si potrebbe essere inclini a considerare molto grande, è piccolo rispetto a 10^{40} . E così via, naturalmente.

Nel linguaggio comune, "molto grande" o "molto piccolo" fa riferimento ad un certo contesto esplicitamente o implicitamente presente. Noi però nella definizione di limite non abbiamo un contesto di questo genere: l'unico contesto esplicito⁵ presente è quello dei numeri reali positivi, dove non ha senso l'idea di "molto piccolo".

Eppure un modo c'è per esprimere questa idea. Almeno la sua essenza. Consideriamo un numero fissato $\sigma > 0$, scelto da noi a piacere. Ebbene, se nella definizione di limite noi consideriamo solo gli $\varepsilon < \sigma$, non cambia nulla.

Cioè, dato $\sigma > 0$, dire che $\lim_{n \rightarrow \infty} a_n = l$ è equivalente a:

$$(*) \quad \forall \varepsilon' \in]0, \sigma[\quad \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow |a_n - l| < \varepsilon')$$

E si noti che σ lo possiamo scegliere noi, "piccolo" quanto ci fa comodo.

La dimostrazione di questo fatto è semplice. Che $\lim_{n \rightarrow \infty} a_n = l$ implichi (*) è ovvio (in (*) richiediamo che la proprietà che sappiamo essere vera $\forall \varepsilon > 0$ valga per gli $\varepsilon' \in]0, \sigma[$). Ma anche il viceversa è facile da ottenere: se ho tra le mani un $\varepsilon \geq \sigma$ e voglio provare che

$$(\blacktriangleright) \quad \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow |a_n - l| < \varepsilon)$$

scelgo $\varepsilon' = \sigma/2$ in (*) ed ottengo

$$\exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow |a_n - l| < \varepsilon')$$

ma è $|a_n - l| < \varepsilon' = \sigma/2 < \sigma \leq \varepsilon$, quindi posso affermare (\blacktriangleright).

Abbiamo così provato che possiamo limitare la scelta degli ε a quelli minori di un σ scelto da noi a piacere (purché positivo): questo è il modo che abbiamo per esprimere in termini precisi l'intuizione che gli unici ε che con-

⁵ E contesti impliciti non ce ne devono essere in matematica.

tano nella definizione di limite sono quelli piccoli.

Già che siamo in argomento, osserviamo che sono equivalenti le seguenti affermazioni:

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon)$$

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n \geq \nu \Rightarrow |a_n - l| < \varepsilon)$$

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n \geq \nu \Rightarrow |a_n - l| \leq \varepsilon)$$

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| \leq \varepsilon)$$

Esercizio 1 Provare la precedente affermazione. \square

Quindi le disuguaglianze le possiamo mettere strette o larghe, fa lo stesso. Ma non quella che riguarda ε : non possiamo sostituire $\varepsilon > 0$ con $\varepsilon \geq 0$.

Esercizio 2 Provare che solo le successioni costanti da un certo indice in poi sono convergenti secondo la definizione seguente:

$$\forall \varepsilon \geq 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon) . \square$$

5. Le successioni convergenti sono limitate

In questo paragrafo vedremo di ottenere una conseguenza molto facile della convergenza di una successione.

Teorema 1 Una successione convergente è limitata. \square

Dimostrazione Data una successione a , ricordiamo che essa è limitata se la sua immagine è limitata, cioè se $a(\mathbb{N})$ è un sottoinsieme limitato di \mathbb{R} . Cioè se esistono $H, K \in \mathbb{R}$ t.c. $H \leq a_n \leq K \quad \forall n \in \mathbb{N}$. Osserviamo che questa proprietà può essere descritta graficamente: significa che il grafico della successione è contenuto nella striscia di piano compresa tra le rette di equazione $y=H$ ed $y=K$. Questa terminologia dovrebbe far venire in mente, per le notevoli analogie, l'interpretazione grafica di cosa sia una successione convergente: invito il lettore a ricordarla prima di continuare (vedi paragrafo 1).

Dunque, se la successione converge ad l , i punti del grafico di a stanno nella striscia individuata da $y=l-\varepsilon$ ed $y=l+\varepsilon$, per lo meno a partire da un certo indice ν .

Ci siamo quasi. Resta solo da risolvere un piccolo problema. E' che solo dall'indice ν in poi i termini della successione soddisfano $l-\varepsilon < a_n < l+\varepsilon$. D'altronde i termini fino a ν sono in numero finito. Cioè, l'insieme $\{a_1, \dots, a_\nu\}$ è finito (ha un numero di elementi minore o uguale a ν): allora avrà minimo e massimo che possiamo chiamare H_1 e K_1 . Abbiamo quindi:

$$a_n \in [H_1, K_1] \quad \text{per } n \in \{1, \dots, \nu\}$$

$$a_n \in]l-\varepsilon, l+\varepsilon[\quad \text{per } n > \nu.$$

Basta prendere $H = \min \{H_1, l-\varepsilon\}$ e $K = \max \{K_1, l+\varepsilon\}$ per ottenere il risultato desiderato. \square

6. Sottosuccessioni

Se uno considera la successione già usata più volte, così definita: $a_n = \begin{cases} 1 & \text{per } n \text{ pari} \\ 1/n & \text{per } n \text{ dispari} \end{cases}$, può agevolmente notare un fatto. E cioè che, se considerassimo solo i termini di indice pari, avremmo una successione costante e quindi convergente. Abbiamo quindi che, per così dire, "dentro" la successione data è "nascosta" una successione convergente. Questa semplice osservazione ci porterà

molto lontano, ad uno dei risultati fondamentali di questo capitolo, che proveremo nel § 12.

Il mio obiettivo, per ora, è molto meno ambizioso. Si tratta di riuscire a capire per bene che cosa intendiamo dicendo che "dentro" ad una successione ce n'è "nascosta" un'altra; ovverossia, dobbiamo capire che cosa vuol dire "sottosuccessione", che è il termine specialistico usato per indicare una successione "nascosta" dentro un'altra. Cos'è una sottosuccessione? E' quello che ci resta se togliamo via un certo numero di indici, magari anche infiniti indici, avendo però cura che ne restino sempre infiniti: tanto per fare un esempio, se buttiamo via tutti gli indici dopo 37, quello che ci resta è troppo poco. Infatti, quello che vogliamo ci rimanga è qualcosa che sia ancora interpretabile come una successione.

Ricordo che una successione a è una applicazione $a : \mathbb{N} \longrightarrow \mathbb{R}$, e pertanto quanto detto sopra equivale a considerare una restrizione della funzione a ad un sottoinsieme infinito di \mathbb{N} . Nell'esempio da cui sono partito, vorrebbe dire restringere a all'insieme dei numeri pari. Questa è l'idea di sottosuccessione.

E' però più conveniente vederla in altro modo, per ragioni per così dire tecniche. Cioè definiremo una sottosuccessione come una composizione di funzioni, privilegiando l'altra idea di carattere intuitivo che sta dietro alla definizione di sottosuccessione, e cioè che si estraggano dei numeri naturali successivi e si considerino i termini ad essi corrispondenti. Ovvero, sempre nel solito esempio, dato n , "estraiamo" $2n$ e consideriamo il termine a_{2n} . Introduciamo pertanto una applicazione crescente $k : \mathbb{N} \longrightarrow \mathbb{N}$ e consideriamo la funzione composta $a \circ k : \mathbb{N} \longrightarrow \mathbb{R}$. Questa nuova funzione ottenuta è, come si vede, una funzione da \mathbb{N} in \mathbb{R} e pertanto è una successione anch'essa: la chiameremo sottosuccessione estratta da a mediante la regola di estrazione k (se non ci sono rischi di confusione, diremo semplicemente che è una sottosuccessione di a).

A livello elementare, il risultato più interessante che possiamo ottenere è il seguente.

Teorema 1 $(a_n \longrightarrow l) \Rightarrow (a_{k_n} \longrightarrow l) . \square$

Per dimostrare questo teorema ci serve il seguente

Lemma 1 Sia $k : \mathbb{N} \longrightarrow \mathbb{N}$ strettamente crescente. Allora è $k(n) \geq n$ per ogni $n \in \mathbb{N}$. \square

Dimostrazione Per induzione.

(BASE) La tesi è vera per $n=1$, in quanto $k(1) \in \mathbb{N}$ e ogni numero naturale è maggiore o uguale di 1.

(PASSO INDUTTIVO) Sia $n \in \mathbb{N}$: proviamo che $k(n) \geq n \Rightarrow k(n+1) \geq n+1$. Per ipotesi sappiamo che $k(n+1) > k(n)$, ma poiché siamo in \mathbb{N} , ciò significa che $k(n+1) \geq k(n)+1$. Pertanto $k(n) \geq n \Rightarrow k(n+1) \geq k(n)+1 \geq n+1$. \square

Dimostrazione (del teorema) Indicando con b la successione $a \circ k$, dobbiamo dimostrare che:

$$\forall \sigma > 0 \exists \mu \in \mathbb{N} \text{ t.c. } \forall m \in \mathbb{N} (m > \mu \Rightarrow |b_m - l| < \sigma).$$

Ovverossia, ricordando che $b_m = b(m) = (a \circ k)(m) = a(k(m)) = a_{k(m)} = a_{k_m}$,
che:

$$(\blacktriangleright) \forall \sigma > 0 \exists \mu \in \mathbb{N} \text{ t.c. } \forall m \in \mathbb{N} (m > \mu \Rightarrow |a_{k_m} - l| < \sigma).$$

Noi sappiamo che

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon).$$

Sia dunque $\sigma > 0$. Fissiamo $\varepsilon = \sigma$. Allora abbiamo che

$$(*) \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \sigma).$$

Scegliamo $\mu = \nu$, dove ν è un numero naturale tra quelli la cui esistenza è garantita da (*). Otteniamo che

$$(**) \forall n \in \mathbb{N} (n > \mu \Rightarrow |a_n - l| < \sigma).$$

Osserviamo poi che, se $m > \mu$, allora $k(m) > \mu$ (in quanto $k(m) \geq m$ per il lemma), cioè $k_m > \mu$. E quindi, grazie a (**), possiamo dire che

$$(***) \forall m \in \mathbb{N} (m > \mu \Rightarrow |a_{k_m} - l| < \sigma).$$

E con ciò abbiamo ottenuto quanto volevamo. Infatti abbiamo visto che, dato $\sigma > 0$, abbiamo trovato $\mu \in \mathbb{N}$ t.c. valga (***), il che è appunto quel che si voleva dimostrare, cioè la validità di (\blacktriangleright) . \square

Esercizio 1 Si considerino le successioni i cui primi elementi sono:

$$a_1 = 0, a_2 = 1/5, a_3 = 2/5, a_4 = 3/5, a_5 = 4/5, a_6 = 1, a_7 = 0, a_8 = 1/5, \dots$$

$$b_1 = 0, b_2 = 1, b_3 = 0, b_4 = 0, b_5 = 1, \dots$$

Questi primi elementi suggeriscono che vi siano semplici espressioni "compatte" per indicare a_n e b_n per ogni $n \in \mathbb{N}$. Trovarle.

La successione $(b_n)_{n \in \mathbb{N}}$ è una sottosuccessione estratta da $(a_n)_{n \in \mathbb{N}}$. Trovare la "regola di estrazione" $k: \mathbb{N} \longrightarrow \mathbb{N}$ t.c. $b = a \circ k$.

La regola di estrazione non è univocamente determinata. Trovarne un'altra $k': \mathbb{N} \longrightarrow \mathbb{N}$ diversa dalla k trovata prima. \square

7. Limiti infiniti

Abbiamo visto cosa vuol dire che una successione è convergente. Ci sono però anche altri "comportamenti di limite" che è interessante codificare. L'idea sottostante è che i termini di una successione tendano a diventare sempre più grandi. Cioè i termini a_n della successione superano ogni numero reale prefissato, pur di considerare indici n sufficientemente grandi. In tal caso diremo che la successione a tende a $+\infty$. Senza troppi giri di parole, vediamo subito la definizione.

Definizione 1 Sia data una successione a . Diciamo che si ha $\lim_{n \rightarrow \infty} a_n = +\infty$ se:

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow a_n > \varepsilon) . \square$$

Come già nel caso di successioni convergenti, useremo anche la notazione $a_n \xrightarrow[n \rightarrow \infty]{} +\infty$ o anche solo $a_n \longrightarrow +\infty$. Ovviamente possiamo modificare la definizione per adattarla al caso in cui i termini della successione "scendano sotto" ad ogni numero reale prefissato.

Definizione 2 Data una successione a , diciamo che $\lim_{n \rightarrow \infty} a_n = -\infty$ se:

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow a_n < -\varepsilon) . \square$$

Un'altra condizione interessante è la seguente:

Definizione 3 Data una successione a , diciamo che $\lim_{n \rightarrow \infty} a_n = \infty$ se:

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n| > \varepsilon) . \square$$

E' subito evidente, però, che la condizione espressa dalla definizione 3 è meno interessante delle precedenti. In effetti, si può facilmente provare che $a_n \longrightarrow \infty \Leftrightarrow |a_n| \longrightarrow +\infty$ (dove con $|a_n|$ intendiamo che stiamo considerando la successione il cui termine generale è, appunto, $|a_n|$). Per sottolineare ulteriormente questa condizione di inferiorità, a volte viene riservato il termine "regolare" alle successioni che sono convergenti o tendono a $+\infty$ oppure a $-\infty$. Non è insensata, questa discriminazione. In effetti, il teorema di unicità del limite può essere esteso dalle successioni convergenti alle successioni regolari, ma non a quelle che tendono a ∞ . La ragione è ovvia: se abbiamo $a_n \longrightarrow \infty$, non possiamo escludere che sia $a_n \longrightarrow +\infty$ oppure $a_n \longrightarrow -\infty$ (anche se non possiamo garantire che valga almeno una delle due condizioni: $a_n = (-1)^n \cdot n$ tende ad ∞ senza tendere né a $+\infty$ né a $-\infty$).

Teorema 1 (di unicità del limite in generale) Siano $l, l' \in \mathbb{R}$, con $l \neq l'$. Allora si può verificare al più una delle condizioni seguenti:

$$a_n \longrightarrow l \qquad a_n \longrightarrow l' \qquad a_n \longrightarrow +\infty \qquad a_n \longrightarrow -\infty \quad .\square$$

Dimostrazione Supponiamo $a_n \longrightarrow l$. Ovviamente, per il teorema 2.1 non può essere $a_n \longrightarrow l'$. D'altro canto, se $a_n \longrightarrow l$, a_n è limitata e quindi è evidente (lo provi il lettore!) che non può soddisfare né la condizione $a_n \longrightarrow +\infty$ né la condizione $a_n \longrightarrow -\infty$. Lascio al lettore la cura di completare la dimostrazione. \square

8. I limiti e le operazioni in \mathbb{R}

Questo paragrafo e il successivo sviluppano un tema molto importante e ricorrente, che cercherò di illustrare. Si tratta della "compatibilità" della definizione di limite rispetto alle operazioni (e, nel successivo paragrafo, la relazione d'ordine) che sono definite su \mathbb{R} .

Per spiegare cosa intendo, comincerò con un problema preciso. Supponiamo di avere due successioni, a e b . A partire da queste, possiamo costruire una terza successione usando l'operazione di somma. Cioè possiamo definire $c: \mathbb{N} \longrightarrow \mathbb{R}$ così: $c_n = a_n + b_n \quad \forall n \in \mathbb{N}$. Supponiamo per un momento che tutte queste tre successioni abbiano limite: $a_n \longrightarrow \alpha$, $b_n \longrightarrow \beta$ e $c_n \longrightarrow \gamma$. Ci si può aspettare ragionevolmente che $\gamma = \alpha + \beta$. D'altronde, basta fare un po' di esempi per farsi venire questa opinione, se uno non l'avesse⁶.

E' importante naturalmente enunciare precisamente tale risultato, nonché dimostrarlo. Ma rinvio a dopo tale compito. Preferisco mettere in rilievo l'importanza di questo risultato.

Un primo punto di vista lo presenta come risultato utile: ci dà informazioni dirette sul limite di una successione a partire dal limite delle due successioni "addende". Detto così forse non risulta chiara l'utilità, però messo assieme ai risultati analoghi relativi al prodotto permette di sapere rapidamente quanto fa il limite di una successione anche "complicata", senza richiedere di ricorrere alla definizione e ai conti laboriosi che essa richiede.

Un altro punto di vista è quello ricorrente della compatibilità tra le strutture già presenti su \mathbb{R} e questa nuova che è la struttura di tipo analitico, cioè di limite, che abbiamo introdotto. E' l'ennesima versione di un leitmotiv già evidenziato e che ricorrerà spesso. Il fatto di avere diverse strutture su un insieme diventa interessante solo nella misura in cui vi è una certa compatibilità tra di loro.

Infine, da un punto di vista "grafico-astratto", possiamo esprimere tale risultato come commutatività di un diagramma:

⁶ Va da sé che questa è una strategia del tutto generale: se non si ha idea di come vanno le cose, si cerca di farsela con qualche esempio.

$$\begin{array}{ccc}
 a_n, b_n & \xrightarrow{\text{LIM}} & \alpha, \beta \\
 \downarrow \text{SOMMA} & & \downarrow \text{SOMMA} \\
 a_n + b_n & \xrightarrow{\text{LIM}} & \alpha + \beta
 \end{array}$$

cioè per andare dall'angolo in alto a sinistra all'angolo in basso a destra abbiamo due strade equivalenti. Va da sé che è comodo sapere di poter scegliere tra due strade: si potrà usare la più comoda di volta in volta. In realtà, questo terzo punto di vista è in un certo senso la riproposizione del primo: entrambi si ricollegano ad un'idea di comodità, od economicità (di pensiero?) che dir si voglia.

Terminato lo "sproloquio" di carattere generale, passiamo ad enunciare il risultato preciso.

Teorema 1 Siano a, b due successioni, convergenti rispettivamente ad α, β . Allora la successione di termine generale $a_n + b_n$ è convergente e converge ad $\alpha + \beta$ □

Dimostrazione Abbiamo che:

$$\forall \varepsilon' > 0 \exists \nu' \in \mathbb{N} \text{ t.c. } \forall n' \in \mathbb{N} (n' > \nu' \Rightarrow |a_{n'} - \alpha| < \varepsilon')$$

$$\forall \varepsilon'' > 0 \exists \nu'' \in \mathbb{N} \text{ t.c. } \forall n'' \in \mathbb{N} (n'' > \nu'' \Rightarrow |b_{n''} - \beta| < \varepsilon'')$$

vogliamo dimostrare che:

$$(*) \quad \forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |(a_n + b_n) - (\alpha + \beta)| < \varepsilon)$$

Sia dunque dato un numero reale positivo ε .

Scegliamo $\varepsilon' = \varepsilon/2$. Allora:

$$\exists \nu' \in \mathbb{N} \text{ t.c. } \underbrace{\forall n' \in \mathbb{N} (n' > \nu' \Rightarrow |a_{n'} - \alpha| < \varepsilon/2)}_{(*)}$$

Scegliamo $\varepsilon'' = \varepsilon/2$. Allora:

$$\exists \nu'' \in \mathbb{N} \text{ t.c. } \underbrace{\forall n'' \in \mathbb{N} (n'' > \nu'' \Rightarrow |b_{n''} - \beta| < \varepsilon/2)}_{(**)}$$

Scegliamo un ν' ed un ν'' soddisfacenti rispettivamente le condizioni (*) e (**). Consideriamo $\nu = \max\{\nu', \nu''\}$. Allora, se $n > \nu$, è anche $n > \nu'$ ed $n > \nu''$, pertanto:

$$\forall n \in \mathbb{N} \left[n > \nu \Rightarrow \begin{cases} |a_n - \alpha| < \varepsilon/2 \\ |b_n - \beta| < \varepsilon/2 \end{cases} \right].$$

Ma da $\begin{cases} |a_n - \alpha| < \varepsilon/2 \\ |b_n - \beta| < \varepsilon/2 \end{cases}$ otteniamo, sommando membro a membro:

$|a_n - \alpha| + |b_n - \beta| < \varepsilon$. D'altronde:

$$\begin{aligned} |(a_n + b_n) - (\alpha + \beta)| &= && \text{riscrivendo in modo diverso} \\ = |(a_n - \alpha) + (b_n - \beta)| &\leq && \text{per la disuguaglianza triangolare} \\ \leq |a_n - \alpha| + |b_n - \beta| &< && \text{per quanto abbiamo appena visto} \\ &< \varepsilon \end{aligned}$$

Abbiamo pertanto ottenuto che, dato $\varepsilon > 0$, possiamo trovare $\nu \in \mathbb{N}$ t.c. per ogni $n > \nu$ si ha $|(a_n + b_n) - (\alpha + \beta)| < \varepsilon$. Abbiamo cioè provato (*). \square

Esercizio 1 Dimostrare che $\lim_{n \rightarrow \infty} a_n = \alpha$ implica $\lim_{n \rightarrow \infty} -a_n = -\alpha$. \square

La "compatibilità" tra limite ed operazioni si ha anche con la moltiplicazione.

Teorema 2 Siano a, b due successioni, convergenti rispettivamente ad α, β . Allora la successione di termine generale $a_n \cdot b_n$ è convergente e converge ad $\alpha \cdot \beta$. \square

Dimostrazione Abbiamo che:

$$\begin{aligned} \forall \varepsilon' > 0 \exists \nu' \in \mathbb{N} \text{ t.c. } \forall n' \in \mathbb{N} (n' > \nu' \Rightarrow |a_{n'} - \alpha| < \varepsilon') \\ \forall \varepsilon'' > 0 \exists \nu'' \in \mathbb{N} \text{ t.c. } \forall n'' \in \mathbb{N} (n'' > \nu'' \Rightarrow |b_{n''} - \beta| < \varepsilon'') \end{aligned}$$

vogliamo dimostrare che:

$$(*) \quad \forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |(a_n \cdot b_n) - (\alpha \cdot \beta)| < \varepsilon)$$

La chiave di tutto sta nella disequazione

$$(**) \quad |(a_n \cdot b_n) - (\alpha \cdot \beta)| < \varepsilon$$

Nel caso della somma, la maggiorazione $|(a_n - \alpha) + (b_n - \beta)| \leq |a_n - \alpha| + |b_n - \beta|$ era talmente ovvia che non mi è parso il caso di spendere parole di troppo. Nel caso del prodotto, può non essere ovvia la strada da seguire. Procediamo così: aggiun-
giamo e togliamo $a_n \cdot \beta$. Otteniamo $|a_n \cdot b_n - a_n \cdot \beta + a_n \cdot \beta - \alpha \cdot \beta|$. Usiamo il "raccolli-
mento a fattore comune" e vediamo così apparire i "pezzi" che ci interessano: $|a_n \cdot (b_n - \beta) + \beta \cdot (a_n - \alpha)|$. Naturalmente useremo la disuguaglianza triangolare, per cui cercheremo di avere $|a_n \cdot (b_n - \beta)| + |\beta \cdot (a_n - \alpha)| < \varepsilon$. A tale scopo sarà sufficiente⁷ avere $|a_n| \cdot |b_n - \beta| < \varepsilon/2$ e $|\beta| \cdot |a_n - \alpha| < \varepsilon/2$. Per avere la seconda disuguaglianza ci occorre $|a_n - \alpha| < (\varepsilon/2 \cdot |\beta|)$. Per la prima, ricordiamo che a è convergente e

⁷ Mi auguro che chi legge si sia accorto che sto seguendo una linea di ragiona-
mento simile a quella che ha concluso la dimostrazione del teorema 1.

quindi c'è A t.c. $|a_n| \leq A$ per ogni $n \in \mathbb{N}$, pertanto $|a_n| \cdot |b_n - \beta| \leq A \cdot |b_n - \beta|$ e quindi ci basterà $|b_n - \beta| < (\varepsilon/2 \cdot A)$.

Questi sono i pezzi principali, nuovi, che ci occorrono per provare il teorema: il resto consiste in una replica del teorema riguardante la somma. Vediamo comunque tutta la dimostrazione, dall'inizio alla fine.

Abbiamo che:

$$\forall \varepsilon' > 0 \quad \exists \nu' \in \mathbb{N} \quad \text{t.c.} \quad \forall n' \in \mathbb{N} \quad (n' > \nu' \Rightarrow |a_{n'} - \alpha| < \varepsilon')$$

$$\forall \varepsilon'' > 0 \quad \exists \nu'' \in \mathbb{N} \quad \text{t.c.} \quad \forall n'' \in \mathbb{N} \quad (n'' > \nu'' \Rightarrow |b_{n''} - \beta| < \varepsilon'')$$

vogliamo dimostrare che:

$$(*) \quad \forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \quad \text{t.c.} \quad \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow |(a_n \cdot b_n) - (\alpha \cdot \beta)| < \varepsilon)$$

Per quanto riguarda la relazione coinvolgente ε' , sia $B = \max\{|\beta|, 1\}$: questo piccolo trucco ci serve per avere che $B > 0$ e quindi poter dividere per B . Scegliamo $\varepsilon' = (\varepsilon/2 \cdot B)$. Abbiamo allora:

$$\exists \nu' \in \mathbb{N} \quad \text{t.c.} \quad \forall n' \in \mathbb{N} \quad (n' > \nu' \Rightarrow |a_{n'} - \alpha| < (\varepsilon/2 \cdot B))$$

e quindi possiamo affermare (si riguardino le considerazioni informali fatte prima) che:

$$(\blacktriangleright) \quad \exists \nu' \in \mathbb{N} \quad \text{t.c.} \quad \forall n' \in \mathbb{N} \quad (n' > \nu' \Rightarrow |\beta| \cdot |a_{n'} - \alpha| < \varepsilon/2)$$

Passiamo ora alla relazione riguardante ε'' . Osserviamo preliminarmente che, essendo la successione a convergente, $\exists C \in \mathbb{R}$ t.c. $\forall n \in \mathbb{N} \quad |a_n| \leq C$. E quindi anche $|a_n| \geq A$, se scegliamo $A = \max\{C, 1\}$: anche in questo caso facciamo così per garantire che $A > 0$. Pertanto, scelto $\varepsilon'' = (\varepsilon/2 \cdot A)$, abbiamo che:

$$\exists \nu'' \in \mathbb{N} \quad \text{t.c.} \quad \forall n'' \in \mathbb{N} \quad (n'' > \nu'' \Rightarrow |b_{n''} - \beta| < (\varepsilon/2 \cdot A))$$

e quindi, per le considerazioni informali fatte in precedenza:

$$(\blacktriangleright\blacktriangleright) \quad \exists \nu'' \in \mathbb{N} \quad \text{t.c.} \quad \forall n'' \in \mathbb{N} \quad (n'' > \nu'' \Rightarrow |a_{n''}| \cdot |b_{n''} - \beta| < \varepsilon/2)$$

Scelto $\nu = \max\{\nu', \nu''\}$ (dove ν' e ν'' sono due numeri naturali che soddisfano le condizioni (\blacktriangleright) e $(\blacktriangleright\blacktriangleright)$), abbiamo:

$$\forall n \in \mathbb{N} \quad \left[n > \nu \Rightarrow \begin{cases} |\beta| \cdot |a_n - \alpha| < \varepsilon/2 \\ |a_n| \cdot |b_n - \beta| < \varepsilon/2 \end{cases} \right].$$

Sommando membro a membro abbiamo:

$$\forall n \in \mathbb{N} \quad \left[n > \nu \Rightarrow |a_n \cdot (b_n - \beta)| + |\beta \cdot (a_n - \alpha)| < \varepsilon \right].$$

Abbiamo pertanto, dalla disuguaglianza triangolare:

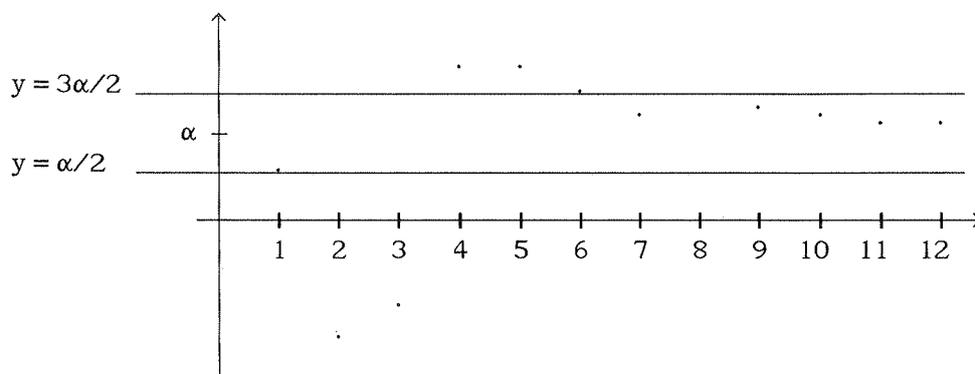
$$\forall n \in \mathbb{N} \left(n > \nu \Rightarrow |a_n \cdot (b_n - \beta) + \beta \cdot (a_n - \alpha)| < \varepsilon \right).$$

Possiamo quindi affermare che, dato $\varepsilon > 0$, $\exists \nu \in \mathbb{N}$ t.c. $\forall n \in \mathbb{N}$ si ha $|(a_n \cdot b_n) - (\alpha \cdot \beta)| < \varepsilon$. \square

In un esercizio precedente si chiedeva di provare che $a_n \longrightarrow \alpha \Rightarrow -a_n \longrightarrow -\alpha$. Cosa si può dire per il reciproco? Cioè, se $a_n \longrightarrow \alpha$, possiamo dire che $1/a_n \longrightarrow 1/\alpha$? Si vede subito che c'è un problema, perché se $\alpha = 0$, allora $1/\alpha$ non ha senso. Siamo quindi obbligati ad escludere questo caso (anche se poi troveremo qualcosa da dire lo stesso: vedi § 10). Si potrebbe quindi pensare di enunciare il seguente:

"Teorema" Sia a successione convergente e sia $\alpha \in \mathbb{R}$ il suo limite. Se $\alpha \neq 0$, allora $1/a_n \longrightarrow 1/\alpha$. \square

In realtà, l'enunciato di questo "teorema" solleva un problema. Chi ci garantisce che $a_n \neq 0$ e quindi che abbia senso fare $1/a_n$? Possiamo fare delle semplici considerazioni di carattere grafico:



Basta prendere una striscia di ampiezza $\varepsilon = |\alpha|/2$! E allora da un certo ν in poi (n, a_n) è obbligato a stare dentro la striscia e quindi non può certo essere $a_n = 0$, proprio per come abbiamo scelto ε . Non possiamo però garantire che tutti i termini siano diversi da zero: solo ad un certo indice in poi.

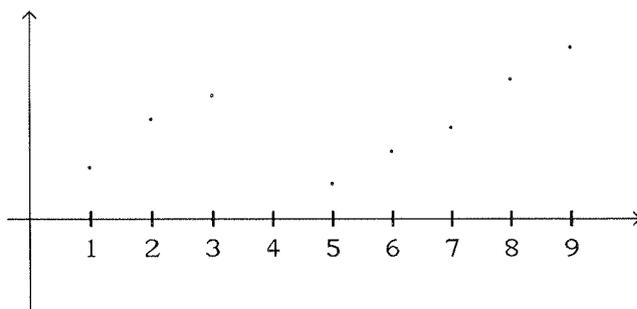
E allora?

Non possiamo far finta di niente. In matematica non si può nascondere la polvere sotto il tappeto. Occorre trovare una soluzione. La più semplice potrebbe essere quella di abbandonare il "teorema". Ma un tale risultato è troppo prezioso per lasciarlo cadere così. Potremmo cercare di trovare una buona idea per conservarlo, per riuscire cioè a dimostrarne una qualche utile (e vera!) versione.

C'è un modo molto drastico di procedere, ed è il seguente: ricominciare tutto

daccapo⁸. Addirittura dalla definizione stessa di successione. Perché? Perché nello svolgersi delle considerazioni è emerso che in fondo ci interessa poco cosa fanno i primi termini di una successione, per tutto quel che riguarda i limiti (ma non, per esempio, per trovare $\max a(N)$). Potremmo allora provare ad "allargare" la definizione di successione: richiedere che sia una applicazione definita su $\mathbb{N} \setminus F$, dove F è un sottoinsieme finito di \mathbb{N} .

Questo è l'approccio giusto. Ma offre delle difficoltà nuove. Per esempio, se ho una successione definita su $\mathbb{N} \setminus F'$ ed un'altra definita su $\mathbb{N} \setminus F''$, la loro somma sarebbe definita su $\mathbb{N} \setminus F$, dove $F = F' \cup F''$. Per dire che una successione è crescente, non v'è più bene dire che $a_n < a_{n+1} \quad \forall n \in \mathbb{N}$, e neanche $a_n < a_{n+1} \quad \forall n \in \mathbb{N} \setminus F$, come mostra l'esempio qui sotto (con $F = \{4\}$), ma occorre richiedere che $\forall n, m \in \mathbb{N} \setminus F \quad (n < m \Rightarrow a_n < a_m)$. E così via.



Quindi non adotterò questo punto di vista, ma continuerò a considerare una successione come una applicazione da \mathbb{N} in \mathbb{R} . Nel caso di $1/a_n$ (o in altri casi in cui si pongano problemi simili), definirò arbitrariamente $1/a_n = 0$ per $n \in F$. Naturalmente questo sarà lecito solo se questa operazione di maquillage non creerà problemi rispetto al contesto cui siamo interessati.

⁸ E' una delle cose che sono avvenute più di frequente nella storia della matematica (l'esempio più noto è quello delle geometrie non euclidee) e che capitano spesso nella ricerca matematica.

9. I limiti e la relazione d'ordine su \mathbb{R}

Il teorema più importante tra quelli di carattere elementare che riguardano la relazione tra limiti e \leq è il teorema di permanenza del segno.

Teorema 1 (di permanenza del segno) Sia $a: \mathbb{N} \rightarrow \mathbb{R}$, convergente ad $\alpha > 0$ (rispettivamente $\alpha < 0$). Allora $\exists \nu \in \mathbb{N}$ t.c. $\forall n > \nu$ è $a_n > 0$ (rispettivamente $a_n < 0$). \square

Dimostrazione E' una conseguenza diretta della definizione di limite. Abbiamo infatti:

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - \alpha| < \varepsilon)$$

se scegliamo $\varepsilon = \alpha/2$, abbiamo che:

$$\exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - \alpha| < \alpha/2)$$

ma $|a_n - \alpha| < \alpha/2 \Leftrightarrow -(\alpha/2) < a_n - \alpha < \alpha/2 \Leftrightarrow \alpha/2 < a_n < (3\alpha/2) \Rightarrow a_n > \alpha/2 > 0$. Quindi abbiamo ottenuto quello che volevamo. \square

Corollario 1 Sia $a: \mathbb{N} \rightarrow \mathbb{R}$, con $a_n \geq 0 \forall n \in \mathbb{N}$. Supponiamo che $a_n \rightarrow \alpha \in \mathbb{R}$. Allora $\alpha \geq 0$. \square

Dimostrazione E' una conseguenza del risultato precedente. Dimostro la contronominale. Supponiamo $\alpha < 0$. Allora per il teorema di permanenza del segno si ha che $\exists \nu \in \mathbb{N}$ t.c. $\forall n > \nu$ $a_n < 0$. Questo è ovviamente in contraddizione con l'ipotesi. \square

Corollario 2 Sia $a: \mathbb{N} \rightarrow \mathbb{R}$. Supponiamo $\exists \nu \in \mathbb{N}$ t.c. $\forall n > \nu$ $a_n \geq 0$. Se $a_n \rightarrow \alpha \in \mathbb{R}$, è $\alpha \geq 0$. \square

Dimostrazione In realtà, dimostrando il corollario 1, abbiamo dimostrato il corollario 2. \square

Ulteriore risultato che esprime il legame tra \leq e limiti è il teorema seguente.

Teorema 2 (del confronto, o dei due carabinieri) Siano date tre successioni a, b, c e supponiamo che sia $a_n \leq b_n \leq c_n \forall n \in \mathbb{N}$. Se $a_n \rightarrow l$ e $c_n \rightarrow l$, allora $b_n \rightarrow l$. \square

Dimostrazione Basta scrivere la definizione di limite. Abbiamo:

$$\forall \varepsilon' > 0 \exists \nu' \in \mathbb{N} \text{ t.c. } \forall n' \in \mathbb{N} (n' > \nu' \Rightarrow |a_{n'} - l| < \varepsilon')$$

$$\forall \varepsilon'' > 0 \exists \nu'' \in \mathbb{N} \text{ t.c. } \forall n'' \in \mathbb{N} (n'' > \nu'' \Rightarrow |c_{n''} - l| < \varepsilon'')$$

e da qui vogliamo ottenere:

$$\forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \quad \text{t.c.} \quad \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow |b_n - l| < \varepsilon) .$$

Sia quindi dato, come al solito, $\varepsilon > 0$. Prendiamo $\varepsilon' = \varepsilon'' = \varepsilon$. E consideriamo $\nu = \max \{ \nu', \nu'' \}$. Per $n > \nu$ abbiamo a disposizione: $|a_n - l| < \varepsilon$ e $|c_n - l| < \varepsilon$. Cioè:

$$\begin{cases} l - \varepsilon < a_n < l + \varepsilon \\ l - \varepsilon < c_n < l + \varepsilon \end{cases} . \quad \text{Ovverossia,} \quad a_n \quad \text{e} \quad c_n \quad \text{appartengono all'intervallo}$$

$]l - \varepsilon, l + \varepsilon[$. Ma allora anche b_n ci sta. \square

10. Miscellanea di risultati sui limiti di successioni

Proverò un paio di risultati che riguardano la relazione tra limiti e la struttura di \mathbb{R} , ma che non si collocano bene nello schema nitido dei due precedenti paragrafi. Sono anche dei risultati che richiedono uno sguardo più da vicino su quella che è la definizione di limite.

Teorema 1 Siano $a, b: \mathbb{N} \longrightarrow \mathbb{R}$ due successioni. Supponiamo che $a_n \longrightarrow +\infty$ e che b sia inferiormente limitata. Allora $a_n + b_n \longrightarrow +\infty$. \square

Dimostrazione Sappiamo che:

$$(*) \quad \forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow a_n > \varepsilon)$$

$$\exists H \in \mathbb{R} \text{ t.c. } \forall m \in \mathbb{N} \quad b_m \geq H$$

Naturalmente possiamo supporre che $H \leq 0$.

Dobbiamo dimostrare che:

$$\forall \varepsilon' > 0 \quad \exists \nu' \in \mathbb{N} \text{ t.c. } \forall n' \in \mathbb{N} \quad (n' > \nu' \Rightarrow a_{n'} + b_{n'} > \varepsilon')$$

L'idea è ovvia: dato ε' , prendiamo $\varepsilon = \varepsilon' - H$ (osserviamo che, avendo preso $H \leq 0$, è garantito che $\varepsilon > 0$). Ovviamente si prende $\nu' = \nu$ (più precisamente si dovrebbe dire che si sceglie come ν' uno degli indici ν la cui esistenza è garantita da (*)). E il gioco è fatto. \square

Corollario 1 $a_n \longrightarrow +\infty$ e $b_n \longrightarrow l \Rightarrow a_n + b_n \longrightarrow +\infty$. \square

Per introdurre il prossimo risultato dobbiamo introdurre prima una definizione. E' quella di limite " l^+ " o " l^- ".

Definizione 1 Data $a: \mathbb{N} \longrightarrow \mathbb{R}$ e dato $l \in \mathbb{R}$, diciamo che $\lim_{n \rightarrow \infty} a_n = l^+$ se:

$$\forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow l \leq a_n < l + \varepsilon) . \square$$

Cosa vuol dire che $a_n \longrightarrow l^+$? Vuol dire due cose:

- 1) $a_n \longrightarrow l$
- 2) $a_n \geq l$ tranne che per un numero finito di indici.

Quindi dire che $a_n \longrightarrow l^+$ ci dà qualche informazione in più rispetto a sapere che $a_n \longrightarrow l$. Cioè ci dice che i termini della successione si avvicinano a l "da destra" o, facendo riferimento al grafico, che la striscia compresa tra le rette $y = l - \varepsilon$ e $y = l$ contiene al più solo un numero finito di termini. Si noti, comunque, che $a_n \longrightarrow l^+$ non implica che a_n sia anche debolmente decrescente: l'esempio fatto nel primo paragrafo della pallina che viene fotografata

mentre rimbalza è un esempio di successione che tende a 0^+ senza essere decrescente.

Esercizio 1 Dare la definizione di $a_n \longrightarrow l^-$. \square

Teorema 2 Sia $a: \mathbb{N} \longrightarrow \mathbb{R}$ t.c. $a_n \longrightarrow 0^+$, con $a_n \neq 0 \quad \forall n \in \mathbb{N}$. Allora $(1/a_n) \longrightarrow +\infty$. \square

Dimostrazione L'idea della dimostrazione è questa: la relazione $0 < a_n < \varepsilon$ è equivalente a $(1/a_n) > (1/\varepsilon)$. Invito il lettore a provare a fare da solo la dimostrazione, prima di guardare le righe seguenti.

Dobbiamo passare da

$$(*) \quad \forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \quad \text{t.c.} \quad \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow 0 \leq a_n < 0 + \varepsilon)$$

a:

$$\forall \varepsilon' > 0 \quad \exists \nu' \in \mathbb{N} \quad \text{t.c.} \quad \forall n' \in \mathbb{N} \quad (n' > \nu' \Rightarrow (1/a_{n'}) > \varepsilon)$$

Pertanto, dato $\varepsilon' > 0$, basta prendere $\varepsilon = 1/\varepsilon'$ in (*): si ottiene (almeno) un ν t.c.:

$$\forall n \in \mathbb{N} \quad (n > \nu \Rightarrow 0 \leq a_n < 0 + \varepsilon),$$

ma per ipotesi è $a_n \neq 0 \quad \forall n \in \mathbb{N}$, quindi in realtà abbiamo:

$$\forall n \in \mathbb{N} \quad (n > \nu \Rightarrow 0 < a_n < 0 + \varepsilon).$$

E quindi scegliamo $\nu' = \nu$. E abbiamo il risultato desiderato, utilizzando l'osservazione iniziale. \square

Esercizio 2 Dimostrare che $a_n \longrightarrow l^+ \Rightarrow -a_n \longrightarrow -l^-$. \square

Esercizio 3 Dimostrare che, se $a_n \longrightarrow 0$ ed $a_n \neq 0 \quad \forall n \in \mathbb{N}$, allora $(1/a_n) \longrightarrow \infty$. \square

Esercizio 4 Dimostrare che, se $a_n \longrightarrow \infty$, allora $(1/a_n) \longrightarrow 0$. \square

Esercizio 5 Dimostrare che, se $a_n \longrightarrow +\infty$ e $b_n \longrightarrow +\infty$, allora $a_n + b_n \longrightarrow +\infty$. \square

11. Forme indeterminate

Abbiamo visto che, se $a_n > 0 \quad \forall n \in \mathbb{N}$, da $a_n \rightarrow 0^+$ possiamo dedurre che $(1/a_n) \rightarrow +\infty$. Questo risultato molto spesso viene sintetizzato con l'espressione $1/0^+ = +\infty$. Quindi tale espressione è come una sorta di resoconto "steno-grafico" dell'enunciato del teorema o, meglio, è un artificio mnemonico per ricordarsi tale teorema. Lo stesso dicasi della scrittura $1/0 = \infty$. Si usa anche dire che "uno fratto zero più fa più infinito". Queste espressioni possono però essere fonte di pericolosi malintesi. Scrivendo $1/0 = \infty$, non si intende dire che il reciproco di 0 c'è ed è ∞ . Non può essere così, per la semplice ragione che 0 non ha reciproco. Oltretutto, " ∞ " non è un simbolo che indica un numero reale. Quindi scrivendo $1/0 = \infty$, non si intende fare dei conti con dei numeri reali ma, ripeto, solo sintetizzare molto rapidamente il contenuto di un teorema. Lo stesso significato hanno scritture come $(+\infty)+(+\infty) = +\infty$, oppure $l(+\infty) = +\infty$, etc.

Si usa dire, invece, che ad esempio $(+\infty)+(-\infty)$ è una "forma indeterminata" (più comunemente si usa la notazione $\infty-\infty$, in questo caso). Cosa vuol dire? Significa che, se sappiamo solo che $a_n \rightarrow +\infty$ e $b_n \rightarrow -\infty$, il comportamento di limite della somma non è determinato. Può addirittura non avere limite.

Esempio 1

$$\begin{aligned} a_n = n \text{ e } b_n = -n & : a_n + b_n = 0 \rightarrow 0 \\ a'_n = 2n \text{ e } b'_n = -n & : a'_n + b'_n = n \rightarrow +\infty \\ a''_n = n \text{ e } b''_n = -2n & : a''_n + b''_n = -n \rightarrow -\infty \\ \hat{a}_n = n + (-1)^n \text{ e } \hat{b}_n = -n & : \hat{a}_n + \hat{b}_n = (-1)^n \text{ non ha limite. } \square \end{aligned}$$

Sono forme indeterminate, oltre a $\infty-\infty$, anche $0 \cdot \infty$, $0/0$, ∞/∞ e 1^∞ . In realtà, quando scrivo $\infty-\infty$, intendo usare una scrittura abbreviata per riferirmi a vari casi possibili: $(+\infty)-(+\infty)$ come in precedenza, ma anche $(+\infty)+(-\infty)$, $(+\infty)+(\infty)$, etc., oltre naturalmente proprio a $(\infty)-(\infty)$. Chiaramente lo stesso discorso vale anche per le altre forme indeterminate.

Vanno evitati due malintesi, che spesso si verificano quando si parla di forme indeterminate.

Il primo malinteso consiste nel ritenere che, se capita che $a_n \rightarrow +\infty$ e $b_n \rightarrow -\infty$, non si possa dire nulla sul limite di $a_n + b_n$. Questo non è vero. Dipende dalla informazione complessiva che abbiamo a disposizione. Ripeto ancora: se sappiamo solo che $a_n \rightarrow +\infty$ e $b_n \rightarrow -\infty$, allora proprio non possiamo dire nulla. Ma se conosciamo qualcosa di più, soprattutto se abbiamo a disposizione,

come spesso capita, l'espressione esplicita delle due successioni a e b , possiamo utilizzare queste informazioni supplementari per cercare di capire cosa fa $a_n + b_n$. Anzi, nell'esempio precedente, non c'era molto da lavorare per decidere cosa fa $a_n + b_n$. Possono però capitare situazioni intermedie, come:

Esempio 2 Sia $a_n = n^2$, $b_n = -n$. Abbiamo $a_n + b_n = n^2 - n$ e non è immediato capire cosa faccia. Però possiamo scrivere $n^2 - n = n(n-1)$ e quindi dedurre che $n^2 - n \longrightarrow +\infty$. \square

In tal caso, pochissimi conti ci hanno permesso di dirimere la questione di cosa facesse $a_n + b_n$, usando un risultato sul prodotto di successioni, senza bisogno di fare ricorso alla definizione di limite (evitando così i conti pesanti che essa usualmente richiede). Quindi la forma indeterminata $(+\infty) - (+\infty)$ non è necessariamente destinata a rimanere tale. Lo è solo se " $a_n \longrightarrow +\infty$ e $b_n \longrightarrow -\infty$ " è l'unica informazione che abbiamo a disposizione.

Il secondo malinteso consiste nel fare confusione con le operazioni tra numeri reali. Questo capita in particolare di fronte alla forma indeterminata $0/0$. Cosa significa questa forma indeterminata? Significa che, se ho due successioni a e b e so che $a_n \longrightarrow 0$ e $b_n \longrightarrow 0$, da questo non posso dedurre alcunché sul limite della successione a_n/b_n (ammesso che sia definita!). Quindi, quando si scrive $0/0$, si intende solo dire questo fatto. Niente di più. Non si sta cercando di trovare il risultato della divisione di 0 per 0 ! Nessun numero è divisibile per 0 , come già notato in precedenza. Ma, a parte questo, come detto non si sta cercando di dividere 0 per 0 . Si sta solo dicendo che "da $a_n \longrightarrow 0$ e $b_n \longrightarrow 0$ non posso dedurre quale sia il comportamento di limite della successione a_n/b_n ".

Spero di avere messo in luce a sufficienza questi due potenziali modi erronei di approcciarsi alle forme indeterminate.

12. Teoremi sulle successioni che richiedono la completezza di \mathbb{R}

In questo paragrafo finale sulle successioni ho raggruppato quei risultati relativi alle successioni che richiedono l'assioma di completezza. In effetti, tutti i teoremi relativi alle successioni di numeri reali avrebbero potuto essere enunciati e dimostrati per le successioni di numeri razionali: non è stato mai usato infatti l'assioma n. 16 o il suo equivalente n. 17. Attenzione: ho usato la proprietà di completezza nell'esempio 1.1, perché ho sfruttato la proprietà di Archimede che avevo provato servendomi dell'assioma 17 nel § 1.8. Ribadisco, però, che non ho mai usato la proprietà di completezza per la dimostrazione dei teoremi.

I risultati che proveremo sono:

- il teorema sulle successioni monotone
- il teorema di Bolzano-Weierstrass
- la convergenza delle successioni di Cauchy

Il primo risultato è interessante di per sé, ma soprattutto è utile perché permette di ottenere ulteriori teoremi: il teorema di Bolzano-Weierstrass (e quindi anche la convergenza delle successioni di Cauchy, come vedremo) e, più avanti, il teorema degli zeri e la "regolarità" delle serie a termini positivi. Viene ad assumere in un certo senso il ruolo di capostipite per tutti i risultati che usano in modo essenziale la completezza di \mathbb{R} . Visto che tutta la fatica che si fa per passare dai razionali ai reali è giustificata dall'avere questa proprietà a disposizione (e le sue conseguenze!), il teorema sulle successioni monotone ha una importanza che è difficile sopravvalutare. In realtà anche gli altri due sono altrettanto significativi: la posizione di privilegio che viene ad assumere il teorema sulle successioni monotone è solo dovuta al particolare percorso argomentativo che ho scelto. Si potrebbe per esempio provare come primo risultato la convergenza delle successioni di Cauchy e da questo ricavare gli altri due. Anzi, è proprio questo di solito il teorema più "considerato", più che altro perché si presta a generalizzazioni molto ampie, in contesti ancora più astratti di quello nel quale stiamo lavorando (a dire il vero, però, neanche il teorema di Bolzano-Weierstrass scherza in quanto a possibilità di generalizzazioni).

Prima di enunciare e dimostrare questi tre teoremi, anticipo che alla fine di questo paragrafo il lettore troverà un esempio di successione in \mathbb{Q} che non "ubbidisce" a nessuno di questi tre teoremi.

Teorema 1 (delle successioni monotone) Sia $a: \mathbb{N} \longrightarrow \mathbb{R}$, debolmente crescente. Allora, se a è superiormente limitata, è convergente e il suo limite è $\sup a(\mathbb{N})$; se a non è superiormente limitata, $a_n \longrightarrow +\infty$. \square

Dimostrazione Supponiamo a superiormente limitata. Consideriamo $a(\mathbb{N})$: esso è non vuoto e superiormente limitato, quindi c'è $\lambda = \sup a(\mathbb{N})$. Sia $\varepsilon > 0$. Poiché λ è il più piccolo dei maggioranti, $\lambda - \varepsilon$ non è un maggiorante per $a(\mathbb{N})$. Quindi $\exists \alpha \in a(\mathbb{N})$ t.c. $\lambda - \varepsilon < \alpha$. Ma se $\alpha \in a(\mathbb{N})$, $\exists \nu \in \mathbb{N}$ t.c. $\alpha = a_\nu$. Quindi abbiamo che $\exists \nu \in \mathbb{N}$ t.c. $\lambda - \varepsilon < a_\nu$. Essendo λ un maggiorante, d'altro canto è $a_n \leq \lambda \quad \forall n \in \mathbb{N}$.

Riassumendo, si ha:

$$\left\{ \begin{array}{l} \exists \nu \in \mathbb{N} \text{ t.c. } \lambda - \varepsilon < a_\nu \\ \forall n \in \mathbb{N} \quad a_n \leq \lambda \end{array} \right. .$$

Ma a è debolmente crescente. Quindi $\lambda - \varepsilon < a_\nu \leq a_n \quad \forall n \geq \nu$. Pertanto posso dire che

$$\exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} \quad (n \geq \nu \Rightarrow \lambda - \varepsilon < a_n \leq \lambda)$$

e quindi

$$\exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} \quad (n \geq \nu \Rightarrow |a_n - \lambda| < \varepsilon).$$

Con ciò abbiamo provato che $a_n \longrightarrow \lambda$.

Vediamo l'altro caso. Cioè supponiamo che $a(\mathbb{N})$ non sia superiormente limitato. Allora, dato $\varepsilon > 0$, $\exists \nu \in \mathbb{N}$ t.c. $a_\nu > \varepsilon$. Per la debole crescita, posso dire che

$$\exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} \quad (n \geq \nu \Rightarrow a_n > \varepsilon).$$

E con ciò abbiamo dimostrato che $a_n \longrightarrow +\infty$. \square

Ovviamente c'è una versione "speculare" di questo teorema che riguarda le successioni debolmente decrescenti. Si può anche notare che per garantire la convergenza della successione è sufficiente che essa sia crescente solo da un certo indice in poi (ma in questo caso non posso più garantire che $a_n \longrightarrow \sup a(\mathbb{N})$ quando a è superiormente limitata).

Passiamo ora al teorema di Bolzano-Weierstrass.

Immaginate di avere a disposizione un tavolo e di dover sistemare su di esso un certo numero di oggetti identici tra loro, per esempio gomme da cancellare. Se le gomme sono poche, riuscirete a sistemarle ben distanziate tra loro, se così vi fa piacere. Immaginate che continuino a passarvi sempre più gomme. Arriverà il momento in cui da qualche parte le dovrete sovrapporre, e poi ammucciarle, via via

che il loro numero aumenta. Provate a pensare cosa può succedere se il loro numero aumenta indefinitamente. Qualunque strategia di posizionamento delle gomme voi decidiate, ci sarà sempre qualche zona del tavolo nella quale l'altezza raggiunta dalle gomme raggiungerà il soffitto. E poi anche il soffitto sarà superato, e così dicasi per ogni altezza prefissata...

Questo problema "concreto", anche se stupido, rende un pò conto dell'intuizione che c'è dietro al teorema di Bolzano-Weierstrass.

Supponiamo di avere una successione a limitata. Quindi tutti i termini della successione stanno in un certo intervallo $[-L, L]$. Ebbene, ci sarà qualche punto dell'intervallo $[-L, L]$ vicino al quale "cadranno" termini a_n della successione per infiniti valori dell'indice n . Questo è ovvio se la successione è convergente, ma vale anche per la successione $(-1)^n$, o per qualsiasi successione che possiate concepire, purché sia limitata. Questo fatto si può esprimere così: dalla successione a_n riusciamo ad estrarre una sottosuccessione a_{k_n} che converge a qualche punto dell'intervallo $[-L, L]$. Ed è proprio questo il teorema di Bolzano-Weierstrass.

Teorema 2 Da ogni successione limitata si può estrarre una sottosuccessione convergente. \square

Dimostrazione Sia $a: \mathbb{N} \rightarrow \mathbb{R}$, limitata. Cioè $\exists L > 0$ t.c. $a_n \in [-L, L]$ $\forall n \in \mathbb{N}$. Utilizzeremo il metodo di bisezione per individuare la sottosuccessione convergente. Per seguire meglio il ragionamento che segue, consiglio di aiutarsi con un disegno per rappresentarne le varie fasi.

Dividiamo $[-L, L]$ in due parti. Abbiamo due intervalli $[-L, 0]$ e $[0, L]$ che, per ragioni di comodità di notazione evidenti nel seguito, indicherò rispettivamente con $[\alpha'_1, \beta'_1]$ e $[\alpha''_1, \beta''_1]$. Indicherò con $N'_1 = \{ n \in \mathbb{N} \text{ t.c. } a_n \in [\alpha'_1, \beta'_1] \}$ e con $N''_1 = \{ n \in \mathbb{N} \text{ t.c. } a_n \in [\alpha''_1, \beta''_1] \}$. Osservo che $[\alpha'_1, \beta'_1] \cup [\alpha''_1, \beta''_1] = [-L, L]$ e pertanto $N'_1 \cup N''_1 = \mathbb{N}$. Poiché \mathbb{N} è infinito, uno dei due insiemi N'_1 o N''_1 sarà anch'esso infinito. Sceglierò l'insieme infinito (se lo sono entrambi, prenderò N'_1) e lo chiamerò N_1 . Analogamente indicherò con $[\alpha_1, \beta_1]$ l'intervallo corrispondente all'insieme scelto. Infine, sceglierò a piacere un elemento di N_1 e lo indicherò con $k(1)$. Non bisogna dimenticare che dobbiamo individuare una sottosuccessione e quindi dobbiamo individuare una applicazione $k: \mathbb{N} \rightarrow \mathbb{N}$.

Per ora ho individuato il valore di k in $1 \in \mathbb{N}$. Procederò ora iterando la

stessa procedura (ma con un indispensabile accorgimento in più) a partire da $[\alpha_1, \beta_1]$. Divido $[\alpha_1, \beta_1]$ a metà e chiamo $[\alpha'_2, \beta'_2]$ l'intervallo $[\alpha_1, (\alpha_1 + \beta_1)/2]$ e con $[\alpha''_2, \beta''_2]$ l'intervallo $[(\alpha_1 + \beta_1)/2, \beta_1]$. Definisco $N'_2 = \{ n \in \mathbb{N} \text{ t.c. } a_n \in [\alpha'_2, \beta'_2] \}$ ed $N''_2 = \{ n \in \mathbb{N} \text{ t.c. } a_n \in [\alpha''_2, \beta''_2] \}$. Anche qui $[\alpha'_2, \beta'_2] \cup [\alpha''_2, \beta''_2] = N_1$. Poiché N_1 ha infiniti elementi, almeno uno tra N'_2 e N''_2 è infinito e sceglierò di indicare con N_2 appunto quello che ha infiniti elementi (come prima, se entrambi hanno infiniti elementi, prenderò $N_2 = N'_2$). Indicherò con $[\alpha_2, \beta_2]$ l'intervallo corrispondente ad N_2 . Infine, sceglierò un elemento di N_2 che sia più grande di $k(1)$ e lo indicherò con $k(2)$. Quello sottolineato è l'accorgimento la cui presenza avevo anticipato. E' ovvio che è possibile soddisfare tale requisito in quanto N_1 è infinito e pertanto conterrà dei numeri naturali più grandi di $k(1)$.

Procedendo in questo modo ottengo una successione di intervalli $[\alpha_n, \beta_n]$ e una applicazione $k: \mathbb{N} \rightarrow \mathbb{N}$ che è strettamente crescente (sempre per quell'accorgimento).

Proverò ora che α_n e β_n sono entrambe convergenti e convergono allo stesso limite. Intanto, poiché $[\alpha_n, \beta_n] \supseteq [\alpha_{n+1}, \beta_{n+1}] \quad \forall n \in \mathbb{N}$, ho che la successione α_n è debolmente crescente e la successione β_n è debolmente decrescente. Poiché $[\alpha_n, \beta_n] \subseteq [-L, L] \quad \forall n \in \mathbb{N}$, la successione α_n è superiormente limitata da L e β_n è inferiormente limitata da $-L$. Pertanto, il teorema sulle successioni monotone ci dice che entrambe convergono: $\exists \alpha, \beta \in \mathbb{R}$ t.c. $\alpha_n \rightarrow \alpha$ e $\beta_n \rightarrow \beta$. Si tratta di dimostrare che $\alpha = \beta$. A tale scopo, osserviamo che $\beta_n = \alpha_n + (\beta_n - \alpha_n)$. So che $\alpha_n \rightarrow \alpha$, mentre $\beta_n - \alpha_n = (2L)/2^n$ (poiché abbiamo usato la procedura di bisezione). Quindi $\beta_n - \alpha_n \rightarrow 0$. Ma allora $\beta_n \rightarrow \alpha$ per il teorema sul limite di una somma. Possiamo così garantire, per l'unicità del limite, che $\alpha = \beta$. Essendo $\alpha \leq L$ e $-L \leq \beta$, abbiamo anche che $\alpha \in [-L, L]$.

A questo punto, la conclusione è in vista. Osserviamo infatti che, per come abbiamo definito k , la sottosuccessione $a \circ k$ soddisfa la condizione $a_{k_n} \in [\alpha_n, \beta_n] \quad \forall n \in \mathbb{N}$. Detto altrimenti, abbiamo che $\alpha_n \leq a_{k_n} \leq \beta_n \quad \forall n \in \mathbb{N}$. Per quanto abbiamo appena visto, il teorema dei due carabinieri ci garantisce la convergenza di $a \circ k$ ad $\alpha \in [-L, L]$. \square

L'ultimo risultato riguarda le successioni "di Cauchy", ovvero quelle che soddisfano la condizione di Cauchy. L'idea base dietro a questo è abbastanza sem-

plice: se abbiamo una successione convergente, cioè se i termini a_n si avvicinano a un numero reale l , allora i termini dovrebbero avvicinarsi tra loro. La condizione di Cauchy esprime appunto questa idea in modo preciso.

Definizione 1 Diremo che una successione a soddisfa la condizione di Cauchy se:

$$\forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n, m \in \mathbb{N} (n, m > \nu \Rightarrow |a_n - a_m| < \varepsilon) . \square$$

E' del tutto immediato che una successione convergente soddisfa la condizione di Cauchy. E' un po' più difficile provare che vale il viceversa: occorre usare la completezza di \mathbb{R} , come già anticipato. Vale comunque la pena di farlo, perché la condizione di Cauchy ha un aspetto particolarmente interessante che intendo sottolineare, confrontandolo con la definizione di successione convergente: non abbiamo bisogno di conoscere a priori il limite l . Cioè: per dire che una successione è convergente, bisogna trovare in qualche modo un numero reale l che soddisfi la nota condizione. Per verificare invece la validità della condizione di Cauchy, ci basta avere a disposizione la successione a : se la condizione di Cauchy è soddisfatta, siamo sicuri (grazie al teorema seguente) che un limite l ci sarà.

Vediamo quindi il seguente:

Teorema 3 (criterio di convergenza di Cauchy) Una successione è convergente se e solo se soddisfa la condizione di Cauchy. \square

Dimostrazione E' abbastanza facile provare il "solo se". Infatti, se a è convergente:

$$\exists l \in \mathbb{R} \text{ t.c. } \forall \varepsilon > 0 \exists \nu \in \mathbb{N} \text{ t.c. } \forall n \in \mathbb{N} (n > \nu \Rightarrow |a_n - l| < \varepsilon)$$

vediamo che è di Cauchy, cioè che:

$$\forall \varepsilon' > 0 \exists \nu' \in \mathbb{N} \text{ t.c. } \forall n', m' \in \mathbb{N} (n', m' > \nu' \Rightarrow |a_{n'} - a_{m'}| < \varepsilon')$$

Dato ε' , prendiamo $\varepsilon = \varepsilon'/2$ e da qui abbiamo a disposizione un ν che possiamo utilizzare anche come ν' : la disuguaglianza triangolare fa il resto, essendo:

$$|a_n - a_m| = |a_n - l + l - a_m| \leq |a_n - l| + |a_m - l| .$$

Il "se" è un po' più laborioso. Farò una dimostrazione divisa in tre parti:

- 1) una successione di Cauchy è limitata
- 2) da una successione limitata posso estrarre una sottosuccessione convergente
- 3) se una sottosuccessione di una successione di Cauchy è convergente, allora converge tutta la successione.

La prima parte si prova facilmente, con una piccola rielaborazione della dimostrazione in base alla quale una successione convergente è limitata. Abbiamo:

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

Scegliamo $\varepsilon = 1$, ed $n = \nu + 1$. Otteniamo che $a_{\nu+1} - \varepsilon < a_m < a_{\nu+1} + \varepsilon$ per ogni $m > \nu$. Abbiamo ottenuto che la successione è limitata da un certo indice in poi. Quindi tutta la successione è limitata. (Se qualcuno ha dei dubbi su come si possa procedere, può rivedersi la dimostrazione del teorema 5.1).

La seconda parte è naturalmente immediata, perché è esattamente l'enunciato del teorema di Bolzano-Weierstrass.

Resta la terza parte, che è nuova ed è un po' delicata. Prima di ingaggiare la solita lotta con ε e ν , vorrei osservare che l'idea è ovvia. Se una sottosuccessione converge ad un certo l , ciò vuol dire che i suoi termini si "avvicinano" ad l . La condizione di Cauchy a sua volta garantisce che tutti i termini si avvicinino tra loro, e quindi anche ai termini della sottosuccessione, pertanto tutti i termini si avvicineranno ad l .

Indicheremo come di consueto con k la regola di estrazione, per cui abbiamo che $\exists l \in \mathbb{R}$ t.c. $a_{k_n} \xrightarrow{n \rightarrow \infty} l$. Cioè:

$$\forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \quad \text{t.c.} \quad \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow |a_{k_n} - l| < \varepsilon).$$

Sappiamo anche che:

$$\forall \varepsilon' > 0 \quad \exists \nu' \in \mathbb{N} \quad \text{t.c.} \quad \forall n', m' \in \mathbb{N} \quad (n', m' > \nu' \Rightarrow |a_{n'} - a_{m'}| < \varepsilon').$$

Vogliamo ottenere che:

$$\forall \varepsilon'' > 0 \quad \exists \nu'' \in \mathbb{N} \quad \text{t.c.} \quad \forall n'' \in \mathbb{N} \quad (n'' > \nu'' \Rightarrow |a_{n''} - l| < \varepsilon'').$$

Sia dunque ε'' dato. Prendiamo $\varepsilon = \varepsilon' = \varepsilon''/2$. Abbiamo a disposizione

$$\left\{ \begin{array}{l} \nu \in \mathbb{N} \quad \text{t.c.} \quad \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow |a_{k_n} - l| < \varepsilon''/2) \\ \nu' \in \mathbb{N} \quad \text{t.c.} \quad \forall n', m' \in \mathbb{N} \quad (n', m' > \nu' \Rightarrow |a_{n'} - a_{m'}| < \varepsilon''/2) \end{array} \right.$$

Definiamo $\nu'' = \max\{\nu, \nu'\}$. Allora per tale ν'' si ha:

$$\left\{ \begin{array}{l} n > \nu'' \Rightarrow |a_{k_n} - l| < \varepsilon''/2 \\ n', m' > \nu'' \Rightarrow |a_{n'} - a_{m'}| < \varepsilon''/2 \end{array} \right.$$

Sia $n'' > \nu''$. Scegliamo $m' = k_{n''}$ (notare che $k_{n''} \geq n''$ per il lemma 6.1 e quindi $k_{n''} > \nu''$). Prendiamo inoltre $n = n''$ ed $n' = n''$. Abbiamo:

$$|a_{k_{n''}} - l| < \varepsilon''/2 \quad \text{e} \quad |a_{n''} - a_{k_{n''}}| < \varepsilon''/2.$$

Si noti che nella relazione precedente si trova espresso in modo preciso quanto detto prima in termini intuitivi: e cioè che i termini della sottosucce-

sione sono "vicini" ad l ($|a_{k_n} - l| < \varepsilon/2$) e che i termini della successione, essendo tra loro "vicini", lo sono anche a quelli della sottosuccessione ($|a_n - a_{k_n}| < \varepsilon/2$).

Allora abbiamo:

$$|a_n - l| = |a_n - a_{k_n} + a_{k_n} - l| \leq |a_n - a_{k_n}| + |a_{k_n} - l| < \varepsilon \quad \square$$

Dopo aver visto la dimostrazione del criterio di convergenza di Cauchy, vorrei osservare un fatto. Quando cercavo di presentare questo criterio, ho messo in connessione il fatto che una successione sia convergente col fatto che i termini della successione si avvicinino tra loro. Può darsi che qualcuno, nel cercare di precisare cosa significhi che "i termini della successione si avvicinano tra loro", abbia pensato al fatto che la distanza tra un termine ed il successivo va a zero. Cioè, che $a_{n+1} - a_n \rightarrow 0$. Notiamo subito che, se a converge, allora $a_{n+1} - a_n \rightarrow 0$ (la successione a_{n+1} è una sottosuccessione estratta da a_n , con $k(n) = n+1$, e quindi se $a_n \rightarrow l$, anche $a_{n+1} \rightarrow l$, pertanto $a_{n+1} - a_n \rightarrow 0$). Quindi questa si presenta come una sensatissima condizione tramite la quale esprimere la convergenza di una successione. Oltretutto più semplice, più familiare di quella espressa dalla condizione di Cauchy. Ma, contrariamente alle apparenze, la condizione $a_{n+1} - a_n \rightarrow 0$ non garantisce la convergenza della successione a . Si potrebbe anche fare un esempio, ma preferisco rinviarlo a un momento più appropriato, cioè a quando parleremo di serie.

Ora che abbiamo visto l'uso della completezza di \mathbb{R} , vediamo l'esempio che avevo promesso: una successione in \mathbb{Q} che non rispetta l'enunciato di questi "grossi" teoremi.

L'esempio è facilissimo. E' la successione delle approssimazioni decimali per difetto di $\sqrt{2}$ (ma qualunque numero irrazionale va bene al posto di $\sqrt{2}$). Più precisamente, con a_n intendo indicare il numero decimale finito che approssima meglio $\sqrt{2}$ per difetto tra tutti quelli che hanno n cifre decimali. E quindi $a_1 = 1.4$, $a_2 = 1.41$, $a_3 = 1.414$, $a_4 = 1.4142$, E' ovviamente una successione monotona ma non ha limite in \mathbb{Q} (se lo avesse, avrei un esempio di successione in \mathbb{R} con due limiti diversi, visto che convergerebbe sia a $\sqrt{2}$ che ad un numero razionale). Poiché è convergente a $\sqrt{2}$, ogni sua sottosuccessione estratta converge anch'essa a $\sqrt{2}$: quindi non posso trovare una sottosuccessione convergente a un numero razionale. Mentre ovviamente è una successione limitata. Infine,

soddisfa la condizione di Cauchy, in quanto in \mathbb{R} è convergente. Eppure non converge in \mathbb{Q} .

13. Conclusioni

In questo capitolo abbiamo incontrato per la prima volta il concetto di limite. Lo si è visto nel contesto più semplice possibile, cioè quello delle successioni. Si è poi passati ad esaminare le principali proprietà elementari dei limiti di successioni. Il primo risultato importante è naturalmente l'unicità del limite. Ma non meno significativi sono i teoremi sul limite della somma e del prodotto di successioni, nonché il teorema di permanenza del segno; essi garantiscono la "compatibilità" tra il concetto di limite e la struttura dei numeri reali. Al di là di varie altre considerazioni presenti in questo capitolo (richiamo l'attenzione in particolare sui paragrafi 2 e 3, nonché 11), un posto di particolare riguardo è occupato dal precedente paragrafo. Vengono qui dimostrati tre risultati la cui portata è difficile sopravvalutare; non a caso, sono accomunati dalla caratteristica di richiedere in modo essenziale l'assioma di completezza. Vedremo in seguito fin dove essi ci permetteranno di giungere: posso anticipare che grazie a questi risultati potremo dimostrare un paio di proprietà delle funzioni continue. L'aspetto curioso della faccenda è che queste proprietà, per la cui dimostrazione è necessario un macchinario così sofisticato, sono proprietà di significato intuitivo immediato! Rinvio naturalmente al capitolo apposito (il III) per considerazioni più dettagliate.

CAPITOLO III

FUNZIONI REALI DI VARIABILE REALE: LIMITI E CONTINUITÀ

1. Funzioni reali di variabile reale

Questo capitolo si occupa di funzioni reali di variabile reale, per quanto riguarda i limiti e la continuità.

Che cosa è una funzione reale di variabile reale? Molto semplice:

Definizione 1 Una applicazione $f:A \longrightarrow B$, dove $\emptyset \neq A \subseteq \mathbb{R}$ e $B = \mathbb{R}$, si dice funzione reale di variabile reale. \square

Osservazione 1 La condizione $A \neq \emptyset$ non verrà d'ora in poi ricordata continuamente, però è da intendersi sempre presente, salvo esplicito avviso contrario. \square

Osservazione 2 Una successione è un caso particolare di funzione reale di variabile reale. \square

Proprio a partire da questa osservazione, ripercorriamo rapidamente le note introduttive del capitolo II: una funzione reale di variabile reale si dirà limitata se $f(A)$ è limitato, $\sup f(A)$ si dirà \sup di f , diremo che f ha max se $f(A)$ ha max, etc. Se $x_0 \in A$ è t.c. $f(x_0) = \max f(A)$, ovvero sia $f(x_0) \geq f(x) \quad \forall x \in A$, diremo che x_0 è punto di massimo per f ; analogamente per il minimo. Anche qui abbiamo poi la definizione di f crescente:

Definizione 2 $f:A \longrightarrow \mathbb{R}$, con $A \subseteq \mathbb{R}$, si dirà

debolmente crescente se $\forall x,y \in A \quad (x < y \Rightarrow f(x) \leq f(y))$

strettamente crescente se $\forall x,y \in A \quad (x < y \Rightarrow f(x) < f(y))$

debolmente decrescente se $\forall x,y \in A \quad (x < y \Rightarrow f(x) \geq f(y))$

strettamente decrescente se $\forall x,y \in A \quad (x < y \Rightarrow f(x) > f(y))$. \square

Anche qui parleremo di funzioni monotone e strettamente monotone, come per le

successioni.

Esercizio 1 Provare che la seguente condizione non garantisce la debole crescita:

$$\forall x \in A \quad f(x) \leq f(x+1) \quad .\square$$

Rispetto alle successioni abbiamo tre importanti novità: la restrizione, composizione e inversa di una funzione. Si tratta di nozioni del tutto generali (di pertinenza dell'"insiemistica"), naturalmente: se però nel contesto delle successioni sono poco interessanti, qui assumono un ruolo ben più importante.

Definizione 3 Data $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$ e dato $B \subseteq A$, diremo "restrizione di f a B " la funzione $g:B \longrightarrow \mathbb{R}$ così definita: $g(x) = f(x) \quad \forall x \in B$. Useremo la notazione $f|_B$ per indicare g . \square

C'è poco da dire, per ora. Casomai ricordo che l'idea di sottosuccessione (secondo il primo punto di vista iniziale, poi non seguito) era proprio quella di considerare la restrizione di una successione a un sottoinsieme infinito di \mathbb{N} .

La definizione di funzione composta è ben nota. Ricordo che, data $f:A \longrightarrow \mathbb{R}$ e $g:B \longrightarrow \mathbb{R}$, la funzione $f \circ g:B \longrightarrow \mathbb{R}$, definita come $(f \circ g)(x) = f(g(x))$, risulta essere definita solo se è soddisfatta la condizione $g(B) \subseteq A$.

Infine, anche la nozione di funzione inversa è ben nota. La tralascio, pertanto. Salvo riprenderla più in là (al § 13) dove sarò obbligato a fare un pò di puntualizzazioni.

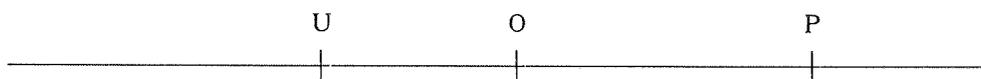
2. Grafico di una funzione

Come è ben noto, in modo del tutto generale, data $f:A \longrightarrow B$, dicesi grafico di f il seguente sottoinsieme di $A \times B$: $\text{gph}(f) = \{ (x,y) \in A \times B \text{ t.c. } y = f(x) \}$.

Anche l'idea di grafico di una funzione reale di variabile reale è in genere ben conosciuta. Esso è naturalmente un sottoinsieme di $A \times \mathbb{R}$ e quindi anche di \mathbb{R}^2 . Quindi si può "disegnare il grafico di f " nel "piano cartesiano". Anche se tutto ciò è, come detto, grosso modo noto a tutti, spesso le idee a questo proposito non sono chiare come dovrebbero. Cercherò quindi di precisare come stanno le cose, invitando fin d'ora il lettore a chiedersi alla fine se davvero aveva le idee chiare in proposito¹.

Tutto è basato sul fatto che i numeri reali sono l'insieme giusto per "descrivere la retta". La situazione è la seguente. Da una parte abbiamo i numeri reali, con i loro assiomi. Dall'altra la "retta euclidea". Cioè un insieme di elementi detti "punti" soddisfacenti gli assiomi della geometria euclidea relativi alla retta. Questi due enti matematici, numeri reali e retta euclidea, risultano essere una sorta di gemelli: infatti è possibile usare i numeri reali per descrivere compiutamente la retta euclidea e viceversa. Il ponte tra i due enti è costituito dalla idea cartesiana di geometria analitica, ovvero dalla possibilità di mettere un sistema di coordinate sulla retta. In realtà, quello che nelle scuole secondarie viene enfatizzato è il rapporto tra il "piano euclideo" e \mathbb{R}^2 : alla base di questo rapporto sta però quello tra retta reale ed \mathbb{R} , che ora mi accingo a descrivere, per poi passare alla geometria analitica del piano.

Come si fa a introdurre un sistema di coordinate cartesiane sulla retta? E' facile. Si fissano due punti distinti, O e U sulla retta. Il punto O svolgerà il ruolo di origine del sistema di riferimento cartesiano ed il punto U invece quello di fissare l'unità di misura (più propriamente, sarà il segmento OU che svolgerà il ruolo di unità di misura). Sia ora dato un generico punto P sulla retta:



¹ Evidentemente sto dando per scontato di essere capace di chiarirle, le idee, il che è tutto da dimostrare!

L'ascissa del punto P sarà il numero reale che esprime la misura del segmento OP in termini del segmento "unità" OU , con il "segno +" se P appartiene alla semiretta di origine O contenente U , con il "segno -" altrimenti (nel caso sopra disegnato, l'ascissa di P è negativa).

Chi garantisce che tutto questo si possa fare? Sono proprio gli assiomi della geometria euclidea che ci permettono di parlare di segmenti, di semirette, e anche di misura di segmenti con l'ausilio degli assiomi di \mathbb{R} . Poiché questa è proprio la parte che di solito è più carente nella preparazione pre-universitaria, traccio a grandi linee quale è la procedura che si segue. Si va a vedere quante volte il segmento OU "sta" nel segmento OP (questo è possibile grazie agli assiomi della geometria euclidea che riguardano l'ordinamento e grazie alla proprietà archimedea della retta euclidea). Fatto questo, si passa ai "sottomultipli" di OU , con una procedura ovvia: dividiamo per esempio OU in dieci parti e andiamo a vedere quante volte questo nuovo segmento dieci volte più piccolo "sta" nel segmento OP . E poi dividiamo ancora per dieci, e così via. Se da una parte mettiamo tutte le misure approssimate per difetto, e dall'altra quelle per eccesso, otteniamo una coppia di classi separate di numeri razionali (e quindi anche reali). Grazie all'assioma di completezza si ha la garanzia che vi è un numero reale elemento separatore di queste due classi. Tale elemento separatore è per di più unico perché le due classi sono contigue. Assumiamo pertanto questo unico numero reale che funge da elemento separatore come misura del segmento OP rispetto al segmento OU . Si può anche vedere C-S a pag. 39, a questo proposito.

Per quanto riguarda la geometria analitica del piano, si tratta di usare due rette incidenti e di sfruttare alcune proprietà della geometria euclidea del piano, tra cui il famoso quinto postulato di Euclide.

Si considera una coppia di rette incidenti, r ed s . Su ognuna si fissa un sistema di ascisse, cioè si fissano un punto O su r ed un punto O' su s , nonché un punto U ed U' rispettivamente su r ed s (vedi figura 1):

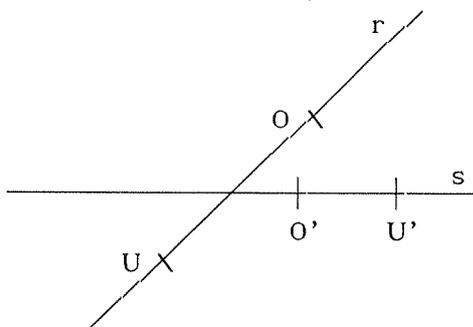


figura 1

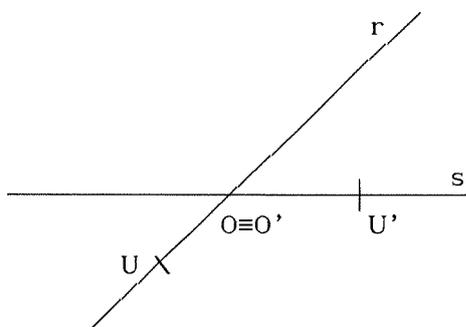


figura 2

Per evitare inutili complicazioni si sceglie come origine delle coordinate su entrambe le rette il punto di incidenza, come in figura 2.

Ebbene, dato un punto P del piano, grazie al "quinto postulato", per esso passano una e una sola retta parallela rispettivamente ad r e ad s . Tali parallele incontrano r ed s in due punti, che indicheremo con P_r e P_s rispettivamente. L'ascissa di P_r sulla retta r rispetto al sistema di riferimento individuato su r da O e U sarà detta ascissa di P e l'ascissa di P_s sarà invece detta ordinata di P (figura 3).

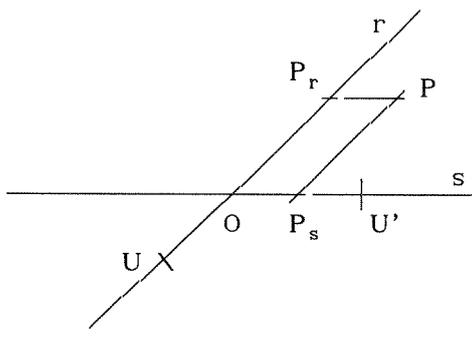


figura 3

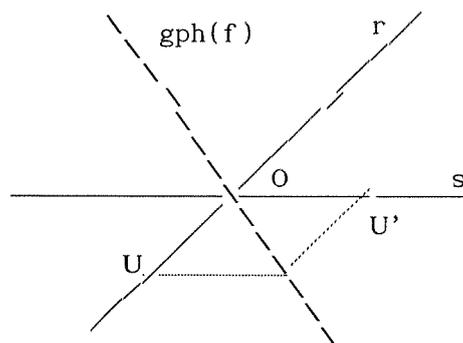


figura 4

Abbiamo pertanto assegnato a P una coppia di numeri reali, cioè la sua ascissa e la sua ordinata. Cioè abbiamo definito una funzione $\phi: \Pi \rightarrow \mathbb{R}^2$, dove Π è il piano euclideo. Tale funzione, che è in realtà una corrispondenza biunivoca, istituisce un sistema di coordinate cartesiane nel piano.

E' grazie alla intermediazione di ϕ che si può "disegnare il grafico" di una funzione reale di variabile reale. Essendo il grafico un sottoinsieme di \mathbb{R}^2 ,

la sua controimmagine mediante ϕ sarà un sottoinsieme del piano euclideo, ed è per l'appunto quello che disegniamo quando "disegniamo il grafico" di una funzione.

Nel nostro caso, per esempio $f(x) = x$ ha il grafico come tratteggiato alla figura 4.

E' ben noto, naturalmente, che spesso si scelgono sistemi di riferimento ortogonali ($r \perp s$) e monometrici (i segmenti OU ed OU' sono uguali).

Per di più si assume che l'angolo orientato $\hat{r}s$ individuato dalle rette "orientate" r ed s sia retto (l'orientamento sulle rette è fissato naturalmente in modo conforme alla scelta di U ed U' , cioè in modo che U segua O nell'ordinamento fissato su r , ed analogamente per s). Di solito si usano anche delle frecce per indicare l'ordinamento scelto. Infine, di solito, nel fare il disegno concreto su un pezzo di carta per convenienza si traccia la retta r in modo che essa sia parallela al bordo inferiore del foglio su cui si lavora. Come nella figura 5. Sarebbe molto utile se chi legge facesse un disegno nel quale le varie condizioni elencate in questo capoverso sono violate.

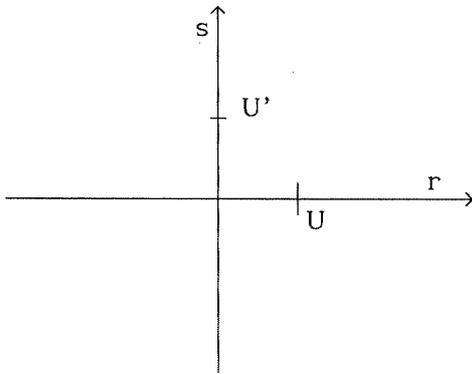


figura 5

L'ultima considerazione di opportunità fatta, relativa al foglio di lavoro, introduce sulla scena un terzo attore: oltre ad \mathbb{R}^2 e al piano euclideo, c'è anche il supporto "concreto" che scegliamo per rappresentare il piano euclideo. Come al solito, abbiamo l'idea che il "foglio di lavoro" sia una realizzazione attendibile del piano euclideo (o, più propriamente, a rovescio: cioè che il piano euclideo sia una sensata astrazione dei tanti possibili fogli di lavoro ...).

Esercizio 1 Disegnare il grafico di $f(x) = x^2$ in un sistema di coordinate cartesiane non ortogonali e non monometriche. \square

Un'ultima cosa. La grande diffusione dei sistemi di coordinate cartesiane ortogonali (e positivamente orientati), è dovuta anche ad un fatto che mi preme sottolineare.

Una funzione $f:A \longrightarrow \mathbb{R}$, esprime spesso una relazione tra due grandezze (per esempio fisiche) tra di loro ben diverse. Esempio classico, x rappresenta la temperatura di una sbarra in base ad un certo sistema di misura ed $f(x)$ ci dà la lunghezza di tale sbarra, sempre rispetto ad un certo sistema di unità di misura (CGS, MKSA, SI o che altro). Il grafico di f non è quindi "di per sé" una figura geometrica piana, al contrario di funzioni la cui origine risiede in una definizione geometrica (per esempio la parabola). E allora, tanto vale scegliere il sistema di riferimento che ci offre il punto di vista più familiare per motivi di consuetudine e che quindi ci dà una idea visiva più immediata delle proprietà di f .

E' interessante notare, a questo proposito, che se x ed $f(x)$ rappresentano misure di due grandezze diverse (per esempio, rispettivamente temperatura e lunghezza), il requisito di essere monometrico perde qualunque significato. Anzi, questo fatto viene a volte usato con scopi volutamente ingannatori da chi detiene potere culturale nei confronti di chi non ne ha, per rappresentare graficamente dei fenomeni sociali od economici in modo tale da ingenerare false impressioni d'origine visiva.

3. Definizione di limite

L'idea di limite "fondamentale" che dobbiamo precisare per le funzioni reali di variabile reale è quella del limite di una funzione quando x tende verso un certo x_0 .

Il punto più importante da capire, in questo contesto, è che noi vogliamo separare completamente due aspetti che ad uno sguardo poco accorto potrebbero anche non apparire scollegati.

Un aspetto è molto semplice da definire: si tratta semplicemente del valore che f assume in x_0 .

L'altro aspetto, anche se è tutto da scoprire in termini rigorosi, in termini intuitivi può però essere enunciato semplicemente: si tratta di sapere quale è il comportamento di f quando x si avvicina a x_0 . Cioè, vogliamo sapere se $f(x)$ "tende" da qualche parte quando x "tende" a x_0 . Questo secondo aspetto, né più né meno, corrisponde proprio al concetto di limite.

Pertanto il nostro scopo è scoprire cosa fa $f(x)$ quando x "si avvicina" ad x_0 , senza però far entrare in gioco il valore di f in x_0 , cioè $f(x_0)$ ². Come facciamo? Non è poi così difficile. In fondo, abbiamo già un sacco di esperienza sulle successioni. Tanto è vero che abbiamo visto come scimmiettare l'aspetto dinamico della cosa (senza parlare di "flussioni" di newtoniana memoria), giocando sulla struttura di una frase in cui sono presenti due quantificatori, \forall ed \exists , in un ben preciso ordine.

Il lettore è vivamente pregato di andare a rileggersi quanto detto a pagina II.4 a proposito della tensione dialettica esistente tra approccio statico e dinamico.

Ripensando appunto alle successioni, ricordo che la chiave sta tutta in: " $\forall \varepsilon > 0 \exists \nu \in \mathbb{N}$ t.c.". Qui ricorriamo a una struttura analoga. Se è plausibile che il " $\forall \varepsilon > 0$ " possa andare ancora bene, certo " $\exists \nu \in \mathbb{N}$ " va in qualche modo sostituito, perché è manifestamente troppo legato al fatto che una successione è definita sui naturali. Con cosa lo sostituiamo? Bisogna capire a che cosa serve ν . Questo ν ci dava una misura di "quanto grandi" dovevano essere gli indici n affinché a_n fosse "abbastanza vicino" ad l . Con un pò di fantasia, e magari anche con un pò di forzatura, e comunque certo in modo non formale, potremmo dire

² Ripeto: perché vogliamo tenere separati i due aspetti sopra menzionati

che ν misura quanto gli indici n devono essere "vicini a $+\infty$ " affinché a_n disti meno di ε da l .

Allora, se ora vogliamo vedere il limite di una funzione per x che tende a x_0 , dovremo sostituire ν con qualcosa che esprima la "vicinanza" di x ad x_0 . Ora, la distanza tra due numeri reali x e x_0 è data da $|x-x_0|$. Potremmo allora considerare un numero reale $\delta > 0$, e usare la disuguaglianza $|x-x_0| < \delta$ per esprimere che x è vicino ad x_0 a meno di δ .

A questo punto, possiamo provare anche ad abbozzare una definizione:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A (|x-x_0| < \delta \Rightarrow |f(x)-l| < \varepsilon)$$

Sembra ragionevole. Una sensata "traduzione" della definizione di limite dal linguaggio delle successioni a quello delle funzioni. Penso sia ragionevole, chiaro, ovvio, perché " $\forall n \in \mathbb{N}$ " è stato sostituito da " $\forall x \in A$ " (ricordo che A è l'insieme su cui è definita f , come \mathbb{N} era l'insieme su cui erano definite le successioni).

Ci siamo però dimenticati una cosa.

E cioè il discorso dei due aspetti che vogliamo tenere separati. Cioè vogliamo che dall'idea di limite sia assente ogni riferimento al comportamento di f nel punto x_0 . Come si fa? Basta una semplice modifica. Scriveremo cioè:

$$(*) \quad \forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A (0 < |x-x_0| < \delta \Rightarrow |f(x)-l| < \varepsilon)$$

Cioè, abbiamo escluso ogni considerazione riguardante il valore di f in x_0 dalla definizione di limite.

Invito caldamente tutti, ed in particolar modo chi sa già qualcosa sui limiti, a chiedersi se davvero siamo arrivati alla definizione di limite o se c'è ancora qualcosa da mettere a posto. Per dare un pò di suspense, proseguirò alla pagina successiva.

No. Non abbiamo finito. C'è ancora un punto un po' "sottile" da sistemare.

Con la speranza di farlo capire meglio, proporrò un esempio:

Esempio 1 Sia $A = [0, +\infty[$. Sia $f(x) = 3 \quad \forall x \in A$. Allora, con $x_0 = -6$, la condizione (*) è soddisfatta da $\ell = 47$. \square

Dettaglio Dato $\varepsilon > 0$, prendiamo $\delta = 6$. Poiché non c'è nessun $x \in A$ t.c. $0 < |x - x_0| < 6$, la implicazione che in (*) è racchiusa in parentesi sarà vera, in quanto la premessa è falsa (ricordare la tavola di verità di \Rightarrow). Quindi la condizione (*) è soddisfatta con $\ell = 47$. \square

A dire il vero, qualunque valore di ℓ avessimo provato, sarebbe andato bene, anche 3.

Cosa c'è che non va? Non è tanto ovvio da capire, in quanto ci sono troppe cose che non quadrano. Perché 47 o un altro numero dovrebbe essere il limite di f per x che tende a -6 ? Cosa c'entra con f che vale sempre 3? Casomai 3 potrebbe essere un candidato più sensato, ed in effetti come detto sopra anch'esso soddisfa la condizione (*). La cosa però che disturba di più in questo esempio è che non si capisce cosa voglia dire fare il limite per x che tende a -6 , visto che f è definita solo per $x \geq 0$.

Forse il guaio è tutto qui. Evidentemente dovevamo stare più attenti. Ma come rimediarvi? Non possiamo certo farlo imponendo come preconditione che sia $x_0 \in A$! Perché abbiamo detto che ci vogliamo disinteressare di quel che succede in x_0 , parlando di limite per x che tende a x_0 .

E allora? Occorre una cosa ovvia, e cioè portare alle estreme conseguenze il "secondo aspetto" di cui ho parlato in precedenza. Se la definizione di limite si deve occupare di quel che fa f "vicino" ad x_0 ma non di quel che capita in x_0 , evidentemente ho bisogno di garantire che A contenga punti "vicini quanto si vuole ad x_0 ". In termini precisi, dovremo richiedere che x_0 sia punto di accumulazione per A .

Definizione 1 Sia $A \subseteq \mathbb{R}$ e sia $x_0 \in \mathbb{R}$. Diremo che x_0 è punto di accumulazione per A se:

$$\forall \delta > 0 \quad \exists x \in A \quad \text{t.c.} \quad 0 < |x - x_0| < \delta \quad \square$$

E' questa la "condizione preliminare" perché abbia senso parlare di limite: per descrivere il comportamento di $f(x)$ per x che tende a x_0 dobbiamo avere a disposizione delle x vicine ad x_0 ma distinte da x_0 in cui f sia defini-

ta. Naturalmente nell'esempio di prima x_0 non era un punto di accumulazione per $[0, +\infty[$.

Osservazione 1 Si noti che nella definizione di punto di accumulazione non è richiesto che $x_0 \in A$. \square

Esercizio 1 Provare che 1 è di accumulazione per $A = [0, 1[$. \square

Esercizio 2 Provare che $\sqrt{3}$ non è punto di accumulazione per \mathbb{N} . \square

Siamo finalmente in grado di dare la definizione di limite. La definizione di limite che adotteremo è la seguente:

Definizione 2 Sia data $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$ e siano dati $x_0 \in \mathbb{R}$ ed $l \in \mathbb{R}$. Supponiamo che x_0 sia di accumulazione per A . Diremo che f ha limite l per x tendente a x_0 se:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A \ (0 < |x - x_0| < \delta \Rightarrow |f(x) - l| < \varepsilon) . \square$$

Se la condizione sopra indicata è soddisfatta, useremo la notazione $\lim_{x \rightarrow x_0} f(x) = l$, oppure scriveremo $f(x) \xrightarrow{x \rightarrow x_0} l$ o anche solo $f(x) \rightarrow l$ se non vi è ambiguità possibile.

Dopo aver visto la definizione, proviamo subito il risultato che ci si aspetta, cioè l'unicità del limite

Teorema 1 Sia $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in \mathbb{R}$, x_0 punto di accumulazione per A , $l, l' \in \mathbb{R}$. Se $\lim_{x \rightarrow x_0} f(x) = l$ e $\lim_{x \rightarrow x_0} f(x) = l'$, allora $l = l'$. \square

Dimostrazione Abbiamo a disposizione entrambe le seguenti relazioni:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A \ (0 < |x - x_0| < \delta \Rightarrow |f(x) - l| < \varepsilon)$$

$$\forall \varepsilon' > 0 \exists \delta' > 0 \text{ t.c. } \forall x' \in A \ (0 < |x' - x_0| < \delta' \Rightarrow |f(x') - l'| < \varepsilon')$$

Sia ε'' un qualsiasi numero reale positivo. Posto $\varepsilon = \varepsilon''/2$, otteniamo un $\delta > 0$; similmente, $\varepsilon' = \varepsilon''/2$ ci offre un $\delta' > 0$. Scelto $\delta'' = \min\{\delta, \delta'\}$, possiamo affermare che:

$$\forall x \in A \ (0 < |x - x_0| < \delta'' \Rightarrow |f(x) - l| < (\varepsilon''/2))$$

$$\forall x' \in A \ (0 < |x' - x_0| < \delta'' \Rightarrow |f(x') - l'| < (\varepsilon''/2))$$

Poiché x_0 è di accumulazione per A , siamo sicuri che esiste un elemento di A , che possiamo indicare con x'' , t.c. $0 < |x'' - x_0| < \delta''$. Allora, per tale x'' si ha:

$$|f(x'') - l| < \varepsilon''/2$$

$$|f(x'')-l'| < \varepsilon''/2 .$$

Da cui, sommando membro a membro e usando la disuguaglianza triangolare:

$$|l-l'| = |f(x'')-l-l'-f(x'')| \leq |f(x'')-l|+|f(x'')-l'| < \varepsilon'' .$$

Abbiamo quindi ottenuto che, dato un qualsiasi numero reale positivo ε'' , si ha che $|l-l'| < \varepsilon''$. Allora non può essere altro che $l = l'$. \square

Osservazione 2 Si noti il ruolo fondamentale avuto nella dimostrazione dal fatto che x_0 fosse punto di accumulazione per A . \square

Concludo con una definizione che fornisce una condizione sufficiente affinché un punto sia di accumulazione.

Definizione 3 Sia $A \subseteq \mathbb{R}$ e sia $x_0 \in A$. Diciamo che x_0 è un punto interno ad A se $\exists \delta > 0$ t.c. $]x_0-\delta, x_0+\delta[\subseteq A$. \square

Osservo che, contrariamente a quanto succede per i punti di accumulazione, un punto per essere interno ad A deve obbligatoriamente appartenere ad A . Ad esempio, l'intervallo $]a, b]$ ha a come punto di accumulazione anche se $a \notin]a, b]$; però a non è un punto interno ad $]a, b]$.

Che un punto interno ad A sia anche di accumulazione per A è "straovvio". Lo formalizzo come teorema solo per richiamarvi l'attenzione.

Teorema 2 Sia $A \subseteq \mathbb{R}$ ed x_0 punto interno ad A . Allora x_0 è di accumulazione per A . \square

Dimostrazione Sia $\delta' > 0$. Per ipotesi $\exists \delta'' > 0$ t.c. $]x_0-\delta'', x_0+\delta''[\subseteq A$. Quindi, detto $\sigma = \min\{\delta'/2, \delta''/2\}$, $x_0+\sigma \in A$ e $0 < |(x_0+\sigma)-x_0| < \delta'$. \square

4. Caratterizzazione del limite mediante successioni e sue conseguenze

Questo paragrafo ci permetterà di sfruttare per le funzioni vari risultati a suo tempo provati per le successioni: costituisce anche una sorta di ricompensa per lo sforzo fatto con le successioni.

Si tratta di ricondurre il limite di una funzione al limite di successioni.

L'idea è abbastanza semplice: se $f(x)$ tende ad l quando x tende a x_0 , se io ho una successione x_n che tende ad x_0 , dovrei ottenere di conseguenza che la successione $f(x_n)$ tenda ad l . E questo dovrebbe valere per qualunque successione x_n che tenda ad x_0 . In effetti così è (con qualche doverosa precisazione, però). La cosa interessante è che rende molto più utile questo discorso è che vale anche il viceversa. Cioè se so che $f(x_n)$ tende ad l per ogni successione x_n che tende a x_0 , posso dedurre che $\lim_{x \rightarrow x_0} f(x) = l$.

Questo tipo di considerazioni sono riassunte nel seguente teorema, molto utile ma con un enunciato e una dimostrazione un pò intricate.

Teorema 1 (caratterizzazione del limite mediante successioni) Siano dati $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in \mathbb{R}$, x_0 punto di accumulazione per A . Allora è:

$$\left(\lim_{x \rightarrow x_0} f(x) = l \right) \quad \text{se e solo se}$$

$$\left(\forall \text{ successione } \underbrace{\left(\begin{array}{l} x_n \in A \quad \forall n \in \mathbb{N}, \\ x_n \rightarrow x_0 \quad \text{e} \\ x_n \neq x_0 \quad \forall n \in \mathbb{N} \end{array} \right)}_{(*)} \Rightarrow f(x_n) \rightarrow l \right) . \square$$

Dimostrazione Dobbiamo dimostrare una proposizione del tipo $p \Leftrightarrow q$. Come si fa spesso, dimostriamo $p \Rightarrow q$ e $q \Rightarrow p$. Dimostro dapprima $p \Rightarrow q$ o, come si usa anche dire, dimostro " \Rightarrow ".

Sia quindi x una successione soddisfacente le condizioni elencate nel teorema. Dobbiamo dimostrare che $f(x_n) \rightarrow l$. Tradotto in "formule", sappiamo che $x_n \neq x_0 \quad \forall n \in \mathbb{N}$, $x_n \in A \quad \forall n \in \mathbb{N}$ e

$$\forall \varepsilon' > 0 \quad \exists \nu' \in \mathbb{N} \quad \text{t.c.} \quad \forall n' \in \mathbb{N} \quad (n' > \nu' \Rightarrow |x_{n'} - x_0| < \varepsilon')$$

Sappiamo inoltre che:

$$\forall \varepsilon'' > 0 \quad \exists \delta'' > 0 \quad \text{t.c.} \quad \underbrace{\left(\forall x'' \in A \quad (0 < |x'' - x_0| < \delta'' \Rightarrow |f(x'') - \ell| < \varepsilon'') \right)}_{(*)}$$

Dobbiamo ottenere che:

$$\forall \varepsilon > 0 \quad \exists \nu \in \mathbb{N} \quad \text{t.c.} \quad \forall n \in \mathbb{N} \quad (n > \nu \Rightarrow |f(x_n) - \ell| < \varepsilon)$$

Sia dunque dato il solito $\varepsilon > 0$. Prendiamo $\varepsilon'' = \varepsilon$. Da questo ε'' ricaviamo (almeno) un δ'' . Scegliamo $\varepsilon' = \delta''$. Allora abbiamo a disposizione un ν' t.c. per $n' > \nu'$ si ha $|x_{n'} - x_0| < \varepsilon'$. Poiché abbiamo anche che $x_n \in A$ e $x_n \neq x_0$, da (*) possiamo dedurre che (sempre per $n' > \nu'$) si ha $|f(x_{n'}) - \ell| < \varepsilon$.

Abbiamo quindi ottenuto quel che volevamo (dato ε , prendiamo $\nu = \nu'$).

Passiamo ora al " \Leftarrow ". Cioè a $q \Rightarrow p$. Dimostriamo la contronominale. Ovverossia proviamo che $\neg p \Rightarrow \neg q$.

Supponiamo di avere $\neg p$, cioè supponiamo non sia $f(x) \xrightarrow{x \rightarrow x_0} \ell$. Allora abbiamo che

$$\exists \sigma > 0 \quad \text{t.c.} \quad \forall \tau > 0 \quad \exists x \in A \quad \text{t.c.} \quad \underbrace{\left(0 < |x - x_0| < \tau \quad \text{e} \quad |f(x) - \ell| \geq \sigma \right)}_{(**)}$$

Fissiamo allora un σ siffatto. Dato $n \in \mathbb{N}$, consideriamo $\tau = 1/n$. Possiamo affermare quindi che $\exists x \in A$ t.c. valga (**). Per ogni $n \in \mathbb{N}$ ne scegliamo uno di questi e lo indichiamo con x_n . In questo modo otteniamo una successione, il cui termine generale è appunto x_n . Tale successione ovviamente è tale che $x_n \in A \quad \forall n \in \mathbb{N}$. Inoltre è $x_n \neq x_0$ (da (**)) otteniamo che $0 < |x_n - x_0| < 1/n$). Da (**)) otteniamo, come appena notato, che $0 < |x_n - x_0| < 1/n$ e quindi $|x_n - x_0| \rightarrow 0$ per il teorema dei due carabinieri, e quindi $x_n \rightarrow x_0$. Quindi la successione x_n è proprio una delle successioni di cui "parla" il teorema. Ciò nonostante, è $|f(x_n) - \ell| \geq \sigma$, il che impedisce ad $f(x_n)$ di tendere ad ℓ . Abbiamo ottenuto una successione t.c. la premessa scritta nella parentesi (*) è vera, ma la conseguenza è falsa. Quindi abbiamo provato che vale $\neg q$. □

E' valsa la pena di dimostrare questo teorema, perché ci permette di estendere facilmente al caso delle funzioni reali di variabile reale molti teoremi dimostrati per le successioni. Vediamo come, su un esempio.

Teorema 2 Siano $f, g: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in \mathbb{R}$, x_0 punto di accumulazione per A . Siano $\ell, m \in \mathbb{R}$. Se $\lim_{x \rightarrow x_0} f(x) = \ell$ e $\lim_{x \rightarrow x_0} g(x) = m$,

allora $\lim_{x \rightarrow x_0} f(x)+g(x) = l+m$.□

Dimostrazione L'ipotesi, assieme al teorema 1, ci garantisce che valgono le seguenti due proposizioni:

$$\left(\begin{array}{l} \forall \text{ successione} \\ x' \end{array} \left(\begin{array}{l} x'_n \in A \quad \forall n \in \mathbb{N}, \\ x'_n \longrightarrow x_0 \quad e \\ x'_n \neq x_0 \quad \forall n \in \mathbb{N} \end{array} \right) \Rightarrow f(x'_n) \longrightarrow l \right)$$

$$\left(\begin{array}{l} \forall \text{ successione} \\ x'' \end{array} \left(\begin{array}{l} x''_n \in A \quad \forall n \in \mathbb{N}, \\ x''_n \longrightarrow x_0 \quad e \\ x''_n \neq x_0 \quad \forall n \in \mathbb{N} \end{array} \right) \Rightarrow g(x''_n) \longrightarrow m \right)$$

Il teorema II.8.1 ci consente di affermare che la tesi è equivalente a:

$$(\bullet) \left(\begin{array}{l} \forall \text{ successione} \\ x \end{array} \left(\begin{array}{l} x_n \in A \quad \forall n \in \mathbb{N}, \\ x_n \longrightarrow x_0 \quad e \\ x_n \neq x_0 \quad \forall n \in \mathbb{N} \end{array} \right) \Rightarrow f(x_n)+g(x_n) \longrightarrow l+m \right)$$

Allora la dimostrazione è ovvia. Data una successione x soddisfacente le condizioni richieste in (\bullet) , prendiamo $x'_n = x''_n = x_n \quad \forall n \in \mathbb{N}$ e abbiamo che $f(x_n) \longrightarrow l$ e $g(x_n) \longrightarrow m$. Per il teorema sul limite di una somma di successioni abbiamo che $f(x_n)+g(x_n) \longrightarrow l+m$, che è appunto quanto ci serviva per ottenere la validità di (\bullet) .□

5. Estensioni della definizione di limite

Le funzioni reali di variabile reale consentono di fare un limite di tipo nuovo rispetto a quelli visti per le successioni, e cioè il limite per $x \longrightarrow x_0$. Per le successioni non abbiamo mai fatto un limite di questo tipo (per esempio per n che tende a 90), ma solo per $n \longrightarrow \infty$.

Esercizio 1 Provare che, qualunque sia $f: \mathbb{N} \longrightarrow \mathbb{R}$, qualunque sia $x_0 \in \mathbb{N}$ e qualunque sia $\ell \in \mathbb{R}$, è soddisfatta la condizione (*) del § 3. Notare anche che \mathbb{N} non ha punti di accumulazione. \square

A proposito dei limiti di successioni, vale la pena di spendere due parole a proposito delle notazioni usate. Abbiamo usato la notazione $n \longrightarrow \infty$ (e la più "seria" $\lim_{n \rightarrow \infty} a_n$). Forse avremmo dovuto essere più precisi, usando la notazione $n \longrightarrow +\infty$, ma proprio perché avevamo considerato solo la possibilità che n diventasse "molto grande" e visto che eravamo all'interno dei numeri naturali, non è sembrato opportuno usare questa precisazione. Nei limiti di funzioni reali di variabile reale dovremo stare attenti, presumibilmente, e distinguere tra i limiti per $x \longrightarrow +\infty$, $x \longrightarrow -\infty$ e $x \longrightarrow \infty$.

Non c'è comunque solo questa novità, per i limiti di funzioni rispetto ai limiti di successioni. Oltre alla possibilità di fare il limite per $x \longrightarrow x_0$, c'è anche quella di fare il limite per $x \longrightarrow x_0$ "da destra" o "da sinistra". Grosso modo, fare il limite per $x \longrightarrow x_0$ da sinistra (ovverossia fare il limite per $x \longrightarrow x_0^-$), consiste nel guardare solo a cosa fa f a sinistra di x_0 , disinteressandosi di quel che succede a destra.

La definizione di questi limiti, da destra e da sinistra, è facile da dare. Occorre però non cadere in un trabocchetto. Consideriamo per esempio $A = [0,1]$. Naturalmente 1 è di accumulazione per A , quindi ha senso domandarsi se $\exists \lim_{x \rightarrow 1} f(x)$, per una qualsiasi f definita su A . Ha anche senso chiedersi cosa faccia f per x che tende a x_0 da sinistra, ma non ha senso evidentemente domandarsi quale possa essere il limite di f per x che tende a x_0 da destra. Detto altrimenti, data $f: A \longrightarrow \mathbb{R}$, il fatto che x_0 sia un punto di accumulazione per A non è sufficiente a garantire che abbia senso considerare il limite di f per x che tende a x_0 da destra (o da sinistra). Ci vuole qualcosa di più.

Definizione 1 Siano dati $A \subseteq \mathbb{R}$ ed $x_0 \in \mathbb{R}$. Diremo che x_0 punto di

accumulazione per A da sinistra (rispettivamente: da destra) se $\forall \delta > 0$
 $\exists x \in A$ t.c. $0 < x_0 - x < \delta$ (rispettivamente: $0 < x - x_0 < \delta$). \square

Possiamo quindi introdurre la seguente:

Definizione 2 Siano dati $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in \mathbb{R}$, x_0 punto di accumulazione per A da sinistra (rispettivamente: da destra), $l \in \mathbb{R}$. Diremo che l è il limite di f per x tendente a x_0 da sinistra (rispettivamente: da destra) se:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A \ (0 < x_0 - x < \delta \Rightarrow |f(x) - l| < \varepsilon),$$

(rispettivamente:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A \ (0 < x - x_0 < \delta \Rightarrow |f(x) - l| < \varepsilon)). \square$$

Useremo la notazione $\lim_{x \rightarrow x_0^-} f(x) = l$ per dire che l è il limite di f da sinistra in x_0 , ovvero sia che l è il limite di $f(x)$ per x tendente a x_0 da sinistra, o anche che l è il limite di $f(x)$ per $x \rightarrow x_0$. Si usa anche la scrittura abbreviata $f(x) \xrightarrow{x \rightarrow x_0^-} l$. Naturalmente cose analoghe valgono da destra.

Queste considerazioni, e la notazione x_0^- , penso abbiano ricordato che per le successioni avevamo dato la definizione di limite l^+ ed l^- . Naturalmente possiamo semplicemente replicare il tutto per funzioni. Diremo che $\lim_{x \rightarrow x_0} f(x) = l^+$ se:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A \ (0 < |x - x_0| < \delta \Rightarrow 0 \leq f(x) - l < \varepsilon).$$

Possiamo anche mettere assieme le cose, definendo $\lim_{x \rightarrow x_0} f(x) = l^+$ se:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A \ (0 < x_0 - x < \delta \Rightarrow 0 \leq f(x) - l < \varepsilon).$$

Non c'è solo questo. C'è una varietà pressoché sterminata di limiti: $\lim_{x \rightarrow +\infty} f(x) = l^-$, $\lim_{x \rightarrow x_0^+} f(x) = -\infty$, $\lim_{x \rightarrow \infty} f(x) = +\infty$, etc., etc.

Non ha senso pensare di dare una definizione per ognuno dei casi! E' più facile trovare una regola "automatica" per generare la definizione.

Tutto è basato sulla osservazione che la definizione di limite per funzioni reali di variabile reale ha la seguente struttura:

$\lim_{x \rightarrow \blacksquare} f(x) = \star$ avrà una definizione del tipo:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A \ (\blacksquare \Rightarrow \star).$$

Dove al posto di " \blacksquare " ci sarà una condizione che esprime (tramite δ) la "vicinanza" di x al "punto" \blacksquare nel quale vogliamo calcolare il limite, ed ana-

logamente al posto di " \star " c'è invece una condizione che esprime (tramite ε) la "vicinanza" di $f(x)$ al limite \star .

Ad esempio, se $\star = +\infty$, come faccio ad esprimere la "vicinanza" di $f(x)$ a $+\infty$? Così: mettiamo $f(x) > \varepsilon$ al posto di " \star ". Questo "algoritmo" va bene sempre. Sia che facciamo il limite in $\blacksquare = x_0$, o in $\blacksquare = -\infty$, o in $\blacksquare = x_0^+$, etc.

Analogamente, se $\blacksquare = x_0^-$, abbiamo che al posto di " \blacksquare " metteremo $0 < x_0 - x < \delta$. E così via ...

Non va dimenticata però una cosa: ogni "punto" che mettiamo al posto di \blacksquare richiede una condizione preliminare soddisfatta, e cioè che vi siano punti di A (distinti da \blacksquare , se ha senso) "vicini" quanto si voglia a \blacksquare . Quindi, per esempio, se $\blacksquare = +\infty$, richiederemo come condizione preliminare che $\forall \delta > 0$ vi sia $x \in A$ che soddisfa la condizione " \blacksquare ", e cioè $x > \delta$: essendo $\blacksquare = +\infty$, e cioè non un numero reale, non abbiamo da aggiungere la condizione che sia $x \neq \blacksquare$.

Esercizio 2 Provare a dare qualche definizione di limite. \square

Quello che abbiamo appena visto, con i \blacksquare e \star , appare come un gioco puramente formale, un "trucco" per trattare in modo unitario vari casi di limite. Come spesso succede, c'è della sostanza dietro alle forme. Vale a dire, ogni particolare definizione di limite (per esempio $\lim_{x \rightarrow +\infty} f(x) = \ell^-$) può essere vista come la specificazione in un dato caso di una idea generale. Ora, l'idea generale è che un "limite" esprime il fatto che, mentre " x tende a \blacksquare ", allora " $f(x)$ tende a \star ". Come si vede, entrambe le frasi tra virgolette sono centrate sull'idea di "tendere" o "avvicinarsi a". E questo sia per la "variabile indipendente" x che per la "variabile dipendente" y , che è uguale a $f(x)$. Allora il concetto fondamentale da esprimere è quello di "tendere a". Se per esempio ci interessa l'idea di tendere a t_0 , ci occorrerà un modo per stimare la "distanza" di un punto t da t_0 : visto che la distanza tra i due punti è $|t - t_0|$, un numero reale positivo σ è tutto quello che ci occorre per poter dire che la distanza è "minore di σ "; cioè la condizione che ci interessa è $|t - t_0| < \sigma$.

Questo vale per la variabile indipendente (la x , come di solito si usa indicarla), e allora ritroveremo la condizione $|x - x_0| < \delta$, usando come è nella tradizione la lettera greca delta; vale anche per la variabile dipendente, nel qual caso si avrà la condizione $|f(x) - \ell| < \varepsilon$ (a noi interessa infatti una stima della distanza tra il numero reale ℓ e il valore assunto dalla f in corrispondenza del punto x).

Le precedenti considerazioni valgono, mutatis mutandis, quando ci interessa esprimere che qualcosa "tende" a $+\infty$ (o a $-\infty$, o a $+\infty$). In tal caso, non possiamo esprimere la distanza tra la t e $+\infty$ come $|t-(+\infty)|$, visto che non sappiamo fare la sottrazione tra t e $+\infty$ (per meglio dire, non ci abbiamo neanche provato). Non è necessario essere aquile, però, per riuscire a trarsi d'impaccio: in fondo, dire che t "tende" a $+\infty$ è esprimibile dicendo che t "tende" a diventare "grosso". Allora, esprimeremo questo fatto con questa semplice relazione: $t > \sigma$, dove σ ci dà, come prima, una misura di quanto "grosso" prendiamo t . Esattamente come prima, possiamo sfruttare questa idea sia per la variabile indipendente che per la variabile dipendente: quindi, se ci interessa il limite per " x che tende a $+\infty$ ", ci imatteremo nella relazione $x > \delta$; se vogliamo parlare invece di " $f(x)$ che tende a $+\infty$ ", avremo a che fare con la relazione $f(x) > \varepsilon$.

6. La proprietà di limite è una proprietà locale

L'idea di limite è fondata sullo studio di cosa fa f "vicino" a x_0 . Può anche essere prevedibile quindi che il limite di f in x_0 non dipenda dai valori che f assume in punti "lontani" da x_0 . Come forse qualcuno ha già cominciato a sospettare, ci troviamo di fronte a una difficoltà non del tutto nuova: si sente un po' di somiglianza con la storia degli " ε piccoli". E, visto che quella questione è stata in qualche modo sistemata, possiamo ragionevolmente sperare che essa ci offra utili insegnamenti anche per questo nuovo problema.

Supponiamo di essere interessati al limite di f in x_0 . Potremmo pensare di fissare $\sigma > 0$ e osservare che una definizione equivalente di $\lim_{x \rightarrow x_0} f(x) = l$ è la seguente:

$$(*) \quad \forall \varepsilon \in]0, \sigma[\quad \exists \delta > 0 \quad \text{t.c.} \quad \forall x \in A \quad (0 < |x - x_0| < \delta \Rightarrow |f(x) - l| < \varepsilon).$$

Ciò tuttavia non rende pienamente conto di quello che stiamo facendo. E' meglio adottare un altro punto di vista. Sia quindi dato sempre $\sigma > 0$. Consideriamo $f|_{]x_0 - \sigma, x_0 + \sigma[\cap A}$, ovvero sia la restrizione di f a $]x_0 - \sigma, x_0 + \sigma[\cap A$. Chiamiamo E l'insieme $]x_0 - \sigma, x_0 + \sigma[\cap A$, per avere una notazione meno pesante. Scriviamo quindi $f|_E$. Questa è una nuova funzione, anche se ovviamente "imparentata" con f : chiamiamola g . Quel che succede è che $\lim_{x \rightarrow x_0} f(x) = l$ è come dire $\lim_{x \rightarrow x_0} g(x) = l$. Perché? Naturalmente proprio perché (*), che è equivalente alla definizione di $\lim_{x \rightarrow x_0} f(x) = l$, è anche evidentemente equivalente a $\lim_{x \rightarrow x_0} g(x) = l$.

A questo punto occorre un momento di attenzione. Cosa ho guadagnato ad introdurre g , cioè la restrizione di f a E ? Sembrerebbe nulla, perché in fondo poi la conclusione è basata sulla osservazione che per f (*) è equivalente alla definizione di limite. In realtà, l'introduzione di $f|_E$ è estremamente utile perché offre un punto di vista molto fruttuoso. Per capirlo, basta osservare il seguente:

Esempio 1 Sia $f(x) = [x]$. Quanto fa $\lim_{x \rightarrow 7/2} f(x)$? Osserviamo che, se prendiamo $\sigma = 1/2$, abbiamo che $E =]3, 4[$ e che $f|_E(x) = 3 \quad \forall x \in E$. Se consideriamo $h: \mathbb{R} \rightarrow \mathbb{R}$, definita come $h(x) = 3 \quad \forall x \in \mathbb{R}$, è ovvio che $\lim_{x \rightarrow 7/2} h(x) = 3$, poiché h è una costante. Ebbene, $h|_E = f|_E$! Quindi abbiamo che:

$$\lim_{x \rightarrow 7/2} h(x) = 3 \Rightarrow \lim_{x \rightarrow 7/2} h|_E(x) = 3 \Rightarrow \lim_{x \rightarrow 7/2} f|_E(x) = 3 \Rightarrow \lim_{x \rightarrow 7/2} f(x) = 3$$

Abbiamo ottenuto in modo molto rapido il valore del limite di f in $7/2$, osservando che sia f che un'altra funzione molto semplice hanno la stessa restrizione a $]x_0 - \sigma, x_0 + \sigma[\cap A$, per un σ opportuno. \square

E' immediato estendere le considerazioni fatte per il limite in x_0 ai limiti da sinistra o da destra (useremo $]x_0 - \sigma, x_0[\cap A$ e rispettivamente $]x_0, x_0 + \sigma[\cap A$) ed anche a $+\infty$ (con $] \sigma, +\infty[\cap A$), etc. Ciò ci consente di esaminare il seguente

Esempio 2 Sia $f(x) = [x]$. E sia $x_0 = 3$. Quanto fa $\lim_{x \rightarrow 3^+} f(x)$? Scelto $\sigma = 1$, abbiamo che $f|_E(x) = 3$ per ogni $x \in E =]3, 4[$. Ma allora

$$\lim_{x \rightarrow 3^+} f(x) = \lim_{x \rightarrow 3^+} f|_E(x) = \lim_{x \rightarrow 3^+} h|_E(x) = \lim_{x \rightarrow 3^+} h(x) = 3,$$

dove h è la funzione costantemente uguale a 3 già usata nell'esempio precedente. \square

Concludo questo paragrafo con una precisa formalizzazione di quanto ho scritto finora, anche perché non bisogna dimenticare che anche il fatto di essere di accumulazione è una proprietà locale, osservazione che in realtà avrebbe dovuto essere fatta fin dall'inizio, riguardando un "prerequisito" affinché si possa parlare di limite.

Teorema 1 Siano dati $A, B \subseteq \mathbb{R}$ ed $x_0 \in \mathbb{R}$. Se $\exists \sigma > 0$ t.c. $(]x_0 - \sigma, x_0 + \sigma[\cap A) \setminus \{x_0\} = (]x_0 - \sigma, x_0 + \sigma[\cap B) \setminus \{x_0\}$, allora x_0 è punto di accumulazione per A se e solo se lo è per B . \square

Esercizio 1 Scrivere l'enunciato di un teorema analogo riguardante i punti di accumulazione da sinistra. \square

Teorema 2 Siano $A, B \subseteq \mathbb{R}$, $f: A \rightarrow \mathbb{R}$, $g: B \rightarrow \mathbb{R}$, $x_0 \in \mathbb{R}$ ed $l \in \mathbb{R}$. Se $\exists \sigma > 0$ t.c.

$$\begin{cases} A' \stackrel{\text{def}}{=} (]x_0 - \sigma, x_0 + \sigma[\cap A) \setminus \{x_0\} = (]x_0 - \sigma, x_0 + \sigma[\cap B) \setminus \{x_0\} \stackrel{\text{def}}{=} B', \\ f|_{A'} = g|_{B'}, \end{cases}$$

allora è

$$\lim_{x \rightarrow x_0} f(x) = l \Leftrightarrow \lim_{x \rightarrow x_0} g(x) = l. \square$$

Le dimostrazioni di questi teoremi sono ovviamente lasciate al lettore, come

il compito di formulare le versioni del teorema 2 per gli altri tipi di limiti.

7. Limiti di funzioni monotone

Concludo la parte che riguarda i limiti, prima di passare alla seconda parte dedicata alla continuità, con un risultato che estende direttamente alle funzioni il teorema II.12.1.

Così come è capitato per le successioni, questa è la prima volta che usiamo in modo essenziale l'assioma di completezza.

Teorema 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in \mathbb{R}$ punto di accumulazione per A . Se f è debolmente crescente su $A \cap]-\infty, x_0[$, allora esiste $\lim_{x \rightarrow x_0} f(x)$, e si ha $\lim_{x \rightarrow x_0} f(x) = \sup(f(A \cap]-\infty, x_0[))$. \square

Dimostrazione Naturalmente l'uguaglianza va intesa nel senso che, se $\sup(f(A \cap]-\infty, x_0[)) = \lambda \in \mathbb{R}$, allora $\lim_{x \rightarrow x_0} f(x) = \lambda$; se invece $f(A \cap]-\infty, x_0[)$ non è superiormente limitato³, allora $\lim_{x \rightarrow x_0} f(x) = +\infty$.

La dimostrazione di questo teorema è un adattamento immediato della dimostrazione del teorema II.12.1.

Supponiamo $f(A \cap]-\infty, x_0[)$ sia superiormente limitato. Sia $\lambda = \sup(f(A \cap]-\infty, x_0[))$. Dato $\lambda - \varepsilon$, esso non è più il sup. Quindi $\exists \bar{x} \in A \cap]-\infty, x_0[$ t.c. $f(\bar{x}) > \lambda - \varepsilon$. Se prendiamo $\delta = x_0 - \bar{x}$, abbiamo che $\forall x \in A$ ($0 < x_0 - x < \delta$) $\Rightarrow \lambda - \varepsilon \leq f(\bar{x}) \leq f(x) \leq \lambda$. E quindi la condizione di limite è soddisfatta. Il caso in cui $f(A \cap]-\infty, x_0[)$ non è superiormente limitato si dimostra in modo analogo. \square

Osservazione 1 In realtà abbiamo dimostrato che $\lim_{x \rightarrow x_0} f(x) = \lambda^-$. \square

³ Come già detto, ciò viene per consuetudine indicato brevemente dicendo che il suo sup è $+\infty$.

8. Definizione di continuità in un punto

Cosa vuol dire che una funzione è continua in un punto x_0 ? Molto semplice. Significa che i due aspetti di cui parlavamo nel § 3 vanno d'accordo. Cioè: da una parte ho $f(x_0)$. Dall'altra $\lim_{x \rightarrow x_0} f(x)$. Che i due aspetti vadano d'accordo, significa molto semplicemente che $\lim_{x \rightarrow x_0} f(x)$ esiste e che è uguale a $f(x_0)$.

Effettivamente questa è l'idea di continuità di una funzione in un punto. Lungi da me dire che si tratta di qualcosa di intuitivo! Anche perché l'idea intuitiva casomai riguarda la continuità di una funzione su un intervallo. E' più naturale parlare di "discontinuità" in un punto. Quindi l'approccio seguito non è "spontaneistico". Ho solo cercato di rendere sensato un punto di vista che è il coronamento (almeno, finora) di uno sforzo durato secoli.

La definizione di continuità di f in un punto x_0 quindi si ottiene modificando opportunamente la definizione di limite per f in x_0 , che qui sotto riporto:

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \text{t.c.} \quad \forall x \in A \quad (0 < |x - x_0| < \delta \Rightarrow |f(x) - \ell| < \varepsilon) .$$

La prima ovvia modifica da fare consiste nel sostituire $f(x_0)$ ad ℓ : in fondo, non vogliamo proprio che il limite in x_0 coincida col valore di f in x_0 ? Naturalmente dobbiamo anche modificare i prerequisiti, richiedendo che $x_0 \in A$, sennò non possiamo considerare $f(x_0)$. E nella definizione possiamo, se ci fa piacere, eliminare la condizione $0 < |x - x_0|$, ovverossia che sia $x \neq x_0$, in quanto per $x = x_0$, $f(x) = f(x_0)$ e quindi è sicuramente soddisfatta la condizione $|f(x) - f(x_0)| < \varepsilon$. Ripeto che si tratta di una modifica del tutto opzionale: in effetti non cambia nulla se nella definizione lasciamo ancora la condizione che sia $0 < |x - x_0|$. La definizione quindi diventa:

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \text{t.c.} \quad \forall x \in A \quad (|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon) .$$

E' opportuno però soffermarsi ancora sui prerequisiti. Ricordo che, per poter "parlare" di limite, avevamo bisogno di avere a disposizione:

$$\left\{ \begin{array}{l} \text{una } f: A \longrightarrow \mathbb{R} , \text{ con } A \subseteq \mathbb{R} \\ x_0 \in \mathbb{R} , \text{ con } x_0 \text{ di accumulazione per } A . \\ \text{un } \ell \in \mathbb{R} \end{array} \right.$$

Ovviamente, il dato ℓ non ci sarà più (è sostituito da $f(x_0)$). Come già notato, abbiamo invece bisogno del fatto che $x_0 \in A$. D'altro canto, ora non ci interessa più che x_0 sia di accumulazione per A , perché la condizione $0 <$

$< |x-x_0|$ è diventata irrilevante, essendo interessati ad $l = f(x_0)$. Pertanto non porremo più la condizione che x_0 sia di accumulazione per A . Riassumendo:

Definizione 1 (di funzione continua in un punto) Sia data $f: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$. Diremo che f è continua in x_0 se:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A (|x-x_0| < \delta \Rightarrow |f(x)-f(x_0)| < \varepsilon) . \square$$

Possiamo ora chiederci se la definizione di f continua in x_0 che abbiamo dato ha davvero la proprietà di fondere i due aspetti molte volte richiamati. Naturalmente per porre la questione in questi termini, deve avere senso il limite in x_0 , quindi una tale verifica si potrà fare solo se x_0 è punto di accumulazione per A . Abbiamo il seguente:

Teorema 1 Sia $f: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$ e x_0 punto di accumulazione per A . Allora:

$$(f \text{ è continua in } x_0) \Leftrightarrow \left(\exists \lim_{x \rightarrow x_0} f(x) \text{ e } \lim_{x \rightarrow x_0} f(x) = f(x_0) \right) . \square$$

Dimostrazione Dimostriamo dapprima " \Rightarrow ".

Per ipotesi abbiamo che:

$$(\bullet) \quad \forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x \in A (|x-x_0| < \delta \Rightarrow |f(x)-f(x_0)| < \varepsilon) .$$

Vogliamo ottenere che:

$$(\blacktriangle) \quad \forall \varepsilon' > 0 \exists \delta' > 0 \text{ t.c. } \forall x' \in A (0 < |x'-x_0| < \delta' \Rightarrow |f(x')-f(x_0)| < \varepsilon') .$$

Il che è iper-evidente (dato ε' , prendiamo $\varepsilon = \varepsilon'$ e $\delta' = \delta$).

Vediamo ora " \Leftarrow ".

Supponiamo che valga (\blacktriangle) . Ma allora è ovvio che si ha (\bullet) (dato ε , prendiamo $\varepsilon' = \varepsilon$ e $\delta = \delta'$; oltre a questo basta notare che, se $x = x_0$, allora $f(x) = f(x_0)$ e quindi $|f(x)-f(x_0)| < \varepsilon$). \square

E nell'altro caso? Cioè quando x_0 non è di accumulazione e quindi non abbiamo a disposizione la definizione di limite, cosa significa che f è continua in x_0 ? La risposta è semplice e un po' drastica: nulla. Per chiarire per bene questo, abbiamo bisogno prima di una definizione.

Definizione 2 Sia $A \subseteq \mathbb{R}$ ed $x_0 \in A$. Se x_0 non è di accumulazione per A , allora x_0 si dice punto isolato di A . \square

Quindi, se x_0 è isolato, vuol dire che $\exists \delta > 0$ t.c. $]x_0-\delta, x_0+\delta[\cap A = \{x_0\}$, cioè non ci sono in A altri punti vicini ad x_0 . Ebbene, si dimostra

facilmente il seguente risultato:

Teorema 2 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 punto isolato di A . Allora f è continua in x_0 . \square

Dimostrazione Basta usare la definizione di continuità:

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \text{t.c.} \quad \forall x \in A \quad (|x-x_0| < \delta \Rightarrow |f(x)-f(x_0)| < \varepsilon).$$

Dato ε , fissiamo δ t.c. $]x_0-\delta, x_0+\delta[\cap A \subseteq \{x_0\}$. Per tale δ è ovvio che vale $\forall x \in A \quad (|x-x_0| < \delta \Rightarrow |f(x)-f(x_0)| < \varepsilon)$. \square

Naturalmente la cosa da notare nell'enunciato di questo teorema è che ogni funzione, qualsiasi valore assuma in un punto isolato x_0 , è continua in tale punto x_0 (indipendentemente dai valori che essa assume negli altri punti). Tale risultato non è per nulla sorprendente. Ritorno alla idea che continuità significa che i due aspetti, valore di f in x_0 e valori che assume nei punti "vicini", vanno d'accordo. Se non ci sono punti vicini ad x_0 , è ovvio che non c'è molto da andare d'accordo. In realtà, uno avrebbe anche potuto dire che, non essendovi punti "vicini" ad x_0 , non ha senso porre il problema della continuità di f in x_0 . E' un punto di vista questo più che rispettabile. La ragione della scelta, e cioè di definire la continuità anche nei punti isolati (e quindi, a questo punto, dire che si ha automaticamente la continuità in tali punti) sta nel fatto che una definizione siffatta è più flessibile, è maneggevole con maggior confidenza. Propongo, dopo il teorema 9.3, un esempio che rende conto di quanto ho affermato: vedasi l'osservazione 9.1.

9. Conseguenze elementari della continuità in un punto

Vi è tutta una serie di risultati (continuità della somma, del prodotto, teorema di permanenza del segno, etc.) la cui dimostrazione è una ovvia conseguenza delle analoghe proprietà dimostrate per i limiti, tenuto conto del teorema 8.1. Naturalmente ciò è valido per i punti che sono di accumulazione per A : per gli altri punti però non c'è alcun problema perché nei punti isolati la continuità è automatica. Comunque, tali risultati si possono ottenere anche dalla caratterizzazione della continuità mediante successioni.

Teorema 1 (caratterizzazione della continuità mediante le successioni) Siano dati $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$. Allora è:

$$\left(f \text{ è continua in } x_0 \right) \quad \text{se e solo se} \\ \left(\forall \text{ successione } \left(\begin{array}{l} x \\ x_n \in A \quad \forall n \in \mathbb{N} \\ x_n \longrightarrow x_0 \quad \text{e} \end{array} \right) \Rightarrow f(x_n) \longrightarrow f(x_0) \right) . \square$$

Tralascio la dimostrazione, che è un semplice adattamento della dimostrazione del teorema 4.1. Vorrei solo vedere, a mo' di esempio, come si possa usare questo teorema per ottenere un interessante risultato.

Teorema 2 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$. Se f è continua in x_0 , allora anche $|f|$ lo è. \square

Dimostrazione Naturalmente $|f|$ è la funzione definita su A nel modo seguente: $|f|(x) = |f(x)| \quad \forall x \in A$. Allora, sia $x_n \longrightarrow x_0$. Per la continuità di f , $f(x_n) \longrightarrow f(x_0)$. Per una proprietà dei limiti di successioni, posso affermare che $|f(x_n)| \longrightarrow |f(x_0)|$. Poiché questo vale per qualsiasi successione x t.c. $x_n \longrightarrow x_0$, posso asserire la continuità di $|f|$ in x_0 . \square

Una conseguenza più importante è il seguente:

Teorema 3 (continuità della funzione composta) Siano dati: $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, $g:B \longrightarrow \mathbb{R}$, $B \subseteq \mathbb{R}$. Supponiamo che $f(A) \subseteq B$, che f sia continua in x_0 e che g sia continua in $f(x_0)$. Allora $g \circ f$ è continua in x_0 . \square

Dimostrazione Assolutamente banale usando la caratterizzazione della continuità mediante successioni. Sia quindi data $x_n \longrightarrow x_0$: dobbiamo provare

che $(g \circ f)(x_n) \longrightarrow (g \circ f)(x_0)$. La continuità di f in x_0 ci garantisce che $f(x_n) \longrightarrow f(x_0)$. Poiché la successione di termine generale $f(x_n)$ converge al punto $f(x_0)$ e g è continua in tale ipotesi, la caratterizzazione della continuità mediante successioni, applicata a g in $f(x_0)$, ci dice che $g(f(x_n)) \longrightarrow g(f(x_0))$, cioè proprio quel che volevamo. \square

Osservazione 1 Consideriamo $f(x) = \sin^2(x) - 1$ e $g(x) = \sqrt{x}$. La funzione composta risulta essere definita solo nei punti in cui $\sin(x) = \pm 1$, cioè i punti del tipo $(\pi/2) + k\pi$. Quindi tutti i punti del suo insieme di definizione sono punti isolati. Grazie al teorema, possiamo essere sicuri che sia continua in tali punti. Se avessimo limitato la definizione di continuità ai punti di accumulazione, avremmo dovuto dare un enunciato del teorema di continuità della funzione composta più restrittivo, meno "user friendly". \square

10. Continuità di una funzione su un insieme

Cosa vuol dire che una funzione è continua su un certo insieme? La definizione è molto semplice

Definizione 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$ e sia $B \subseteq A$. Diciamo che f è continua su B (o, indifferentemente, in B) se f è continua in ogni $x_0 \in B$. \square

Esempio 1 $f(x) = |x|$ è continua in \mathbb{R} ; $g(x) = 1/x$ è continua su tutto il suo insieme di definizione, che è $\mathbb{R} \setminus \{0\}$; $h(x) = \begin{cases} 1 & \text{se } x > 0 \\ 0 & \text{se } x = 0 \\ -1 & \text{se } x < 0 \end{cases}$, definita su \mathbb{R} , è continua su $B = \mathbb{R} \setminus \{0\}$; $k(x) = [x]$, definita su \mathbb{R} , è continua su $B = \mathbb{R} \setminus \mathbb{Z}$. \square

Bisogna stare attenti a non cadere in un trabocchetto. Sia data $f:A \longrightarrow \mathbb{R}$, e sia $B \subseteq A$. Possiamo considerare la restrizione di f a B , $f|_B$. Che legame c'è tra la continuità di f su B e la continuità invece di $f|_B$ sul suo insieme di definizione, che naturalmente è B ? Un po' sorprendentemente si ha che la continuità di f su B implica la continuità di $f|_B$ sul suo insieme di definizione, ma non vale il viceversa. In realtà questo fatto può essere sorprendente se si ragiona in termini "analitici", ma non lo è affatto se si fa un semplice esempio, magari corredato da un disegno.

Esempio 2 Sia $f(x) = [x]$. E sia $B = [1,2[$. Abbiamo $f|_B(x) = 1 \quad \forall x \in B$. Quindi $f|_B$ è costante e vale 1, pertanto è continua su tutto il suo insieme di definizione, cioè B . Ma f non è continua su tutto B , perché non lo è in 1! Se si guarda la figura 6 si capisce subito perché.

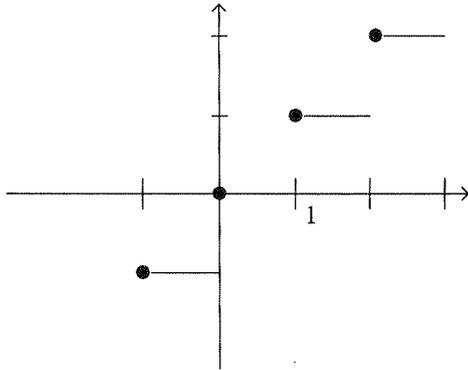


figura 6

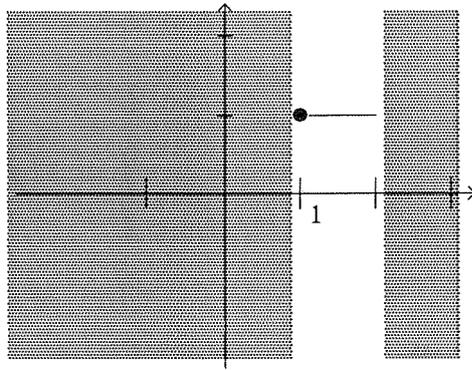


figura 7

La funzione f non è continua in 1 perché non c'è il limite in 1 (il limite da destra e quello da sinistra sono diversi). Invece, se ci limitiamo al solo insieme B , il punto 1 non ha punti "vicini a sinistra" e quindi abbiamo la continuità di $f|_B$ nel punto 1 . È come se guardassimo al grafico di f ristretto solo alla striscia $[1, 2[\times \mathbb{R}$: se "copriamo" il resto del grafico che non ci interessa, come in figura 7, "vediamo" una funzione manifestamente continua in quanto costante. Se invece rimuoviamo la "copertura", la continuità "salta" in 1 . □

Osservazione 1 Nell'enunciato del teorema 9.3 ho supposto che fosse $f(A) \subseteq B$, in modo da poter considerare la composizione delle funzioni f e g . Quindi una condizione molto naturale. Ma in realtà spesso il quadro che ci si trova dinnanzi è diverso. Ad esempio, consideriamo $\sqrt{x^2-1}$. Cioè $f(x) = x^2-1$, $g(x) = \sqrt{x}$. E' $A = \mathbb{R}$, e $f(A) = [-1, +\infty[\not\subseteq [0, +\infty[= B$. E allora? Non possiamo usare il teorema? Certo che sì. Tutti sanno che in un caso come questo ci si riferisce alla funzione composta come funzione definita su A' , "il più grosso insieme t.c. $f(A') \subseteq B$ ". Ovverossia $A' = f^{-1}(B)$. Quindi si applica il teorema ad $f|_{A'}$, la quale sarà continua. Questo è quanto ho fatto, per esempio, nell'osservazione 9.1, senza farne esplicita menzione. □

11. Completezza di \mathbb{R} e continuità

In questo paragrafo proverò due importantissimi teoremi relativi alle funzioni continue su un intervallo. Quel che è particolarmente interessante di questi teoremi è che hanno un enunciato "intuitivamente evidente". Uno dice che una funzione f la quale sia continua su $[a,b]$ ha massimo e minimo su $[a,b]$: e sembra ragionevole. Ancora più ragionevole è l'altro teorema (detto degli zeri). Esso garantisce che una funzione continua su $[a,b]$, con $f(a) < 0$ ed $f(b) > 0$, si deve annullare in qualche punto di $[a,b]$. Cioè, deve esistere $c \in]a,b[$ t.c. $f(c) = 0$. Si guardi la figura 8 . Si provi ad unire i due punti $(a, f(a))$ e $(b, f(b))$ con una "linea continua" (più precisamente, una linea che sia grafico di una funzione continua su $[a,b]$): si vedrà che, comunque si tracci il disegno, c'è per forza un punto in cui la "linea attraversa l'asse delle x ". Chiaramente, senza il requisito della continuità, questo non è affatto detto: si

veda la funzione $g(x) = \begin{cases} -1 & \text{per } x \in [a, (a+b)/2] \\ 1 & \text{per } x \in](a+b)/2, b] \end{cases}$ il cui grafico è disegnato alla figura 9.

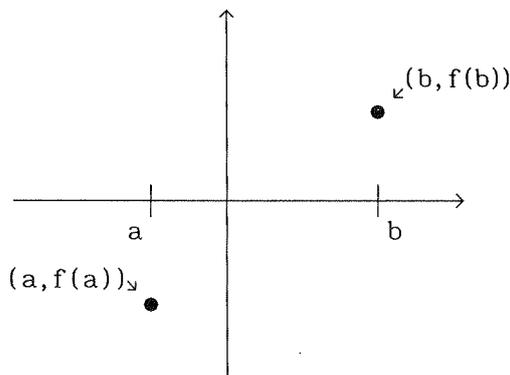


figura 8

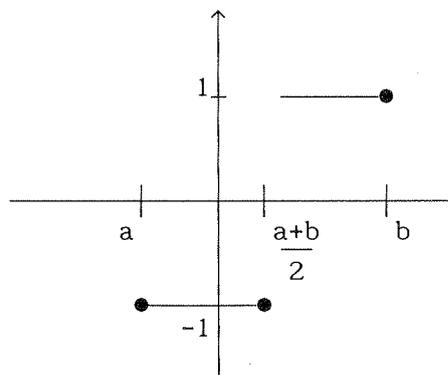


figura 9

Quindi il teorema degli zeri esprime una proprietà "intuitivamente evidente" delle funzioni continue. Di più. Invito a riflettere su come questa proprietà sia strettamente inerente all'idea stessa di continuità. Come vedremo, per ottenerla dovremo usare l'assioma 16, e cioè la completezza di \mathbb{R} . Anzi, di più. Vedremo poi un esempio che mostra come in \mathbb{Q} il "teorema degli zeri" sia falso: quindi senza la proprietà di completezza non riusciamo ad ottenere questo risultato. Morale, il teorema degli zeri ci dà una doppia conferma della giustezza della strut-

tura matematica che abbiamo elaborato per descrivere i fenomeni continui. Dico doppia perché essa dà una conferma a due aspetti distinti. Il primo aspetto riguarda i numeri reali: sono davvero, da questo punto di vista, la struttura matematica giusta da utilizzare (i razionali non vanno bene). Il secondo aspetto riguarda la continuità. Se andate a rileggere cosa dicevo nel § 8, vedrete che non avevo dato una giustificazione intuitiva della definizione di continuità adottata: avevo anzi dato delle ragioni per evitare questo approccio. Naturalmente ciò lasciava aperto il problema se la definizione matematica precisa data della continuità corrispondesse alla nostra idea intuitiva di continuità. E' proprio il teorema degli zeri che dà una importantissima conferma, a posteriori, della giustezza della scelta fatta per quel che riguarda la definizione di continuità. Già che siamo in argomento, dirò che i meriti principali della definizione di continuità che abbiamo dato stanno, oltre che nel riuscire a dimostrare questo teorema, nella grande praticità d'uso della nozione di continuità (somma, prodotto, composizione e, come vedremo, inversa di una funzione continua sono ancora funzioni continue). Non bisogna però pensare che la definizione di funzione continua sia l'"ideale". Vi sono degli esempi di funzioni continue che non rispondono minimamente all'idea intuitiva che noi abbiamo (dirò qualcosa a questo proposito nel capitolo IV dopo la definizione di derivata).

Passo ora, dopo tanti discorsi, alla parte formale. Comincio col teorema di Weierstrass.

Teorema 1 (di Weierstrass) Sia $f:[a,b] \longrightarrow \mathbb{R}$, continua su $[a,b]$. Allora f ha massimo e minimo su $[a,b]$. \square

Dimostrazione Consideriamo $f([a,b])$. Dobbiamo dimostrare che questo sottoinsieme di \mathbb{R} ha massimo e minimo. Dimostrerò che ha massimo, in quanto il caso del minimo è perfettamente analogo (può anche essere notato che il massimo di f è l'opposto del minimo di $-f$: questo suggerisce un'altra strada per provare l'esistenza del minimo).

Due casi possono verificarsi, a priori: $f([a,b])$ è superiormente limitato o non lo è. Nel primo caso, $\exists \lambda \in \mathbb{R}$ t.c. $\lambda = \sup f([a,b])$, in quanto $f([a,b]) \neq \emptyset$, essendo $[a,b] \neq \emptyset^4$. Nel secondo caso non c'è il \sup : si usa dire, per convenzione, che $\sup f([a,b]) = +\infty$.

⁴ Ricordo quanto detto nell'osservazione I.4.1: salvo esplicito avviso contrario, assumo che sia $a < b$ quando parlo di intervalli.

Ebbene, proveremo che in entrambi i casi possiamo determinare una successione $x_n \in [a, b] \quad \forall n \in \mathbb{N}$ e con $f(x_n) \longrightarrow \sup f([a, b])$. Cioè $f(x_n) \longrightarrow \lambda$ se $f([a, b])$ è superiormente limitato, $f(x_n) \longrightarrow +\infty$ altrimenti.

Nel primo caso, essendo $\lambda = \sup f([a, b])$, abbiamo che, dato $n \in \mathbb{N}$, $\exists y \in f([a, b])$ t.c. $\lambda - (1/n) < y \leq \lambda$. Ma allora c'è $x \in [a, b]$ t.c. $\lambda - (1/n) < f(x) \leq \lambda$. Scegliamo uno di questi x e indichiamolo con x_n . Abbiamo così la successione desiderata. Infatti, da $\lambda - (1/n) < f(x_n) \leq \lambda$ deduciamo subito per il teorema dei due carabinieri che $f(x_n) \longrightarrow \lambda$. Nel secondo caso, cioè se $f([a, b])$ non è superiormente limitato, vorrà dire che $\forall n \in \mathbb{N} \exists y \in f([a, b])$ t.c. $y > n$. E quindi otteniamo una successione $x_n \in [a, b]$ t.c. $f(x_n) > n$, per cui $f(x_n) \longrightarrow +\infty$, come era stato anticipato.

Abbiamo quindi a disposizione la successione $f(x_n)$ che è regolare. Soffermiamoci ora sulla successione di termine generale x_n . Non c'è nessuna ragione per cui possiamo affermare che converga. Però il teorema di Bolzano-Weierstrass ci assicura che ha una sottosuccessione estratta convergente. Cioè $\exists x_0 \in \mathbb{R}$ ed $\exists k: \mathbb{N} \longrightarrow \mathbb{N}$ strettamente crescente t.c. $x_{k_n} \longrightarrow x_0$.

E' immediato notare che $x_0 \in [a, b]$: infatti è $a \leq x_{k_n} \leq b$ e quindi $a \leq x_0 \leq b$ per il teorema di permanenza del segno.

Usiamo finalmente la continuità di f per notare che, essendo $x_{k_n} \longrightarrow x_0$, sarà $f(x_{k_n}) \longrightarrow f(x_0)$. Notiamo anche che, essendo $f(x_{k_n})$ una sottosuccessione estratta da $f(x_n)$, avremo che $f(x_{k_n})$ avrà lo stesso limite di $f(x_n)$. Ma allora dovremo avere:

nel 1° caso $f(x_{k_n}) \longrightarrow f(x_0)$ e $f(x_{k_n}) \longrightarrow \lambda$, da cui per l'unicità del limite $f(x_0) = \lambda$, vale a dire che $\lambda \in f([a, b])$, e quindi che λ non solo è il sup, ma è in realtà il massimo di f su $[a, b]$ (e x_0 è il punto di massimo).

nel 2° caso, $f(x_{k_n}) \longrightarrow f(x_0)$ e $f(x_{k_n}) \longrightarrow +\infty$: queste sono due affermazioni tra loro contraddittorie, per cui il 2° caso in realtà non si può verificare. \square

Vorrei fare una osservazione sul "come si usa" questo teorema. Il problema è che quasi nessuna delle funzioni che ci si trova davanti è definita su un intervallo chiuso e limitato, però si è interessati all'esistenza o meno, per quella funzione, di un massimo o minimo su un certo intervallo $[a, b]$. Come fare? Il "trucco" è semplice. Sia data una funzione reale di variabile reale

$g: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$. Se g è continua su un intervallo $[a,b]$, allora $f = g|_{[a,b]}$ sarà anch'essa continua su $[a,b]$ e quindi basta applicare il teorema di Weierstrass a questa f , che rientra perfettamente nelle ipotesi del teorema. Vorrei osservare che astuzie di questo genere ricorrono spesso per colmare il divario che c'è tra l'enunciato di un teorema e quanto invece serve nella specifica sua applicazione.

Passiamo ora alla dimostrazione dell'altro importante teorema sulle funzioni continue.

Teorema 2 (degli zeri) Sia $f: [a,b] \longrightarrow \mathbb{R}$, con $f(a) < 0 < f(b)$. Allora $\exists c \in [a,b]$ t.c. $f(c) = 0$. \square

Dimostrazione La dimostrazione sfrutta un'idea molto semplice. Considero il punto di mezzo dell'intervallo $[a,b]$, e cioè $(a+b)/2$. Se $f((a+b)/2) = 0$, ho finito. Se non sono così fortunato, sarà $f((a+b)/2) < 0$ oppure $f((a+b)/2) > 0$. Nel primo caso, considero l'intervallo $[(a+b)/2, b]$, nel secondo l'intervallo $[a, (a+b)/2]$: in entrambi i casi mi ritrovo di nuovo nella stessa condizione di partenza. Cioè, la f assume valore strettamente minore di zero nel primo estremo e maggiore di zero nel secondo. Se ho fiducia nella validità di questo teorema, mi aspetto che f si annulli in qualche punto di questo intervallo, visto che mi ritrovo con le ipotesi di partenza soddisfatte con questo nuovo intervallo. Solo che ora la lunghezza dell'intervallo è metà di quello di partenza. Allora l'idea è: itero la procedura. Così facendo, dovrei riuscire ad "intrappolare" un punto in cui la f si annulla, "stringendolo" dentro intervalli sempre più piccoli. In effetti, funziona. Vediamo quindi la dimostrazione formale.

Intendo formalizzare completamente la dimostrazione a scopi esplicativi (per far capire che bisogna stare attenti a dire: "e così via..."). In effetti, l'idea è semplice, ma ci accorgeremo che l'eventualità "fortunata" che f si annulli in uno dei punti della suddivisione rende molto complicata la formalizzazione precisa.

Potrei procedere nel modo seguente.

Definisco per induzione due successioni $\alpha, \beta: \mathbb{N} \longrightarrow \mathbb{R}$, tali da soddisfare le seguenti proprietà:

- (1) $\alpha_1 = a$, $\beta_1 = b$
- (2) $\alpha_n < \beta_n \quad \forall n \in \mathbb{N}$
- (3) $\beta_n - \alpha_n = (b-a)/2^{n-1}$

$$(4) \quad f(\alpha_n) < 0 < f(\beta_n) \quad \forall n \in \mathbb{N} .$$

Proverò inoltre che:

$$(5) \quad \alpha \text{ è debolmente crescente e } \beta \text{ è debolmente decrescente.}$$

L'idea è che chiamerò $[\alpha_1, \beta_1]$ l'intervallo di partenza $[a, b]$, poi $[\alpha_2, \beta_2]$ quello ottenuto dopo la prima suddivisione in due, e così via. Per

esempio, se $f(a) < 0 < f(\frac{a+b}{2})$, sarà $[\alpha_2, \beta_2] = [\alpha_1, \frac{\alpha_1 + \beta_1}{2}] = [a, \frac{a+b}{2}]$.

Se poi $f(\frac{\alpha_2 + \beta_2}{2}) < 0 < f(\beta_2)$, avremo:

$$[\alpha_3, \beta_3] = \left\{ \frac{\alpha_2 + \beta_2}{2}, \beta_2 \right\} = \left[\frac{\alpha_1 + \frac{\alpha_1 + \beta_1}{2}}{2}, \frac{\alpha_1 + \beta_1}{2} \right] = \left[\frac{a + \frac{a+b}{2}}{2}, \frac{a+b}{2} \right] .$$

E così via...

Se uno riflette un attimo sulle successioni α e β così ottenute, si rende conto che effettivamente la proprietà (1)+(5) sembrano essere ragionevoli.

Riprendiamo quindi il filo del discorso formale.

Naturalmente ho bisogno di definire α e β per $n=1$ e poi di definire il "passo induttivo".

BASE: definiamo $\alpha_1 = a$ e $\beta_1 = b$; è immediato constatare che la proprietà (1) è soddisfatta e che le proprietà (2)+(4) valgono per $n=1$.

PASSO INDUTTIVO: per ogni $n \in \mathbb{N}$, avendo a disposizione α_n e β_n e la validità della proprietà (2)+(4) per quel $n \in \mathbb{N}$, devo definire α_{n+1} e β_{n+1} e constatare che le proprietà (2)+(4) valgono anche per $n+1$. Considero il punto $(\alpha_n + \beta_n)/2$.

Se $f((\alpha_n + \beta_n)/2) < 0$, definisco $\alpha_{n+1} = (\alpha_n + \beta_n)/2$ e $\beta_{n+1} = \beta_n$.

Se $f((\alpha_n + \beta_n)/2) > 0$, definisco $\alpha_{n+1} = \alpha_n$ e $\beta_{n+1} = (\alpha_n + \beta_n)/2$.

Ma se $f((\alpha_n + \beta_n)/2) = 0$, cosa faccio? Invito lo studente diligente a riflettere su come si possa procedere. La dimostrazione prosegue nella pagina seguente.

Spero che sia chiaro dove è la difficoltà: se sono "fortunato" e trovo un punto in cui f vale zero durante la procedura di bisezione, non posso continuare a definire le successioni α e β in modo da rispettare le proprietà (1)÷(5). Uno potrebbe pensare che si possa "ammorbidire" il requisito (5), richiedendo solo che sia $f(\alpha_n) \leq f(\beta_n)$. Sfortunatamente questa facile via d'uscita non funziona: basti pensare al caso in cui $f(x) = x-1$ ed $a = 0$, $b = 2$: in tal caso, $f(1) = 0$ e, qualunque sia la metà dell'intervallo che scegliamo per continuare a definire α e β , lo "zero" di f ci scappa.

Ci troviamo di fronte ad una struttura molto comune in programmazione: abbiamo un ciclo iterativo con all'interno una condizione che prevede una uscita "laterale" dal ciclo.

Possiamo rimediare a questo con una migliore formalizzazione del processo induttivo, la quale tenga conto di questa possibilità di "uscita prematura dal ciclo".

Avviso per lo studente che sta per stufarsi: d'ora in poi inizia una formalizzazione corretta della dimostrazione.

Per dimostrare il teorema degli zeri, proviamo innanzi tutto la seguente affermazione:

$$\star \left\{ \begin{array}{l} \text{O definiamo per induzione due successioni soddisfacenti le proprietà} \\ \text{(1)÷(4), oppure troviamo un punto } d \in [a,b] \text{ t.c. } f(d) = 0 . \end{array} \right.$$

Procediamo naturalmente per induzione, come già avevamo iniziato a fare.

BASE: definiamo $\alpha_1 = a$ e $\beta_1 = b$; è immediato constatare che la proprietà (1) è soddisfatta e che le proprietà (2)÷(4) valgono per $n = 1$.

PASSO INDUTTIVO: per ogni $n \in \mathbb{N}$, avendo a disposizione α_n e β_n e la validità delle proprietà (2)÷(4) per quel $n \in \mathbb{N}$, devo definire α_{n+1} e β_{n+1} e constatare che le proprietà (2)÷(4) valgono anche per $n+1$. Considero il punto $(\alpha_n + \beta_n)/2$.

Se $f((\alpha_n + \beta_n)/2) < 0$, definisco $\alpha_{n+1} = (\alpha_n + \beta_n)/2$ e $\beta_{n+1} = \beta_n$.

Se $f((\alpha_n + \beta_n)/2) > 0$, definisco $\alpha_{n+1} = \alpha_n$ e $\beta_{n+1} = (\alpha_n + \beta_n)/2$.

Se $f((\alpha_n + \beta_n)/2) = 0$, allora poniamo $d = \frac{\alpha_n + \beta_n}{2}$.

Vediamo che la condizione (\star) è soddisfatta.

I casi sono due: o per qualche $n \in \mathbb{N}$ si ha $f((\alpha_n + \beta_n)/2) = 0$, e allora

sarà sufficiente chiamare $d = \frac{\alpha_n + \beta_n}{2}$, oppure è $f((\alpha_n + \beta_n)/2) \neq 0 \quad \forall n \in \mathbb{N}$.

Si tratta ora di verificare che in questo secondo caso abbiamo a disposizione due successioni soddisfacenti (1)÷(4).

Intanto, la procedura descritta fornisce due successioni, cioè due applicazioni da \mathbb{N} in \mathbb{R} . Infatti α è definita in $1 \in \mathbb{N}$ ($\alpha_1 = a$) e, se α è definita in $n \in \mathbb{N}$, allora è definita in $n+1 \in \mathbb{N}$ (come $\alpha_{n+1} = \alpha_n$ oppure $\alpha_{n+1} = \frac{\alpha_n + \beta_n}{2}$ a seconda che sia $f(\frac{\alpha_n + \beta_n}{2}) < 0$ oppure $f(\frac{\alpha_n + \beta_n}{2}) > 0$). Analogamente dicasi per β .

La condizione (1) è ovviamente soddisfatta.

Proviamo, per induzione, che la (2) è soddisfatta. Per $n=1$ è vero. Sia $n \in \mathbb{N}$: se $\alpha_n < \beta_n$, allora abbiamo due casi: $\alpha_{n+1} = (\alpha_n + \beta_n)/2$ e $\beta_{n+1} = \beta_n$ oppure $\alpha_{n+1} = \alpha_n$ e $\beta_{n+1} = (\alpha_n + \beta_n)/2$. Nel primo caso, possiamo asserire che $\alpha_n < \beta_n$ implica $\alpha_n + \beta_n < 2 \cdot \beta_n$ e quindi $(\alpha_n + \beta_n)/2 < \beta_n$, cioè $\alpha_{n+1} < \beta_{n+1}$. Il secondo caso è analogo.

Per la (3), essa è vera per $n=1$. Dato $n \in \mathbb{N}$, se è $\beta_n - \alpha_n = (b-a)/2^{n-1}$, per come sono definite α_{n+1} e β_{n+1} , si ha ovviamente che $\beta_{n+1} - \alpha_{n+1} = (\beta_n - \alpha_n)/2$, cioè $\beta_{n+1} - \alpha_{n+1} = (b-a)/2^n$.

La (4) è ovvia.

Abbiamo quindi provato la validità dell'affermazione (*).

Ricordo che l'obbiettivo è quello di provare l'esistenza di uno "zero" per f : ovviamente, se abbiamo trovato un d t.c. $f(d) = 0$, siamo a posto. Se così non è, abbiamo a disposizione le due successioni soddisfacenti le proprietà (1)÷(4). Vedremo di ricavare anche in questo caso uno "zero" per f .

Proviamo che vale la proprietà (5), cioè la monotonia di α e β (naturalmente, nel caso in cui non si abbia mai che $f(\frac{\alpha_n + \beta_n}{2}) = 0$). Anche questa la otteniamo per induzione. Per $n=1$, abbiamo che $\alpha_2 = \alpha_1$ oppure $\alpha_2 = \frac{\alpha_1 + \beta_1}{2}$ a seconda dei casi; ma in ogni caso otteniamo $\alpha_2 \geq \alpha_1$. Se poi è dato $n \in \mathbb{N}$ e si ha $\alpha_n \leq \alpha_{n+1}$, si deduce che $\alpha_{n+1} \leq \alpha_{n+2}$ perché a seconda dei casi si avrà $\alpha_{n+2} = \alpha_{n+1}$ oppure $\alpha_{n+2} = \frac{\alpha_{n+1} + \beta_{n+1}}{2} \geq$ (per la (2)) $\geq \frac{\alpha_{n+1} + \alpha_{n+1}}{2} = \alpha_{n+1}$. E' così provato che α è debolmente crescente. Analogamente si ottiene la debole decre-

scenza di β .

Avere la monotonia di α e β ci serve per poter asserire che sono convergenti, purché siano limitate. Ma limitate lo sono: dalla (2) e dalla (5) otteniamo che $\forall n \in \mathbb{N} \quad \alpha_1 \leq \alpha_n < \beta_n \leq \beta_1$; vale a dire che α è superiormente limitata da β_1 e che β è inferiormente limitata da α_1 .

Grazie al teorema II.12.1, possiamo affermare che $\exists \hat{\alpha} = \lim_{n \rightarrow \infty} \alpha_n$ e $\hat{\beta} = \lim_{n \rightarrow \infty} \beta_n$.

Ma è $\beta_n - \alpha_n = (b-a)/2^{n-1}$, per la (3). Quindi $\lim_{n \rightarrow \infty} (\beta_n - \alpha_n) = 0$. Ne segue che $\hat{\beta} = \lim_{n \rightarrow \infty} \beta_n = \lim_{n \rightarrow \infty} (\alpha_n + (\beta_n - \alpha_n)) = \hat{\alpha}$. Cioè: le due successioni hanno lo stesso limite.

Per la continuità di f , $\lim_{n \rightarrow \infty} f(\alpha_n) = f(\hat{\alpha}) = f(\hat{\beta}) = \lim_{n \rightarrow \infty} f(\beta_n)$. Essendo $f(\alpha_n) < 0$ e $f(\beta_n) > 0 \quad \forall n \in \mathbb{N}$ (per la (4)), il teorema di permanenza del segno ci permette di affermare che $f(\hat{\alpha}) \leq 0$ e $f(\hat{\beta}) \geq 0$. Ma è $\hat{\alpha} = \hat{\beta}$, quindi $f(\hat{\alpha}) = f(\hat{\beta})$. E quindi $f(\hat{\alpha}) = 0$.

Abbiamo quindi ottenuto lo "zero" che cercavamo per f . \square

12. Conseguenze del teorema degli zeri

La prima ovvia generalizzazione del teorema degli zeri che si può fare, consiste nell'osservare che in generale ogni valore reale compreso tra $f(a)$ e $f(b)$ sarà assunto da una funzione continua. Se abbiamo $y \in]f(a), f(b)[$ (supponiamo di essere nel caso in cui $f(a) < f(b)$), basterà considerare $g(x) = f(x) - y$ e applicare il teorema degli zeri a g . Suggestivo di fare un disegno per capire meglio l'idea. Si può anche dire di meglio: ogni numero reale y che sia compreso tra due valori u, v assunti dalla funzione, starà esso pure nell'immagine. Ma ancora di più si può enunciare una estensione del teorema degli zeri che ne coglie la "vera essenza". E cioè: se abbiamo una funzione continua definita su un intervallo, la sua immagine è ancora un intervallo. Dimostrare questo teorema è facilissimo, se però abbiamo capito davvero che cosa è un intervallo.

Definizione 1 Un sottoinsieme E di \mathbb{R} si dice intervallo se gode della seguente proprietà:

$$\forall x, y, z \in \mathbb{R} \quad ((x, z) \in E \text{ e } x < y < z) \Rightarrow y \in E) . \square$$

Osservazione 1 L'uno o l'altro dei " $<$ " o entrambi possono essere sostituiti con " \leq ". \square

Spero che qualche attento lettore si sia accorto che c'è qualcosa che non va. E infatti abbiamo già dato la definizione di intervallo! Vedi definizione I.4.2. Ora, non è lecito dare diverse definizioni della stessa cosa, senno' si crea confusione (a dir poco...). E' tollerabile solo se si dimostra che le due definizioni date sono equivalenti. E così è.

Teorema 1 I sottoinsiemi di \mathbb{R} che soddisfano a definizione 1 sono i seguenti:

$$\emptyset, \mathbb{R}, [a, b],]a, b[,]a, b],]a, b[,]-\infty, b],]-\infty, b[, [a, +\infty[, [a, +\infty[. \square$$

Dimostrazione Che tutti gli insiemi sopra elencati soddisfino la definizione 1 è immediato. Occorre dimostrare che non ce ne sono altri.

Sia allora E soddisfacente la definizione 1. Se $E = \emptyset$, non abbiamo nulla da dimostrare. Altrimenti, noto che si possono verificare quattro casi, esaustivi e mutuamente escludentesi.

- i) E è limitato
- ii) E non è limitato inferiormente ma è limitato superiormente

iii) E non è limitato superiormente ma è limitato inferiormente

iv) E non è limitato né superiormente né inferiormente

Dimostreremo allora che:

nel caso i) , \exists due numeri reali a, b , con $a \leq b$, t.c. $E = [a, b]$ oppure $E = [a, b[$ oppure $E =]a, b]$ oppure $E =]a, b[$;

nel caso ii) , \exists un numero reale b t.c. $E =]-\infty, b]$ oppure $E =]-\infty, b[$;

nel caso iii) , \exists un numero reale a t.c. $E = [a, +\infty[$ oppure $E =]a, +\infty[$;

nel caso iv) . $E = \mathbb{R}$.

Vedrò in dettaglio la dimostrazione del caso iii). La dimostrazione degli altri casi è analoga. Come preannunciato, si possono presentare due alternative: $E = [a, +\infty[$, oppure $E =]a, +\infty[$. La prima cosa da scoprire è che sia a . Poiché $E \neq \emptyset$ ed E è inferiormente limitato, esiste $\inf E$. Ed è proprio $\inf E$ il numero reale a che ci serve. Ovverossia, le due alternative sono: o $E =]\inf E, +\infty[$, oppure $E =]\inf E, +\infty[$. Da cosa dipende quale delle due si presenta? Da una circostanza: se $\inf E \in E$, in tal caso $E =]\inf E, +\infty[$; se $\inf E \notin E$, allora $E =]\inf E, +\infty[$.

Vediamo la prima alternativa. Ricapitoliamo un pò cosa abbiamo a disposizione. C'è $E \neq \emptyset$, che è inferiormente limitato, è $\inf E \in E$ ed E non è superiormente limitato. Voglio ottenere che $E =]\inf E, +\infty[$. E' ovvio che $E \subseteq]\inf E, +\infty[$. Per provare il viceversa, prendiamo $y \in]\inf E, +\infty[$ e proviamo che $y \in E$. Poiché E non è superiormente limitato, $\exists z \in E$ t.c. $y < z$. Ma allora possiamo applicare la definizione di intervallo: essendo $\inf E \leq y \leq z$, possiamo garantire che $y \in E$.

Anche nella seconda alternativa c'è solo da dimostrare che $] \inf E, +\infty[\subseteq E$. Dato $y \in] \inf E, +\infty[$, poiché $y > \inf E$, c'è $x \in E$ t.c. $x < y$. Come nella prima alternativa, ci sarà anche $z \in E$ t.c. $y < z$. Poiché $x < y < z$ ed $x, z \in E$, ne segue dalla definizione di intervallo che $y \in E$. \square

Il teorema degli zeri allora diventa:

Teorema 2 Sia $f: I \longrightarrow \mathbb{R}$, I intervallo di \mathbb{R} . Se f è continua su I , allora $f(I)$ è un intervallo. \square

Dimostrazione Siano $u, w \in f(I)$ e sia v t.c. $u < v < w$. Devo dimostrare che $v \in f(I)$. Poiché $u, w \in f(I)$, $\exists x, z \in I$ t.c. $u = f(x)$ e $w = f(z)$. Applico le considerazioni fatte all'inizio del paragrafo ad $f|_{[x, z]}$ (se $x < z$, sennò a $[z, x]$). Poiché $u = f(x) < v < f(z) = w$, c'è $y \in [x, z]$ t.c.

$f(y) = v$. Ma $[x,z] \subseteq I$ poiché $x,z \in I$ ed I è un intervallo. Quindi ho trovato $y \in I$ t.c. $f(y) = v$. Cioè $v \in f(I)$, come volevasi dimostrare. \square

13. Inversione di funzioni

Lo scopo principale di questo paragrafo è di dare conto di una terminologia usata in analisi che si discosta da quella standard. Si tratta delle funzioni invertibili. A livello "insiemistico", è noto che una funzione $f:A \longrightarrow B$ è invertibile, per definizione, se $\exists g:B \longrightarrow A$ t.c. $f \circ g = \text{Id}_B$ e $g \circ f = \text{Id}_A$ (naturalmente g si chiama inversa di f e si indica con f^{-1}). E' anche noto che una condizione necessaria e sufficiente affinché una funzione sia invertibile è che sia bigettiva⁵. Invece in analisi è consuetudine dire che una funzione reale di variabile reale $f:A \longrightarrow \mathbb{R}$ è invertibile se f è iniettiva. C'è quindi un contrasto di terminologia, non trascurabile in quanto riguarda una nozione fondamentale. La ragione di questa divergenza terminologica sta nel fatto che la nozione standard di funzione invertibile è poco "compatibile" con la nozione di funzione reale di variabile reale.

Una funzione reale di variabile reale è una $f:A \longrightarrow \mathbb{R}$: quindi, se invertibile "insiemisticamente", la sua inversa dovrebbe essere $g:\mathbb{R} \longrightarrow A$, e quindi g non sarebbe una funzione reale di variabile reale, a meno che $A = \mathbb{R}$. Una tale restrizione è eccessiva, per gli usi dell'analisi. Ad esempio, il logaritmo non sarebbe definibile come inversa dell'esponenziale.

Cosa si fa, quindi? Sia $f:A \longrightarrow \mathbb{R}$, iniettiva. Naturalmente, se come codominio prendiamo $f(A)$, otteniamo una funzione⁶, $\tilde{f}:A \longrightarrow f(A)$ che è anche suriettiva, e quindi bigettiva. Allora \tilde{f} è invertibile. Abbiamo quindi $(\tilde{f})^{-1}:f(A) \longrightarrow A$. A meno che non sia $A = \mathbb{R}$, questa funzione $(\tilde{f})^{-1}$ non sarà una funzione reale di variabile reale. Allora la modifico, semplicemente "allargando" il codominio, cioè considerando $(\tilde{f})^{-1}:f(A) \longrightarrow \mathbb{R}$, definita ovviamente come $(\tilde{f})^{-1}(x) = (\tilde{f})^{-1}(x) \quad \forall x \in f(A)$ (vedi nota precedente).

Notiamo che la funzione ottenuta alla fine di tutto questo rigiro, e cioè $(\tilde{f})^{-1}$, è una funzione reale di variabile reale. E che, anche se non è l'inversa di f in termini rigorosi, ne è però una parente assai prossima. Infatti è una

⁵ Si usa anche, come sinonimo, il termine "corrispondenza biunivoca".

⁶ Ovviamente $\tilde{f}(a) = f(a) \quad \forall a \in A$. Le indico però diversamente in quanto f ed \tilde{f} hanno un diverso codominio, e pertanto in termini rigorosi sono due funzioni diverse.

"gemella" di $(\tilde{f})^{-1}$, che è l'inversa di \tilde{f} , a sua volta "gemella" di f . Le uniche modifiche fatte sono state quelle di restringere il codominio nel passare da f a \tilde{f} e di allargare⁷ il codominio per passare da $(\tilde{f})^{-1}$ ad $\tilde{\tilde{f}}^{-1}$. Quindi abbiamo una funzione $\tilde{\tilde{f}}^{-1}$, che è una funzione reale di variabile reale che assomiglia molto ad una "inversa" di f . Ecco perché in analisi una funzione f , che sia solo iniettiva, viene chiamata invertibile, ed ecco perché la funzione $\tilde{\tilde{f}}^{-1}$ viene chiamata inversa di f . E, naturalmente, di consueto poi si usa semplicemente la notazione f^{-1} , anche se imprecisa, anziché la troppo pesante $\tilde{\tilde{f}}^{-1}$.

Tutta questa premessa per enunciare il seguente:

Teorema 1 Sia $f: I \longrightarrow \mathbb{R}$, I intervallo, f continua su I e strettamente crescente. Allora f è invertibile ed $f^{-1}: f(I) \longrightarrow \mathbb{R}$ è strettamente crescente e continua. \square

Dimostrazione Fatte le premesse, doverose, riguardo alla terminologia, tutte le affermazioni di questo teorema sono molto facili da provare (e sono lasciate come esercizio al lettore), tranne la continuità. Per questa, un argomento molto convincente consiste nel ricordare che il grafico di f^{-1} si ottiene da quello di f usando come intermediari un paio di assi cartesiani e facendo una riflessione rispetto alla bisettrice del I e III quadrante. Questa però non è una dimostrazione. Non abbiamo provato un teorema che garantisca questa ovvia constatazione "grafica". Quindi occorre fare una dimostrazione, per la quale mi limito a rinviare al C-S. \square

⁷ Proprio perché si usano due specie di operazioni (restringimento/allargamento) tra di loro "rovesciate", ho usato l'accorgimento di mettere il segno \sim sopra in un caso e sotto nell'altro.

14. Conclusioni

Oggetto di studio di questo capitolo sono le funzioni reali di variabile reale. Dopo aver precisato con la definizione 1.1 che cosa siano, ho lasciato ampio spazio all'idea di grafico di una funzione e alla sua rappresentazione. Naturalmente, in questo capitolo l'attenzione si è soffermata sui limiti per funzioni reali di variabile reale. Ho dapprima visto il caso del limite finito in un punto x_0 , cercando di mettere in evidenza come il limite sia definito in modo da non avere nulla a che fare con il valore che f assume in x_0 .

Dopo aver provato vari risultati, ed avere inoltre esteso la definizione di limite in modo da coprire vari altri casi oltre a quello "standard" di " $l-x_0$ ", abbiamo visto la definizione di continuità.

L'idea di continuità in un punto è racchiusa semplicemente nel tentativo di conciliare il comportamento di limite con il valore assunto in quel punto: a costo di essere noioso, ribadisco che si tratta di dare concetti del tutto indipendenti tra loro.

Ho dimostrato varie conseguenze della proprietà di continuità in un punto e su un insieme, mostrando la maneggevolezza della nozione di continuità.

Il teorema di Weierstrass e il teorema degli zeri, infine, costituiscono un paio di risultati di evidente contenuto intuitivo (si veda quanto detto nelle conclusioni del precedente capitolo): è un successo per l'umanità essere stata in grado di elaborare una struttura matematica quale quella dei numeri reali, che permette di provare la validità di tali risultati.

CAPITOLO IV

CALCOLO DIFFERENZIALE

1. L'idea di derivata

Questo capitolo ed il successivo costituiscono la parte più tipica, più specifica, dell'analisi. In realtà l'analisi è essenzialmente calcolo differenziale e integrale. L'idea stessa di limite può essere considerata ancillare rispetto a loro.

Occupiamoci quindi di calcolo differenziale. Il primo passo è capire la definizione di derivata. Si potrebbe anche dare subito la definizione e iniziare ad enunciare e dimostrare teoremi su teoremi. Dopotutto, chi è sopravvissuto fin qui non dovrebbe avere soverchi problemi ad assimilare quanto verrebbe proposto. Tuttavia, l'importanza dell'argomento rende opportuna una discussione preliminare.

La prima cosa da mettere in chiaro è che l'idea di derivata appare all'inizio come utile solo "in funzione di". E cioè per poter dare la definizione di retta tangente. In effetti, il calcolo differenziale nasce proprio dal tentativo di trovare la retta tangente a una "generica curva". Volendo essere onesti, vi è anche un'altra importante sorgente da cui si sprigiona l'idea di derivata: è il tentativo di trovare la velocità istantanea di un punto che si muove di moto "arbitrario". Dovendo fare una scelta, privilegerò il primo aspetto: ritornerò in seguito, però, sull'idea di velocità.

Quindi: come definire la retta tangente a una curva? Prima dell'analisi, uno conosce la definizione di retta tangente a una circonferenza, come retta che ha in comune con la circonferenza un solo punto. Ha anche visto qualche problema connesso all'idea di retta tangente alle coniche¹, con l'ausilio della geometria analitica. In questo caso il trucco era: mettere $\Delta = 0$, per avere due radici coincidenti. Parlando francamente, il tutto era ben lontano dall'essere giustificato.

¹ Parabola, ellisse (e quindi anche cerchio) ed iperbole.

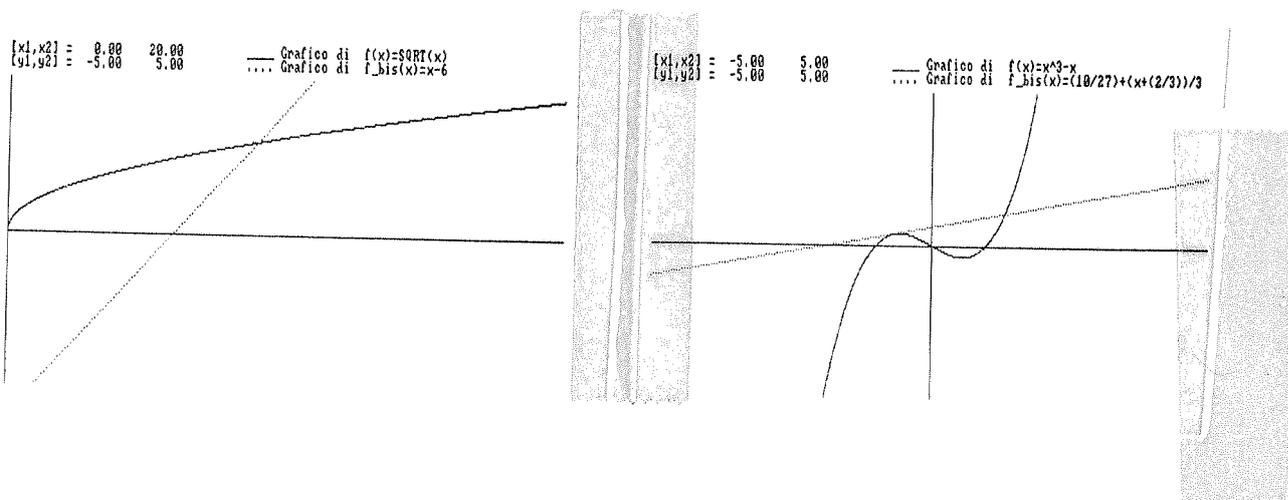
Come piccolo test, chi legge saprebbe rispondere alle seguenti domande?

1. Chi garantisce che per un punto (o per quali punti) c'è una (una sola?) retta tangente alla parabola?
2. Perché la retta tangente viene fuori dalla procedura di porre $\Delta = 0$?
3. Due "radici coincidenti" sono per caso due radici distinte "infinitamente vicine"?

Sorvolando sulle lacune esistenti in merito alla giustificazione delle procedure adottate, emerge comunque l'idea di retta tangente ad una curva come retta che ha in comune un punto con quella curva.

Possiamo adottare questa come definizione generale? Ovviamente no.

Ad esempio, $y = x-6$ ha un solo punto in comune con $y = \sqrt{x}$, ed è $x = 9$. Ma non sembra proprio una retta tangente! D'altronde, il disegno a destra mostra una retta che sembra "ragionevolmente tangente" a una curva, eppure ha più di un punto in comune con quella curva.



Non sembra il caso di insistere con il fatto di avere un solo punto in comune. Sarebbe meglio utilizzare un'altra idea, che oltretutto sembra fatta apposta per essere tradotta in termini rigorosi nel contesto dell'analisi. E cioè, vedere la retta tangente come "posizione limite" di rette secanti. Come si vede, è presente la parola magica dell'analisi: "limite".

Se proviamo a perseguire quest'idea, ci accorgiamo tuttavia che vi sono delle

² Se per caso qualcuno è tentato di rispondere sì a questa domanda, è meglio che faccia un po' di ripasso, prima di andare avanti. In particolare si riveda il § 1.8.

difficoltà. Cosa vuol dire "posizione limite"? Sembra che dobbiamo fare un limite di rette, ma noi sappiamo fare solo limiti di funzioni (e, come caso particolare, di successioni). E' vero che una retta può essere vista come grafico di una funzione³, ma il guaio è che qui vogliamo fare il limite non al variare delle x , per così dire, ma al variare delle rette, ovverossia delle funzioni! Tutto quanto abbiamo fatto finora non ci dà nessuna indicazione.

E allora bisogna farsi furbi e cercare di cavarsela lo stesso. Cominciamo con l'osservare una cosa.

Potremmo interessarci di retta tangente a una curva in un punto P_0 appartenente alla curva. Questo non è proprio il punto di vista consueto, in quanto nella geometria euclidea il problema che si pone è quello della retta tangente ad una circonferenza passante per un punto esterno ad essa. Comunque, già riuscire a risolvere il problema di trovare la retta tangente a una curva in un suo punto P_0 è un bel passo in avanti.

E' davvero un passo avanti decisivo, se scegliamo in modo furbo le rette secanti di cui la retta tangente è la posizione limite. Non è affatto chiaro in generale, quando si immagina la retta tangente come posizione limite di rette secanti, come queste "si muovano" per "avvicinarsi" alla retta tangente. Ebbene, un modo molto semplice per farle "muovere" è di prenderle tutte passanti per P_0 . Naturalmente esse passeranno anche per un altro punto P della curva (sennò, che secanti sono?). Allora potremmo tradurre l'idea di "posizione limite" nel fatto che questa "2^a intersezione" con la curva, e cioè P , tenda a P_0 .

Purtroppo, non sembra che l'idea di "fissare" il punto P_0 e di fare "muovere il punto P verso P_0 " ci abbia risolto un granché. In effetti, non sappiamo fare una cosa del tipo " $\lim_{P \rightarrow P_0}$ ". E, a dire il vero, se anche lo sapessimo fare, sapremmo cosa sostituire ai puntini nella frase "il limite per P che tende a P_0 di ..."?

Ma, allora, siamo ancora al punto di partenza?

Neanche per idea.

Tanto per cominciare, abbiamo che tutte le rette che stiamo considerando passano per il punto P_0 e quindi sono univocamente determinate dal loro coefficiente angolare. Quindi potremmo mettere il coefficiente angolare al posto dei "puntini". Cioè, fare il "limite per P che tende a P_0 del coefficiente angola-

³ A parte un caso scalognato.

re della retta passante per P e P_0 ". La "retta limite" quindi sarebbe semplicemente la retta passante per il punto P_0 il cui coefficiente angolare ha il valore limite.

Resta ancora la questione del limite per $P \longrightarrow P_0$. Anche questa riusciamo a sistemarla, se solo riduciamo un po' le nostre ambizioni. Noi infatti ci siamo posti il problema della retta tangente ad una generica curva. Il guaio è che non sappiamo neanche cos'è una curva. Nel senso che non abbiamo una definizione di carattere generale di una curva. Potremmo magari cercare di darla. Ma possiamo anche accontentarci per ora di curve particolari. E cioè: grafici di funzioni. Così facendo, abbiamo una chiara definizione dell'oggetto di cui stiamo cercando la retta tangente. Ma non è solo questo il vantaggio che ci offre l'occuparci di grafici di funzioni!

Infatti, trattandosi di grafici di funzioni, abbiamo che un punto del grafico è univocamente individuato dalla sua ascissa. Quindi, anche fare il limite per $P = (x, f(x))$ che tende a $P_0 = (x_0, f(x_0))$ lo potremmo tradurre con il fare il limite per x che tende a x_0 ⁴.

Quindi, se vogliamo trovare la retta tangente al grafico di una funzione f in un punto di ascissa x_0 , dobbiamo fare $\lim_{x \rightarrow x_0} m(x)$. Dove $m(x)$ è il coefficiente angolare della retta passante per $(x_0, f(x_0))$ ed $(x, f(x))$. Vale a dire, dalla ben nota formula di geometria analitica, $m(x) = \frac{f(x) - f(x_0)}{x - x_0}$. Quindi dobbiamo fare $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$. Questo limite (se ci sarà, s'intende) ci darà il coefficiente angolare della retta tangente ad f nel punto $(x_0, f(x_0))$.

Sembra proprio che siamo riusciti a capire come dare una risposta rigorosa e rispettosa dell'intuizione al problema della retta tangente a una curva (o, almeno, alle curve che sono grafici di funzioni). Si tratta ora di sistemare per bene le cose dal punto di vista formale.

Definizione 1 Sia $f: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 punto di accumulazione per A . Se $\exists \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \ell \in \mathbb{R}$, diremo che f è derivabile in x_0 , e chiameremo ℓ "derivata di f in x_0 ", denotandolo con $f'(x_0)$. \square

⁴ In realtà ci sarebbero delle obiezioni da fare, ma per non appesantire troppo la discussione e per stimolare la curiosità e lo spirito critico di chi legge, le rimando a un momento più opportuno (vedasi il prossimo paragrafo, nella discussione che segue l'esempio 2.1).

Osservazione 1 Nella definizione sono presenti due richieste, ovvie, su x_0 : deve appartenere ad A , sennò non possiamo calcolare $f(x_0)$; deve essere di accumulazione per A , sennò non possiamo fare il limite.□

Osservazione 2 Si richiede che il limite esista e sia reale. Mentre il requisito dell'esistenza del limite pare scontato da imporre, così non si può dire della richiesta che tale limite sia reale. Perché non accettiamo anche un limite

uguale a ∞ ? Per esempio, $\lim_{x \rightarrow 0} \frac{\sqrt[3]{x}}{x} = +\infty$, il che sembra conforme all'intuizione

che il grafico di $\sqrt[3]{x}$ abbia una retta tangente verticale nell'origine. La ragione per cui scartiamo casi come questo la vedremo tra un po' (vedasi l'esempio 2.1).□

Ora che abbiamo visto come si possa arrivare alla definizione di derivata cercando di sistemare in maniera rigorosa l'idea di retta tangente, vorrei spendere un po' di parole sull'altra sorgente dell'idea di derivata, e cioè il problema della velocità di un punto che si muova di un moto arbitrario.

Più precisamente si tratta di definire la velocità istantanea di un punto come valore limite della velocità media. Per semplicità ci occuperemo di un punto che si muova su una retta. Indicheremo con $x(t)$ il valore dell'ascissa del punto all'istante t . Se ci interessa la velocità media del punto nell'intervallo tra t_0 e t , la sua misura⁵ è data da $\frac{x(t)-x(t_0)}{t-t_0}$. E non ci vuole davvero molto a pensare di fare $\lim_{t \rightarrow t_0} \frac{x(t)-x(t_0)}{t-t_0}$, definendo il valore di questo limite come velocità istantanea all'istante t_0 .

E' abbastanza evidente che questa problematica ci ha condotto molto più direttamente all'idea di derivata di quanto non sia stato per la retta tangente. Avrei anche potuto sorvolare su questa motivazione, visto che non dà poi molto di nuovo e interessante. C'è però un aspetto che voglio sottolineare. L'idea di velocità (media o istantanea) è solo un caso particolare di un'idea ben più generale. Si tratta dell'idea di "saggio di variazione" di una grandezza rispetto ad un'altra, cosa che possiamo definire ogni qual volta una grandezza vari in funzione di un'altra. Quindi non c'è solo la velocità (spazio/tempo), ma anche, per esempio:

- calore specifico (calore/temperatura)

⁵ Supponiamo di avere scelto una unità di misura per il tempo e una per lo spazio.

- baud
- utilità marginale ("soddisfazione"/quantità consumata di un dato bene)
- produttività marginale del lavoro (quantità prodotta di un certo bene/lavoro erogato)
- inflazione (prezzi/tempo)

2. Definizione di derivata

Comincio a ricordare la definizione, già data.

Definizione 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 punto di accumulazione per A . Se $\exists \lim_{x \rightarrow x_0} \frac{f(x)-f(x_0)}{x-x_0} = \ell \in \mathbb{R}$, diremo che f è derivabile in x_0 , e chiameremo ℓ "derivata di f in x_0 ", denotandolo con $f'(x_0)$. \square

Ricordo soltanto che il rapporto $\frac{f(x)-f(x_0)}{x-x_0}$ di solito, per ovvie ragioni, viene chiamato "rapporto incrementale". La notazione $f'(x_0)$ non è l'unica usata:

altre sono $\left(\frac{df}{dx}\right)(x_0)$ o $\left(\frac{df}{dx}\right)_{x_0}$ o ancora $Df(x_0)$ e altre ancora.

Visto che la derivata è definita mediante un limite, possiamo anche introdurre facilmente l'idea di derivata sinistra e destra. Naturalmente, avremo bisogno di richiedere qualcosa di più ad x_0 , e cioè di essere punto di accumulazione da sinistra o da destra rispettivamente.

Definizione 2 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 punto di accumulazione da sinistra (rispettivamente: da destra) per A . Se $\exists \lim_{x \rightarrow x_0^-} \frac{f(x)-f(x_0)}{x-x_0} = \ell \in \mathbb{R}$ (rispettivamente: $\lim_{x \rightarrow x_0^+} \frac{f(x)-f(x_0)}{x-x_0} = \ell \in \mathbb{R}$), diremo che f è derivabile in x_0 da sinistra (rispettivamente: da destra), e chiameremo ℓ "derivata sinistra (rispettivamente: destra) di f in x_0 ", denotandolo con $f'_-(x_0)$ (rispettivamente: $f'_+(x_0)$). \square

Va da sé che, qualora x_0 sia punto di accumulazione per f sia da sinistra che da destra (quindi, in particolare, se x_0 è punto interno ad A), l'essere derivabile per f è equivalente alla esistenza ed uguaglianza delle derivate destre e sinistre. Ricordo che si può, infatti, dimostrare facilmente il seguente teorema (che avrebbe dovuto trovarsi nel terzo capitolo, ma mi son dimenticato...):

Teorema 1 Sia $g:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 di accumulazione da sinistra e da destra per A . Allora:

$$\exists \lim_{x \rightarrow x_0} g(x) \Leftrightarrow \left(\exists \lim_{x \rightarrow x_0^-} g(x), \exists \lim_{x \rightarrow x_0^+} g(x) \text{ e } \lim_{x \rightarrow x_0^-} g(x) = \lim_{x \rightarrow x_0^+} g(x) \right)$$

Inoltre, se $\lim_{x \rightarrow x_0} g(x) = \ell$ (oppure $+\infty$, $-\infty$, ∞), anche i due limiti da

sinistra e da destra sono uguali ad l (oppure $+\infty$, $-\infty$, ∞), e viceversa. \square

Dimostrazione Ovvvia conseguenza delle definizioni. \square

Definizione 3 Diremo che f è derivabile in $B \subseteq A$ se è derivabile in ogni $x_0 \in B$. \square

Sia data $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$. Prendiamo l'insieme dei punti in cui f è derivabile e chiamiamolo A_1 . Se ad ogni punto $x \in A_1$ associamo il numero reale $f'(x)$, otteniamo una nuova funzione reale di variabile reale, definita su A_1 , che potremmo indicare con f' e che chiameremo funzione derivata prima di f . Abbiamo quindi $f': A_1 \rightarrow \mathbb{R}$, una nuova funzione reale di variabile reale.

Il primo risultato da provare, particolarmente significativo, è che la derivabilità in un punto x_0 implica la continuità.

Teorema 2 Sia $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 di accumulazione per A . Se f è derivabile in x_0 , allora f è continua in x_0 . \square

Dimostrazione Abbiamo che $\exists l \in \mathbb{R}$ t.c. $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = l$. Poiché $\lim_{x \rightarrow x_0} (x - x_0) = 0$, dal teorema sul limite di un prodotto, segue che $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \cdot (x - x_0) = 0$, cioè $\lim_{x \rightarrow x_0} f(x) - f(x_0) = 0$. Cioè $\lim_{x \rightarrow x_0} f(x) = f(x_0)$. E quindi f è continua in x_0 . \square

Ora che abbiamo a disposizione questo risultato, possiamo capire perché nella definizione di derivata si pretende che il limite del rapporto incrementale sia finito. Come già notato, tenendo conto dell'interpretazione geometrica, poteva sembrare invece più sensato accettare anche il caso in cui il limite era ∞ (o, magari, almeno il caso in cui il limite era $+\infty$ o $-\infty$). In effetti, l'esempio fornito nell'osservazione 1.2 sembrava portare buone ragioni per fare così. Non è però l'unico esempio che si possa fare.

Esempio 1 Sia $f(x) = \begin{cases} 1 & \text{se } x > 0 \\ 0 & \text{se } x = 0 \\ -1 & \text{se } x < 0 \end{cases}$. E' $\lim_{x \rightarrow 0^+} \frac{f(x) - f(0)}{x - 0} = \lim_{x \rightarrow 0^+} \frac{1}{x} = +\infty$ e analogamente $\lim_{x \rightarrow 0^-} \frac{f(x) - f(0)}{x - 0} = +\infty$. Quindi $\exists \lim_{x \rightarrow 0} \frac{f(x) - f(0)}{x - 0} = +\infty$. Si noti però che f non è continua in 0 . \square

Mi pare che l'insegnamento offerto dall'esempio 1 sia chiaro. Se vogliamo riuscire a dimostrare un risultato come il teorema 1, non possiamo accettare come derivata una funzione per la quale il limite del rapporto incrementale sia ∞ .

Con ciò, sia chiaro, non intendo dire che l'accettare una "derivata infinita" in un punto sia contro l'intuizione che abbiamo di retta tangente. Da questo punto di vista l'esempio 1 non è così dirimente. Infatti, tutto sommato, la posizione limite delle rette secanti è proprio l'asse delle y (farsi un disegnino!), cioè una retta verticale, a cui rimanda l'immaginazione nel caso di una "pendenza" infinita. Uno però può restare interdetto da una situazione come questa, in cui il grafico di f offre l'esempio di una "curva spezzata in vari pezzi": non è scontato che in una situazione come questa la nostra intuizione accetti di parlare di "retta tangente" in un punto quale è $(0,0)$ che è "staccato" dal resto della curva.

Quindi, si tratta di fare una scelta tra due posizioni. Una è quella di insistere sull'idea di retta tangente come posizione limite di rette secanti, nonostante qualche perplessità creata da esempi come quello appena visto. L'altra è di privilegiare la validità del teorema 1, tenendo anche conto del fatto che volere mantenere come derivabile una funzione il cui limite del rapporto incrementale sia ∞ non dà comunque piena soddisfazione all'intuizione. E' evidente che ho scelto la seconda posizione (non è una scelta personale, è la scelta che viene sempre fatta in analisi, in quanto vi sono molte altre ragioni che inducono a privilegiare questa scelta rispetto all'altra).

Approfitto dell'esempio 1 anche per sistemare una questione lasciata aperta nel paragrafo precedente.

In effetti, come mostra l'esempio 1, non è affatto vero che $x \rightarrow x_0$ garantisca $P \rightarrow P_0$ (sto ragionando in termini grafico-intuitivi, evidentemente, visto che non abbiamo nessuna definizione per $P \rightarrow P_0$).

Il viceversa è vero: la distanza tra x ed x_0 è minore o uguale della distanza tra P e P_0 (la lunghezza di un cateto è minore o uguale di quella dell'ipotenusa), pertanto la convergenza di P a P_0 dovrebbe sensatamente garantire la convergenza di x a x_0 . Nella precedente frase è già contenuta un'idea illuminante su quando sia vero che $x \rightarrow x_0$ garantisce $P \rightarrow P_0$. Il teorema di Pitagora ci dice che la distanza tra P e P_0 è pari a $\sqrt{(x-x_0)^2 + (f(x)-f(x_0))^2}$. Quindi, se $x \rightarrow x_0$ ed f è continua, si ha $f(x) \rightarrow f(x_0)$ e quindi anche $P \rightarrow P_0$ (ribadisco: sto sempre parlando in termini intuitivi). Morale: il passaggio "salvifico" che avevamo fatto a pagina 4 da $P \rightarrow P_0$ ad $x \rightarrow x_0$, per stare in piedi richiede qualche condizione

extra. Una condizione che "funziona" è proprio la continuità di f in x_0 . Ma allora il discorso fatto a suo tempo nel § 1 per formalizzare l'idea di derivata era legato alla continuità della funzione. Pertanto la validità del teorema 1 è preziosa per conservare una migliore assonanza tra definizione rigorosa e considerazioni intuitive. E da qui ecco quindi un'altra ragione "a favore" del restringere l'idea di derivabilità al caso in cui il limite del rapporto incrementale sia reale.

Poiché sono fermamente convinto che un atteggiamento critico sia essenziale in ogni circostanza, vorrei presentare anche un esempio il quale mostra come non tutte le cose sono limpide come sembrerebbe.

Esempio 2 Sia $f(x) = \begin{cases} x^2 & \text{se } x \in \mathbb{Q} \\ 0 & \text{se } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$. Si noti che f è derivabile (e

quindi anche continua) in 0 . □

Dettaglio E' $0 \leq f(x) \leq x^2 \quad \forall x \in \mathbb{R}$, quindi, per $x > 0$,
 $0 \leq \frac{f(x)-f(0)}{x-0} \leq \frac{x^2-0}{x-0}$ e pertanto per il teorema dei due carabinieri
 $\lim_{x \rightarrow 0^+} \frac{f(x)-f(0)}{x-0} = 0$. Analogamente $\lim_{x \rightarrow 0^-} \frac{f(x)-f(0)}{x-0} = 0$. Quindi f è derivabile in 0 con derivata uguale a zero. E pertanto f è anche continua. □

Il grafico di f è fatto di punti sparpagliati sull'asse x (quelli di ascissa irrazionale) e sulla parabola di equazione $y = x^2$ (con ascissa razionale). Chi legge ritiene questo esempio conforme all'intuizione di funzione continua o derivabile?

E' anche possibile fare un esempio di funzione che è continua su tutto un intervallo $[a,b]$ senza essere derivabile in alcun punto. Si tratta di un famoso esempio, dovuto a Weierstrass: esso contrasta piuttosto severamente in particolare con la nostra idea intuitiva di continuità.

3. Compatibilità tra la derivata e le operazioni

Per l'enunciato e la dimostrazione dei teoremi relativi alla derivabilità di una somma, prodotto, etc, rinvio a C-S. Dò solo un esempio di come si possano adattare gli enunciati dei teoremi al contesto nel quale ho dato la definizione di derivata.

Teorema 1 Siano $f, g: A \longrightarrow \mathbb{R}$, $x_0 \in A$, x_0 di accumulazione per A . Se f e g sono derivabili in x_0 , allora $f+g$ è derivabile anch'essa in x_0 e si ha: $(f+g)'(x_0) = f'(x_0) + g'(x_0)$. \square

Vorrei far notare che la tesi del teorema ha una struttura tipica di questo tipo di teoremi relativi alla derivata. Cioè si compone di due parti. La prima afferma che la derivata della somma c'è, la seconda permette di trovarne il valore a partire dal valore delle derivate delle funzioni addende.

4. Compatibilità tra la derivata e la relazione d'ordine

Se sappiamo che $f \leq g$, cioè se abbiamo due funzioni $f, g: A \longrightarrow \mathbb{R}$ e sappiamo che $f(x) \leq g(x) \quad \forall x \in A$, non possiamo dire nulla rispetto alle loro derivate.

Ad esempio, se c'è la derivata di f in un $x_0 \in A$, non è detto che ci sia quella di g in x_0 . E, nel caso in cui ci sia, non è detto che sia $f'(x_0) \leq g'(x_0)$.

Invito il lettore a fare degli esempi per convincersi della giustezza di quanto affermato, e a chiedersi come mai ci sono questi "guai".

In realtà, come mi auguro che il lettore abbia dedotto dagli esempi che certamente avrà avuto cura di fare, non è questo il punto di vista "giusto" per guardare ai rapporti tra derivata e " \leq ". Vedremo dopo, nel § 7, quale è l'angolo visuale migliore.

5. Differenziabilità

Supponiamo di avere una funzione f , magari un pò difficile da calcolare. Possiamo utilizzare "al suo posto" un'altra funzione g , più facile da calcolare? O, per lo meno, possiamo fare questo "vicino" ad un certo punto x_0 ? Questo ci porta a focalizzare l'attenzione sull'incremento di f : non è troppo difficile da notare che $f(x) = f(x_0) + (f(x) - f(x_0))$. Quindi, se conosciamo $f(x_0)$ (o, almeno, una sua ragionevole approssimazione), possiamo pensare di "sostituire" all'incremento $f(x) - f(x_0)$ qualche altra espressione più facile da calcolare.

Per esempio, potremmo pensare di approssimare l'incremento $f(x) - f(x_0)$ con qualcosa del tipo $k \cdot (x - x_0)$, dove $k \in \mathbb{R}$ è una costante da individuare in qualche modo per ottenere una buona approssimazione (magari si può anche pensare di trovare k t.c. la approssimazione sia la "migliore possibile"). Per essere un po' più "concreti", si può guardare al disegno di un grafico con annessa retta tangente nel punto $(x_0, f(x_0))$. Basta poco per convincersi che $k = f'(x_0)$ sembra proprio essere la scelta migliore. In effetti, tra tutte le rette passanti per $(x_0, f(x_0))$, la retta di coefficiente angolare $f'(x_0)$ è quella che più resta "attaccata" al grafico di f vicino ad x_0 . Ovverossia, visto altrimenti, la migliore approssimazione che possiamo fare per $f(x) - f(x_0)$ è data da $f'(x_0) \cdot (x - x_0)$.

Abbiamo già colto i due elementi essenziali di questo paragrafo. Il primo è l'idea di sostituire all'incremento di f qualcosa di più semplice (in particolare: un polinomio di 1° grado nella variabile $x - x_0$). Il secondo è la connessione presente tra differenziabilità e derivabilità.

Definizione 1 (di differenziabilità) Sia $f: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 di accumulazione per A . Diciamo che f è differenziabile in x_0 se $\exists k \in \mathbb{R}$ ed $\exists \omega: A \longrightarrow \mathbb{R}$ t.c.:

$$(\star) \quad \begin{cases} f(x) - f(x_0) = k \cdot (x - x_0) + \omega(x) \cdot (x - x_0) \\ \lim_{x \rightarrow x_0} \omega(x) = 0 \\ \omega(x_0) = 0 \end{cases} \quad .\square$$

E' chiaro che la condizione $\omega(x) \longrightarrow 0$ fa sì che l'addendo $\omega(x) \cdot (x - x_0)$ tenda a zero "più rapidamente" dell'altro addendo. La condizione $\omega(x_0) = 0$ è chiaramente superflua: qualsiasi sia il valore che essa assume in x_0 è irrile-

vante, visto che il valore di ω non influenza la prima uguaglianza ($\omega(x_0)$ è moltiplicato per $x_0 - x_0$, cioè per zero) e ancor meno la richiesta sul limite di ω . Quindi serve solo per poter dire che ω è continua in x_0 , cosa che non ci interessa minimamente. Ma conviene porre la condizione $\omega(x_0) = 0$ in (\star) , perché essa risulta essere comoda (non può essere indispensabile, visto che è irrilevante!): per esempio, la dimostrazione del teorema 47.6 del C-S non funziona, così com'è, senza la condizione $\omega(x_0) = 0$. Senza questa condizione si può fare un'altra dimostrazione, che però è più complicata.

Come già anticipato, vi è una connessione tra derivabilità e differenziabilità. In effetti, sono due concetti tra di loro equivalenti.

Teorema 1 Sia $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 di accumulazione per A . Si ha che f è derivabile in x_0 se e solo se è differenziabile in x_0 . \square

Dimostrazione

$$\Rightarrow \text{ Sia } k = f'(x_0) \text{ e } \omega(x) = \begin{cases} \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) & \text{se } x \neq x_0 \\ 0 & \text{se } x = x_0 \end{cases}.$$

E' evidente che la condizione (\star) è soddisfatta: per la prima uguaglianza, basta sostituire e si ha che

$$k \cdot (x - x_0) + \omega(x) \cdot (x - x_0) = f'(x_0) \cdot (x - x_0) + \left(\frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right) \cdot (x - x_0) = f(x) - f(x_0)$$

se $x \neq x_0$ (se $x = x_0$, la prima uguaglianza di (\star) è automaticamente soddisfatta, indipendentemente dalla scelta fatta per k e ω). E' inoltre

$$\lim_{x \rightarrow x_0} \omega(x) = \lim_{x \rightarrow x_0} \left(f'(x_0) - \frac{f(x) - f(x_0)}{x - x_0} \right) = f'(x_0) - \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = 0$$

" \Leftarrow " Dalla prima uguaglianza di (\star) otteniamo, per $x \neq x_0$: $\frac{f(x) - f(x_0)}{x - x_0} = k + \omega(x)$. Da cui $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = k$ (perché $\lim_{x \rightarrow x_0} \omega(x) = 0$). E quindi f è derivabile, essendo $k \in \mathbb{R}$. \square

6. Derivata di funzioni inverse e composte

Già sappiamo che se una funzione, definita in un intervallo, è continua e strettamente crescente, allora essa è invertibile. Ci si può chiedere cosa si possa dire della derivabilità. E' evidente che, a livello grafico-intuitivo, possiamo

fare riferimento al fatto che l'inversione di una funzione corrisponde all'effettuare una simmetria rispetto alla bisettrice del I e III quadrante. Sembra più che ragionevole che, se una funzione è derivabile in un certo punto, anche la sua inversa lo sia, naturalmente nel punto corrispondente. Non solo, ma la retta tangente ad f^{-1} in $f(x_0)$ dovrebbe essere la simmetrica della retta tangente in x_0 : così almeno appare dal disegno. Ora, se abbiamo una retta con coefficiente angolare k , la sua simmetrica rispetto alla bisettrice del I e III quadrante ha coefficiente angolare $1/k$.

Esercizio 1 Data una retta di equazione $y = kx + m$, scrivere l'equazione della retta simmetrica ad essa rispetto alla bisettrice del I e III quadrante. □

E' sperabile che il lettore abbia scorto un problema, e cioè che se il coefficiente angolare k è uguale a zero, allora otteniamo $1/0$, ovvero non otteniamo nulla. In effetti, la retta simmetrica (sempre rispetto alla bisettrice del I e III quadrante) di una retta orizzontale è una retta verticale. Questo ci fa intuire che, se vogliamo enunciare e dimostrare un preciso risultato, probabilmente dovremo imporre la condizione che sia $f'(x_0) \neq 0$. Detto questo, è più che ragionevole aspettarsi che sia $(f^{-1})'(f(x_0)) = \frac{1}{f'(x_0)}$. Se usiamo la variabile y

come "variabile indipendente" per la funzione f^{-1} , la formula precedente assuma la forma $(f^{-1})'(y_0) = \frac{1}{f'(f^{-1}(y_0))}$. Infatti, trattandosi di funzioni inverse,

abbiamo che $y_0 = f(x_0) \Leftrightarrow f^{-1}(y_0) = x_0$. Vorrei far notare che c'è una certa consuetudine ad usare la variabile y come "variabile indipendente" quando si ha a che fare con una funzione inversa. Ma non va dimenticato che f^{-1} è una funzione "come tutte le altre". Pertanto useremo spesso la variabile x quando vorremo indicare il generico punto nel quale calcoliamo f^{-1} e, per analogia, il simbolo x_0 per indicare un punto "fissato" (ovverossia, scelto in precedenza). In tal

caso la formula diventa: $(f^{-1})'(x_0) = \frac{1}{f'(f^{-1}(x_0))}$. Quest'ultima è la forma che

appare più frequentemente. Vorrei comunque fare un esempio dei tre aspetti, che sono del tutto equivalenti tra loro, assunti dalla formula di derivazione della funzione inversa.

Esempio 1 Sia $f: [0, +\infty[\longrightarrow \mathbb{R}$ così definita: $f(x) = x^2$. Allora $f^{-1}(y) = \sqrt{y}$, con $f^{-1}: [0, +\infty[\longrightarrow \mathbb{R}$. Ed abbiamo (ricordo che $f'(x) = 2x$):

$$\begin{aligned} (f^{-1})'(f(x_0)) &= \frac{1}{f'(x_0)} & \text{cioè} & \quad D(\sqrt{\cdot})(f(x_0)) = \frac{1}{2x} \\ (f^{-1})'(y_0) &= \frac{1}{f'(f^{-1}(y_0))} & \text{cioè} & \quad D(\sqrt{\cdot})(y_0) = \frac{1}{2\sqrt{y_0}} \\ (f^{-1})'(x_0) &= \frac{1}{f'(f^{-1}(x_0))} & \text{cioè} & \quad D(\sqrt{\cdot})(x_0) = \frac{1}{2\sqrt{x_0}} \quad .\square \end{aligned}$$

Come si vede, la formula più utile è l'ultima (in realtà la penultima e l'ultima sono la stessa formula, solo che uso due diversi simboli per indicare il punto cui sono interessato).

Per l'enunciato e la dimostrazione rigorosa di un risultato che riguarda la derivazione della funzione inversa, vedasi C-S.

Passiamo alla derivazione delle funzioni composte. Anche qui, rinvio la parte formale al C-S.

Mi limito a due sole osservazioni.

La prima riguarda la formula: la derivata di una funzione composta è il prodotto delle derivate. Bene, questo fatto è il più naturale che ci si possa aspettare. Supponiamo di avere tre grandezze A, B e C tra loro proporzionali e di usare le lettere x, y e z per indicarne le rispettive misure. Supponiamo che la grandezza B dipenda dalla grandezza A e vari del doppio di quel che varia A : cioè che $y = g(x)$, con $g(x) = 2x + \alpha$ ⁶ (qui α è una costante il cui valore non ci interessa). Analogamente, supponiamo che la grandezza C a sua volta dipenda dalla grandezza B , variando del triplo: stavolta avremo che $z = f(y)$ con $f(y) = 3y + \beta$. Allora la grandezza C varierà del sestuplo della grandezza A , cioè $z = 6x + \alpha'$ (infatti $z = f(g(x)) = f(2x + \alpha) = 3 \cdot (2x + \alpha) + \beta = 6x + (3\alpha + \beta)$: ho indicato con α' la costante $3\alpha + \beta$, il cui valore non mi interessa⁷).

Questo è quel che accade nei caso più semplice, cioè nel caso di grandezze tra loro proporzionali la variazione della funzione composta è proprio il prodotto delle variazioni. Ma allora in casi più complicati, questo è ciò che ci dobbiamo aspettare a livello di derivate!

Perché? Cercherò di spiegarlo.

Ciò che entra in gioco, a dire il vero, è piuttosto l'idea di

⁶ Ma il lettore è convinto di questa affermazione?

⁷ Per chi non l'avesse ancora capito, il valore di queste costanti non è rilevante perché in questo discorso siamo interessati alle variazioni delle grandezze in esame

differenziabilità anziché quella di derivabilità: sappiamo però che questi due concetti sono equivalenti tra loro per il teorema 5.1. Ricordo che l'idea sottostante alla differenziabilità è quella di sostituire all'incremento di f un incremento approssimato, il più semplice possibile: si tratta proprio dell'incremento proporzionale. Infatti, sostituiamo ad $f(x)-f(x_0)$ un opportuno $k \cdot (x-x_0)$. Quindi è prevedibile che la regola di derivazione delle funzioni composte si possa inferire (a livello intuitivo) da quel che succede nel caso di grandezze che variano in modo proporzionale.

La seconda osservazione riguarda la formula che è $(f \circ g)'(x_0) = f'(g(x_0)) \cdot g'(x_0)$. Come nel caso della derivazione della funzione inversa, è importante notare dove vengono calcolate le derivate. In particolare, la derivata di f viene calcolata nel punto $g(x_0)$, non nel punto x_0 (che magari non appartiene neanche all'insieme di definizione di g).

7. Monotonia di f e segno di f'

Questo paragrafo recupera l'apparente sconnessione tra derivate e \leq , vista al § 4, dove sembrava che le cose non andassero granché bene. In realtà, avevamo guardato le cose da un punto di vista sbagliato. D'altronde, che qualche connessione vi possa essere tra derivata e \leq è ragionevole per le seguenti considerazioni di carattere puramente formale:

- la derivata è definita come un limite
- c'è il teorema di permanenza del segno per i limiti.

Se teniamo conto di ciò, si capisce che dobbiamo concentrare l'attenzione sulla funzione di cui facciamo il limite. Tale funzione è il rapporto incrementale: quindi possiamo immediatamente dedurre che, se il rapporto incrementale è maggiore o uguale a zero, allora anche la derivata lo è (e abbiamo anche una sorta di viceversa, in cui intervengono le disuguaglianze strette, sempre per il teorema di permanenza del segno). Ma cosa vuol dire che il rapporto incrementale è maggiore o uguale di zero? E' una proprietà significativa? Perbacco se lo è! $\frac{f(x)-f(x_0)}{x-x_0} \geq 0$ significa che:

- quando $x > x_0$, allora $f(x) \geq f(x_0)$
- quando $x < x_0$, allora $f(x) \leq f(x_0)$.

Come si vede, assomiglia parecchio al fatto che f sia debolmente crescente! Se così è, il segno del rapporto incrementale ha a che fare con una importante proprietà delle funzioni. Proviamo allora a formalizzare il tutto:

Teorema 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 punto di accumulazione per A . Supponiamo che f sia derivabile in x_0 . Se f è debolmente crescente (rispettivamente: decrescente) su A , allora $f'(x_0) \geq 0$ (rispettivamente: $f'(x_0) \leq 0$). □

Dimostrazione Naturalmente vediamo solo un caso. Se f è debolmente crescente, allora $\frac{f(x)-f(x_0)}{x-x_0} \geq 0 \quad \forall x \in A \setminus \{x_0\}$. E quindi, per il teorema di permanenza del segno, $\lim_{x \rightarrow x_0} \frac{f(x)-f(x_0)}{x-x_0} \geq 0$. Cioè, $f'(x_0) \geq 0$. □

Osservazione 1 Evidentemente basta che esista $\delta > 0$ t.c. f sia monotona su $A \cap]x_0 - \delta, x_0 + \delta[$. Se f soddisfa una tale proprietà, si usa dire che f è localmente monotona in x_0 . □

Osservazione 2 Come è ben noto, "al limite le disuguaglianze si indeboliscono". Pertanto, se anche f fosse strettamente crescente, non potremmo dedurre che $f'(x_0) > 0$. Come esempio si prenda $f(x) = x^3$ ed $x_0 = 0$. \square

Abbiamo quindi visto come usare il teorema di permanenza del segno per ricavare, dalla monotonia di f , informazioni sul segno di $f'(x_0)$. Come il lettore ben sa, in realtà quello usato non è propriamente il teorema di permanenza del segno, bensì il suo rovescio. Il teorema "diretto" di permanenza del segno offre informazioni sul segno di una funzione dalla conoscenza del segno del suo limite. Vedi per esempio il teorema 33.8 di C-S oppure, nel contesto delle successioni, il teorema II.9.1 e corollario II.9.2 delle dispense.

Teorema 2 Sia $f: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 punto di accumulazione per A . Supponiamo che f sia derivabile in x_0 . Se $f'(x_0) > 0$ (rispettivamente: < 0), allora $\exists \delta > 0$ t.c. $\frac{f(x)-f(x_0)}{x-x_0} > 0$ (rispettivamente: < 0) per ogni $x \in (A \cap]x_0 - \delta, x_0 + \delta[) \setminus \{x_0\}$. \square

Dimostrazione E' una conseguenza immediata del teorema di permanenza del segno (applicato alla funzione "rapporto incrementale", cioè alla funzione $x \mapsto \frac{f(x)-f(x_0)}{x-x_0}$). \square

Credo che ad un attento lettore l'enunciato del teorema 2 sia apparso un po' deludente. In effetti, la tesi del teorema non parla di monotonia. Certo solo un ingenuo potrebbe aspettarsi una tesi in cui si afferma che f è monotona su A , visto che il limite è una proprietà locale (vedi § III.6). Però almeno la locale monotonia in x_0 (vedi osservazione 1), uno se la poteva anche ragionevolmente aspettare. Cioè, si poteva immaginare che fosse possibile dimostrare il seguente:

"Teorema" Sia $f: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 punto di accumulazione per A . Supponiamo che f sia derivabile in x_0 . Se $f'(x_0) > 0$ (rispettivamente: < 0), allora $\exists \delta > 0$ t.c. f è strettamente crescente (rispettivamente: strettamente decrescente) su $A \cap]x_0 - \delta, x_0 + \delta[$. \square

Ebbene, contrariamente alle attese e all'intuizione, così non è. Il teorema tra virgolette sopra riportato è falso. Evidentemente, se dico questo, ci sarà un esempio di funzione che soddisfa le ipotesi del "teorema" ma non la sua tesi. Vedremo in un prossimo paragrafo un tale esempio. Per ora, quel che è più importante capire è cosa esattamente dice il teorema 2 e perché non dice che f è localmente

monotona in x_0 . La tesi del teorema è che (vediamo il caso $f'(x_0) > 0$):

$$(\star) \quad \exists \delta > 0 \text{ t.c. } \frac{f(x) - f(x_0)}{x - x_0} > 0 \quad \forall x \in (A \cap]x_0 - \delta, x_0 + \delta[) \setminus \{x_0\}.$$

Se f fosse strettamente crescente su $A \cap]x_0 - \delta, x_0 + \delta[$, si avrebbe:

$$(\star) \quad \frac{f(x_2) - f(x_1)}{x_2 - x_1} > 0 \quad \forall x_1, x_2 \in A \cap]x_0 - \delta, x_0 + \delta[\text{ e t.c. } x_1 \neq x_2.$$

Dovrebbe essere evidente che l'affermazione (\star) è più forte di (\star) . Infatti, basta prendere $x_1 = x_0$ e $x_2 = x$ in (\star) per ottenere (\star) . Ma il viceversa non funziona. Nessun problema ad usare la variabile x_2 al posto di x , ma x_0 è un dato, quindi non lo posso sostituire a mio piacimento con un punto $x_1 \in A \cap]x_0 - \delta, x_0 + \delta[$!

Quindi la tesi del teorema 2 non è la locale stretta crescita. Si tratta di una proprietà che soltanto sembra che abbia a che fare con la stretta crescita, in quanto ci dice che $f(x) > f(x_0)$ se $x > x_0$ e $f(x) < f(x_0)$ se $x < x_0$. Tanto è vero che nel C-S la proprietà espressa dalla tesi del teorema 2 viene chiamata "crescenza nel punto x_0 "⁸.

Ma, allora, continuano ad esserci dei guai nel rapporto tra derivata e \leq ? Vedremo in seguito come rimediare a questo problema. Avremo però bisogno nientedimeno che del teorema fondamentale del calcolo differenziale, vale a dire il teorema di Lagrange.

Per ora, dobbiamo accontentarci di questi risultati. I quali, ad onore del vero, non sono proprio da buttare via.

8. Massimi e minimi locali

Chi abbia visto un po' di analisi nelle scuole secondarie, solitamente ritiene di sapersela cavare a proposito dei massimi e minimi relativi per una funzione. La terminologia usata nel titolo di questo paragrafo si riferisce proprio ai minimi e massimi relativi. Non uso la terminologia più consueta per due motivi. Il primo è che preferisco uniformarmi all'uso del tutto generale in matematica dell'attributo "locale". Il secondo motivo è di carattere didattico: usando l'aggettivo "locale" spero di attirare l'attenzione dello studente "già esperto di

⁸ Si tratta di una terminologia curiosa, in quanto la crescita è una proprietà di una funzione su un insieme. Si tratta in effetti di una terminologia inventata ad hoc per "dare un nome" alla tesi del teorema 2. Non va confusa con la crescita locale in x_0 .

analisi" su questo paragrafo. Altrimenti, potrebbe avere la tentazione di leggerlo superficialmente, senza quindi accorgersi che magari non sa cosa vuol dire massimo o minimo relativo⁹.

Definizione 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$. Il punto x_0 si dice punto di minimo (rispettivamente: massimo) locale se $\exists \delta > 0$ t.c. x_0 sia punto di minimo (rispettivamente: massimo) per $f|_{A \cap]x_0 - \delta, x_0 + \delta[}$. \square

Cioè, il punto x_0 è un punto di minimo per f ristretta ad un certo intorno di x_0 . Per distinguere senza equivoci i due concetti, si usa la contrapposizione locale/globale od anche quella relativo/assoluto. In formule, x_0 è punto di minimo (globale) per f se $f(x) \geq f(x_0) \quad \forall x \in A$; è invece punto di minimo locale per f se $f(x) \geq f(x_0) \quad \forall x \in A \cap]x_0 - \delta, x_0 + \delta[$. Volendo, possiamo quindi riscrivere la definizione 1 così:

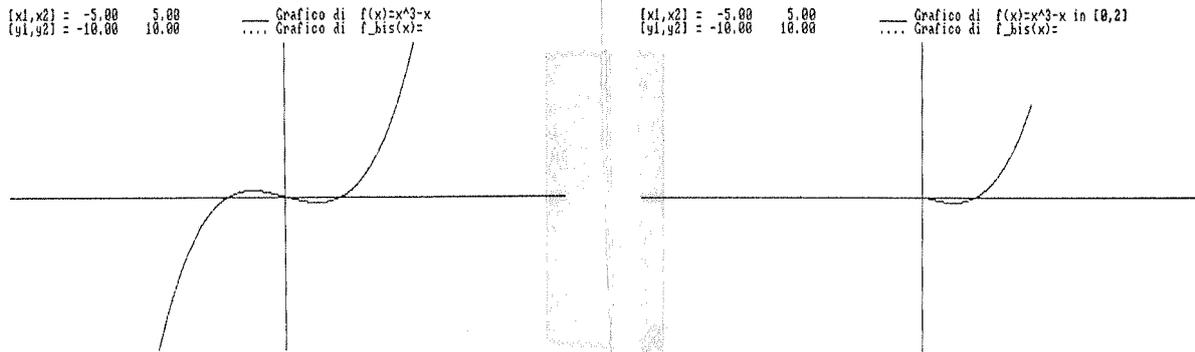
Definizione 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$. Il punto x_0 si dice punto di minimo (rispettivamente: massimo) locale se $\exists \delta > 0$ t.c. $f(x) \geq f(x_0)$ (rispettivamente: $f(x) \leq f(x_0)$) $\forall x \in A \cap]x_0 - \delta, x_0 + \delta[$. \square

Osservazione 1 Per amor di sinteticità si usa dire che x_0 è un punto di estremo locale (rispettivamente: globale) se è un punto di massimo o di minimo locale (rispettivamente: globale). \square

Quindi, evidentemente, un punto di minimo globale è anche un punto di minimo locale. Il viceversa non è garantito.

L'idea di punto di minimo locale può anche essere colta graficamente. Significa che ci disinteressiamo di quel che succede fuori dalla striscia $]x_0 - \delta, x_0 + \delta[\times \mathbb{R}$: vedasi la figura sotto a sinistra che mostra il grafico di f e la figura sotto a destra che mostra il grafico di f ristretta a $A \cap]x_0 - \delta, x_0 + \delta[=]0, 2[$. In questo esempio, il punto $x_0 = 1$ è un punto di minimo locale ma non globale.

⁹ Parlo per esperienza...



Chi ha un po' di esperienza di analisi, magari non conosce la definizione 1 o l'ha dimenticata, ma di certo si ricorda che "gli estremi locali (pardon: relativi) si trovano imponendo la condizione $f'(x_0) = 0$ ". Il che spiega, per il novizio dell'analisi, come mai ci si interessa di questi strani punti nel capitolo dedicato alle derivate. In realtà, l'affermazione riportata tra virgolette è solo parzialmente vera¹⁰. Cerchiamo comunque di capire dove sia il legame tra minimi locali e derivata, provando a dimostrare che in un punto di minimo locale la derivata prima si annulla. Per ora, ometterò i dettagli (importanti!) perché mi interessa di più cogliere l'idea principale.

Se x_0 è punto di minimo relativo, allora $f(x) - f(x_0) \geq 0$ per x vicini ad x_0 . Ma allora, $x > x_0$ implica $\frac{f(x) - f(x_0)}{x - x_0} \geq 0$ e quindi (facendo il limite per $x \rightarrow x_0^+$) $f'_+(x_0) \geq 0$; invece $x < x_0$ implica $\frac{f(x) - f(x_0)}{x - x_0} \leq 0$ e quindi $f'_-(x_0) \leq 0$. Orbene, se f è derivabile in x_0 , allora abbiamo $f'_-(x_0) = f'(x_0) = f'_+(x_0)$. E quindi $0 \leq f'_-(x_0) = f'(x_0) = f'_+(x_0) \leq 0$, cioè $f'(x_0) = 0$. Restano, come dicevo, i dettagli da sistemare, cosa che vedremo nella dimostrazione vera e propria. Insisto sul fatto che questi cosiddetti dettagli sono molto importanti, perché dalla loro incomprendione seguono molti degli errori tipici che si fanno in questo contesto.

Teorema 1 Sia $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, $x_0 \in A$, x_0 punto di accumulazione per A da sinistra e da destra. Se x_0 è punto di minimo (rispettivamente: massimo) locale per f e se f è derivabile in x_0 , allora è $f'(x_0) = 0$. □

Dimostrazione Vediamo il caso in cui x_0 sia punto di massimo locale. Sia quindi $\delta > 0$ t.c. $f(x) \leq f(x_0) \forall x \in A \cap]x_0 - \delta, x_0 + \delta[$. Allora:

$$(*) \quad \forall x \in A \cap]x_0, x_0 + \delta[\quad , \quad \frac{f(x) - f(x_0)}{x - x_0} \leq 0$$

¹⁰ Il che significa, in matematica, che è falsa. Non esistono sfumature o mezze misure. Per lo meno, non ai bassi livelli del discorso ai quali ci stiamo muovendo.

$$(**) \quad \forall x \in A \cap]x_0 - \delta, x_0[, \quad \frac{f(x) - f(x_0)}{x - x_0} \geq 0 .$$

Poiché f è derivabile in x_0 , ed x_0 è di accumulazione sia da destra che da sinistra, allora $\exists f'_+(x_0)$, $f'_-(x_0)$ e sono uguali a $f'(x_0)$ (teorema 2.1).

Ma da (*) otteniamo, per la permanenza del segno, che

$$\lim_{x \rightarrow x_0^+} \frac{f(x) - f(x_0)}{x - x_0} = f'_+(x_0) \leq 0 .$$

Da (**) si ha $f'_-(x_0) \geq 0$. Ma allora $0 \leq f'_-(x_0) = f'(x_0) = f'_+(x_0) \leq 0$, e quindi $f'(x_0) = 0$. \square

Il dettaglio importante consiste nella richiesta che x_0 sia di accumulazione da destra e da sinistra. Senza questa ipotesi, il teorema non è vero: non basta che x_0 sia di accumulazione soltanto.

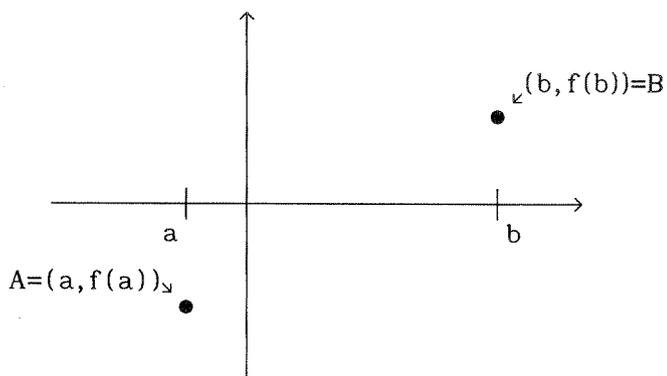
Osservazione 2 Se $\exists \delta > 0$ t.c. $]x_0 - \delta, x_0 + \delta[\subseteq A$, vale a dire se x_0 è un punto interno ad A , evidentemente x_0 è un punto di accumulazione per A sia da destra che da sinistra. \square

Esercizio 1 Notare che 1 non è punto di accumulazione da destra per $[0,1]$ e fornire un esempio di funzione derivabile in 1, per la quale 1 è un punto di massimo locale, senza però che sia $f'(1) = 0$. \square

9. Il teorema fondamentale del calcolo differenziale

Siamo arrivati evidentemente a un punto importante. Si tratta di provare il teorema di Lagrange. Come avviene spesso (non sempre) per i grandi teoremi, esso traduce un'idea molto semplice.

L'idea è questa. Se parto con l'auto da Genova e dopo un'ora ho percorso 90 Km , ci sarà stato almeno un momento in cui il tachimetro segnava esattamente 90 Km/h . Detto altrimenti, se nell'intervallo di tempo $[t_1, t_2]$ la velocità media è \bar{v} , ci deve essere almeno un istante t in cui la velocità istantanea è proprio uguale alla velocità media \bar{v} . La ragione è ovvia: se la velocità istantanea fosse sempre minore di \bar{v} , la velocità media non potrebbe essere \bar{v} ! Ugualmente la velocità istantanea non può essere sempre maggiore, e quindi... Conoscendo il parallelo che c'è tra velocità media/istantanea e retta tangente/secante, non ci si stupirà se c'è una interpretazione grafica analoga alle precedenti considerazioni cinematiche. Comunque si provi a disegnare il grafico di una funzione (derivabile) che congiunge i due punti A e B , ci sarà un punto nell'intervallo $[a, b]$ in cui la pendenza della retta tangente è uguale alla pendenza della retta secante passante per A e B . Invito chi legge a provare davvero a disegnare qualche grafico, per una concreta verifica.



Come si fa a dimostrare il teorema di Lagrange, che esprime in modo preciso queste "intuizioni"? La strada tradizionale consiste nel provare dapprima un caso particolare, detto teorema di Rolle. Nel disegno, se la retta passante per i due punti A e B fosse orizzontale, allora ci sarebbe un punto in $[a, b]$ in cui la retta tangente alla funzione è anch'essa orizzontale. Vale a dire, ci sarebbe un

punto in cui la derivata prima è uguale a zero.

Teorema 1 (di Rolle) Sia $f:[a,b] \longrightarrow \mathbb{R}$, continua su $[a,b]$ e derivabile su $]a,b[$. Se $f(a) = f(b)$, allora $\exists \xi \in]a,b[$ t.c. $f'(\xi) = 0$. \square

Dimostrazione Poiché f è continua in $[a,b]$, il teorema di Weierstrass ci permette di affermare che f ha massimo e minimo su $[a,b]$. Sia m il minimo e M il massimo. Se per caso $m = M$, allora da $m \leq f(x) \leq M \quad \forall x \in [a,b]$, si deduce che f è costante su $[a,b]$ e quindi di punti $\xi \in]a,b[$ t.c. $f'(\xi) = 0$ ne troviamo a bizzeffe.

Se $m < M$, allora ci sarà almeno un punto di minimo o di massimo interno ad $]a,b[$. Infatti, se così non fosse, detto x_m un punto di minimo ed x_M un punto di massimo, essi sarebbero obbligatoriamente punti estremi dell'intervallo. Ma allora $f(x_m) = m = M = f(x_M)$, in quanto f assume uguali valori negli estremi dell'intervallo, in contrasto con l'ipotesi che sia $m < M$. Sia allora $\xi \in]a,b[$ un punto di massimo o di minimo interno ad $]a,b[$, la cui esistenza abbiamo garantito. Ovviamente esso, essendo un punto di massimo o di minimo assoluto, è anche un punto di massimo o di minimo locale: essendo interno ad $]a,b[$, possiamo affermare che $f'(\xi) = 0$, grazie al teorema 8.1. \square

Come si può vedere, la dimostrazione del teorema di Rolle non è particolarmente complessa. Non va dimenticato, però, che essa utilizza il teorema di Weierstrass. Quindi la dimostrazione che abbiamo usato richiede la proprietà di completezza. Inoltre si tratta di una dimostrazione non costruttiva in quanto fa uso del teorema di Weierstrass la cui dimostrazione non è costruttiva. Si badi bene che le precedenti affermazioni non dicono che per il teorema di Rolle occorre l'assioma di completezza o che non sia possibile darne una dimostrazione costruttiva. Tanto è vero che di queste due ultime affermazioni la prima è vera mentre la seconda è falsa. Si può fare un esempio di funzione $f:A \longrightarrow \mathbb{Q}$, con $A \subseteq \mathbb{Q}$, che soddisfa le ipotesi del teorema di Rolle¹¹ senza che però vi sia un punto in cui la derivata prima è zero: questo prova che senza l'assioma di completezza (più precisamente: solo con gli assiomi 1+15) non possiamo dimostrare questo teorema. Si può invece dare una diversa dimostrazione, costruttiva, del teorema di Rolle.

Passiamo al teorema di Lagrange

¹¹ Modificato in modo ovvio per tenere conto del fatto che siamo in \mathbb{Q} e non in \mathbb{R} .

Teorema 2 (di Lagrange, o teorema fondamentale del calcolo differenziale) Sia $f: [a, b] \rightarrow \mathbb{R}$, continua su $[a, b]$ e derivabile su $]a, b[$. Allora $\exists \xi \in]a, b[$ t.c. $f(b) - f(a) = f'(\xi) \cdot (b - a)$. \square

Dimostrazione Rinvio al C-S. Poiché però su questo testo il teorema di Lagrange è provato come conseguenza del teorema di Cauchy, faccio notare che c'è un semplice trucco per ricavare il teorema di Lagrange dal teorema di Rolle. L'idea è ovvia: modifichiamo f aggiungendo $m \cdot x$ ad $f(x)$, scegliendo m opportunamente di modo che la funzione $g(x) = f(x) + m \cdot x$ soddisfi le ipotesi del teorema di Rolle, ed il gioco è fatto. \square

Una estensione del teorema di Lagrange è il teorema di Cauchy. Esso dice (sotto opportune ipotesi) che $\exists \xi \in]a, b[$ t.c. $\frac{f'(\xi)}{g'(\xi)} = \frac{f(b) - f(a)}{g(b) - g(a)}$. Rinvio al C-S per enunciato e dimostrazione. Preferisco osservare che esso costituisce una estensione non triviale del teorema di Lagrange, anche se per dimostrarlo basta un semplice trucco. Vediamo quindi come esso dica effettivamente di più del teorema di Lagrange. Se applicassimo Lagrange a f e a g , otterremmo che $\exists \xi_1$ t.c. $f(b) - f(a) = f'(\xi_1) \cdot (b - a)$ e che $\exists \xi_2$ t.c. $g(b) - g(a) = g'(\xi_2) \cdot (b - a)$. Da cui: $\frac{f'(\xi_1)}{g'(\xi_2)} = \frac{f(b) - f(a)}{g(b) - g(a)}$. Come si vede, direttamente dal teorema di Lagrange si può ottenere una formula simile ma con una importante differenza: nel teorema di Cauchy il punto ξ è uno solo, lo stesso sia per f che per g .

Si può dare anche per il teorema di Cauchy una interpretazione cinematica. Se un'auto percorre 120 Km in un'ora ed un'altra 60, cioè se il rapporto tra le due velocità medie è pari a 2, allora ci sarà un istante (ξ) in cui il rapporto tra le velocità istantanee è anch'esso pari a 2. Ci sarà anche un istante ξ_1 in cui la velocità della prima auto è di 120 Km/h e un istante ξ_2 in cui la velocità istantanea della seconda macchina sarà pari a 60 Km/h. Si noti però che questi istanti possono essere tutti diversi tra loro: anzi, provare a fare un esempio in cui si veda che sono effettivamente tutti distinti.

Più significative del teorema di Cauchy, che è un risultato per così dire "tecnico", sono alcune conseguenze del teorema di Lagrange.

Il primo risultato è il

Teorema 3 Sia $f: I \rightarrow \mathbb{R}$, I intervallo, f continua su I . Supponiamo che per ogni x_0 interno ad I , f sia derivabile in x_0 con $f'(x_0) = 0$.

Allora f è costante su I . \square

Dimostrazione Ricordo che dire che f è costante significa:

$$(\star) \quad \exists k \in \mathbb{R} \text{ t.c. } \forall x \in I \text{ si ha } f(x) = k.$$

Ciò è equivalente a:

$$(\star) \quad \forall x_1, x_2 \in I \quad f(x_1) = f(x_2).$$

Che (\star) implichi (\star) è ovvio. Per il viceversa, sia $x_1 \in I$: prenderemo $k = f(x_1)$. Allora abbiamo che $f(x_2) = k \quad \forall x_2 \in I$, cioè proprio quanto affermato in (\star) .

Evidentemente ho introdotto la condizione (\star) perché dimostrerò che f è costante proprio provando che vale (\star) .

Siano allora $x_1, x_2 \in I$. Applico il teorema di Lagrange all'intervallo $[x_1, x_2]$ (se $x_1 < x_2$, sennò¹² considero l'intervallo $[x_2, x_1]$). Posso applicarlo, perché f è certamente continua su $[x_1, x_2]$ ed è anche derivabile in $]x_1, x_2[$, in quanto tutti i punti di questo intervallo sono interni ad $]x_1, x_2[$ e quindi, a fortiori, ad I . Allora $\exists \xi \in]x_1, x_2[$ t.c. $f(x_2) - f(x_1) = f'(\xi) \cdot (x_2 - x_1)$. Ma $f'(\xi) = 0$, quindi $f(x_2) = f(x_1)$. \square

Si ottiene subito il seguente:

Corollario 1 Sia $f: I \rightarrow \mathbb{R}$, I intervallo, f derivabile in ogni $x_0 \in I$, con $f'(x_0) = 0$. Allora f è costante su I . \square

Dimostrazione Basta osservare che, negli eventuali estremi di I , f è derivabile e pertanto continua, per cui si ricade nelle ipotesi del teorema 3. \square

Credo che gli enunciati del teorema 3 e del suo corollario meritino un po' di commenti.

Tanto per cominciare, più spesso si applica il corollario. In effetti, succede una cosa strana: sono la stessa cosa... Voglio cioè dire che, se sono soddisfatte le ipotesi del teorema 3, allora f è costante su I e quindi soddisfa anche le ipotesi del corollario. Sia ben chiaro, però, che questo nulla toglie all'importanza del teorema 3: può benissimo capitare che a priori uno non sappia se la funzione data è derivabile negli estremi e quindi il corollario non possa essere utilizzato.

L'altro commento da fare è che nel teorema 3 non possiamo eliminare la ipotesi di continuità agli estremi.

¹² Se $x_1 = x_2$, non c'è niente da dimostrare, ovviamente!

Esempio 1 $f(x) = \begin{cases} 0 & \text{per } x \in]0,1[\\ 1 & \text{per } x = 0,1 \end{cases} . \square$

Il commento più importante però è che è assolutamente essenziale per la validità del teorema 3 che I sia un intervallo.

Esercizio 1 Scoprire dove è stato usato, nella dimostrazione del teorema 3, il fatto che I sia un intervallo. \square

Se l'insieme su cui è definita f non è un intervallo, può succedere quanto segue:

Esempio 2 Sia $f(x) = \begin{cases} 1 & \text{se } x > 0 \\ -1 & \text{se } x < 0 \end{cases}$. f è derivabile su $\mathbb{R} \setminus \{0\}$, cioè sul suo insieme di definizione, ed è $f'(x) = 0 \quad \forall x \in \mathbb{R} \setminus \{0\}$, però f non è costante. Infatti assume due valori distinti. \square

Vorrei far notare che i commenti relativi al teorema 3 si applicano anche al seguente teorema 4, anch'esso conseguenza del teorema di Lagrange.

Teorema 4 Sia $f: I \longrightarrow \mathbb{R}$, I intervallo, f continua su I . Supponiamo che per ogni x_0 interno ad I , f sia derivabile in x_0 con $f'(x_0) \geq 0$ (rispettivamente: $\leq, >, <$). Allora f è debolmente crescente (rispettivamente: debolmente decrescente, strettamente crescente, strettamente decrescente) su I . \square

Dimostrazione Solo nel caso del " \geq ".

Siano $x_1, x_2 \in I$, con $x_1 < x_2$. Si applica il teorema di Lagrange ad $[x_1, x_2]$. E si ha: $f(x_2) - f(x_1) = f'(\xi) \cdot (x_2 - x_1)$, ma $f'(\xi) \geq 0$ e quindi $f(x_2) \geq f(x_1)$. \square

Esercizio 2 Dare l'enunciato di un corollario che stia al teorema 4 come il corollario 1 sta al teorema 3. \square

Che commento fare a questo teorema? Abbiamo trovato quel che ci mancava prima, nel § 7. Cioè, abbiamo finalmente uno strumento che ci permette di passare da informazioni sul segno della derivata a informazioni sulla monotonia della funzione. Attenzione, però. Il teorema 7.1 è un teorema di carattere locale, cioè coinvolge solo informazioni sulla funzione in un intorno di x_0 (in effetti, per il teorema 7.1 basta solo che f sia debolmente crescente in un intorno di x_0 , come è detto nella osservazione 7.1) Invece il teorema 4 è un teorema di carattere

globale. Cioè, non ci basta sapere che $f'(x_0) \geq 0$ (e neanche $f'(x_0) > 0$): d'altronde era già stato visto a suo tempo che il teorema della permanenza del segno non bastava per ottenere la monotonia di f ; abbiamo bisogno di sapere che è $f'(x) \geq 0$ su un insieme I per ottenere la monotonia. Per di più, questo insieme deve avere una struttura particolare, cioè deve essere un intervallo.

A questo proposito, vorrei notare che siamo di fronte ad un leitmotiv tipico delle conseguenze del teorema di Lagrange. Se uno si dimentica l'essenziale requisito di essere su un intervallo, ottiene risultati come questo:

Esempio 3 Sia $f(x) = \operatorname{tg}(x)$. La funzione tangente è definita su $A = \mathbb{R} \setminus \{ x \in \mathbb{R} : \exists k \in \mathbb{Z} \text{ t.c. } x = (\pi/2) + k\pi \}$. E' noto che f è derivabile su A e che $f'(x) = 1 + \operatorname{tg}^2(x) \quad \forall x \in A$. Quindi $f'(x) > 0 \quad \forall x \in A$. Ma la funzione tangente non si sogna minimamente di essere strettamente crescente sul suo insieme di definizione. Se qualcuno la pensa diversamente, è bruscamente invitato a rileggersi la definizione di funzione strettamente crescente. \square

10. I teoremi di de l'Hôpital

Questo paragrafo sarà breve in quanto rinvierò per gli enunciati e le dimostrazioni dei teoremi al C-S.

Mi limito solo a qualche breve considerazione.

Essenzialmente il teorema di de l'Hôpital mette in relazione $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)}$ con $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$. Quando e perché ci dobbiamo aspettare una relazione tra questi due limiti? Facciamo un esempio cinematico, cioè supponiamo che x rappresenti il tempo ed $f(x)$ e $g(x)$ la posizione all'istante x di due automobili che si muovono su una strada rettilinea. Se all'istante x_0 le due auto si trovavano entrambe al chilometro "zero", allora il rapporto $\frac{f(x)}{g(x)}$ non dovrebbe essere molto diverso dal rapporto $\frac{f'(x)}{g'(x)}$, almeno per x non troppo discosto da x_0 . Infatti, se la prima macchina parte con una velocità che è pari a k volte la velocità con cui parte la seconda, anche la distanza tra la prima macchina e la linea di partenza sarà pari a k volte la distanza tra la seconda auto e la linea di partenza (almeno per i "primi istanti").

C'è un altro modo per immaginare come mai possa esserci un rapporto tra $\frac{f(x)}{g(x)}$ e $\frac{f'(x)}{g'(x)}$. Supponiamo che f e g siano derivabili in x_0 e che sia $g'(x_0) \neq 0$. Supponiamo ancora che sia $f(x_0) = g(x_0) = 0$ (questa ipotesi non è particolarmente restrittiva: la faccio più che altro per evitare di appesantire l'argomentazione con questioni inessenziali). Allora abbiamo che $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{g(x) - g(x_0)} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{g(x) - g(x_0)} \cdot \frac{x - x_0}{x - x_0} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \cdot \frac{x - x_0}{g(x) - g(x_0)} = \frac{f'(x_0)}{g'(x_0)}$. Se supponiamo inoltre che le due funzioni derivate prime f' e g' siano continue in x_0 , abbiamo che $f'(x_0) = \lim_{x \rightarrow x_0} f'(x)$ e analogamente per la g' : pertanto possiamo dire che $\frac{f'(x_0)}{g'(x_0)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$. Abbiamo cioè in conclusione che $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{g(x) - g(x_0)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$. Si noti che per la derivabilità di f e g abbiamo che f e g sono continue e quindi $\lim_{x \rightarrow x_0} f(x) - f(x_0) = 0$ e $\lim_{x \rightarrow x_0} g(x) - g(x_0) = 0$.



= 0 . Cosa abbiamo ottenuto di interessante? Avevamo $\lim_{x \rightarrow x_0} \frac{f(x)-f(x_0)}{g(x)-g(x_0)}$ che, come abbiamo appena visto, è una forma indeterminata e abbiamo visto come esso sia uguale a $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$. Nelle ipotesi oltremodo forti che abbiamo (continuità di f' e di g' , nonché $g'(x_0) \neq 0$), ciò ci permette addirittura di "determinare" la forma indeterminata, in quanto abbiamo anche che $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)} = \frac{f'(x_0)}{g'(x_0)}$; l'importanza dei teoremi di de l'Hôpital è che essi permettono di asserire l'equivalenza tra $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)}$ e $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$ in ipotesi meno severe. In particolare, sarà irrilevante il fatto che sia $\lim_{x \rightarrow x_0} g'(x) = g'(x_0) = 0$: questo è di estrema importanza in quanto consentirà di applicare più volte il teorema di de l'Hôpital, e quindi serbare la speranza che derivando e derivando si riesca ad avere che il denominatore abbia messo giudizio e si sia rassegnato ad avere un limite diverso da zero.

Come detto all'inizio del paragrafo, per enunciati e dimostrazioni rinvio a C-S.

11. Le derivate successive

Sia data $f: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$. Prendiamo l'insieme dei punti in cui f è derivabile e chiamiamolo A_1 . Come abbiamo già notato nel §2, se ad ogni punto $x \in A_1$ associamo il numero reale $f'(x)$, otteniamo una nuova funzione reale di variabile reale, definita su A_1 , che indichiamo con f' e che chiamiamo funzione derivata prima di f .

Abbiamo quindi $f': A_1 \longrightarrow \mathbb{R}$. Questa è una funzione reale di variabile reale. Da un punto di vista astratto, formale, nessuno ci può impedire di rifare con f' tutto quello che già avevamo fatto per la funzione reale di variabile reale f . E quindi di vedere se, per esempio, f' sia derivabile in un punto $x_0 \in A_1$ il quale sia anche di accumulazione per A_1 . Se così è, cioè se $\exists \lim_{x \rightarrow x_0} \frac{f'(x)-f'(x_0)}{x-x_0} = \ell \in \mathbb{R}$, avremo che f' è derivabile in x_0 . Ma ci interessa di più continuare a fare riferimento ad f : diremo pertanto, per ragioni che mi sembrano ovvie, che f è derivabile due volte in x_0 . E indicheremo ℓ con $f''(x_0)$ (notazione sensata: dopotutto si tratta di $(f')'(x_0)$).

Possiamo ancora andare avanti. E cioè considerare la funzione derivata

seconda $f'' : A_2 \rightarrow \mathbb{R}$, dove A_2 è evidentemente l'insieme dei punti di A_1 nei quali la funzione derivata prima, f' , è derivabile a sua volta. Da qui, si avrà $f'''(x_0)$, etc., etc.

Ha qualche interesse questo gioco formale? Sì. Eccome, se ne ha. Vediamo, un po' alla rinfusa, alcune buone ragioni per prendere sul serio le derivate seconde, terze, etc. Tanto per cominciare, probabilmente tutti sanno che in cinematica la derivata prima rappresenta la velocità, mentre la derivata seconda ci dà l'accelerazione. Le derivate prime, seconde e terze hanno dei riscontri di carattere grafico, visivamente percepibili (crescenza, convessità, flessi sono parole in parte note e che comunque impareremo a conoscere più in là, nel § 13). Ma il significato "applicativo" non si limita alle prime due o tre derivate. Per esempio, nello studio della flessione di travi soggette a sforzo, è di fondamentale importanza una equazione che coinvolge la derivata quarta della funzione la quale descrive lo scostamento della trave dalla posizione di "riposo". Vi sono poi delle ulteriori ragioni, che si vedranno nel paragrafo successivo, che rendono utile considerare derivate di ordine anche molto alto.

Detto questo, vorrei però anche soffermarmi brevemente su un aspetto di carattere più astratto-formale. E cioè vorrei osservare che gli insiemi A_1 , A_2 , etc. possono anche essere molto "complicati", in generale. In pratica, però, quasi sempre si usano degli intervalli o unioni di un numero finito o infinito di intervalli¹³. Avranno quindi quasi sempre una struttura molto semplice. Non voglio nascondere il fatto che ciò rende discutibile la mia scelta di voler affrontare la derivabilità (e la continuità e i limiti) in un contesto di carattere più generale rispetto al C-S. In effetti, quasi sempre gli insiemi di definizione delle funzioni che "capitano tra i piedi" hanno una struttura semplice come quella descritta sopra (intervalli o unione di intervalli): quindi il guadagno ottenuto in generalità, rispetto al C-S, parlando di punti di accumulazione è assai modesto. L'ho fatto comunque perché ritengo che non sia modesto invece dal punto di vista concettuale (e a volte si lavora più facilmente in un contesto di carattere più generale che non in un caso particolare).

12. La formula di Taylor

I risultati di questo paragrafo sono straordinariamente utili ai fini dell'applicazione dell'analisi. Forniscono infatti nuovi strumenti per il "calcolo

¹³ Ricordo sempre che gli intervalli sono intesi non degeneri. Sennò, dato che tra gli intervalli degeneri ci sono anche i "singleton", ogni sottoinsieme di \mathbb{R} lo potrei vedere evidentemente come unione di un numero infinito di intervalli!

dei limiti", danno condizioni necessarie e condizioni sufficienti per i massimi e minimi locali, permettono di effettuare valutazioni approssimate dei valori assunti da funzioni anche "complicate".

Dietro a tutto questo sta la formula di Taylor. Ci si arriva insistendo testardamente con l'idea che stava già dietro alla definizione di differenziabilità. Chi legge dovrebbe ricordarsi che cercavamo di approssimare una funzione con un'altra "più semplice". Più precisamente, cercavamo di trovare, fra tutti i polinomi di primo grado, quello il cui grafico si accostasse meglio al grafico della funzione data, almeno vicino al punto $(x_0, f(x_0))$.

La scelta dei polinomi di primo grado era motivata da un'ovvia considerazione di semplicità: stavamo cercando qualcosa da sostituire alla funzione data, in modo da rendere il più possibile agevoli i conti da fare. Ora, questo va bene fino a un certo punto. Può benissimo darsi che uno sia disponibile a prendere in considerazione funzioni approssimanti più complicate che non i polinomi di primo grado, se può però ottenere una migliore approssimazione. Per esempio, potremmo pensare di usare polinomi di 2° grado (che, dopotutto, sono ancora abbastanza maneggevoli) per ottenere una approssimazione migliore. Più precisamente, tenendo conto di quanto visto a proposito della differenziabilità, possiamo pensare di cercare un polinomio g di 2° grado t.c. $f(x)-g(x)$ sia minore di quanto non lo fosse $f(x)-[f(x_0)+f'(x_0)\cdot(x-x_0)]$. E' chiaro che solo se si ottiene veramente una approssimazione migliore vale la pena di passare dai polinomi di 1° grado a quelli di 2°. Effettivamente le cose stanno così. Anzi, in generale si ottiene una approssimazione che migliora via via di più ogni qual volta si aumenta il grado del polinomio approssimante: tutto ciò, però, purché la funzione data sia derivabile un numero conveniente di volte. Si ha addirittura una regola di questo genere: se una funzione è derivabile n volte, è vero che aumentando il grado del polinomio migliora l'approssimazione, ma solo fino al grado n . E' inutile spingersi oltre.

Ci sono molte altre considerazioni da fare a proposito della formula di Taylor, ma mi sembra più furbo farle dopo aver visto come stanno veramente le cose da un punto di vista formale. Vediamo quindi subito il primo risultato, che costituisce una diretta generalizzazione delle considerazioni fatte a proposito della differenziabilità.

Teorema 1 (formula di Taylor con resto nella forma di Peano) Sia data $f:I \longrightarrow \mathbb{R}$, I intervallo, $x_0 \in I$. Supponiamo f sia derivabile $n-1$ volte in I e derivabile n volte in x_0 . Allora, per ogni $x \in I$ si ha:

$$f(x) = f(x_0) + f'(x_0) \cdot (x-x_0) + f''(x_0) \cdot \frac{(x-x_0)^2}{2!} + \dots + f^{(n)}(x_0) \cdot \frac{(x-x_0)^n}{n!} + \omega_n(x) \cdot (x-x_0)^n,$$

dove $\omega_n: I \rightarrow \mathbb{R}$ soddisfa le condizioni $\lim_{x \rightarrow x_0} \omega_n(x) = \omega_n(x_0) = 0$.□

Dimostrazione vedi C-S.□

Vi è però anche un'altra strada che conduce ad un'altra versione della formula di Taylor. Si tratta di portare avanti, di generalizzare, l'idea che sta dietro il teorema di Lagrange, seguendo lo stesso itinerario che ci ha condotti dalla differenziabilità alla formula di Taylor col resto di Peano: e cioè, banalmente, considerare un polinomio di grado generico n .

Teorema 2 (formula di Taylor con resto nella forma di Lagrange) Sia $f: I \rightarrow \mathbb{R}$, I intervallo, $x_0 \in I$. Supponiamo f sia derivabile n volte in I , con derivata n -esima continua. Supponiamo inoltre che in ogni punto di $I \setminus \{x_0\}$ f sia derivabile $n+1$ volte. Allora, per ogni $x \in I$ esiste $\xi \in]x_0, x[$ (se $x_0 < x$, senno $\xi \in]x, x_0[$; se $x = x_0$ il tutto ha ben scarso interesse...) t.c.:

$$f(x) = f(x_0) + f'(x_0) \cdot (x-x_0) + \dots + f^{(n)}(x_0) \cdot \frac{(x-x_0)^n}{n!} + f^{(n+1)}(\xi) \cdot \frac{(x-x_0)^{n+1}}{(n+1)!} . \square$$

Dimostrazione vedi C-S.□

Quel che bisogna assolutamente capire è che le due formule di Taylor servono per scopi del tutto distinti tra loro.

La formula di Taylor col resto di Peano ci permette di ricavare dei risultati di carattere locale relativamente ad x_0 . Per esempio, ci consente di calcolare dei limiti, oppure ci offrirà condizioni necessarie e condizioni sufficienti per massimi e minimi locali. Essa, tuttavia, non ci dice nulla sul valore che f assume in uno specifico punto \bar{x} dato. Infatti abbiamo $f(\bar{x}) = f(x_0) +$

$f'(\bar{x}) \cdot (\bar{x}-x_0) + \dots + f^{(n)}(x_0) \cdot \frac{(\bar{x}-x_0)^n}{n!} + \omega_n(\bar{x}) \cdot (\bar{x}-x_0)^n$: ebbene, noi non sappiamo assolutamente nulla su $\omega_n(\bar{x})$ ¹⁴. Quindi, se $x_0 = 0$ e voglio sapere quanto fa f in $\bar{x} = 0.1$, non posso fare alcun affidamento sulla formula di Taylor con il

¹⁴ Salvo il fatto, ovviamente, che :

$$\omega_n(\bar{x}) = \left[f(\bar{x}) - \left[f(x_0) + f'(x_0) \cdot (\bar{x}-x_0) + \dots + f^{(n)}(x_0) \cdot \frac{(\bar{x}-x_0)^n}{n!} \right] \right] / (\bar{x}-x_0)^n .$$

Da cui si vede che, se conosciamo $f(\bar{x})$, possiamo sapere quanto vale $\omega_n(\bar{x})$: ma il nostro problema era proprio quello di saper il valore di $f(\bar{x})$.

resto di Peano.

Completamente diverso è il discorso per la versione con il resto di Lagrange.

Ho infatti
$$f(\bar{x}) = f(x_0) + f'(x_0) \cdot (\bar{x} - x_0) + \dots + f^{(n)}(x_0) \cdot \frac{(\bar{x} - x_0)^n}{n!} + f^{(n+1)}(\xi) \cdot \frac{(\bar{x} - x_0)^{n+1}}{(n+1)!} .$$

Non bisogna farsi ingannare dalle apparenze. E' vero che non so assolutamente chi sia ξ . Però ho una informazione su ξ ! So che $\xi \in]x_0, \bar{x}[$ ¹⁵. Se riesco a trovare $M \in \mathbb{R}$ t.c. $|f^{(n+1)}(x)| \leq M \quad \forall x \in]x_0, \bar{x}[$, questo mi garantisce che è anche $|f^{(n+1)}(\xi)| \leq M$: da qui ottengo una stima dell'errore commesso se ad f sostituiamo il polinomio di Taylor di ordine n . Cioè:

$$|f(\bar{x}) - [f(x_0) + f'(x_0) \cdot (\bar{x} - x_0) + \dots + f^{(n)}(x_0) \cdot \frac{(\bar{x} - x_0)^n}{n!}]| \leq M \cdot \frac{|\bar{x} - x_0|^{n+1}}{(n+1)!} .$$

Quindi occorrerà sincerarsi che sia $M \cdot \frac{|\bar{x} - x_0|^{n+1}}{(n+1)!} < E$, dove E è l'errore assoluto che non vogliamo superare, per poter dire che la stima ottenuta è utile.

¹⁵ Se $x_0 < \bar{x}$, altrimenti $\xi \in]\bar{x}, x_0[$: è il solito guaio con le notazioni che si usano per gli intervalli, che obbliga ogni volta a fare queste frustranti precisazioni

13. "Disegnare il grafico di f "

Vorrei fare alcune considerazioni relative a quella idea piuttosto sfuggente cui si fa riferimento quando si parla di disegnare il grafico di una funzione.

Evidentemente si vuole fare riferimento al problema di rappresentare su un foglio di carta (o altro supporto: lavagna, monitor, lastra di acciaio, etc.) "qualcosa" che assomigli il più possibile al grafico di f .

Vorrà dire che sul nostro foglio di carta avremo posto un sistema di coordinate e che cercheremo di fare un disegno che rappresenti in maniera adeguata il grafico di f nel dato sistema di coordinate.

E' chiaro che, se uno vuole un buon risultato quantitativo e qualitativo, un'idea è di disegnare un grafico "per punti" scegliendo punti "molto vicini" tra loro. In effetti questo è quello che si fa praticamente sempre quando questo compito è fattibile tenendo conto della difficoltà di calcolare $f(x)$ e degli strumenti di calcolo a disposizione per fare questi calcoli. Oggi, grazie alla microelettronica, è ora possibile avere a disposizione una grande capacità di calcolo¹⁶: è del tutto realistico disegnare il grafico, per esempio, di $f(x) = e^{\arctg \sqrt{x}} \cdot (|\sin(x)| + 1)^{x+3} \sqrt{x}$ con $x \in [0, 100]$ calcolando f in $x_k = k \cdot 10^{-3}$, per $k = 0, 1, \dots, 10^5$. E' abbastanza evidente che in questo modo si ottiene una rappresentazione del grafico di f piuttosto fedele (non solo: ma è in genere anche facile eventualmente variare intervalli, scala, finezza della suddivisione, etc., qualora interessassero particolari che non si è riusciti a mettere subito in chiara luce).

A cosa serve allora tutto il macchinario sulla ricerca di crescita, decrescenza, concavità, flessi, massimi, asintoti, etc.? Possiamo dire tranquillamente che serve a ben poco. Naturalmente la terminologia, le definizioni e i teoremi chiave che riguardano i termini sopra citati sono importanti. Non solo: spesso può essere utile fare una accurata verifica di proprietà che sembrano essere presenti nel disegno. Di più: senza una buona padronanza di questi concetti (che corrispondono a proprietà molto significative nelle applicazioni) si rischia di non saper interpretare ciò che il calcolatore (o calcolatrice) sforna, di non essere cioè in grado di fare una buona lettura dell'output ottenuto. Quello che è

¹⁶ In realtà, dipende anche da dove si vive: nel "Sud" del mondo, questa "capacità di calcolo" è ancora poco diffusa.

meno importante è una sofisticata padronanza dei metodi di calcolo utili a trovare per via analitica la presenza di flessi, estremi relativi, intervalli di crescita, etc. E' un po' lo stesso discorso che vale per le tecniche di derivazione: a cosa serve saper fare le derivate se i programmi di calcolo simbolico richiedono solo di scrivere l'espressione di f e poi di premere un tasto per ottenere l'espressione di f' ? E' quindi importante curare di più la comprensione effettiva dei concetti e dei teoremi (l'ambito delle ipotesi, in particolare) che non una capacità di calcolo sofisticata (naturalmente, se uno è anche bravo a fare i conti, tanto meglio).

Pertanto, l'utilità di saper "disegnare il grafico" di una funzione, cioè di sapere trovare massimi e minimi, crescita, flessi, etc., è più che altro da vedersi quale una sorta di "esercizio di riepilogo", che richiede per la soluzione di sapere utilizzare opportunamente i vari strumenti del calcolo differenziale. L'obiettivo è quindi rovesciato: non bisogna imparare i teoremi del calcolo differenziale per disegnare dei bei grafici, ma viceversa il problema di disegnare grafici va visto come una utile palestra per allenarsi nella comprensione e collegamento dei principali risultati e concetti del calcolo differenziale.

Va anche detto che, tradizionalmente, quando si chiede di "disegnare il grafico di f " (nel senso di tracciare una buona approssimazione del grafico di f mediante i risultati del calcolo differenziale) viene privilegiato oltre misura l'aspetto per così dire "qualitativo". E ciò può anche portare a distorsioni nel disegno rispetto ad una accurata tabulazione ottenuta con l'ausilio di un efficiente mezzo di calcolo. Come conferma e delucidazione di quanto ho appena detto, invito lo studente a considerare attentamente il seguente esercizio. Per la definizione di flesso, chi non la conosce può vedere la successiva definizione 1.

Esercizio 1 Disegnare il grafico di $f(x) = |x+1| \cdot e^{-(1/x)}$ su $]0, +\infty[$, mettendo in rilievo la presenza di un punto di flesso. Confrontare il disegno ottenuto con un grafico ottenuto da calcolatore¹⁷. Riflettere sull'evidenza "grafica" o meno di tale punto di flesso. Cercare di capire cosa significhi un punto di flesso così poco "appariscante". □

Esercizio 2 Simile al precedente. Osservare che $f(x) = x^4 - 2 \cdot 10^{-50} \cdot x^2$ ha tre

¹⁷ In ogni caso un grafico da calcolatore per f lo si trova alla fine del capitolo.

punti di estremo locale e che in particolare il punto 0 è punto di massimo locale. Confrontare questi risultati con un grafico ottenuto da calcolatore. Riflettere. □

Vorrei comunque dare un paio di definizioni relative ad elementi di rilievo per i grafici di funzioni, perché spesso attorno a tali caratteristiche vi è confusione.

Definizione 1 Sia $f:A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, x_0 interno ad A , f derivabile in x_0 . Diremo che x_0 è un punto di flesso ascendente (rispettivamente: discendente) per f se:

$$\exists \delta > 0 \text{ t.c. } \begin{cases} \forall x \in]x_0 - \delta, x_0[& f(x) \leq f(x_0) + f'(x_0) \cdot (x - x_0) \\ \forall x \in]x_0, x_0 + \delta[& f(x) \geq f(x_0) + f'(x_0) \cdot (x - x_0) \end{cases}$$

(rispettivamente: $\begin{cases} \forall x \in]x_0 - \delta, x_0[& f(x) \geq f(x_0) + f'(x_0) \cdot (x - x_0) \\ \forall x \in]x_0, x_0 + \delta[& f(x) \leq f(x_0) + f'(x_0) \cdot (x - x_0) \end{cases}$) . □

Osservazione 1 Ho richiesto che x_0 sia un punto interno per A . Sarebbe stato sufficiente richiedere che x_0 fosse di accumulazione per A sia da sinistra che da destra. □

Osservazione 2 Graficamente, dire che x_0 è punto di flesso ascendente significa dire che f a sinistra di x_0 sta sotto la retta tangente in x_0 , mentre a destra sta sopra. □

Osservazione 3 Si noti che, anche se x_0 è punto di flesso ascendente, la funzione f può tranquillamente essere strettamente decrescente. Vedasi per esempio $f(x) = -\arctg(x)$! Quindi attenzione a non farsi trarre in inganno dal significato che le parole usate hanno nel linguaggio comune. □

Osservazione 4 Vedremo nel prossimo paragrafo che un punto di flesso può essere definito in altro modo (non equivalente!). Occorre quindi prestare attenzione a quale definizione viene usata sul libro che si sta leggendo. □

Si può dimostrare (usando la formula di Taylor con il resto di Peano) il seguente:

Teorema 1 Se f soddisfa le ipotesi del teorema 12.1 per $n=2$, allora CN affinché x_0 sia un punto di flesso è che si abbia $f''(x_0) = 0$. Se f soddisfa le ipotesi del teorema 12.1 con $n=3$, allora CS affinché x_0 sia punto di

flesso è che $f''(x_0) = 0$ e $f'''(x_0) \neq 0$ (se $f'''(x_0) > 0$, abbiamo un flesso ascendente, se $f'''(x_0) < 0$ il flesso è discendente).□

Un'altra proprietà d'interesse grafico per una funzione è quella di avere asintoti.

Definizione 2 Dato $A \subseteq \mathbb{R}$, diremo che $+\infty$ è "di accumulazione" per A se: $\forall \delta > 0, A \cap]\delta, +\infty[\neq \emptyset$. □

Definizione 3 Sia $f: A \rightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$, e supponiamo che $+\infty$ sia "di accumulazione" per A . Se $\exists k, m \in \mathbb{R}$ t.c. $\lim_{x \rightarrow +\infty} [f(x) - (kx + m)] = 0$, diremo che f ha la retta di equazione $y = kx + m$ come asintoto "a $+\infty$ ". □

Osservazione 5 Ovviamente, considerazioni assolutamente identiche possono essere fatte per $x \rightarrow -\infty$. Il lettore interessato può anche provare a dare una definizione per $x \rightarrow \infty$. □

Osservazione 6 Se $k = 0$, diremo che la retta $y = m$ è un asintoto orizzontale per f (se necessario, si preciserà "a $+\infty$ " se siamo nel caso della definizione 3; "a $-\infty$ " oppure "a ∞ " nei casi indicati nell'osservazione 5). Se $k \neq 0$, parleremo di asintoto obliquo. □

Osservazione 7 Anche qui, come nel caso dei flessi ascendenti e discendenti, il linguaggio corrente (in questo caso anche l'etimologia) può indurre in equivoci. Se si riflette sulla definizione di asintoto, si noterà che non è per nulla proibito ad f di "toccare" la retta asintotica. Se il grafico di f è per esempio una retta, allora esso "toccherà" dappertutto l'asintoto! Si consideri poi anche il caso di $f(x) = (\sin(x))/x$. □

Definizione 4 Sia $f: A \rightarrow \mathbb{R}$, x_0 di accumulazione per A . Se $\lim_{x \rightarrow x_0} f(x) = \infty$, diremo che la retta di equazione $x = x_0$ è un asintoto verticale per f . □

Osservazione 8 Ovviamente il grafico di f potrà "toccare" un asintoto verticale al più nel punto x_0 stesso (e ciò si verifica se e solo se $x_0 \in A$, per definizione stessa di funzione). □

Per quanto riguarda gli asintoti obliqui, c'è un semplice trucco per trovarli. Per trovare k , occorre fare $\lim_{x \rightarrow +\infty} \frac{f(x)}{x}$ e poi per (l'eventuale) m si

deve guardare se esiste in \mathbb{R} $\lim_{x \rightarrow +\infty} [f(x) - kx]$. E' lasciato al lettore enunciare un preciso risultato. Si noti che l'esistenza (in \mathbb{R}) di $\lim_{x \rightarrow +\infty} \frac{f(x)}{x}$ non garantisce che vi sia un asintoto obliquo. Vedasi il caso di $f(x) = x + \log(x)$. \square

14. Funzioni convesse

Per proprietà e teoremi relativi alle funzioni convesse rinvio senz'altro a C-S.

Vorrei solo fare un po' di considerazioni.

Comincio dall'idea "grafica" della convessità, dalla quale si ottiene la definizione formale. L'idea è che una funzione è convessa se il suo grafico sta sotto ad ogni sua "corda". Ovverossia, presa $f: I \rightarrow \mathbb{R}$, I intervallo di \mathbb{R} , e presi due punti $x_1, x_2 \in I$, vogliamo che il segmento congiungente i punti $(x_1, f(x_1))$ ed $(x_2, f(x_2))$ stia "sopra" al grafico di f . E' molto semplice esprimere analiticamente questo fatto. Supponiamo che sia $x_1 < x_2$: basta considerare un punto $x \in [x_1, x_2]$ e richiedere che $f(x) \leq r(x)$: qui $r(x)$ è la funzione che ha come grafico la retta passante per i punti $(x_1, f(x_1))$ ed $(x_2, f(x_2))$. L'equazione di tale retta è $y - f(x_1) = \frac{f(x_2) - f(x_1)}{x_2 - x_1} \cdot (x - x_1)$. Ovveros-

sia, $r(x) = f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} \cdot (x - x_1)$. Quindi la condizione di convessità è $f(x) \leq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} \cdot (x - x_1) \quad \forall x \in [x_1, x_2]$.

Forse non si vede immediatamente la connessione tra questa relazione e la definizione usuale di convessità. Ci si può giungere facilmente sfruttando la seguente osservazione: al variare di μ in $[0, 1]$, il punto $x_1 + \mu \cdot (x_2 - x_1)$ descrive l'intervallo $[x_1, x_2]$ (la verifica è ovvia: $x_1 \leq x_1 + \mu \cdot (x_2 - x_1) \leq x_2 \Leftrightarrow 0 \leq \mu \cdot (x_2 - x_1) \leq x_2 - x_1 \Leftrightarrow 0 \leq \mu \leq 1$). Allora, dire

$$f(x) \leq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} \cdot (x - x_1) \quad \forall x \in [x_1, x_2]$$

è equivalente a:

$$f(x_1 + \mu \cdot (x_2 - x_1)) \leq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} \cdot \mu \cdot (x_2 - x_1) \quad \forall \mu \in [0, 1],$$

ovverossia:

$$(\blacktriangleright) \quad f(x_1 + \mu \cdot (x_2 - x_1)) \leq f(x_1) + \mu \cdot (f(x_2) - f(x_1)) \quad \forall \mu \in [0,1] .$$

Questa relazione non è ancora quella che tradizionalmente descrive la convessità, e cioè:

$$(*) \quad f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2) \quad \forall \lambda \in [0,1] .$$

Basta però prendere $\lambda = 1 - \mu$, osservare che μ varia in $[0,1]$ se e solo se λ varia in $[0,1]$ ¹⁸, e fare i conti con un po' di algebratta.

E' senza dubbio una mia mania, ma a me piace sottolineare come la relazione (\blacktriangleright) sia una espressione più diretta della convessità, ed abbia un pregio rispetto alla più consueta $(*)$: al variare di μ in $[0,1]$, il segmento $[x_1, x_2]$ viene percorso da $x_1 + \mu \cdot (x_2 - x_1)$ "nel verso giusto" (ad esempio, per $\mu = 0$ otteniamo x_1 , per $\mu = 1$, abbiamo x_2); quando il parametro λ varia tra 0 e 1, il punto $\lambda x_1 + (1-\lambda)x_2$ si sposta da x_2 a x_1 .

Detto questo, rinvio a C-S per i risultati più importanti relativi alla convessità.

Mi limito a due parole sulla questione dei flessi, come avevo promesso nella osservazione 13.4.

E' facile provare che, se f è derivabile in x_0 ed inoltre $\exists \delta > 0$ t.c. f è convessa su $[x_0 - \delta, x_0[$ e concava su $]x_0, x_0 + \delta[$, allora f ha in x_0 un punto di flesso ascendente. In effetti questa è di solito la situazione che si ha in mente quando si parla di flesso (assieme a quella in cui concavo/convesso sono rovesciati, e cioè per i flessi discendenti). Non solo, talvolta sui libri viene data questa come definizione di flesso (vedasi per esempio le dispense di Zolezzi). Occorre allora tenere ben presente che le due definizioni non sono affatto equivalenti. Può cioè capitare che x_0 sia un punto di flesso nel senso della mia definizione 13.1, senza però che si possa trovare δ in modo che f sia convessa su $[x_0 - \delta, x_0[$ e concava su $]x_0, x_0 + \delta[$. E' una situazione simile a quella che si ha con la "crescenza in un punto", per usare la terminologia di C-S: vedasi il § 7. Mi limiterò a mostrare un esempio di questa situazione patologica.

$$\text{Esempio 1} \quad \text{Sia } f(x) = \begin{cases} |x^2 \cdot \text{sen}(1/x)| & \text{se } x > 0 \\ 0 & \text{se } x = 0 \\ -|x^2 \cdot \text{sen}(1/x)| & \text{se } x < 0 \end{cases} \quad \text{Provare che } 0 \text{ è un punto}$$

di flesso ascendente. Provare che non esiste $\delta > 0$ t.c. f sia convessa su $]x_0 - \delta, x_0[$ (suggerimento: studiare il segno di f''). □

¹⁸ Anche se "a rovescio".

15. Conclusioni

Un lettore attento avrà percepito un cambiamento progressivo nel proseguire dei capitoli.

Il grado di dettaglio delle dimostrazioni è andato diminuendo. Inoltre, mentre ho continuato a riservare grande spazio all'introduzione dei concetti nuovi e alla illustrazione dei risultati chiave, ho invece prestato scarsa attenzione ai risultati per così dire secondari. I rinvii a C-S sono diventati in questo caso la norma. Mi pare superfluo dare delle ragioni per questo modo mio di procedere. Ho però voluto menzionare questo fatto per mettere in guardia il lettore: è assolutamente necessario integrare la lettura di questi appunti con il C-S (o altro testo equivalente). Non pensi il lettore di poter avere una immagine compiuta dell'analisi¹⁹ solo da questi appunti: se ne ricava invece una immagine parziale e molto deformata.

Detto questo, passo brevemente in rassegna i punti salienti di questo capitolo. Ovviamente tutto comincia con la definizione di derivata e le ragioni che portano ad introdurre questa idea. Vi sono, poi, come al solito, risultati per la cui dimostrazione non è richiesto l'assioma di completezza (in particolare, i risultati sulla compatibilità tra derivazione da un lato e le operazioni algebriche e "insiemistiche"²⁰ dall'altro).

Si osservi che il ruolo dell'assioma di completezza diventa sempre più preminente. Non solo esso è essenziale per il teorema fondamentale del calcolo differenziale, o teorema di Lagrange: è anche cruciale in situazioni nelle quali poteva non essere così scontato che lo fosse. Un esempio paradigmatico è dato dalla relazione tra monotonia e derivate: per ottenere informazioni utili, significative, sulla monotonia di una funzione a partire dal segno della derivata prima, abbiamo bisogno del teorema di Lagrange (vedasi teorema 9.4). Ancor di più! Per ottenere un risultato che può sembrare banale da dimostrare, e cioè che se una funzione ha sempre la derivata nulla allora è costante, abbiamo dovuto metterci su un intervallo e abbiamo avuto bisogno del teorema di Lagrange (e quindi della completezza). Chi abbia voglia di dedicare un po' di tempo aggiuntivo, diciamo un

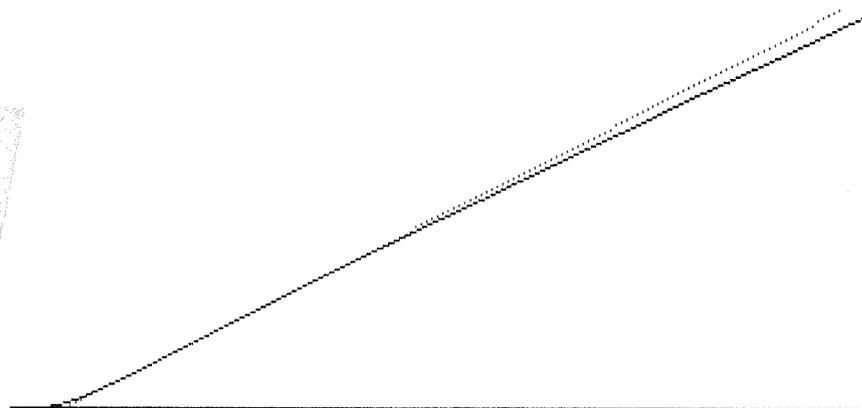
¹⁹ Neanche dei risultati elementari nel semplice contesto delle funzioni reali di una sola variabile reale!

²⁰ Penso all'inversione e composizione di funzioni.

paio di ore, a questa questione, è invitato a cercare di ottenere una dimostrazione diretta del fatto che " $f' = 0 \Rightarrow f$ costante", senza usare il teorema di Lagrange.

Mi pare opportuno sottolineare come questo capitolo offra, nettamente più dei precedenti, risultati interessanti per una accurata descrizione di proprietà significative delle funzioni (come si trovano massimi e minimi, flessi, l'idea di convessità). Offre, infine, anche un risultato di grande valore per la approssimazione delle funzioni: la formula di Taylor ci dà infatti la possibilità di sostituire ad una funzione "complicata" un polinomio, avendo una stima dell'errore che si commette²¹.

```
[x1,x2] = 0.02    5.00    — Grafico di f(x)=(abs(x+1))*exp(-(1/x))
[y1,y2] = -1.00   5.00    .... Grafico di f_bis(x)=(3*x-1)/exp(1)
```



²¹ Va da sé che, senza avere una stima dell'errore, dire che qualcosa approssima qualcos'altro, è una espressione verbale priva di significato reale.

CAPITOLO V

CALCOLO INTEGRALE

1. La definizione di integrale

Tra le varie funzioni che ha il calcolo integrale, quella che meglio si presta ad essere descritta riguarda le aree delle figure piane. Infatti, il calcolo integrale permette di dare una buona risposta al problema di definire (e calcolare) l'area di una classe la più ampia possibile di figure. Il calcolo integrale ha anche altre importanti applicazioni e motivazioni, sulle quali non mi soffermo (ricordo solo, dalla fisica, l'idea del lavoro compiuto da una forza).

Dunque occupiamoci di aree. Questo è un argomento che viene visto nella scuola secondaria nell'ambito della geometria euclidea. Spesso il livello di comprensione è sconsolatamente basso, in quanto questo approccio richiede preliminarmente la teoria delle grandezze (geometriche), che non è facile da digerire. Resta così un po' sospeso a mezz'aria, a metà strada tra l'approccio euristico-intuitivo della scuola media ed un approccio più rigoroso. Comunque, qualsiasi sia il livello di rigore cui si è giunti, di solito è noto come trovare l'area dei poligoni e del cerchio. E basta. Di figure non particolarmente difficili, come quella che è racchiusa da un'ellisse, non si sa cosa dire¹.

Come fare ad estendere (e ampiamente) la classe di figure cui si riesce ad assegnare un'area? Molto semplice: con la geometria analitica e il calcolo integrale. Osserviamo che, per esempio, una ellisse di semiassi a e b ha una equazione nel piano cartesiano² che è $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Se siamo interessati alla figura

¹ Non certo da parte della geometria euclidea studiata nelle secondarie.

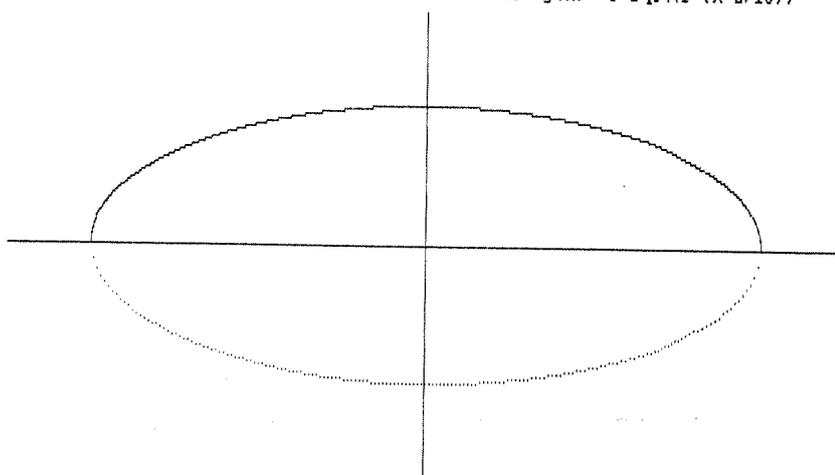
² Scelto convenientemente il sistema di riferimento!

"racchiusa" dentro, essa sarà descritta dalla relazione $\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1$: ovvero, ossia,

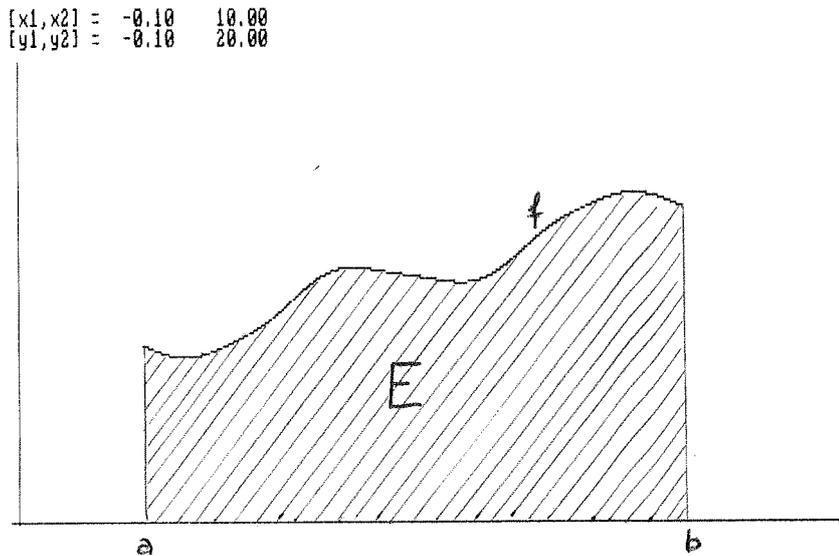
è la porzione di piano "compresa" tra il grafico di $g(x) = -b \cdot \sqrt{1 - \frac{x^2}{a^2}}$ e

$f(x) = b \cdot \sqrt{1 - \frac{x^2}{a^2}}$. Siamo quindi ricondotti al problema di definire l'area del sottinsieme di \mathbb{R}^2 che è compreso tra i grafici di due funzioni. Anzi, considerazioni di simmetria ci inducono a ritenere che ci basti trovare l'area compresa tra la funzione positiva f e l'asse delle x : una volta trovata quest'area, basterà moltiplicare per due.

```
[x1,x2] = -5.00    5.00
[y1,y2] = -5.00    5.00
— Grafico di f(x)=3*sqrt(1-(x^2/16))
... Grafico di g(x)=-3*sqrt(1-(x^2/16))
```



Il campo d'indagine è quindi ben recintato. Data $f:[a,b] \rightarrow \mathbb{R}$, con $f(x) \geq 0 \quad \forall x \in [a,b]$, occorre riuscire a definire e, possibilmente, calcolare l'area di $E = \{ (x,y) \in \mathbb{R}^2 : a \leq x \leq b \text{ e } 0 \leq f(x) \}$. Ovvero, della figura tratteggiata alla pagina seguente, dove E è rappresentato nel piano col solito tramite della geometria euclidea.

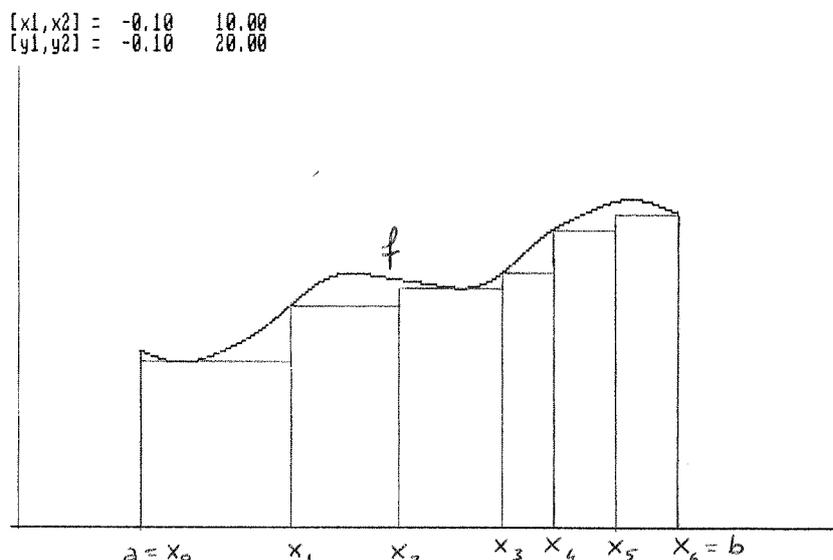


Come fare per definire l'area di E ? Come già detto in altre occasioni, introducendo discorsivamente i concetti principali dell'analisi, non è affatto mia intenzione ripercorrere il cammino storico, né adottare un punto di vista pseudo ingenuo, andando a tentoni. C'è una idea buona, anche se non molto semplice da sfruttare: viene da lontano, trattandosi del metodo di esaustione di Archimede.

L'idea è veramente semplice: "approssimiamo E per difetto e per eccesso" mediante delle figure elementari di cui sappiamo quant'è l'area. Dopodiché, raffinando sempre più l'approssimazione, con una qualche procedura di limite o suo parente stretto, potremo arrivare all'area di E .

Quindi dobbiamo trovare figure semplici di cui conosciamo l'area; per di più, se vogliamo che approssimino per difetto l'area di E , una scelta sensata è di prenderle contenute dentro E (e contenenti E per avere approssimazioni per eccesso). D'ora in avanti, mi concentrerò sulle approssimazioni per difetto: per quelle per eccesso, le considerazioni sono del tutto simili.

Osserviamo che l'insieme E ha una struttura particolare, per cui una scelta abbastanza naturale è quella di usare come figure contenute dentro E insiemi costituiti da più rettangoli accostati l'uno a fianco dell'altro, come indicato in figura.



Questi rettangoli hanno come base un sottointervallo di $[a, b]$ e una altezza tale da non superare il minimo valore di f su tale sottointervallo (di modo che la figura ottenuta mettendo assieme questi rettangoli sia contenuta dentro E).

Volendo precisare le cose, potremmo procedere nel modo seguente. Fissiamo $n+1$ punti x_0, x_1, \dots, x_n nell'intervallo $[a, b]$ t.c. $a = x_0 < x_1 < \dots < x_n = b$. Così otteniamo n intervalli $[x_{i-1}, x_i]$, per $i = 1, \dots, n$, che saranno le basi dei rettangoli in questione.

Per quanto riguarda le altezze, se vogliamo che il grafico di f stia "sopra" al rettangolo, dovrà essere soddisfatta la condizione:

$$h_i \leq \min \{ f(x) : x \in [x_{i-1}, x_i] \},$$

dove con h_i indichiamo appunto l'altezza dell' i -esimo rettangolino. Ma allora tanto vale scegliere come h_i proprio $\min \{ f(x) : x \in [x_{i-1}, x_i] \}$.

Resta solo un "piccolo problemino". L'insieme $f([x_{i-1}, x_i])$, sarà in generale un sottoinsieme di \mathbb{R} con infiniti elementi. Chi ci garantisce che abbia minimo, di modo che la scelta che abbiamo fatto per h_i sia sensata?

Si può rispondere a queste obiezioni seguendo due strade diverse.

La prima è abbastanza ovvia: per essere sicuri che il minimo ci sia, basterà assumere che f sia continua su $[a, b]$. Se così è, f è anche continua sugli intervalli $[x_{i-1}, x_i]$ e quindi ha certamente minimo grazie al teorema di Weierstrass.

La seconda è un po' meno ovvia, ma non particolarmente difficile. Se il lettore prova a riguardare quanto detto a pagina I.14, vi troverà scritto che il \sup è una sorta di surrogato per il massimo, quando questo non c'è. E allora, visto che a noi interessa il minimo, potremmo pensare di prendere

$$h_1 = \inf \{ f(x) : x \in [x_{i-1}, x_i] \} .$$

C'è questo \inf ? Sì, grazie alla completezza di \mathbb{R} , purché $\{ f(x) : x \in [x_{i-1}, x_i] \}$ sia inferiormente limitato³. Sarà quindi sufficiente richiedere ad f di essere inferiormente limitata su $[x_{i-1}, x_i]$ per poter avere a disposizione gli h_1 (nel caso in esame non ci sono problemi, perché abbiamo addirittura supposto che fosse $f(x) \geq 0 \quad \forall x \in [a, b]$: quindi f è certamente inferiormente limitata). Ma va bene la scelta dell'estremo inferiore? Ricordiamo che a noi serviva h_1 t.c. fosse $h_1 \leq f(x) \quad \forall x \in [x_{i-1}, x_i]$. Cioè a noi serviva un minorante per $f([x_{i-1}, x_i])$. E quindi la scelta fatta va benissimo, poiché l'estremo inferiore è proprio "il migliore dei minoranti possibili", essendo il massimo dei minoranti. Quindi, nessun problema: scegliendo $h_1 = \inf \{ f(x) : x \in [x_{i-1}, x_i] \}$, otteniamo che la figura complessivamente ottenuta è contenuta dentro E , quindi ci fornisce una approssimazione per difetto.

Abbiamo quindi due strade percorribili, a priori. C'è quella, un po' più facile, di assumere che f soddisfi la condizione di essere continua, e quella di richiedere ad f solo di essere inferiormente limitata (ma in realtà servirà limitata, perché dovremo trovare anche le approssimazioni per eccesso).

Come il lettore smaliziato avrà ormai capito, scelgo la seconda.

Resta ancora da fare un importante passo.

Per ora ci siamo limitati a trovare una figura semplice (un "plurirettangolo") contenuta in E e, per ovvia analogia, una figura semplice contenente E ⁴. Abbiamo pertanto ottenuto una approssimazione per difetto e una per eccesso per l'area di E .

Come migliorarla?

L'idea è abbastanza scontata. Suddividiamo $[a, b]$ in un maggior numero di intervallini, più piccoli. E speriamo che in questo modo la approssimazione migliori. Potremmo per esempio pensare di dividere $[a, b]$ in n parti uguali: otterremmo così due successioni di aree approssimanti l'area di E , una per difetto

³ Non vuoto lo è, perché immagine mediante f di un insieme non vuoto.

⁴ Basta prendere $k_i = \sup \{ f(x) : x \in [x_{i-1}, x_i] \}$.

e l'altra per eccesso, che sperabilmente convergono entrambe all'area di E . Oppure possiamo seguire la strada indicata da C-S. Cioè, considerare una coppia di classi separate. Mettiamo in A tutte le aree dei plurirettangoli contenuti in E e in B le aree dei plurirettangoli contenenti E , al variare di tutte le possibili partizioni di $[a,b]$ in sottointervalli.

E' sensato aspettarsi che:

- (A,B) sia una coppia di classi separate
- l'area di E sia l'elemento separatore per (A,B) .

Potrebbe sembrare che, a questo punto, non si tratti altro che di mettere in bella copia questi discorsi euristici e dare le definizioni formali. Prima di passare ad esse, vorrei però prima sottolineare che nelle ultime affermazioni sono nascosti due problemi.

Il primo non è complicato da risolvere, però va notato che non è proprio scontato (pur se "evidente dal disegno") che le classi A e B siano effettivamente separate: con un pizzico di inventiva comunque riusciremo a provarlo.

Il secondo è invece un problema di carattere fondamentale.

Ho parlato dell'elemento separatore. Ma perché posso parlarne al singolare? L'assioma di completezza mi garantisce solo che esista un elemento separatore, non che sia unico.

Questo non è solo un problema tecnico. E' un problema di fondo. Il discorso che abbiamo fatto finora partiva infatti dall'ipotesi (data per scontata) che E avesse un'area. Ora, questo può andare bene proprio per un discorso euristico come quello che abbiamo fatto. Però, il nostro problema non è solo quello di calcolare l'area di E : c'è prima da sistemare la questione preliminare, e cioè sapere se ha senso parlare di area per E .

In altre parole, il problema è di vedere se si può estendere in qualche modo la definizione di area da figure "semplici" come poligoni e cerchi figure di carattere più generale come è l'insieme E , cioè $\{ (x,y) \in \mathbb{R}^2 : a \leq x \leq b \text{ e } 0 \leq f(x) \}$.

Sarà quindi opportuna una pausa di riflessione. Abbiamo messo in piedi un sacco di cose e argomentazioni, ma c'è tutto il discorso da riorganizzare.

Ricomincio quindi daccapo, ripetendo in modo più compatto quanto detto finora, per meglio evidenziare la questione da risolvere.

E' data $f: [a,b] \longrightarrow \mathbb{R}$, limitata (abbiamo visto perché ci fa comodo

questa ipotesi) e t.c. $f(x) \geq 0 \quad \forall x \in [a, b]$. Dato $E = \{ (x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ e } 0 \leq f(x) \}$, suddividiamo $[a, b]$ in intervallini come già visto, ed osserviamo che il plurirettangolo costituito dai rettangoli aventi per base $[x_{i-1}, x_i]$ e per altezza $\inf \{ f(x) : x \in [x_{i-1}, x_i] \}$ è contenuto dentro E , così come quello avente come altezza $\sup \{ f(x) : x \in [x_{i-1}, x_i] \}$, contiene E .

Ma allora, se vogliamo dare una definizione sensata di area per E , essa dovrà verificare la condizione

$$\begin{aligned} & \sum_i (\inf \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}) \leq \\ & \leq \text{area}(E) \leq \\ & \leq \sum_i (\sup \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}), \end{aligned}$$

per qualsiasi scelta della suddivisione di $[a, b]$ in sottointervalli. Ora, facendo riferimento alle classi A e B già menzionate in precedenza, osservo che si riesce a dimostrare che sono separate anche senza alcun bisogno di fare riferimento all'area di E (per fortuna, visto che l'area di E la dobbiamo ancora definire!).

Quindi A e B sono due classi separate e quindi hanno elemento separatore, non necessariamente unico; se è unico, siamo a posto: definiremo l'area di E proprio come quell'unico numero reale che funge da elemento di separazione. Se l'elemento separatore non è unico, ciò vuol dire che tra le approssimazioni per difetto e quelle per eccesso c'è un "gap": in tal caso, che cosa fare? Oltretutto gli esempi grafico-intuitivi non prefigurano situazioni di questo genere. Sfido chiunque a disegnare il grafico di una f per il quale appaia evidente la presenza di questo "gap". E' allora inutile continuare con le chiacchiere: in caso di non unicità dell'elemento separatore, adotteremo una linea di condotta molto drastica, dicendo che E non ha area. Cioè, non cercheremo di definire in alcun modo l'area di E .

Finalmente le chiacchiere sono finite e possiamo deliziarci con un bel po' di definizioni, teoremi, etc.

Solo una considerazione ancora, un po' pignola, lo ammetto. Finora ho considerato funzioni f definite su $[a, b]$, l'intervallo che mi interessa. Può capitare benissimo, però, di avere una funzione $f: A \rightarrow \mathbb{R}$, con A s.i. qualsiasi di \mathbb{R} , e di essere interessati a quel che succede su un intervallo $[a, b] \subseteq A$. Per esempio, potrei prendere $f(x) = \sqrt{x}$ e voler sapere quale è l'area

tra il grafico di f , l'asse delle x e la retta $x=1$: in altre parole, sono interessato ad f su $[0,1]$; tuttavia, f è definita su $A=[0,+\infty[$, non solo su $[0,1]$. Quindi, il punto di partenza più naturale è costituito da due dati: la funzione f (con il suo insieme di definizione A) e l'intervallo $[a,b]$ sul quale concentro l'attenzione, con la ovvia condizione di compatibilità che $[a,b] \subseteq A$. Se così stanno le cose, possiamo però considerare $f|_{[a,b]}$ ed applicare ad essa le considerazioni che faremo. Morale: quando d'ora in poi assumerò che f sia definita su $[a,b]$, deve rimanere inteso che, se la funzione che ci interessa è in realtà definita su un insieme che contiene $[a,b]$, le considerazioni che faremo vanno intese come riferite a $f|_{[a,b]}$.

D'ora in poi, abolirò la restrizione che sia $f \geq 0$. Naturalmente, se f non è più maggiore o uguale a zero, andrà perduta l'interpretazione dell'integrale come area. Come ho già detto, vi sono però altri motivi per introdurre l'integrale: in molti casi, una restrizione del genere sarebbe eccessiva.

Sia quindi data $f:[a,b] \longrightarrow \mathbb{R}$, limitata. Considero una partizione $P = \{x_0, x_1, \dots, x_n\}$ di $[a,b]$, con $a = x_0 < x_1 < \dots < x_n = b$. Definiamo:

$$s(P,f) = \sum_1 (\inf \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}),$$

$$S(P,f) = \sum_1 (\sup \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}).$$

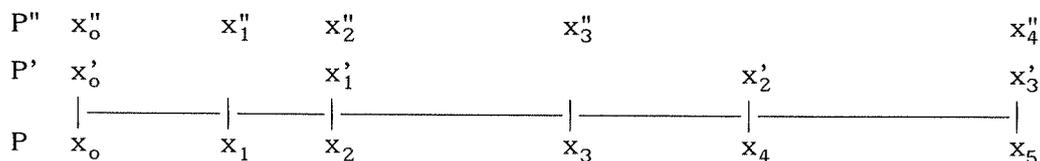
E' evidente che $s(P,f) \leq S(P,f)$, in quanto $\inf \{ f(x) : x \in [x_{i-1}, x_i] \} \leq \sup \{ f(x) : x \in [x_{i-1}, x_i] \}$ per ogni i , e quindi

$$\begin{aligned} & (\inf \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}) \leq \\ & \leq (\sup \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}) \end{aligned}$$

per ogni i , da cui, sommando su i , si ottiene

$$\begin{aligned} s(P,f) &= \sum_1 (\inf \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}) \leq \\ &\leq \sum_1 (\sup \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}) = S(P,f). \end{aligned}$$

In realtà, si può far meglio. E cioè si può dimostrare che, date due partizioni P' e P'' di $[a,b]$, si ha comunque $s(P',f) \leq S(P'',f)$. Vediamo con un esempio come mai ciò sia vero. Consiglio di fare riferimento al disegno qua sotto nel leggere quel che segue.



Supponiamo che sia $P' = \{ x_0', x_1', x_2', x_3' \}$ e $P'' = \{ x_0'', x_1'', x_2'', x_3'', x_4'' \}$. Ovviamente è $x_0' = x_0'' = a$, $x_3' = x_4'' = b$. Supponiamo che sia poi, per esempio, $x_1' = x_2''$ e che $x_3' < x_2''$. Mettendo assieme i punti di P' e di P'' si ottiene $P' \cup P'' = \{ x_0', x_1'', x_1', x_3', x_2'', x_3' \}$, dove ho elencato i punti in modo crescente. Chiamiamo allora $x_0 = x_0'$, $x_1 = x_1''$, $x_2 = x_1'$, $x_3 = x_3''$, $x_4 = x_2'$, $x_5 = x_3'$: otteniamo così la partizione $P = \{ x_0, x_1, x_2, x_3, x_4, x_5 \}$. Credo che sia chiaro quale sia la buona idea che cerchiamo di sfruttare: basterà dimostrare che $s(P', f) \leq s(P, f)$ e che $S(P, f) \leq S(P'', f)$. Dopo di che, tenendo conto che abbiamo già provato che $s(P, f) \leq S(P, f)$, otteniamo proprio $s(P', f) \leq S(P'', f)$. Mi limito a far vedere che $s(P', f) \leq s(P, f)$: l'altra relazione si ottiene specularmente.

Abbiamo:

$$\begin{aligned}
 s(P', f) &= \\
 &= \sum_{i=1}^3 (\inf \{ f(x) : x \in [x_{i-1}', x_i'] \}) \cdot (x_i' - x_{i-1}') = \\
 &= (\inf \{ f(x) : x \in [x_0', x_1'] \}) \cdot (x_1' - x_0') + (\inf \{ f(x) : x \in [x_1', x_2'] \}) \cdot (x_2' - x_1') + \\
 &\quad + (\inf \{ f(x) : x \in [x_2', x_3'] \}) \cdot (x_3' - x_2') = \\
 &= (\inf \{ f(x) : x \in [x_0', x_1'] \}) \cdot [(x_1 - x_0) + (x_2 - x_1)] + \\
 &\quad + (\inf \{ f(x) : x \in [x_1', x_2'] \}) \cdot [(x_3 - x_2) + (x_4 - x_3)] + \\
 &\quad + (\inf \{ f(x) : x \in [x_2', x_3'] \}) \cdot (x_5 - x_4) = \\
 &= (\inf \{ f(x) : x \in [x_0', x_1'] \}) \cdot (x_1 - x_0) + (\inf \{ f(x) : x \in [x_0', x_1'] \}) \cdot (x_2 - x_1) + \\
 &\quad + (\inf \{ f(x) : x \in [x_1', x_2'] \}) \cdot (x_3 - x_2) + \\
 &\quad + (\inf \{ f(x) : x \in [x_1', x_2'] \}) \cdot (x_4 - x_3) + \\
 &\quad + (\inf \{ f(x) : x \in [x_2', x_3'] \}) \cdot (x_5 - x_4) \leq \\
 &\leq (\inf \{ f(x) : x \in [x_0, x_1] \}) \cdot (x_1 - x_0) + (\inf \{ f(x) : x \in [x_1, x_2] \}) \cdot (x_2 - x_1) + \\
 &\quad + (\inf \{ f(x) : x \in [x_2, x_3] \}) \cdot (x_3 - x_2) + \\
 &\quad + (\inf \{ f(x) : x \in [x_3, x_4] \}) \cdot (x_4 - x_3) + \\
 &\quad + (\inf \{ f(x) : x \in [x_4, x_5] \}) \cdot (x_5 - x_4) = \\
 &= s(P, f) .
 \end{aligned}$$

Ovviamente il passaggio importante è quello relativo alla disuguaglianza, la

quale è conseguenza del fatto che $\inf S \leq \inf T$ se $\emptyset \neq S \subseteq T$.

Naturalmente quanto osservato in questo caso particolare si generalizza a due generiche partizioni P' e P'' ; per la dimostrazione rinvio a C-S.

Avendo $s(P',f) \leq S(P'',f)$ per ogni scelta delle partizioni P' e P'' , possiamo affermare che A e B sono due classi separate, dove:

$$A = \{ s(P',f) : P' \text{ partizione di } [a,b] \}$$

$$B = \{ S(P'',f) : P'' \text{ partizione di } [a,b] \}.$$

Infatti, se $a \in A$, allora $\exists P'$ t.c. $a = s(P',f)$; se $b \in B$, $\exists P''$ t.c. $b = S(P'',f)$. Grazie a quanto abbiamo appena visto, possiamo affermare che $a \leq b$, per cui le classi A e B sono effettivamente separate.

Esercizio 1 Siano $f, g: X \longrightarrow \mathbb{R}$, dove X è un qualsiasi insieme. Mostrare con un controesempio che la seguente implicazione è falsa:

$$(f(x) \leq g(x) \quad \forall x \in X) \Rightarrow \left[(f(X), g(X)) \text{ è una coppia di classi separate} \right].$$

Provare invece che la seguente implicazione è vera:

$$(f(x_1) \leq g(x_2) \quad \forall x_1, x_2 \in X) \Rightarrow \left[(f(X), g(X)) \text{ è una coppia di classi separate} \right]. \square$$

E' facile provare che $\sup A$ e $\inf B$ sono elementi separatori per (A, B) . E' altrettanto ovvio dimostrare che l'elemento separatore è unico se e solo se $\sup A = \inf B$.

Diremo quindi che f è integrabile se $\sup A = \inf B$. Indicheremo con $\int_a^b f(x) dx$ questo unico elemento separatore e lo chiameremo integrale di f su $[a, b]$. Nel caso in cui $f(x) \geq 0 \quad \forall x \in [a, b]$, assumeremo tale numero come valore dell'area di E ($E = \{ (x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ e } 0 \leq f(x) \}$).

Abbiamo quindi finalmente la definizione di integrale; vediamo un esempio di come si possa utilizzare questa definizione.

Esempio 1 Sia $f(x) = x$ e sia $[a, b] = [0, 1]$. Allora f è integrabile su $[0, 1]$ e si ha $\int_0^1 f(x) dx = 1/2$. \square

Dettaglio Consideriamo le partizioni P_n di $[0, 1]$ ottenute dividendolo in n parti di uguale lunghezza. Cioè $P_n = \{ i/n : i \in \{0, 1, \dots, n\} \}$. Osserviamo che sull' i -esimo intervallo, cioè su $[(i-1)/n, i/n]$ è

$$\inf \{ f(x) : x \in [(i-1)/n, i/n] \} =$$

$$\min \{ f(x) : x \in [(i-1)/n, i/n] \} = f((i-1)/n) = (i-1)/n$$

e si ha anche:

$$\sup \{ f(x) : x \in [(i-1)/n, i/n] \} = i/n .$$

Quindi

$$\begin{aligned} s(P_n, f) &= \\ &= \sum_{i=1}^n ((i-1)/n) \cdot (1/n) = \\ &= (1/n^2) \cdot \sum_{i=1}^n (i-1) = \\ &= (1/n^2) \cdot \sum_{i=0}^{n-1} i = \frac{1}{n^2} \cdot \frac{n(n-1)}{2} . \end{aligned}$$

$$\text{Analogamente } S(P_n, f) = (1/n^2) \cdot \sum_{i=1}^n i = \frac{1}{n^2} \cdot \frac{(n+1)n}{2} .$$

Poiché $\lim_{n \rightarrow \infty} s(P_n, f) = \frac{1}{2} = \lim_{n \rightarrow \infty} S(P_n, f)$, le due classi A e B hanno un unico elemento separatore ed esso è $1/2$. Siano infatti λ e μ due elementi separatori con $\lambda \leq \mu$. Si ha $s(P_n, f) \leq \lambda$ e $S(P_n, f) \geq \mu \quad \forall n \in \mathbb{N}$ (ricordo che un elemento separatore è un maggiorante per A e un minorante per B). Pertanto, per il teorema di permanenza del segno:

$$1/2 = \lim_{n \rightarrow \infty} s(P_n, f) \leq \lambda \leq \mu \leq \lim_{n \rightarrow \infty} S(P_n, f) = 1/2 ,$$

da cui otteniamo in un sol colpo l'unicità dell'elemento separatore e il fatto che esso valga $1/2$. \square

Come si vede, la definizione adottata ci dà dei risultati confortanti, conformi all'intuizione. Infatti per fortuna il risultato ottenuto è $1/2$, cioè quel che già "si sapeva". Abbiamo quindi una definizione che ci ridà, per i poligoni, lo stesso valore dell'area che ci era già noto per altre vie⁵. Quindi la definizione di integrale, dal punto di vista del fornire una definizione per l'area delle figure piane, soddisfa una delle condizioni minimali che vanno rispettate ogni qual volta si cerca di estendere una definizione: la nuova definizione non è in contrasto con la vecchia per le figure di cui già ci era nota l'area.

⁵ In realtà l'esempio si occupa di un particolare triangolo. Ma non ci vuole molto a considerare un generico poligono, effettuare una triangolazione e ripetere conti analoghi, per vedere che mediante gli integrali riotteniamo il valore dell'area che già conoscevamo.

D'altro canto, però, l'esempio fa anche emergere un problema: non sarà facile calcolare l'area di figure un po' complicate, visto il tipo di calcoli necessari già con una funzione semplice come $f(x) = x$! Per nostra fortuna, vedremo che c'è una strada completamente diversa per calcolare gli integrali: in molti casi si tratterà di fare conti molto semplici.

Mi occuperò dopo di questo problema. Per ora vorrei ancora discutere della definizione.

Dopo aver visto un esempio di funzione integrabile, è doveroso fornire almeno un esempio di funzione non integrabile.

Esempio 2 Sia $f: [0,1] \longrightarrow \mathbb{R}$ così definita: $f(x) = \begin{cases} 1 & \text{se } x \in \mathbb{Q} \\ 0 & \text{se } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$.

Allora f non è integrabile. \square

Dettaglio Consideriamo una qualsiasi partizione P di $[0,1]$. Considero il generico intervallino $[x_{i-1}, x_i]$. Poiché $x_i < x_{i-1}$, in tale intervallino "cadrà" un elemento $x' \in \mathbb{Q}$ ed anche un elemento $x'' \in \mathbb{R} \setminus \mathbb{Q}$, grazie ai risultati sulla densità di razionali ed irrazionali in \mathbb{R} ⁶. Allora

$$\inf \{ f(x) : x \in [x_{i-1}, x_i] \} \leq f(x'') = 0$$

ed, analogamente:

$$\sup \{ f(x) : x \in [x_{i-1}, x_i] \} \geq f(x') = 1.$$

Quindi $s(P, f) = \sum_{i=1}^n (\inf \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}) \leq 0$, mentre $S(P, f) = \sum_{i=1}^n (\sup \{ f(x) : x \in [x_{i-1}, x_i] \}) \cdot (x_i - x_{i-1}) \geq \sum_{i=1}^n (x_i - x_{i-1}) = 1$.

Quindi abbiamo che, comunque si scelgano partizioni P e Q di $[0,1]$, si ha $s(P, f) \leq 0$ e $1 \leq S(Q, f)$, pertanto sia 0 che 1 sono elementi separatori per le classi A e B . Quindi l'elemento separatore non è unico e pertanto f non è integrabile. \square

Che commento fare a questo esempio? Sembra riguardare una funzione un po' curiosa, inventata apposta per fare questo controesempio. Non è che io sono volutamente andato alla ricerca di una funzione strana per impressionare il lettore: come vedremo dopo, ogni funzione continua, anche "a pezzi", purché limitata, è integrabile. Per trovare un controesempio occorre quindi lavorare con funzioni molto irregolari. Benissimo! Questo significa che la definizione che abbiamo dato

⁶ Vedi Corollario 9.5 a pagina 36 e Problemi 4 e 5 a pagina 37 di C-S.

ci permette di definire l'area per una classe molto vasta di figure piane. E, più in generale, visto che il problema dell'area non è l'unico cui cerca di dare risposta il calcolo integrale, abbiamo uno strumento con un ampio spettro di applicabilità.

Aggiungo ancora che, con la loro fervida fantasia, i matematici sono in realtà andati ben oltre. Sono riusciti a fare due cose non proprio ovvie. Sono riusciti ad estendere ancora la definizione di integrale: oltre a quella che abbiamo visto e che è legata al nome di Riemann, ce n'è un'altra, dovuta essenzialmente a Lebesgue, grazie alla quale risultano essere integrabili molte più funzioni (tra cui anche quella dell'esempio 2). Inoltre, i matematici sono stati in grado di provare addirittura che non è possibile dare una sensata definizione di area per le figure piane in modo da poter assegnare un'area a tutti i sottoinsiemi di \mathbb{R}^2 . Tutto questo va ben al di là di quanto valga la pena di faticare in questo contesto, per cui mi limito solo a questi brevi cenni. Si noti, però, che non si tratta di fare delle speculazioni intellettuali gratuite: l'integrale nel senso di Lebesgue è uno strumento utilissimo per le concrete applicazioni della matematica.

Giunti alla fine del paragrafo, è opportuno ricapitolare un po' tutto il discorso. Data una funzione $f:[a,b] \longrightarrow \mathbb{R}$, limitata, abbiamo detto che f è integrabile (sottinteso: su $[a,b]$) se le classi A e B hanno un unico elemento separatore. Tale elemento lo interpreteremo come area di $E = \{ (x,y) \in \mathbb{R}^2 : a \leq x \leq b \text{ e } 0 \leq f(x) \}$, nel caso in cui f sia maggiore o uguale di zero su tutto $[a,b]$. In generale chiameremo tale elemento separatore integrale di f su $[a,b]$ e lo indicheremo come $\int_a^b f(x) dx$.

2. Una divagazione sulle notazioni

Approfitto dell'occasione offerta dalla notazione scelta per indicare l'integrale di f su $[a,b]$ per alcune puntualizzazioni sull'uso delle notazioni, sia in generale che in questo contesto particolare.

La notazione scelta fa comparire una lettera, la x , la quale non c'entra nulla. Infatti, l'integrale dipende solo da f e da $[a,b]$. Una notazione più sensata sarebbe, per esempio, $I(f,[a,b])$. In questa notazione sono coinvolte la f e l'intervallo $[a,b]$, nonché la lettera I la cui funzione è evidentemente solo mnemonica. Nessuno penso ritiene che la I sia una variabile, in questo contesto. La notazione tradizionale, e cioè $\int_a^b f(x)dx$, assegna invece alla lettera x un ruolo poco chiaro. Non solo, ma non si capisce affatto cosa possa voler dire quel "dx": è un simbolo con un significato a sè stante, oppure la "d" davanti alla x significa qualcosa di particolare, o che altro?

Trascurando per ora il problema di quella "d" messa davanti alla x , occupiamoci del ruolo di questa lettera x che non è una variabile. Il caso del simbolo $\int_a^b f(x)dx$ non è l'unico in cui compare una variabile cosiddetta "muta". Altri casi già si sono presentati. Per esempio $\lim_{x \rightarrow x_0} f(x)$ o $\lim_{n \rightarrow \infty} a_n$ o ancora $\sum_{k=1}^n c_k$. In tutte queste notazioni compaiono delle lettere (rispettivamente x, n, k) che non sono delle variabili, cioè non possiamo sostituire a loro delle costanti. Non avrebbe nessun senso scrivere $\sum_{7=1}^n c_7$. Si usa allora dire che sono delle variabili "mute", perché non possono "parlare", ovvero sia ad esse non possiamo sostituire dei valori particolari.

Perché si usano le variabili mute? Perché sono molto comode, rispetto a notazioni che non le usino. Ad esempio, per i limiti di funzioni si potrebbe usare la notazione $\lim(f, x_0)$ la quale fa intervenire, oltre al simbolo "lim" ad uso mnemonico, i simboli f ed x_0 i quali sono delle vere variabili, nel senso che il valore del limite dipende ovviamente dalla specifica funzione considerata e dal punto nel quale lo si calcola: se ad f sostituiamo la funzione "radice quadrata" e ad x_0 sostituiamo 36, otteniamo un ben preciso valore per il limite. Avremmo cioè $\lim(\text{radice quadrata}, 36) = 6$. Credo che già si intraveda la scarsa maneggevolezza di questa notazione che non usa le variabili mute. Peggio ancora sarebbe

se volessimo scegliere come f la funzione che in un generico punto x assume il valore $\frac{x^5-x^3+6x}{x^2+3x-4}$. Essa è la "funzione quoziente tra, al numeratore, la funzione ottenuta sommando la funzione 5^a potenza con sei volte la funzione identità e sottraendo la funzione cubo e, al denominatore, la funzione quadrato cui aggiungiamo il triplo della funzione identità e sottraiamo la funzione costante 4". Una cosa mostruosa. Ma non è facile far meglio se si vuole indicare una funzione senza fare riferimento ai valori che essa assume in un "generico x ".

Insomma, l'uso della lettera x come variabile "muta" in una espressione come $\lim_{x \rightarrow x_0} f(x)$ sta nel fatto che molto frequentemente si ha a disposizione un modo succinto per indicare $f(x)$ mentre non si ha a disposizione un modo altrettanto conciso per indicare f .

Fatte queste considerazioni di carattere generale, ritorniamo a concentrare la nostra attenzione sulla notazione $\int_a^b f(x)dx$.

La notazione $\int_a^b f(x)dx$ è, semplicemente, un retaggio storico (un relitto del passato, se il lettore preferisce). Anche se uno non è interessato al perché "anticamente" si usasse questa notazione, costui comunque dovrebbe chiedersi come mai questo "fossile" sia sopravvissuto alla selezione naturale, la quale opera anche nell'ambito delle notazioni. Vedremo dopo, anche in successivi paragrafi, il perché di questa sopravvivenza.

3. Proprietà delle funzioni integrabili e dell'integrale

Come al solito, andiamo a vedere quanta compatibilità con la struttura dei numeri reali abbia il nuovo concetto.

Enuncio i risultati che mi interessano senza dimostrazione, Chi è interessato, può utilizzare il C-S (§ 77, non il § 70!).

Prima, una notazione per abbreviare: indicheremo con $\mathcal{R}([a,b])$ l'insieme delle $f: [a,b] \rightarrow \mathbb{R}$ limitate e integrabili nel senso di Riemann su $[a,b]$. Inoltre f^+ indicherà la funzione così definita: $f^+(x) = \max\{f(x), 0\}$ E $f^-(x) = -\min\{f(x), 0\}$. Si noti che $f^+, f^- \geq 0$ e che $f = f^+ - f^-$.

Teorema 1 Siano $f, g \in \mathcal{R}([a,b])$ e $k \in \mathbb{R}$. Siano inoltre m e M rispettivamente un minorante ed un maggiorante per f su $[a,b]$. Allora:

$$f+g \in \mathcal{R}([a,b]) \quad \text{e} \quad \int_a^b f(x)+g(x)dx = \int_a^b f(x)dx + \int_a^b g(x)dx$$

$$k \cdot f \in \mathcal{R}([a,b]) \quad \text{e} \quad \int_a^b k \cdot f(x)dx = k \cdot \int_a^b f(x)dx$$

$$\text{Se } f \geq 0 \text{ su } [a,b], \text{ allora } \int_a^b f(x)dx \geq 0$$

$$f^+, f^- \in \mathcal{R}([a,b])$$

$$m \cdot (b-a) \leq \int_a^b f(x)dx \leq M \cdot (b-a) \quad (\text{"teorema della media"})$$

Infine, se $c \in]a,b[$, allora f è integrabile su $[a,c]$ e su $[c,b]$ e si ha:

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \quad (\text{"additività sul campo"}). \square$$

Corollario 1 Siano $f, g \in \mathcal{R}([a,b])$. Allora:

$$f-g \in \mathcal{R}([a,b]) \quad \text{e} \quad \int_a^b f(x)-g(x)dx = \int_a^b f(x)dx - \int_a^b g(x)dx .$$

$$\text{Se } f \leq g \text{ su } [a,b], \text{ allora } \int_a^b f(x)dx \leq \int_a^b g(x)dx .$$

Infine,

$$|f| \in \mathcal{R}([a,b]) \quad \text{e} \quad \left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)| dx . \square$$

4. Interludio: funzioni uniformemente continue

Nel prossimo paragrafo proveremo che le funzioni continue sono integrabili sugli intervalli chiusi e limitati. Per arrivarci abbiamo però bisogno di introdurre una nuova condizione di "continuità", detta uniforme continuità.

Abbiamo quindi la seguente

Definizione 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$. Diremo che f è uniformemente continua su $B \subseteq A$ se:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x_1, x_2 \in B \left(|x_1 - x_2| < \delta \Rightarrow |f(x_1) - f(x_2)| < \varepsilon \right) . \square$$

Visto il nome usato e la struttura della definizione, si impone immediatamente un confronto tra l'uniforme continuità di f su B e la continuità di f su B . Ricordiamo che f è continua su B se f è continua in ogni $x_1 \in B$, cioè se:

$$\forall x_1' \in B, \forall \varepsilon' > 0 \exists \delta > 0 \text{ t.c. } \forall x_2' \in B \\ \left(|x_1' - x_2'| < \delta' \Rightarrow |f(x_1') - f(x_2')| < \varepsilon' \right) .$$

Invece uniforme continuità vuol dire che:

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ t.c. } \forall x_1 \in B, \forall x_2 \in B \\ \left(|x_1 - x_2| < \delta \Rightarrow |f(x_1) - f(x_2)| < \varepsilon \right) .$$

La differenza c'è, e balza agli occhi anche solo con un esame puramente formale della struttura delle due proposizioni. Infatti risultano invertiti i due quantificatori sottolineati⁷. Possiamo quindi tranquillamente affermare che l'uniforme continuità su B implica la continuità su B . Per il viceversa non possiamo dirlo con certezza, anche se probabilmente non sarà vero (in linea di massima, una proposizione del tipo "esiste... t.c. per ogni..." esprime una condizione più forte che non la proposizione "per ogni... esiste..."). Per essere certi che "il viceversa non vale", il modo migliore è, al solito, quello di esibire un controesempio: grazie ad esso, possiamo dire che in generale la continuità non garantisce l'uniforme continuità.

Esempio 1 Sia $f(x) = x^2$ e sia $B = \mathbb{R}$. Ovviamente f è continua su \mathbb{R} : proviamo che non è uniformemente continua. Cioè proviamo che si ha:

$$\exists \varepsilon > 0 \forall \delta > 0 \exists x_1, x_2 \in \mathbb{R} \text{ t.c. } \left(|x_1 - x_2| < \delta \text{ e } |f(x_1) - f(x_2)| \geq \varepsilon \right) .$$

⁷ C'è anche un'altra inversione, a dire il vero, ma questa riguarda due quantificatori entrambi universali e quindi è irrilevante.

Prendiamo $\varepsilon = 1$. Dato $\delta > 0$, scegliamo $x_1 = 1/\delta$ e $x_2 = (1/\delta) + (\delta/2)$.
 E' $|x_1 - x_2| < \delta$, mentre $|x_1^2 - x_2^2| = |x_1 - x_2| \cdot |x_1 + x_2| = (\delta/2) \cdot ((2/\delta) + (\delta/2)) > (2/\delta) \cdot (\delta/2) = 1$. \square

Osservazione 1 Bene, l'esempio l'abbiamo visto. Ed è chiaro che il trucco è consistito nello sfruttare abilmente il famoso prodotto notevole $(x_1^2 - x_2^2) = (x_1 - x_2) \cdot (x_1 + x_2)$. Ma quali sono le idee che stanno dietro a tutto questo? Basta leggere la definizione. L'uniforme continuità ci dice che, dato $\varepsilon > 0$, c'è un $\delta > 0$ t.c., se prendiamo due punti x_1 e x_2 che distino tra loro meno di δ , allora la distanza tra $f(x_1)$ e $f(x_2)$ è minore di ε . Si noti che sui punti x_1 e x_2 , l'unica condizione posta è che distino tra loro meno di δ (a parte quella, ovvia, che devono appartenere ad A): ciò significa che possiamo prendere, per così dire, la coppia x_1, x_2 dove vogliamo. Nel caso di $f(x) = x^2$, basta dare un'occhiata al grafico per rendersi conto che, a pari distanza fra x_1 e x_2 , la distanza tra $f(x_1)$ ed $f(x_2)$ tende a diventare sempre maggiore, più prendiamo i punti x_1 e x_2 lontano da zero! Abbiamo sfruttato proprio questo. \square

Osservazione 2 Mi auguro che qualcuno, leggendo la precedente osservazione, abbia pensato che forse c'è un legame tra l'uniforme continuità e la derivata di una funzione. Dopotutto, $\frac{f(x_2) - f(x_1)}{x_2 - x_1}$ può essere visto come un rapporto incrementale. Se sapessimo che questo rapporto incrementale è limitato, avremmo l'uniforme continuità. Infatti, se $\exists L > 0$ t.c. $\left| \frac{f(x_2) - f(x_1)}{x_2 - x_1} \right| \leq L$, allora $|f(x_2) - f(x_1)| \leq L \cdot |x_2 - x_1|$ e quindi f è uniformemente continua (effettivamente, dato $\varepsilon > 0$, basta prendere $\delta = \varepsilon/L$...). Funzioni di questo tipo si chiamano funzioni lipschitziane. Grazie al teorema di Lagrange si ottiene facilmente una condizione sufficiente di lipschitzianità. \square

Esercizio 1 Sia $f: A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$ e sia $I \subseteq A$, con I intervallo. Se f è derivabile su I ed $\exists L > 0$ t.c. $|f'(x)| \leq L \quad \forall x \in I$, provare che f è lipschitziana su I . \square

Il risultato più importante di questo paragrafo è il seguente.

Teorema 1 (di Heine-Cantor) Sia $f: [a, b] \longrightarrow \mathbb{R}$, continua su $[a, b]$. Allora f è uniformemente continua su $[a, b]$. \square

Dimostrazione Vedi C-S a pagina 172 (teorema 42.3). \square

Mi limito a far notare che si tratta di un "grosso" teorema sulle funzioni continue (come il teorema di Weierstrass ed il teorema degli zeri). Guarda caso, anch'esso usa in modo essenziale la proprietà di completezza di \mathbb{R} .

5. Integrabilità delle funzioni continue

Il risultato principale che proveremo è il seguente.

Teorema 1 Sia $f:[a,b] \longrightarrow \mathbb{R}$, continua su $[a,b]$. Allora f è integrabile. \square

Dimostrazione Basterà dimostrare che le classi A e B sono contigue, per ottenere che f è integrabile. Cioè che $\forall \varepsilon > 0 \exists a \in A$ ed $\exists b \in B$ t.c. $b-a < \varepsilon$. Ovverossia, sarà sufficiente trovare una partizione P t.c. $S(P,f) - s(P,f) < \varepsilon$.

Come fare? E come si usa l'uniforme continuità per avere questo risultato? Consideriamo una partizione. Se su ogni intervallino della partizione si ha che

$$(\star) \quad \sup \{ f(x) : x \in [x_{i-1}, x_i] \} - \inf \{ f(x) : x \in [x_{i-1}, x_i] \} \leq \frac{\varepsilon}{2(b-a)},$$

allora si ha

$$\begin{aligned} S(P,f) - s(P,f) &= \\ &= \sum_{i=1}^n \left(\sup \{ f(x) : x \in [x_{i-1}, x_i] \} - \inf \{ f(x) : x \in [x_{i-1}, x_i] \} \right) \cdot (x_i - x_{i-1}) \leq \\ &= \sum_{i=1}^n \frac{\varepsilon}{2(b-a)} \cdot (x_i - x_{i-1}) = \frac{\varepsilon}{2(b-a)} \cdot \sum_{i=1}^n (x_i - x_{i-1}) = \frac{\varepsilon}{2(b-a)} \cdot (b-a) = \varepsilon/2 < \varepsilon. \end{aligned}$$

Come facciamo ad ottenere la disuguaglianza (\star) ? Ovviamente sfruttando l'uniforme continuità. Grazie ad essa, sappiamo che $\exists \delta > 0$ t.c. se $|x_2 - x_1| < \delta$, allora $|f(x_2) - f(x_1)| < \frac{\varepsilon}{2(b-a)}$.

Ma, allora, sarà sufficiente prendere una partizione $P = \{x_0, x_1, \dots, x_n\}$ con $|x_i - x_{i-1}| < \delta$ per ogni $i \in \{1, \dots, n\}$.

Notiamo infatti che si ha proprio la disuguaglianza richiesta, e cioè la (\star) . Infatti, per il teorema di Weierstrass, esiste $\max \{ f(x) : x \in [x_{i-1}, x_i] \}$ e quindi $\exists \bar{x} \in [x_{i-1}, x_i]$ t.c. $f(\bar{x}) = \max \{ f(x) : x \in [x_{i-1}, x_i] \}$. Analogamente $\exists \bar{x} \in [x_{i-1}, x_i]$ t.c. $f(\bar{x}) = \min \{ f(x) : x \in [x_{i-1}, x_i] \}$. Ma allora

$$\begin{aligned} \sup \{ f(x) : x \in [x_{i-1}, x_i] \} - \inf \{ f(x) : x \in [x_{i-1}, x_i] \} &= \\ &= f(\bar{x}) - f(\bar{x}) < \frac{\varepsilon}{2(b-a)} \end{aligned}$$

in quanto $|\bar{x} - \bar{x}| \leq |x_i - x_{i-1}| < \delta$. \square

Una conseguenza dell'integrabilità delle funzioni continue è una interessante estensione del teorema della media.

Teorema 2 Sia $f: [a, b] \longrightarrow \mathbb{R}$, continua. Allora $\exists \xi \in [a, b]$ t.c.
 $\int_a^b f(x) dx = f(\xi) \cdot (b-a)$. \square

Dimostrazione Il teorema di Weierstrass garantisce che esistono $m = \min f([a, b])$ ed $M = \max f([a, b])$. L'integrabilità di f ci permette di applicare la versione già nota del teorema della media (teorema 3.1) e quindi ottenere che $m \cdot (b-a) \leq \int_a^b f(x) dx \leq M \cdot (b-a)$. Dividendo per $b-a$, abbiamo:

$$(\star) \quad m \leq \frac{\int_a^b f(x) dx}{b-a} \leq M .$$

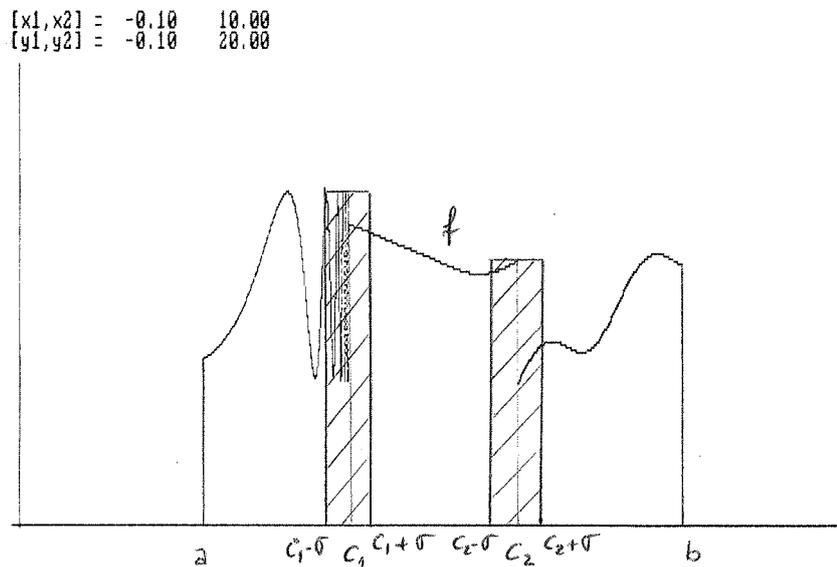
Poiché m, M sono valori assunti da f , il teorema dei valori intermedi (teorema III.12.2) ci dice che f assume anche il valore $\frac{\int_a^b f(x) dx}{b-a}$. Cioè,

$\exists \xi \in [a, b]$ t.c. $f(\xi) = \frac{\int_a^b f(x) dx}{b-a}$. Da qui la tesi. \square

Si noti che il teorema 1 sull'integrabilità delle funzioni continue può ancora essere migliorato. In effetti, f può anche avere "qualche discontinuità". Cioè, si può dimostrare il seguente:

Teorema 3 Sia $f: [a, b] \longrightarrow \mathbb{R}$, limitata su $[a, b]$ e continua su $[a, b] \setminus F$, dove F è un s.i. finito di $[a, b]$. Allora f è integrabile. \square

Dimostrazione La dimostrazione di questo teorema è un pochino complicata: vale però la pena di cercare di capirla, perché se ci si riesce si può ritenere di avere davvero capito che cosa sia un integrale. Sia $F = \{c_1, \dots, c_n\}$. Possiamo naturalmente supporre che i punti di F siano "messi in ordine", cioè che $c_1 < c_2 < \dots < c_n$. Creiamo delle piccole "fasce di sicurezza" attorno a questi punti, di modo che la mancanza di continuità di f in tali punti non infastidisca. Vedasi il disegno seguente.



Ovviamente su un intervallo del tipo $[c_i + \sigma, c_{i+1} - \sigma]$ ⁸, f è certamente integrabile perché è continua. Restano quindi da "sistemare" gli intervalli $[c_i - \sigma, c_i + \sigma]$. Come trattiamo f su $[c_i - \sigma, c_i + \sigma]$? Sfruttando la limitatezza di f possiamo vedere agevolmente che lo scarto tra somme superiori ed inferiori è piccolo, purché σ sia scelto "opportunamente piccolo".

Quelle sopra indicate sono le idee di fondo per dimostrare il teorema. Precisiamole meglio. Dato $\varepsilon > 0$, dobbiamo trovare una partizione P di $[a, b]$ t.c. $S(P, f) - s(P, f) < \varepsilon$. Dividiamo ε per 2: grazie alla citata continuità di f sugli intervalli del tipo $[c_i + \sigma, c_{i+1} - \sigma]$ (ai quali si vanno ad aggiungere eventualmente $[a, c_1 - \sigma]$ (se $a < c_1$) e $[c_n + \sigma, b]$ (se $c_n < b$)), sarà possibile piazzare un numero sufficiente di punti in tali intervalli in modo che complessivamente lo scarto tra somme inferiori e superiori sia minore di $\varepsilon/2$. Ci avanza ancora un $\varepsilon/2$: poiché f è limitata, se prendiamo σ sufficientemente piccolo abbiamo che $(n \cdot (\sup f([a, b]) - \inf f([a, b]))) \cdot 2\sigma < \varepsilon/2$ (abbiamo n intervallini di ampiezza 2σ e su ognuno di loro lo scarto massimo che possono avere i valori di f è proprio $\sup f([a, b]) - \inf f([a, b])$) e quindi con ciò "si-

⁸ Preciserò alla fine come va scelto σ : per ora basti pensare che σ è "abbastanza piccolo".

stemiamo" anche gli intervallini del tipo $[c_i - \sigma, c_i + \sigma]$.

Tutto bene? Ovverossia, non si tratta altro che della tecnica ormai ben nota di "distribuire" la "tolleranza" ε tra tutti i fattori di possibile disturbo? Essenzialmente è così, però non possiamo far finta di non vedere un problema che è presente. Non è indifferente l'ordine col quale "sistemiamo" gli intervallini $[c_i + \sigma, c_{i+1} - \sigma]$ e $[c_i - \sigma, c_i + \sigma]$. Per evitare grane, dobbiamo prima "sistemare" gli intervallini $[c_i - \sigma, c_i + \sigma]$ (ovvero: determinare quanto "piccolo" debba essere σ) e poi, dopo, ci possiamo occupare degli intervallini del tipo $[c_i + \sigma, c_{i+1} - \sigma]$.

Non c'è altro. Ora, non ci resta altro che formalizzare (cioè precisare per bene) i discorsi fatti sin qui.

Sia $K \in \mathbb{R}$ t.c. $|f(x)| \leq K \quad \forall x \in [a, b]$. Allora abbiamo $\sup \{ f(x) : x \in [a, b] \} - \inf \{ f(x) : x \in [a, b] \} \leq 2K$. Sia quindi σ t.c. $\sigma < \varepsilon/4K$. Per esempio, possiamo prendere $\sigma = \varepsilon/8K$. Tolti gli intervalli $]c_i - \sigma, c_i + \sigma[$, $[a, b]$ resta suddiviso in $n+1$ intervalli su ognuno dei quali f è continua⁹. Grazie alla continuità di f su questi intervallini, potremo trovare una partizione su ognuno di loro in modo che lo scarto tra somme superiori ed inferiori sia minore di $\frac{\varepsilon}{2(n+1)}$. Raccattando assieme tutti i punti che individuano ognuna di queste partizioni degli intervallini, otteniamo una partizione P di $[a, b]$. A questo punto, è scontato, per come l'abbiamo costruita, che $S(P, f) - s(P, f) < \varepsilon$: lascio volentieri al lettore il compito di fare quest'ultimo passo. \square

Esercizio 1 Sia $\bar{f}: [a, b] \rightarrow \mathbb{R}$, continua. Sia $f: [a, b] \rightarrow \mathbb{R}$ una funzione ottenuta modificando \bar{f} in un numero finito di punti. Provare che f soddisfa le ipotesi del teorema 3. Fare un esempio di f che soddisfa le ipotesi di tale teorema, ma che non si può ottenere a partire da una \bar{f} continua modificandola in un numero finito di punti. \square

Osservazione 1 Il teorema 2 (della media) non può essere generalizzato alle funzioni con punti di discontinuità (ancorché in numero finito). Esempio:

⁹ Per la precisione (e non si fa matematica se non si è precisi!), quanto è stato appena detto è giusto solo nel caso in cui i punti c_i siano tutti interni ad $[a, b]$. Se così non è, vanno fatte alcune lievi modifiche. Vediamo quali, nel caso in cui sia, per esempio, $c_1 = a$. Bisognerà considerare l'intervallo $[a, a + \sigma]$, visto che $[a - \sigma, a + \sigma]$ "deborda" dall'intervallo $[a, b]$. E allora $[a, b]$ resta suddiviso in n intervallini, anziché $n+1$, se $c_n < b$; addirittura in $n-1$ se $c_n = b$.

$$f(x) = \begin{cases} x & \text{se } x \in [0, 1] \setminus \{1/2\} \\ 0 & \text{se } x = 1/2 \end{cases} \quad .\square$$

6. Integrali orientati

Fino ad ora abbiamo visto la definizione di integrale di una funzione su un insieme (si trattava di insiemi particolari, cioè intervalli).

Possiamo facilmente estendere questa definizione in modo da considerare, per così dire, l'integrazione su insiemi "orientati". In effetti, la notazione stessa di integrale è già concepita per rendere facile questa estensione. Una notazione più precisa di $\int_a^b f(x)dx$ sarebbe stata $\int_{[a,b]} f(x)dx$: questa notazione avrebbe messo in evidenza che stiamo integrando sull'insieme $[a,b]$ (più in generale, si potrebbe scrivere $\int_A f(x)dx$ se l'insieme sul quale si integra f è A : questo, d'altronde, è proprio quel che si fa nel caso delle funzioni di più variabili, cioè per i cosiddetti integrali multipli). Invece, la notazione scelta ci permette di introdurre subito la seguente definizione.

Definizione 1 (di integrale definito orientato) Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$ e siano $x_1, x_2 \in A$. Definiamo

$$\int_{x_1}^{x_2} f(x)dx = \begin{cases} \int_{x_1}^{x_2} f(x)dx & \text{se } x_1 < x_2, [x_1, x_2] \subseteq A \text{ e } f \text{ è integrabile su } [x_1, x_2] \\ 0 & \text{se } x_1 = x_2 \\ -\int_{x_2}^{x_1} f(x)dx & \text{se } x_1 > x_2, [x_2, x_1] \subseteq A \text{ e } f \text{ è integrabile su } [x_2, x_1] \end{cases} \quad .\square$$

Osservazione 1 Abbandoneremo subito la notazione $\int_{x_1}^{x_2} f(x)dx$ per la notazione solita $\int_a^b f(x)dx$. Le ragioni per fare così sono ovvie: se abbiamo $f:[a,b] \longrightarrow \mathbb{R}$, (e quindi con $a < b$!), e prendiamo $x_1 = a$ e $x_2 = b$, otteniamo, dalla definizione testé data, che $\int_{x_1}^{x_2} f(x)dx = \int_a^b f(x)dx$. Ovverossia, la nuova definizione ci ridà il nostro buon vecchio integrale quando integriamo su $[a,b]$, con $a < b$. Quindi la definizione 1 estende la definizione di integrale, senza creare "conflitti di competenza" nei casi di pertinenza di entrambe le de-

finizioni. Per questa ragione non c'è obbligo di introdurre una nuova notazione, non essendoci rischi di confusione.□

Osservazione 2 Leggeremo il simbolo $\int_{x_1}^{x_2} f(x)dx$ "integrale orientato di f fra x_1 e x_2 ", ovvero "integrale di f da x_1 a x_2 ".□

7. Teorema fondamentale del calcolo integrale

Ci occuperemo in questo paragrafo di un risultato che ha avuto una assoluta rilevanza nel determinare il successo del calcolo integro-differenziale.

Prima di poter enunciare questo risultato abbiamo bisogno di introdurre una definizione: quella di funzione integrale.

L'idea di funzione integrale è molto semplice: avendo $f:A \longrightarrow \mathbb{R}$, fissato $x_0 \in A$, posso considerare l'integrale definito orientato da x_0 a x_1 , per ogni $x_1 \in A$ per il quale ciò abbia senso (vedasi la definizione 6.1¹⁰). Sia B l'insieme degli x_1 per i quali le condizioni poste dalla definizione 6.1 sono soddisfatte: abbiamo che ad ogni $x_1 \in B$ possiamo associare il numero reale

$\int_{x_0}^{x_1} f(x) dx$. Ma allora vuol dire che abbiamo definito su B una funzione reale

di variabile reale! Infatti, come appena detto, ad ogni $x_1 \in B$ abbiamo associato uno ed un solo numero reale. Quindi, sempre assumendo che x_0 sia fissato, ab-

biamo la funzione $x_1 \mapsto \int_{x_0}^{x_1} f(x) dx$, definita su B , che per ovvie ragioni, si

chiama funzione integrale di f (se necessario, si specificherà che è la funzione integrale di f essendo fissato x_0).

Prima di passare ai teoremi centrali di questo paragrafo, vorrei insistere sul problema della corretta individuazione di B , poiché l'esperienza insegna che troppo spesso lo studente è distratto su questo punto e non presta la dovuta attenzione ai problemi che ci sono e che ho già cercato di sottolineare nella precedente nota a piè pagina.

Osservazione 1 Sia $A =]-\infty, 0[\cup]1, +\infty[$ e sia $f:A \longrightarrow \mathbb{R}$. Supponiamo che f sia continua su A . Sia $x_0 = -1 \in A$. Allora la funzione integrale è definita per $x_1 \in]-\infty, 0[$ (infatti la continuità di f garantisce l'integrabilità di f da x_0 a x_1), ma non è definita per $x_1 \in]1, +\infty[$. Infatti non posso considerare $\int_{x_0}^{x_1} f(x) dx$ se $x_1 \in]1, +\infty[$, poiché f non è definita su tutto l'inter-

¹⁰ E' assolutamente fondamentale capire quali restrizioni pone la definizione 6.1! Supponiamo, per fissare le idee, che sia $x_1 < x_0$. Sarà necessario che sia $[x_1, x_0] \subseteq A$ (non basta che $x_0, x_1 \in A$!!) e che f sia integrabile su $[x_1, x_0]$.

vallo $[x_0, x_1]$. Tanto per fare un esempio, se $x_1 = 3$, non posso fare $\int_{-1}^3 f(x)dx$ perché f non è definita su tutto $[-1, 3]$. \square

Osservazione 2 Abbiamo ottenuto la funzione integrale come funzione che ad x_1 associa $\int_{x_0}^{x_1} f(x)dx$. Quindi essa si presenta, con la scelta che abbiamo fatto per indicare la variabile "indipendente", come una funzione della variabile x_1 . Poiché siamo abituati ad usare il simbolo x_1 per indicare costanti o dati, sarebbe comodo poter usare "la solita x " come variabile. E' chiaro che si creano, però, dei guai. Non possiamo sostituire direttamente alla variabile x_1 la variabile x , in quanto la lettera x è già da noi usata come pseudo-variabile (variabile muta) nella notazione adottata per l'integrale. Dobbiamo procedere così. Prima facciamo quel che si usa chiamare "cambio alfabetico": usiamo un'altra lettera dell'alfabeto (t ad esempio) al posto della lettera x come variabile muta. Cioè scriviamo $\int_{x_0}^{x_1} f(t)dt$ anziché $\int_{x_0}^{x_1} f(x)dx$. Ora possiamo effettuare la sostituzione, e cioè usare la variabile x (o ogni altra che ci faccia comodo, tranne t) al posto di x_1 e scrivere $\int_{x_0}^x f(t)dt$ anziché $\int_{x_0}^{x_1} f(t)dt$, ottenendo così una funzione della variabile x anziché della variabile x_1 , conformemente alla tradizione. E $\int_{x_0}^x f(t)dt$ sarà la notazione che adotterò d'ora in poi. \square

Siamo arrivati al "dunque". Per evitare di appesantire il discorso con difficoltà accessorie, non mi porrò nelle condizioni "più generali possibili". Assumerò di avere f definita e continua su un intervallo I . L'interesse delle condizioni sottolineate sta nel fatto che allora la funzione integrale risulta essere essa stessa definita su tutto l'intervallo I , qualunque sia il "punto di partenza" $x_0 \in I$ scelto. Infatti, poiché I è un intervallo, qualunque sia $x \in I$ ho che $[x_0, x] \subseteq I$ (o $[x, x_0] \subseteq I$, qualora $x < x_0$); non solo, la continuità di f su $[x_0, x]$ (ovvero $[x, x_0]$, quando $x < x_0$) mi garantisce che esiste l'integrale da x_0 ad x .

Fatte queste premesse, abbiamo:

Teorema 1 (teorema fondamentale del calcolo integrale) Sia $f: I \longrightarrow \mathbb{R}$, I intervallo, f continua su I . Sia $x_0 \in I$ fissato. Allora la funzione inte-

grale è derivabile su tutto I e la sua derivata è f . \square

Dimostrazione Basta scrivere la definizione di derivata e usare il teorema della media. Ma vediamo i dettagli, trattandosi di un teorema fondamentale. Tanto per cominciare, la funzione integrale è definita su tutto I , come già notato prima dell'enunciato del teorema.

Sia $\bar{x} \in I$. Devo provare che la funzione integrale è derivabile in \bar{x} . Per comodità indicherò con F la funzione integrale, cioè $F(x) = \int_{x_0}^x f(t)dt$. Allora

dobbiamo provare che $\exists \lim_{x \rightarrow \bar{x}} \frac{F(x) - F(\bar{x})}{x - \bar{x}} \in \mathbb{R}$, ovvero sia che

$$\exists \lim_{x \rightarrow \bar{x}} \frac{\int_{x_0}^x f(t)dt - \int_{x_0}^{\bar{x}} f(t)dt}{x - \bar{x}} \in \mathbb{R}.$$

Ma abbiamo $\frac{\int_{x_0}^x f(t)dt - \int_{x_0}^{\bar{x}} f(t)dt}{x - \bar{x}} = \frac{\int_{\bar{x}}^x f(t)dt}{x - \bar{x}}$ per il teorema di additività

sul campo.

Il teorema della media assicura che $\exists \xi \in [\bar{x}, x]$ ¹¹ t.c. $\frac{\int_{\bar{x}}^x f(t)dt}{x - \bar{x}} = f(\xi)$.

Osserviamo che avere $\xi \in [\bar{x}, x]$ (o anche $\xi \in [x, \bar{x}]$) è equivalente ad avere $\theta \in [0, 1]$ t.c. $\xi = \bar{x} + \theta \cdot (x - \bar{x})$.

Infatti, per $\bar{x} < x$ si ha:

$$\bar{x} \leq \bar{x} + \theta \cdot (x - \bar{x}) \leq x \Leftrightarrow 0 \leq \theta \cdot (x - \bar{x}) \leq x - \bar{x} \Leftrightarrow \theta \in [0, 1]$$

e per $x < \bar{x}$ si ha:

$$x \leq \bar{x} + \theta \cdot (x - \bar{x}) \leq \bar{x} \Leftrightarrow x - \bar{x} \leq \theta \cdot (x - \bar{x}) \leq 0 \Leftrightarrow \theta \in [0, 1].$$

Il vantaggio di sostituire ξ con $\bar{x} + \theta \cdot (x - \bar{x})$ non consiste solo nel fatto che questa seconda espressione non ci obbliga a fare ogni volta la distinzione dei casi tra $\bar{x} < x$ e $x < \bar{x}$. Per sfruttare appieno le potenzialità offerte da θ , dobbiamo ancora fare un passo.

Ricordo che siamo arrivati ad ottenere che $\forall x \in I, \exists \theta \in [0, 1]$ t.c.

¹¹ Oppure $\xi \in [x, \bar{x}]$...

$$(\star) \quad \frac{F(x)-F(\bar{x})}{x-\bar{x}} = f(\bar{x}+\theta \cdot (x-\bar{x})) .$$

Naturalmente, dato $x \in I$, di θ che "vanno bene" ce ne possono essere tanti. Bene, noi ne sceglieremo uno. Cioè, per ogni $x \in I$, sceglieremo uno ed un solo $\theta \in [0,1]$ tale da soddisfare la (\star) . Ma così facendo abbiamo individuato una funzione $x \mapsto \theta(x)$, definita su I e a valori in \mathbb{R} , ma ovviamente limitata, assumendo di fatto valori in $[0,1]$.

Abbiamo quindi che $\forall x \in I$
$$\frac{F(x)-F(\bar{x})}{x-\bar{x}} = f(\bar{x}+\theta(x) \cdot (x-\bar{x}))$$
, essendo $\theta(x) \in [0,1]$.

Ricordiamo che dobbiamo fare il limite del rapporto incrementale. Naturalmente faremo il limite usando l'espressione che abbiamo ottenuto a destra. Bene: se $x \rightarrow \bar{x}$, abbiamo che $\lim_{x \rightarrow \bar{x}} (x-\bar{x}) = 0$ e che $\theta(x)$ è limitata, pertanto

$$\lim_{x \rightarrow \bar{x}} \theta(x) \cdot (x-\bar{x}) = 0 ; \quad \text{quindi} \quad \lim_{x \rightarrow \bar{x}} \bar{x} + \theta(x) \cdot (x-\bar{x}) = \bar{x} .$$

Ovverossia, la funzione $g(x) = \bar{x} + \theta(x) \cdot (x-\bar{x})$ è continua in \bar{x} . Ma f è continua per ipotesi, e quindi il teorema sulla continuità delle funzioni composte ci dice che
$$\lim_{x \rightarrow \bar{x}} f(\bar{x} + \theta(x) \cdot (x-\bar{x})) = f(\lim_{x \rightarrow \bar{x}} (\bar{x} + \theta(x) \cdot (x-\bar{x}))) = f(\bar{x}) .$$

Abbiamo quindi ottenuto quanto affermato nell'enunciato del teorema, e cioè che la funzione integrale è derivabile in ogni $\bar{x} \in I$ e che $F'(\bar{x}) = f(\bar{x})$, cioè che la derivata di F è proprio f . \square

8. Primitive

Sia chiaro che questo paragrafo non dovrebbe trovarsi qui, perché tratta un argomento di calcolo differenziale, non integrale. Tuttavia, la forza della tradizione e delle notazioni tradizionali, nonché la stretta connessione che vi è tra quel che diremo e il teorema fondamentale del calcolo integrale, suggeriscono questa curiosa collocazione.

Sia dunque data $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$. Un problema abbastanza naturale che ci si può porre è il seguente: esiste una funzione $\Phi:A \longrightarrow \mathbb{R}$ t.c. $\forall x \in A$ $\Phi'(x) = f(x)$? O, più in generale, $\exists \Phi:B \longrightarrow \mathbb{R}$ t.c. $\Phi'(x) = f(x) \quad \forall x \in B$, essendo $B \subseteq A$? Da questa domanda scaturisce una definizione.

Definizione 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$ e sia $B \subseteq A$. Una funzione $\Phi:B \longrightarrow \mathbb{R}$ t.c. $\Phi'(x) = f(x) \quad \forall x \in B$, si dice primitiva di f in B (o su B). Se $B = A$, diremo più semplicemente che Φ è una primitiva di f . \square

E' evidente che vi è una interpretazione cinematica di tutto ciò: se il cronotachigrafo mi dà istante per istante la velocità di un camion, posso grazie a questo risalire alla posizione istante per istante del camion? Molti altri esempi si potrebbero fare, ma preferisco sottolineare che dalla definizione data sorgono molti problemi e molte domande.

Data f , essa ha primitive? O, magari, sotto quali condizioni su f ne ha? (f deve essere limitata? continua? derivabile? altro?). Se ne ha, quante ne ha? Sappiamo "calcolare" queste primitive? Abbiamo cioè una "formula" che ce le fornisce o in quale modo dobbiamo procedere?

Boh. Però, con un piccolo "lampo di genio" potremmo pensare di leggere una tabella contenente le derivate delle funzioni "elementari" da destra a sinistra. Et voilà, una tabella di primitive! Per esempio, una primitiva di 1 è x , di $2x$ è x^2 , di e^x è e^x , di $1/x$ è $\log(x)$ (almeno per $x > 0$), di $\cos(x)$ è $\sin(x)$, di $-\sin(x)$ è $\cos(x)$...

Sembrerebbe quasi fin troppo facile! Certo, vi sono un po' di stranezze. Per esempio, a noi magari interessa una primitiva di $\sin(x)$, non di $-\sin(x)$. Però è ovvio (dalle regole di derivazione) che, se $(\cos(x))' = -\sin(x)$, allora $(-\cos(x))' = \sin(x)$. Vi sono però problemi un po' più difficili da risolvere. Per esempio: in una tabella di derivate, "a destra" non compare $\log(x)$ o $\arctg(x)$. Per cui, se vogliamo trovarne le primitive, dobbiamo farci venire qualche altra

buona idea.

Anzi, meglio ancora: sarà opportuno costruire un pezzo di solida teoria.

Cominciamo col risolvere i problemi "fondamentali" relativi alle primitive: esistenza e unicità.

Per quanto riguarda l'esistenza di primitive, abbiamo a disposizione un comodo risultato (anche se a prima vista sembra essere poco interessante a fini pratici oltre che a quelli teorici).

Teorema 1 Sia $f:A \longrightarrow \mathbb{R}$, $A \subseteq \mathbb{R}$; sia I intervallo t.c. $I \subseteq A$. Se f è continua su I , allora f ha primitive su I . \square

Dimostrazione Banale, grazie al teorema fondamentale del calcolo integrale: fissiamo $x_0 \in I$ a nostro piacimento e possiamo affermare che $\Phi(x) = \int_{x_0}^x f(t)dt$, certamente definita $\forall x \in I$, è t.c. $\Phi'(x) = f(x)$ $\forall x \in I$. \square

Interessante, certamente, teoricamente. Ma praticamente? Se voglio trovare una primitiva di $\log(x)$, devo calcolare per ogni $x > 0$ il valore di $\int_1^x \log(t)dt$ (avendo scelto $x_0 = 1$, per esempio). In linea di principio lo posso fare, almeno con l'approssimazione voluta, costruendo partizioni sufficientemente "fini": è chiaro, però, che si tratta di una fatica non da poco. Sarebbe meglio riuscire a trovare un'altra strada. Così, in effetti, sarà: pur tuttavia ci troveremo in seguito a rivalutare alquanto il teorema 1 a fini pratici, anche se per scopi diversi che quello della ricerca di primitive.

Continuiamo per ora la nostra strada. Abbiamo un risultato che garantisce, sotto opportune ipotesi, l'esistenza di primitive. "Quante" ce ne sono? E' facile vedere che non c'è speranza di avere un'unica primitiva. Infatti:

Osservazione 1 Sia $f:A \longrightarrow \mathbb{R}$, e sia $\Phi:B \longrightarrow \mathbb{R}$ una primitiva di f su B , con $B \subseteq A$. E' immediato osservare che, $\forall c \in \mathbb{R}$, la funzione $\Psi:B \longrightarrow \mathbb{R}$, definita come $\Psi(x) = \Phi(x) + c$, è anch'essa una primitiva di f (basta usare il fatto che la derivata di una somma è la somma delle derivate e che una funzione costante è derivabile con derivata uguale a zero). \square

Si noti che, pensando al problema dell'unicità delle primitive in termini cinematici, si sarebbe potuto immaginare subito che per l'unicità non c'era speranza: anche in un caso banale, se voglio sapere dove si trova un'auto che si muo-

ve alla velocità costante di 130 Km/h, devo anche sapere qual'era la sua posizione almeno ad un certo istante! D'altronde, non sembra che occorra informazione addizionale per sapere la posizione dell'auto. Tradotto in termini generali, verrebbe da pensare che, se fissiamo il valore di una primitiva in un certo $\bar{x} \in I$, allora la primitiva dovrebbe essere univocamente determinata!

Possiamo provare questo risultato? Sì. Ma è molto interessante sapere come. In effetti, il teorema seguente è conseguenza del teorema IV.9.3. Quindi è una conseguenza del teorema fondamentale del calcolo differenziale (teorema di Lagrange). Il che penso possa apparire strano, visto che il risultato che si intende stabilire non è poi un risultato stratosferico o particolarmente sconvolgente. Rinvio a quanto già detto nelle conclusioni del capitolo IV per sottolineare che non ci sono alternative semplici percorribili.

Esempio 1 Sia $f: \mathbb{Q} \longrightarrow \mathbb{Q}$ così definita: $f(x) = \begin{cases} -1 & \text{per } x < \sqrt{2} \\ 1 & \text{per } x > \sqrt{2} \end{cases}$.

Allora f è derivabile su tutto \mathbb{Q} (come "funzione razionale di variabile razionale") e ha derivata nulla in ogni punto di \mathbb{Q} . Però f non è costante. Per curiosità, consideriamo $\hat{f}: \mathbb{Q} \longrightarrow \mathbb{R}$, definita come f (cioè: $\hat{f}(x) = f(x)$ per ogni $x \in \mathbb{Q}$), ma avente come codominio \mathbb{R} . Si noti che \hat{f} , come funzione reale di variabile reale, è derivabile e ha derivata nulla su tutto \mathbb{Q} : ma poiché \mathbb{Q} non è un intervallo, ciò non è in contrasto con il teorema IV.9.3. \square

Teorema 2 Sia $f: A \longrightarrow \mathbb{R}$ e sia $I \subseteq A$, I intervallo. Siano $\Phi, \Psi: I \longrightarrow \mathbb{R}$, due primitive di f su I . Allora $\exists c \in \mathbb{R}$ t.c. $\Psi(x) = \Phi(x) + c$ $\forall x \in I$.

Dimostrazione Basta osservare che $(\Psi - \Phi)'(x) = \Psi'(x) - \Phi'(x) = 0$ $\forall x \in I$. Vale a dire, la funzione $\Psi - \Phi: I \longrightarrow \mathbb{R}$ ha derivata nulla su I , quindi è costante. Cioè: $\exists c \in \mathbb{R}$ t.c. $(\Psi - \Phi)(x) = c$ $\forall x \in I$. Da qui la tesi. \square

Abbiamo così risolto i problemi relativi all'esistenza ed unicità di primitive per una data f . Resta tuttora aperto il problema della ricerca "pratica" di primitive per una data f . Per esempio, il logaritmo... Rinvio questo problema al paragrafo 10 (vedi esempio 1, in particolare, per le primitive del logaritmo).

Dedicherò, invece, l'attenzione ad un altro problema, relativo alle notazioni: finora, per indicare una primitiva di f ho fatto ricorso a simboli

quali Φ o Ψ . Ma, visto che una f ha infinite primitive, avrei bisogno di un simbolo per indicarle tutte. Per esempio, potrei indicare con $\mathcal{P}(f,B)$ l'insieme di tutte le primitive di $f:A \longrightarrow \mathbb{R}$, essendo $B \subseteq A$. Magari, per semplicità, potrei anche omettere B nel caso frequente in cui mi interessino le primitive su tutto l'insieme di definizione di f , cioè quando $B = A$. Ovverossia, scrivere $\mathcal{P}(f)$ anziché $\mathcal{P}(f,A)$. Si noti che il simbolo $\mathcal{P}(f,B)$ (o quello, abbreviato, $\mathcal{P}(f)$) indica un insieme di funzioni: è vero che, per lo meno nel caso in cui B sia un intervallo, esse si assomigliano molto tra di loro, grazie al teorema 2 (infatti la differenza tra due qualsiasi funzioni che stanno nell'insieme $\mathcal{P}(f,B)$ è costante), ma ciò non toglie che si tratta, appunto, di un insieme di funzioni.

La notazione tradizionale usata è, invece, $\int f(x)dx$. Di per sé non ci sarebbe niente di male ad usare questa notazione: osservo che in essa compare la variabile "muta" x : come già notato nel § 2, ci sono buone ragioni per l'uso di tali variabili "mute". Per di più, il simbolo di integrale sta a ricordarci che, dopotutto, una primitiva per f è data proprio dalla "funzione integrale".

Sfortunatamente, si vedono scritte usualmente cose tipo:

$$(\star) \quad \int 3 \cdot x^2 dx = x^3 + c .$$

La scrittura di sopra presenta molteplici incongruenze. La più macroscopica è che il simbolo x è usato contemporaneamente come variabile "muta" a sinistra e come vera variabile a destra. Come minimo, c'è il rischio di confusione se non si presta attenzione. Sarebbe meglio scrivere

$$\int 3 \cdot t^2 dt = x^3 + c ,$$

almeno. Ma anche in questo modo si hanno incongruenze: a sinistra c'è un insieme di funzioni, mentre a destra viene indicato il valore che una generica di queste funzioni assume in un punto x . Si potrebbe rimediare scrivendo:

$$(\star) \quad \left(\int 3 \cdot t^2 dt \right) (x) = x^3 + c .$$

Giova comunque ricordare che a sinistra vi è ancora qualcosa di sconclusionato: poiché il simbolo $\int f(x)dx$ indica un insieme di funzioni, non ha molto senso "calcolare il valore che un insieme di funzioni assume in un punto x ". In effetti, l'uguaglianza (\star) dovrebbe essere interpretata così: il valore che una generica primitiva assume in un punto x è dato da $x^3 + c$ (si sottintende la frase: "per un opportuno $c \in \mathbb{R}$ "). Se così stanno le cose, possiamo allora anche ca-

pire perché si usi (\star) invece che (\star) : a scapito di inesattezze formali, alleggeriamo un po' l'espressione. Basta avere le idee chiare e l'uso di (\star) non dovrebbe creare problemi (al più, se uno ha dei dubbi, si ricorda che in realtà la formula giusta sarebbe (\star)).

Abbiamo messo a posto tutto? Avevo notato che la formula (\star) è una formula sconclusionata ed avevo cercato di ovviare a questo problema dandone una "interpretazione" che sembrava sensata. Così è. Vorrei però cercare di scrivere precisamente quel che abbiamo "cercato di far dire" alla formula (\star) . In modo che non vi siano dubbi residui sul significato preciso della relazione (\star) che viene comunemente usata. Ripeto, (\star) può essere vista come una sorta di abbreviazione della relazione (\star) . Questa, a sua volta, è una abbreviazione: la formula corretta sarebbe:

$$\forall \phi \in \left(\int 3 \cdot t^2 dt \right) \exists c \in \mathbb{R} \text{ t.c. } (\forall x \in A (\phi(x) = x^3 + c)) .$$

Meglio ancora sarebbe toglierci dalla vista quel simbolo ambiguo dell'integrale indefinito. Indicando con f la funzione definita come $f(s) = 3 \cdot s^2$ per ogni $s \in \mathbb{R}$:

$$\forall \phi \in \mathcal{P}(f) \exists c \in \mathbb{R} \text{ t.c. } (\forall x \in A (\phi(x) = x^3 + c)) .$$

Concluderò del tutto il discorso sulle notazioni nel § 10, facendo vedere un esempio classico: si "dimostra" facilmente che $0 = 1$ se si usa la formula (\star) senza riflettere con attenzione sul suo significato.

9 Formula fondamentale del calcolo integrale

Ho appena finito di ribadire, all'inizio del paragrafo precedente, che l'argomento "primitive" è un argomento di calcolo differenziale. Ciò non toglie, però, che anche il calcolo integrale si sia intrufolato in modo essenziale in quel paragrafo col teorema 1¹². Nel caso di questo teorema appena citato, abbiamo che il calcolo integrale dà una mano per la soluzione di un problema di calcolo differenziale. Vedremo ora, invece, un risultato che va nella direzione opposta: il calcolo differenziale ci dà una regoletta comoda per calcolare gli integrali definiti (orientati).

Teorema 1 (formula fondamentale del calcolo integrale) Sia $f:A \rightarrow \mathbb{R}$ e sia $I \subseteq A$, I intervallo, con f continua su I . Allora, dati $a, b \in I$, e data $\Phi:I \rightarrow \mathbb{R}$, Φ primitiva per f su I , si ha:

$$\int_a^b f(x) dx = \Phi(b) - \Phi(a) \quad \square$$

Dimostrazione Osserviamo come prima cosa che l'integrale fra a e b certamente esiste, grazie alla continuità di f .

Sempre per la continuità di f , il teorema fondamentale del calcolo integrale mi garantisce che $F(x) = \int_a^x f(t) dt$ è una primitiva per f .

Ma allora, per il teorema 8.2, $\exists c \in \mathbb{R}$ t.c. $F(x) = \Phi(x) + c \quad \forall x \in I$. Cioè, $\exists c \in \mathbb{R}$ t.c. $\Phi(x) = \int_a^x f(t) dt + c \quad \forall x \in I$.

$$\text{Ma allora } \Phi(b) - \Phi(a) = \left(\int_a^b f(t) dt + c \right) - \left(\int_a^a f(t) dt + c \right) = \int_a^b f(x) dx \quad \square$$

Osservazione 1 Una notazione consueta per indicare $\Phi(b) - \Phi(a)$ è $\Phi(x) \Big|_a^b$ o, similmente, $\Phi(x) \Big|_{x=a}^{x=b}$ \square

Esempio 1 Sia dato $\int_0^1 x dx$. Poiché $x^2/2$ è una primitiva di x , allora $\int_0^1 x dx = \frac{x^2}{2} \Big|_0^1 = 1/2$. Confrontare con l'esempio 1.1 per ammirare la sensazionale semplificazione dei calcoli data dall'uso della formula fondamentale del calcolo integrale. \square

¹² Ecco perché il paragrafo 8 si trova nel capitolo V anziché IV.

10 Le "regole di integrazione"

La strana commistione tra derivate e integrali che emerge grazie al teorema fondamentale del calcolo integrale e alle sue conseguenze, dà come sottoprodotti alcuni teoremi di estrema utilità ai fini pratici di calcolo. Si tratta dei teoremi che vanno sotto il nome di "regole d'integrazione" per parti e per sostituzione (per integrali definiti e indefiniti).

Tanto per chiarire la situazione, osservo che le regole d'integrazione per parti e per sostituzione negli integrali indefiniti non sono altro che teoremi già noti di calcolo differenziale "letti a rovescio". Infatti, la regola di integrazione per parti è conseguenza della regola di derivazione di un prodotto e la regola di integrazione per sostituzione si ricava immediatamente dalla regola di derivazione delle funzioni composte.

Per quanto riguarda le regole di integrazione per parti e per sostituzione per integrali definiti, il loro "status" è ancora più modesto: sono semplicemente i loro omologhi "indefiniti", applicati agli integrali definiti col tramite della formula fondamentale del calcolo integrale.

Vediamo dapprima un risultato molto semplice:

Teorema 1 Siano $f, g: I \longrightarrow \mathbb{R}$, I intervallo, ed $\alpha, \beta \in \mathbb{R}$. Allora, se Φ e Ψ sono primitive rispettivamente di f e g su I , allora $\alpha\Phi + \beta\Psi$ è una primitiva per $\alpha f + \beta g$ su I . \square

Dimostrazione Ovvio conseguenza delle proprietà delle derivate: dalla derivabilità di Φ e Ψ segue la derivabilità di $\alpha\Phi + \beta\Psi$ e $(\alpha\Phi + \beta\Psi)'(x) = \alpha\Phi'(x) + \beta\Psi'(x) = \alpha f(x) + \beta g(x) \quad \forall x \in I$. \square

Si noti che per questo teorema non serve a nulla che le funzioni siano definite su un intervallo. Avrei potuto avere $f, g: A \longrightarrow \mathbb{R}$, con $A \subseteq \mathbb{R}$ e Φ, Ψ primitive su B , con $B \subseteq A$: allora $\alpha\Phi + \beta\Psi$ è una primitiva su B ...

Passiamo ora ai due teoremi un po' più "importanti", cominciando con la regola di integrazione per sostituzione.

Teorema 2 (regola di integrazione per sostituzione) Siano $I, J \subseteq \mathbb{R}$, I, J intervalli. Siano $f: I \longrightarrow \mathbb{R}$ e $g: J \longrightarrow \mathbb{R}$, t.c. $g(J) \subseteq I$. Supponiamo g derivabile su J . Si consideri la funzione composta $h: J \longrightarrow \mathbb{R}$ ($h = f \circ g$). Allora, se Φ è una primitiva di f su I , $(\Phi \circ g)$ è una primitiva di $(\Phi \circ g) \cdot g'$. \square

Dimostrazione Osservo che Φ è derivabile (per forza, è una primitiva!) e che anche g lo è. Allora posso applicare il teorema di derivazione delle funzioni composte e quindi, $\forall x \in J$ $(\Phi \circ g)'(x) = \Phi'(g(x)) \cdot g'(x) = f(g(x)) \cdot g'(x) = ((f \circ g)' \cdot g')(x)$. \square

Usando la notazione tradizionale, possiamo esprimere la tesi così:

$$\int f(g(t)) \cdot g'(t) dt = \Phi(g(t)) + c.$$

Ma Φ è una primitiva di f . Vale a dire: $\int f(x) dx = \Phi(x) + c$. Ma allora:

$$\left(\int f(x) dx \right)_{x=g(t)} = \Phi(g(t)) + c = \int f(g(t)) \cdot g'(t) dt \quad {}^{13}.$$

Ovverossia

$$\left(\int f(x) dx \right)_{x=g(t)} = \int f(g(t)) \cdot g'(t) dt.$$

Teorema 3 (regola di integrazione per parti) Siano $\Phi, g: I \rightarrow \mathbb{R}$, I intervallo, Φ e g derivabili in I . Allora, se Ξ è una primitiva di $\Phi \cdot g'$, una primitiva per $\Phi' \cdot g$ è data da $\Phi \cdot g - \Xi$. \square

Dimostrazione Osservo che $\Phi \cdot g - \Xi$ è derivabile $\forall x \in I$, per le ipotesi fatte. E che $(\Phi \cdot g - \Xi)'(x) = (\Phi \cdot g)'(x) - \Xi'(x) = \Phi'(x) \cdot g(x) + \Phi(x) \cdot g'(x) - \Phi(x) \cdot g'(x) = \Phi'(x) \cdot g(x)$. \square

Osservazione 1 Naturalmente, grazie al teorema fondamentale del calcolo integrale, la continuità di g' ci dà una condizione sufficiente affinché $\Phi \cdot g'$ abbia primitive: infatti $\Phi \cdot g'$ risulta in tal caso continua in quanto prodotto di funzioni continue (ricordare che Φ è derivabile per ipotesi). \square

Anche qui abbiamo una formulazione con la notazione tradizionale:

$$\int \Phi'(x) \cdot g(x) dx = \Phi(x) \cdot g(x) - \int \Phi(x) \cdot g'(x) dx.$$

Osservazione 2 E' ovvio che, quando uno deve "calcolare" un integrale indefinito, uno ha di fronte una espressione del tipo $\int h(x) dx$: evidentemente sta a chi deve "calcolare" questo integrale riuscire a scrivere $h(x)$ nella forma

¹³ Sto usando la notazione tradizionale $\int f(x) dx$ nell'accezione tradizionale (e fortemente ambigua! vedi § 8), cioè "valore di una generica primitiva nel punto x ". Pertanto uso la notazione $\left(\int f(x) dx \right)_{x=g(t)}$ per indicare il valore della "generica primitiva" di f nel punto $g(t)$.

$\Phi'(x) \cdot g(x)$. Ovverossia, scrivere $h(x) = f(x) \cdot g(x)$, con f di cui si conosca una primitiva Φ . A volte si usa anche il "trucco" di scrivere $h(x) = 1 \cdot h(x)$. Cioè $g(x) = h(x)$ e $\Phi(x) = x$. \square

Esempio 1 Calcolare $\int \log(x) dx$. \square

Dettaglio Usiamo il trucco suggerito nell'osservazione precedente. Si ha $\int \log(x) dx = \int 1 \cdot \log(x) dx = x \cdot \log(x) - \int x \cdot \frac{1}{x} dx = x \cdot \log(x) - x + c$. \square

Esempio 2 Vediamo un interessante modo di sfruttare la regola di integrazione per parti, per calcolare $\int \text{sen}^2(x) dx$. Si ha:

$$\begin{aligned} \int \text{sen}^2(x) dx &= \int \text{sen}(x) \cdot \text{sen}(x) dx = -\cos(x) \cdot \text{sen}(x) + \int \cos^2(x) dx = \\ &= -\cos(x) \cdot \text{sen}(x) + \int 1 - \text{sen}^2(x) dx = -\cos(x) \cdot \text{sen}(x) + x - \int \text{sen}^2(x) dx . \end{aligned}$$

Cioè:

$$(\star) \quad 2 \cdot \int \text{sen}^2(x) dx = x - \cos(x) \cdot \text{sen}(x) + c .$$

Da cui

$$(\star) \quad \int \text{sen}^2(x) dx = \frac{x - \cos(x) \cdot \text{sen}(x)}{2} + c . \square$$

Esercizio 1 Da dove esce fuori la costante c che compare nella relazione (\star) ? Perché non c'è $c/2$ nella successiva relazione (\star) ? \square

Esempio 3 Sfruttando l'idea dell'esempio precedente, calcoliamo un altro integrale.

$$\begin{aligned} \int \frac{\cos(x)}{\text{sen}(x)} dx &= \int \cos(x) \cdot \frac{1}{\text{sen}(x)} dx = \\ &= \text{sen}(x) \cdot \frac{1}{\text{sen}(x)} - \int \text{sen}(x) \cdot \frac{-\cos(x)}{\text{sen}^2(x)} dx = 1 + \int \frac{\cos(x)}{\text{sen}(x)} dx . \end{aligned}$$

Quindi $\int \frac{\cos(x)}{\text{sen}(x)} dx = 1 + \int \frac{\cos(x)}{\text{sen}(x)} dx$. Semplificando, otteniamo $0 = 1$.

Abbiamo ottenuto una contraddizione con l'assioma 10. Pertanto l'analisi è una teoria contraddittoria (e quindi inutile). D'altronde, il famoso teorema di incompletezza di Gödel sta lì a ricordarci che non abbiamo né possiamo avere garanzie che l'analisi (e non solo essa: anche l'aritmetica) sia esente da contraddizioni. Peccato aver ottenuto questa contraddizione solo ora, che siamo praticamente alla fine: se l'avessimo saputo prima, avremmo potuto risparmiarci un bel po' di fatica. \square

Esercizio 2 Trovare dove è l'errore nell'esempio 3. □

Chiudo il paragrafo con i teoremi di integrazione per parti e per sostituzione per integrali definiti orientati.

Teorema 4 Siano $f: I \rightarrow \mathbb{R}$, $g: J \rightarrow \mathbb{R}$, con I e J intervalli e t.c. $g(J) \subseteq I$. Supponiamo g derivabile su J , con g' continua su J , ed f continua su I . Siano $\alpha, \beta \in J$. Allora $\int_{g(\alpha)}^{g(\beta)} f(x) dx = \int_a^b f(g(t)) \cdot g'(t) dx$. □

Dimostrazione Le ipotesi fatte ci garantiscono che f abbia primitiva su I e che possiamo applicare la formula fondamentale del calcolo integrale per il calcolo dell'integrale di f da $g(\alpha)$ a $g(\beta)$. Allora $\int_{g(\alpha)}^{g(\beta)} f(x) dx = \Phi(g(\beta)) - \Phi(g(\alpha))$, dove con Φ indichiamo una primitiva di f su I .

Analogamente, le ipotesi garantiscono che $(f \circ g) \cdot g'$ abbia primitive su J e che possiamo applicare la formula fondamentale del calcolo integrale per calcolare l'integrale orientato di $(f \circ g) \cdot g'$ da α a β . Si ha: $\int_a^b f(g(t)) \cdot g'(t) dx = (\Phi \circ g)(\beta) - (\Phi \circ g)(\alpha)$. Ma $(\Phi \circ g)(\beta) - (\Phi \circ g)(\alpha)$ non è altro che $\Phi(g(\beta)) - \Phi(g(\alpha))$. Pertanto il teorema è provato. □

Teorema 5 Siano $\Phi, g: I \rightarrow \mathbb{R}$, I intervallo, Φ e g derivabili in I con derivata prima continua su I . Allora, dati $a, b \in I$, si ha:

$$\int_a^b \Phi'(x) \cdot g(x) dx = \Phi(x) \cdot g(x) \Big|_a^b - \int_a^b \Phi(x) \cdot g'(x) dx \quad \square$$

Dimostrazione Conseguenza immediata del teorema 3 e della formula fondamentale del calcolo integrale. □

INDICE

Introduzione	p	I
CAPITOLO I I NUMERI REALI		
1. I numeri razionali: proprietà formali delle operazioni e della relazione d'ordine	p	1
2. I numeri reali: gli assiomi	p	5
3. Conseguenze elementari degli assiomi	p	7
4. Relazione d'ordine e intervalli	p	13
5. Estremo superiore	p	15
6. I naturali, interi e razionali come sottoinsiemi di \mathbb{R}	p	20
7. Alcune domande di fondo sull'approccio assiomatico seguito	p	23
8. La proprietà archimedeica	p	24
9. Un modello per \mathbb{R}	p	26
10. Confronto con gli assiomi di C-S	p	30
11. I numeri complessi	p	32
12. Sommario e conclusioni	p	36
Il principio di induzione		appendice
Tabella degli assiomi di \mathbb{R}		alla fine del capitolo
CAPITOLO II LIMITI DI SUCCESSIONI		
1. Definizione di limite per successioni	p	1
2. Unicità del limite e successioni convergenti	p	9
3. Microcorso di logica	p	14
4. Bastano gli epsilon piccoli	p	18
5. Le successioni convergenti sono limitate	p	20
6. Sottosuccessioni	p	20
7. Limiti infiniti	p	23
8. I limiti e le operazioni in \mathbb{R}	p	25
9. I limiti e la relazione d'ordine su \mathbb{R}	p	31
10. Miscellanea di risultati sui limiti di successioni	p	33
11. Forme indeterminate	p	35
12. Teoremi sulle successioni che richiedono la completezza di \mathbb{R}	p	37
13. Conclusioni	p	45
CAPITOLO III FUNZIONI REALI DI VARIABILE REALE: LIMITI E CONTINUITÀ		
1. Funzioni reali di variabile reale	p	1
2. Grafico di una funzione	p	3
3. Definizione di limite	p	8
4. Caratterizzazione del limite mediante successioni e sue conseguenze	p	13
5. Estensioni della definizione di limite	p	16
6. La proprietà di limite è una proprietà locale	p	20
7. Limiti di funzioni monotone	p	23
8. Definizione di continuità in un punto	p	24
9. Conseguenze elementari della continuità in un punto	p	27
10. Continuità di una funzione su un insieme	p	29
11. Completezza di \mathbb{R} e continuità	p	31
12. Conseguenze del teorema degli zeri	p	39

13. Inversione di funzioni	p 42
14. Conclusioni	p 44

CAPITOLO IV CALCOLO DIFFERENZIALE

1. L'idea di derivata	p 1
2. Definizione di derivata	p 7
3. Compatibilità tra la derivata e le operazioni	p 10
4. Compatibilità tra la derivata e la relazione d'ordine	p 11
5. Differenziabilità	p 12
6. Derivata di funzioni inverse e composte	p 13
7. Monotonia di f e segno di f'	p 17
8. Massimi e minimi locali	p 19
9. Il teorema fondamentale del calcolo differenziale	p 23
10. I teoremi di de l'Hôpital	p 29
11. Le derivate successive	p 30
12. La formula di Taylor	p 31
13. "Disegnare il grafico di f "	p 35
14. Funzioni convesse	p 39
15. Conclusioni	p 41

CAPITOLO V CALCOLO INTEGRALE

1. La definizione di integrale	p 1
2. Una divagazione sulle notazioni	p 14
3. Proprietà delle funzioni integrabili e dell'integrale	p 16
4. Interludio: funzioni uniformemente continue	p 17
5. Integrabilità delle funzioni continue	p 20
6. Integrali orientati	p 25
7. Teorema fondamentale del calcolo integrale	p 27
8. Primitive	p 31
9. Formula fondamentale del calcolo integrale	p 36
10. Le "regole di integrazione"	p 37

Indice