Diario del Corso di Matematica di base I/II

Corso TANDEM
Docente: Sisto Baldo

ATTENZIONE: Il presente Diario del Corso Tandem di Matematica di Base I/II vuole essere un riassunto abbastanza stringato di quello che è stato detto in aula (senza la parte più euristica e alcune delle chiacchere...), e come tale può essere un utile sussidio per chi voglia sistemare i propri appunti, o per chi sia stato assente e voglia ricostruire i contenuti di una lezione.

Per Matematica di Base I, ci si riferisca alle ottime dispense di Enrico Gregorio e Francesca Mantese:
`http://profs.sci.univr.it/~gregorio/MatematicaDiBase.pdf`*,*
che contengono anche parecchi esercizi!

Dalla Home Page di Enrico Gregorio potete scaricare anche degli esempi di compiti d'esame degli anni precedenti... ai quali fatalmente ci ispireremo per preparare il vostro esame!

Indice

1 Lezione del 12/2/2014

Nelle lezioni precedenti, svolte a scuola, abbiamo imparato cosa è una *relazione* tra due insiemi A e B: si tratta semplicemente di un sottinsieme $\mathcal{R} \subset A \times B$ del prodotto cartesiano

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

In altre parole, $\mathcal{R}$ è un insieme di *coppie ordinate*, il primo elemento delle quali appartiene ad A ed il secondo a B. Si dice che gli elementi $a \in A$ e $b \in B$ *sono in relazione* o *si corrispondono* secondo $\mathcal{R}$ se e soltanto se $(a, b) \in \mathcal{R}$.

Ricordiamo anche che A si chiama *dominio* e B *codominio* della relazione. Dentro A e B, possiamo individuare i seguenti due importanti sottinsiemi, che si chiamano rispettivamente *insieme di definizione* e *immagine* della relazione:

$$\begin{aligned} \mathrm{Def}(\mathcal{R}) &= \{a \in A : \ \textit{esiste } b \in B \textit{ t.c. } (a, b) \in \mathcal{R}\}. \\ \mathrm{Im}(\mathcal{R}) &= \{b \in B : \ \textit{esiste } a \in A \textit{ t.c. } (a, b) \in \mathcal{R}\}. \end{aligned}$$

Abbiamo anche imparato che vi sono classi importanti e significative di relazioni, come le relazioni di equivalenza e le relazioni d'ordine (largo e stretto). Oggi vogliamo però occuparci di un'altra classe di relazioni che riveste particolare interesse, quella delle *funzioni*:

DEFINIZIONE: Dati due insiemi A e B, una *funzione* tra A e B è una relazione $f \subset A \times B$ che gode della seguente proprietà (detta *univocità*): se $(a, b) \in f$ e $(a, c) \in f$ allora $b = c$.

In altre parole, ad ogni elemento a dell'insieme di definizione $\mathrm{Def}(f)$ della funzione, la relazione f fa corrispondere *esattamente un elemento* del codominio B: quest'unico elemento del codominio si indica con $f(a)$

Per indicare che f è una funzione tra A e B scriviamo anche $f : A \to B$, mentre per dire che $b = f(a)$ scriviamo anche $a \mapsto f(a)$.

Per esempio, si considerino le seguenti relazioni:

$$\begin{aligned} f &= \{(0, 1), (4, 5), (5, 1)\} \subset \mathbf{N} \times \mathbf{N} \\ g &= \{(0, 1), (3, 4), (3, 6)\} \subset \mathbf{N} \times \mathbf{N} \\ h &= \{(x, y) : \ x \in \mathbf{R}, \ y = \sin x\} \subset \mathbf{R} \times \mathbf{R} \\ k &= \{(x, y) : \ y \in \mathbf{R}, \ x = y^2\} \subset \mathbf{R} \times \mathbf{R} \end{aligned}$$

Abbiamo che f e h sono funzioni (rispettivamente, $f : \mathbf{N} \to \mathbf{N}$, $h : \mathbf{R} \to \mathbf{R}$, mentre g e k non lo sono (a g appartengono le coppie $(3, 4)$ e $(3, 6)$, che hanno uguale il primo elemento e diverso il secondo e analogamente a k appartengono le coppie $(1,1)$ e $(1,-1)$).

Abbiamo osservato che nel caso di relazioni tra $\mathbf{R}$ e $\mathbf{R}$, possiamo attribuire un ovvio significato geometrico alla definizione di funzione. Infatti, gli elementi del prodotto cartesiano $\mathbf{R} \times \mathbf{R}$ possono essere interpretati come le *coordinate cartesiane di punti del piano euclideo*(in cui, ovviamente, dobbiamo aver scelto l'origine, i due assi cartesiani ortogonali x e y e l'unità di misura!). A questo punto, un sottinsieme f del piano cartesiano (cioè una relazione tra $\mathbf{R}$ e $\mathbf{R}$ è una funzione se e solo se *ogni retta parallela all'asse delle* y incontra f *al più in un punto.*

Tra gli esempi visti sopra, la sinusoide h ha questa proprietà, mentre la parabola con asse orizzontale k non ce l'ha!

Alcune importanti definizioni sono le seguenti:

DEFINIZIONE: Una funzione $f : A \to B$ si dice *totale* se $\mathrm{Def}(f) = A$, ossia *è definita su tutto* A[1].

Si dice *iniettiva* o *uno a uno* se ogni volta che $(a,c) \in f$, $(b,c) \in f$ si ha $a = b$. Questo significa che in punti distinti del suo insieme di definizione la funzione assume valori distinti...ovvero che ogni valore assunto da f viene assunto esattamente una volta.

Si dice *suriettiva* se $\mathrm{Im}(f) = B$, ossia se per ogni $b \in B$ esiste $a \in A$ tale che $b = f(a)$.

Infine, f si dice *biiettiva* se è totale, iniettiva e suriettiva. Questo significa che a *ogni elemento di* A la funzione f associa *uno ed un solo* elemento di B. Una funzione biiettiva tra A e B si chiama anche *corrispondenza biunivoca* tra A e B.

Si noti che per funzioni da $\mathbf{R}$ in $\mathbf{R}$, l'iniettività ha una semplice interpretazione geometrica: una funzione è iniettiva se e solo se ogni retta orizzontale taglia la funzione al più in un punto!

[1] A questo proposito, è importante che vi riveli un triste fatto della vita: a volte, persino i matematici usano notazioni in parziale disaccordo tra di loro! In effetti, nella grande maggioranza dei testi di matematica (comprese, ahimè, le mie dispense di Analisi I, Analisi II, Analisi Funzionale...), la scrittura $f : A \to B$ significa che l'insieme di definizione di f coincide con A: una funzione è *totale per definizione, e non c'è distinzione tra dominio ed insieme di definizione.* Leggendo un testo, è quindi importante capire quale convenzione adotta l'autore... In questo corso, ci atterremo però religiosamente alle definizioni che vi ho dato, che sono quelle della dispensa di Enrico Gregorio e Francesca Mantese!

Veniamo al concetto di *funzione inversa*. Data una funzione $f : A \to B$, consideriamo la relazione tra B e A

$$f^{-1} = \{(b, a) : \ (a, b) \in f\}$$

che si ottiene "scambiando i ruoli di dominio e codominio".

In generale, la relazione f^{-1} non è più una funzione: ad esempio se $f = \{(x, y) \in \mathbf{R}\times\mathbf{R} : x \in \mathbf{R},\ y = x^2\}$ allora f^{-1} è la parabola con asse orizzontale vista sopra, che non è una funzione!

Quando possiamo dire che f^{-1} è una funzione? Se ripensiamo alla sostanziale simmetria dei ruoli giocati da dominio e codominio nella *definizione di funzione* e in quella di *funzione iniettiva*... capiamo subito che f^{-1} *è una funzione se e soltanto se f è iniettiva.*

DEFINIZIONE: Se $f : A \to B$ è una *funzione iniettiva*, la sua *funzione inversa* $f^{-1} : B \to A$ è la relazione sopra definita.

In soldoni, data una funzione iniettiva f la funzione inversa f^{-1} associa ad ogni $b \in \mathrm{Im}(f)$ l'unico elemento $a \in \mathrm{Def}(f)$ tale che $f(a) = b$. Abbiamo quindi, per ogni $a \in \mathrm{Def}(f)$, che $f^{-1}(f(a)) = a$... e analogamente per ogni $b \in \mathrm{Im}(f)$ si ha $b = f(f^{-1}(b))$.

A questo punto è facile convincersi che $\mathrm{Def}(f^{-1}) = \mathrm{Im}(f)$ e $\mathrm{Im}(f^{-1}) = \mathrm{Def}(f)$: in particolare f^{-1} è *totale* se e solo se f è suriettiva, è *suriettiva* se e solo se f è totale. Invece, f^{-1} è sempre iniettiva... perché f è una funzione e quindi gode della proprietà di univocità.

In particolare, f è biiettiva se e solo se lo è f^{-1}.

Ovviamente, se $f : A \to B$ non è iniettiva, la relazione inversa f^{-1} non è una funzione e non è possibile *invertire* f. Per farlo, possiamo però restringere artificialmente il dominio di f ad un suo sottinsieme su cui la funzione sia iniettiva. Ad esempio, la funzione $f(x) = x^2$ (modo stenografico per scrivere $f = \{(x, y) : \ x \in \mathbf{R},\ y = x^2\}$ non è iniettiva se la consideriamo come funzione $f : \mathbf{R} \to \mathbf{R}$, mentre lo è come funzione $f : \mathbf{R}_{\geq 0} \to \mathbf{R}$. Addirittura, diventa biiettiva come funzione $f : \mathbf{R}_{\geq 0} \to \mathbf{R}_{\geq 0}$: la sua funzione inversa $f^{-1} : \mathbf{R}_{\geq 0} \to \mathbf{R}_{\geq 0}$ è... ben nota ed è data da $f^{-1}(y) = \sqrt{y}$.

Analogamente, la funzione esponenziale $f(x) = 10^x$ è una funzione totale e iniettiva di $\mathbf{R}$ in $\mathbf{R}$, con $\mathrm{Im}(f) = \mathbf{R}_{>0}$. La sua inversa è una funzione biiettiva $f^{-1} : \mathbf{R}_{>0} \to \mathbf{R}$ ed è data da $f^{-1}(y) = \log_{10}(y)$.

Infine, la funzione periodica $f(x) = \sin x$ è... "assai poco iniettiva" come funzione da $\mathbf{R}$ in $\mathbf{R}$, essendo 2π-periodica. È però biiettiva se la consideriamo come funzione $f : [-\pi/2, \pi/2] \to [-1, 1]$. La sua inversa $f^{-1} : [-1, 1] \to [-\pi/2, \pi/2]$ è la funzione *arcoseno*: $f^{-1}(y) = \arcsin(y)$.

¡¡¡WARNING!!!! DANGER!!! Da quanto appena discusso, dovreste avere ormai capito che con la notazione $f^{-1}(x)$ si denota la *funzione inversa*, che è cosa *ben diversa dal reciproco di* $f(x)$, ossia $f^{-1}(x) \neq 1/f(x)$!!!!

Per concludere questa lezione forse un po' tecnica (ma vi ho promesso qualcosa di più divertente per il prossimo incontro: ci occuperemo infatti di cardinalità!), vogliamo parlare un istante di funzione composta.

Date due funzioni f, g, con $g \circ f$ vogliamo indicare la funzione che si ottiene applicando prima f e poi g: $(g \circ f)(a) = g(f(a))$. Abbiamo fatto già qualcosa di simile componendo una funzione iniettiva con la sua funzione inversa!

Più precisamente

DEFINIZIONE: Siano $f : A \to B$, $g : C \to D$, la funzione composta

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

è definita dalla regola che $(a, d) \in g \circ f$ se e solo se esiste $c \in C$ tale che $(a, c) \in f$ e $(c, d) \in g$.

Ovviamente, $g \circ f$ avrà insieme di definizione non vuoto se e solo se l'intersezione tra l'immagine di f e l'insieme di definizione di g è non vuoto.

Ci siamo divertiti a scrivere delle condizioni per garantire che la composizione di due funzioni sia totale, iniettiva o suriettiva... Per esempio, è quasi ovvio che la composizione di due funzioni iniettive è iniettiva! Sulle dispense di Enrico e Francesca trovate una gran copia di risultati di questo tipo...

La lezione si è conclusa svolgendo un po' di esercizi, tratti anch'essi dalle dispense di Enrico e Francesca! Altri ne faremo la prossima volta... prima di passare a divertirci con le cardinalità!

2 Lezione del 19/2/2014

Se A e B sono insiemi *finiti* con lo stesso numero n di elementi, allora esiste una funzione biiettiva da A in B. Infatti, se A ha n elementi esiste una funzione biiettiva f tra l'insieme $\{0, 1, 2, \ldots, n-1\}$ e A (la funzione che "conta" gli elementi di A). Analogamente, se B ha n elementi c'è una biiezione $g : \{0, 1, 2, \ldots, n-1\} \to B$. Ma allora la funzione $g \circ f^{-1} : A \to B$ è una biiezione.

Viceversa, se due insiemi finiti sono in biiezione, allora hanno lo stesso numero di elementi: supponiamo che A e B siano in biiezione tramite f e che A abbia n elementi. C'è quindi una funzione biiettiva $g : \{0, 1, 2, \ldots, n-1\} \to A$. Ma allora $f \circ g : \{0, 1, 2, \ldots, n-1\} \to B$ è una biiezione e anche g ha n elementi.

Possiamo usare la stessa idea per "confrontare" insiemi infiniti:

DEFINIZIONE: Diremo che due insiemi A e B hanno la stessa cardinalità (e scriveremo $|A| = |B|$) se e soltanto se esiste una funzione biiettiva $f : A \to B$.

La relazione di "stessa cardinalità" è evidentemente riflessiva e transitiva (anche se purtroppo non sarebbe del tutto esatto dire che è una relazione di equivalenza perché non può esistere l'insieme di tutti gli insiemi...). In ogni caso, è certamente vero che se $|A| = |B|$ e $|B| = |C|$, allora $|A| = |C|$.

Per gli insiemi finiti, avere la stessa cardinalità è la stessa cosa che avere lo stesso numero di elementi: in particolare, nessun insieme finito può essere in biiezione con un suo sottinsieme proprio.

Questo non è più vero per gli insiemi infiniti! A questo proposito, vi ho raccontato la simpatica storiella dell'*albergo di Hilbert*, che può essere distillata come segue: i seguenti 4 insiemi, che diventano sempre più piccoli da sinistra a destra, hanno tutti la stessa cardinalità

$$|\mathbf{Z}| = |\mathbf{N}| = |\mathbf{N} \setminus \{0\}| = |2\mathbf{N}|$$

(con $2\mathbf{N} = \{2n : n \in \mathbf{N}\}$ denotiamo l'insieme dei numeri naturali pari). Una biiezione tra $\mathbf{N}$ e $\mathbf{N} \setminus \{0\}$ è data da $n \mapsto n + 1$, una tra $\mathbf{N}$ e $2\mathbf{N}$ è data da $n \mapsto 2n$. Infine, una biiezione tra $\mathbf{Z}$ e $\mathbf{N}$ è data da

$$n \mapsto \begin{cases} 2n & \textit{se } n \geq 0, \\ -2n - 1 & \textit{se } n < 0. \end{cases}$$

C'è anche un modo molto ragionevole per esprimere il fatto che un insieme è "più piccolo" di un altro:

DEFINIZIONE: Scriviamo $|A| \leq |B|$ se e soltanto se esiste una funzione *totale* e *iniettiva* $f : A \to B$.

In effetti, notiamo subito che se $C \subset B$, allora $|C| \leq |B|$: la cardinalità di un sottinsieme è sempre più piccola di quella dell'insieme di partenza (basta prendere la funzione totale e iniettiva $c \mapsto c$). Inoltre, una funzione totale e iniettiva $f : A \to B$ è una biiezione tra A e $\mathrm{Im}(f) \subset B$: A ha quindi la stessa cardinalità di un sottinsieme di B.

Siccome la composizione di due funzioni totali e iniettive $f : A \to B$, $g : B \to C$ è totale e iniettiva, abbiamo subito la proprietà transitiva: $|A| \leq |B|$, $|B| \leq |C|$ implica $|A| \leq |C$. Meno ovvio è che valga la proprietà antisimmetrica: è la tesi del seguente teorema...grazie al quale la nostra disuguaglianza tra cardinalità gode di tutte le proprietà che ci attendiamo!

TEOREMA (Cantor-Schröder-Bernstein): Siano A, B due insiemi tali che $|A| \leq |B|$ e $|B| \leq |A|$. Allora $|A| = |B|$.

DIM.: Abbiamo due funzioni totali e iniettive $f : A \to B$ e $g : B \to A$. Dobbiamo usarle per costruire una funzione biiettiva! L'idea consiste nel seguire la... genealogia degli elementi di A e B: dato $b \in B$, diciamo che $a \in A$ è il suo *genitore* se $b = f(a)$. Il genitore potrebbe non esserci (cioè b potrebbe essere orfano), altrimenti è unico perché f è iniettiva.

Analogamente, se $a \in A$, diremo che un elemento $b \in B$ è il suo genitore se $a = g(b)$: anche stavolta, se c'è un genitore esso è unico!

Dato $a \in A$, andiamo ad analizzare la sua genealogia: potrebbe avere un genitore in B, che a sua volta potrebbe essere orfano oppure avere un genitore in A, che a sua volta potrebbe essere orfano o avere un genitore in B... Questo processo di "ricerca degli antenati" potrebbe continuare all'infinito, oppure fermarsi in un elemento orfano (che chiamiamo il *capostipite* di a). Suddividiamo dunque A in tre insiemi due a due disgiunti

- A_0 è costituito dagli elementi di A che hanno una successione infinita di antenati.
- A_A è costituito dagli elementi di A che hanno un capostipite in A.
- A_B è costituito dagli elementi di A che hanno un capostipite in B.

In maniera del tutto analoga, andando ad analizzare gli antenati degli elementi di B, riusciamo a partizionare B nei tre sottinsiemi B_0, B_A e B_A definiti analogamente a prima.

Consideriamo la funzione $f : A_0 \to B_0$. Essa è ovviamente totale e iniettiva. È anche suriettiva: se $b \in B_0$, esso per definizione non è orfano, ed il suo genitore appartiene ad A_0. Quindi f è una biiezione tra A_0 e B_0. Anche $f : A_A \to B_A$ è una biiezione: è totale e iniettiva, inoltre il genitore di qualunque $b \in B_A$ deve necessariamente appartenere ad A_A. Infine, per l'analogo motivo $g : B_B \to A_B$ è una biiezione.

Possiamo quindi definire una biiezione da $h : A \to B$ nel modo seguente: sugli elementi di A_0 e A_A, h coincide con f, mentre sugli elementi di A_B essa coincide con g^{-1}. Q.E.D.

Il teorema di Cantor-Schröder Bernstein (dovuto essenzialmente a Cantor nonostante i 3 nomi...) è un risultato piuttosto profondo: si provi per esempio a costruire "a mano" una biiezione tra gli intervalli $[0, 1] = \{x \in \mathbf{R} : 0 \le x \le 1\}$ e $(0, 1] = \{x \in \mathbf{R} : 0 < x \le 1\}$. Vedrete che non è facilissimo (se non ci riuscite posso dirvi come fare...).

Invece, con il teorema di Cantor Schröder Bernstein, si vede subito che la cardinalità dei due intervalli è la stessa. Infatti, $|(0, 1]| \le |[0, 1]|$ perché il primo intervallo è un sottinsieme del secondo. Invece, una funzione totale e iniettiva da $[0, 1]$ in $(0, 1]$ è data per esempio da $x \mapsto \frac{1}{2}x + \frac{1}{4}$, da cui $|[0, 1]| \le |(0, 1]|$.

Gli insiemi che hanno la stessa cardinalità di $\mathbf{N}$ si chiamano *numerabili* (in quanto una biiezione $f : \mathbf{N} \to A$ non fa altro che "contare" o "enumerare" gli elementi di A). Essi giocano un ruolo speciale perché, tra gli insiemi infiniti, sono i più piccoli:

PROPOSIZIONE: Se A è infinito, allora $|\mathbf{N}| \le |A|$.

DIM.: Scegliamo $x_0 \in A$ (c'è perché A, essendo infinito, non è vuoto). Scegliamo poi $x_1 \in A \setminus \{x_0\}$ (c'è perché altrimenti A avrebbe 1 elemento). Segliamo poi $x_2 \in A \setminus \{x_0, x_1\}$, $x_3 \in A \setminus \{x_0, x_1, x_2\}$,... Questo processo non si arresta perchè togliendo ad A un numero finito di elementi rimane sempre qualcosa! La funzione $f : \mathbf{N} \to A$ data da $n \mapsto x_n$ è chiaramente totale e iniettiva. Q.E.D.

Abbiamo visto (si pensi all'albergo di Hilbert) che $\mathbf{N}$ ha la stessa cardinalità di diversi suoi sottinsiemi propri. Questo fatto caratterizza gli insiemi infiniti (numerabili e non):

PROPOSIZIONE: Un insieme A è infinito se e soltanto se esiste un suo sottinsieme proprio B tale che $|B| = |A|$.

DIM.: Abbiamo già visto che se A è finito allora non c'è alcun sottinsieme proprio con tale proprietà.

Se invece A è infinito, prendiamo $f : \mathbf{N} \to A$ totale e iniettiva (c'è perché $|\mathbf{N}| \leq |A|$). Sia $B = A \setminus f(0)$ e costruiamo la seguente biiezione $h : A \to B$

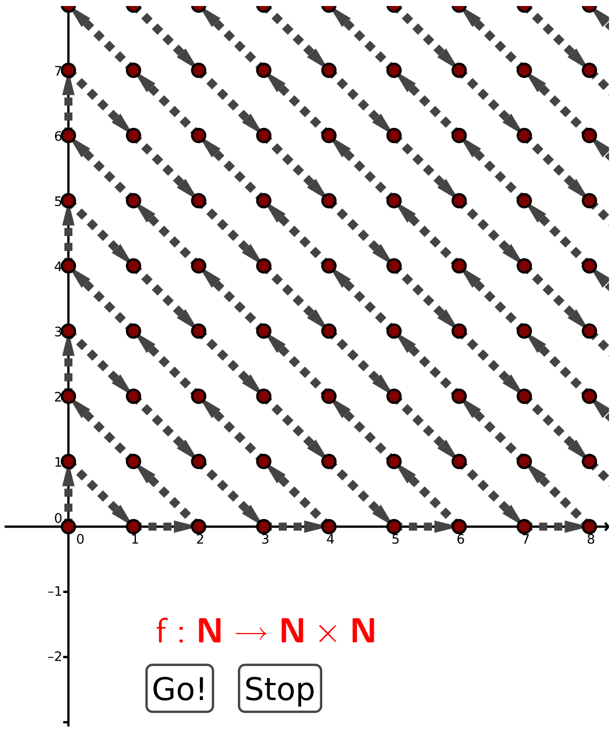
$$h(a) = \begin{cases} a & \textit{se } a \notin \mathrm{Im}(f), \\ f(n+1) & \textit{se } a = f(n) \textit{ per un } n \in \mathbf{N}. \end{cases}$$

In sostanza, se poniamo $a_n = f(n)$, la funzione h lascia fermi gli elementi di A che non appartengono alla successione a_n, mentre sposta ogni elemento di questa in quello che lo segue: si trata di un trucco alla "albergo di Hilbert" che libera il punto $a_0 = f(0)$! Q.E.D.

Anche insiemi apparentemente molto più "grossi" di $\mathbf{N}$ sono numerabili: abbiamo già visto che $|\mathbf{Z}| = |\mathbf{N}|$. Però vale anche che $|\mathbf{Q}| = |\mathbf{N}|$, fatto un bel po' meno ovvio!

PROPOSIZIONE: Si ha $|\mathbf{N}| = |\mathbf{N} \times \mathbf{N}| = |\mathbf{Q}|$.

DIM.: Una funzione biiettiva $f : \mathbf{N} \to \mathbf{N} \times \mathbf{N}$ può essere costruita utilizzando l'idea in figura.



Per vedere una versione animata della costruzione, che ho realizzato con GeoGebra, si clicchi sulla figura o si vada a
`http://www.geogebratube.org/student/m86268`

Poiché $|\mathbf{N}| = |\mathbf{Z}|$, dal fatto che $\mathbf{N} \times \mathbf{N}$ è numerabile segue che anche $\mathbf{Z} \times \mathbf{Z}$ è numerabile. Si ha poi, ovviamente, $|\mathbf{N}| \leq |\mathbf{Q}|$, perché i numeri naturali sono un sottinsieme dei razionali.

D'altra parte, possiamo costruire una funzione totale e iniettiva g da $\mathbf{Q}$ in $\mathbf{Z} \times \mathbf{Z}$ nel modo seguente: se $q \in \mathbf{Q}$, possiamo scrivere in modo unico $q = m/n$ ove $m \in \mathbf{Z}$, $n \in \mathbf{N}_{>0}$ e $MCD(m, n) = 1$ (ossia la frazione è ridotta ai minimi termini). Poniamo allora $g(q) = (m, n)$. Ne segue che $|\mathbf{Q}| \leq |\mathbf{Z} \times \mathbf{Z}| = |\mathbf{N}|$ e grazie al teorema di Cantor-Schröder-Bernstein $|\mathbf{Q}| = |\mathbf{N}|$. Q.E.D.

Esistono però insiemi infiniti che hanno cardinalità più grande di quella di $\mathbf{N}$: un esempio tipico è l'insieme $\mathbf{R}$ dei numeri reali! Vedremo la prossima volta come dimostrarlo.

3 Lezione del 26/2/2014

Dimostriamo subito che, anche nel senso delle cardinalità infinite, i numeri reali sono *più* dei numeri naturali!

PROPOSIZIONE: $|\mathbf{N}| < |\mathbf{R}|$ *(ossia* $|\mathbf{N}| \leq |\mathbf{R}|$ *e i due insiemi non hanno la stessa cardinalità).*

DIM.: Innanzitutto osserviamo che l'intervallo $(0,1)$ ha la stessa cardinalità di $\mathbf{R}$: una biiezione tra questi due insiemi è data per esempio da $x \mapsto \tan(\pi(x - \frac{1}{2}))$ o da $x \mapsto \frac{x-\frac{1}{2}}{x(1-x)}$.

Per far vedere che $(0,1)$ non ha la stessa cardinalità di $\mathbf{N}$, basta far vedere che una *qualunque* funzione totale $f : \mathbf{N} \to (0,1)$ *non può essere suriettiva.*

Sia dunque f una tale funzione: scriviamo i numeri reali $f(0)$, $f(1)$, $f(2)$,... come *numeri decimali* con infinite cifre dopo la virgola. Per rendere unica la scrittura, conveniamo di evitare le scritture che terminano con una "coda" di infiniti 9: grazie a questo accorgimento, due numeri decimali sono uguali se e solo se tutte le loro cifre coincidono!

Avremo dunque

$$
\begin{aligned}
f(0) &= 0, a_0 a_1 a_2 a_3 a_4 a_5 \ldots \\
f(1) &= 0, b_0 b_1 b_2 b_3 b_4 b_5 \ldots \\
f(2) &= 0, c_0 c_1 c_2 c_3 c_4 c_5 \ldots \\
f(3) &= 0, d_0 d_1 d_2 d_3 d_4 d_5 \ldots \\
f(4) &= 0, e_0 e_1 e_2 e_3 e_4 e_5 \ldots
\end{aligned}
$$

Costruiamo un nuovo numero decimale $x_0 = 0, \alpha_0 \alpha_1 \alpha_2 \alpha_3 \alpha_4 \alpha_5 \ldots$ in cui le cifre sono scelte in modo che α_k sia *diversa* dalla k-esima cifra di $f(k)$ e da 9: per esempio, se la k-esima cifra di $f(k)$ è pari, scegliamo $\alpha_k = 5$, se la k-esima cifra di $f(k)$ è dispari scegliamo $\alpha_k = 2$.

x_0 è un numero reale compreso tra 0 e 1 che non appartiene ad $\mathrm{Im}(f)$ perché per come è stato costruito differisce da tutti gli $f(k)$. Q.E.D.

Si può dimostrare in maniera abbastanza semplice che la cardinalità di $\mathbf{R}$ è uguale a quella dell'*insieme delle parti* di $\mathbf{N}$, ossia a quella dell'insieme $\mathcal{P}(\mathbf{N})$ di tutti i sottinsiemi di $\mathbf{N}$:

PROPOSIZIONE: Esiste una funzione biiettiva da $\mathbf{R}$ *a* $\mathcal{P}(\mathbf{N})$*.*

DIM.: Grazie al teorema di Cantor-Schröder-Bernstein, ci basta costruire due funzioni totali e iniettive $f : [0, 1) \to \mathcal{P}(\mathbf{N})$ e $g : \mathcal{P}(\mathbf{N}) \to [0, 1)$. Al posto di $\mathbf{R}$, usiamo l'intervallo $[0, 1)$ che tanto ha la stessa cardinalità[2].

Osserviamo preliminarmente che i numeri reali tra 0 e 1 si possono rappresentare come *decimali binari* o *numeri decimali in base* 2 ossia con una scrittura del tipo $x = 0, a_0 a_1 a_2 a_3 a_4 \ldots$, ove le cifre a_i $(i = 0, 1, 2, 3, \ldots)$ possono essere soltanto 0 o 1. Questa scrittura va letta nel modo seguente:

$$x = \frac{a_0}{2} + \frac{a_1}{2^2} + \frac{a_2}{2^3} + \frac{a_3}{2^4} + \ldots + \frac{a_k}{2^{k+1}} + \ldots,$$

in perfetta analogia con l'usuale interpretazione dei numeri decimali in base 10. Come per i decimali in base 10, ci sono dei casi in cui la scrittura decimale binaria non è unica: questo succede precisamente per i numeri razionali che a denominatore hanno una potenza di due, che si possono scrivere usando un numero finito di cifre binarie diverse da 0, oppure con una coda infinita di tutti 1... In questi casi, conveniamo di scegliere la rappresentazione che *non* termina con una coda di infiniti 1.

A questo punto, dato $x \in [0, 1)$, lo rappresentiamo come sopra come decimale binario e definiamo

$$f(x) = \{n \in N : \ a_n = 1\}$$

(ossia $f(x)$ è l'insieme dei numeri naturali che corrispondono al numero d'ordine delle cifre binarie diverse da 0 di x). La funzione f è evidentemente totale e iniettiva da $[0, 1)$ in $\mathcal{P}(N)$. L'iniettività viene proprio dalla nostra convenione che ci ha permesso di eliminare ogni ambiguità nella rappresentazione decimale binaria.

Non è purtroppo suriettiva: non contiene nessun sottinsieme di $\mathbf{N}$ che contenga tutti i numeri naturali "da un certo punto in poi", per esempio non contiene $\mathbf{N}$ stesso.

Poco male: definiamo $g : \mathcal{P}(\mathbf{N}) \to [0, 1)$ totale e iniettiva come segue. Dato $A \subset \mathbf{N}$, prendiamo il numero decimale binario $f(A) = 0, b_0 b_1 b_2 b_3 b_4 \ldots$ tale che ha *tutte le cifre di posto dispari pari a* 0, mentre per le cifre pari scegliamo $b_{2k} = 1$ se $k \in A$, 0 altrimenti $(k = 0, 1, 2, 3, \ldots)$. Q.E.D.

Che l'insieme delle parti sia "più grosso" dell'insieme di partenza è un fatto del tutto generale: dato un qualunque insieme A, l'insieme $\mathcal{P}(A)$ dei suoi sottinsiemi (detto insieme delle parti di A) ha cardinalità strettamente

[2] Abbiamo visto esplicitamente che $[0, 1]$ e $(0, 1]$ hanno la stessa cardinalità... In maniera del tutto analoga si dimostra che $[0, 1)$ e $(0, 1)$ hanno la stessa cardinalità e sappiamo già che l'intervallo aperto ha la stessa cardinalità di $\mathbf{R}$! Si mostri più in generale per esercizio che tutti gli intervalli non banali di $\mathbf{R}$ hanno cardinalità uguale a quella di $\mathbf{R}$.

maggiore di quella di A. In particolare, esistono cardinalità infinite arbitrariamente grandi!

TEOREMA (Cantor): Se A è un insieme e $\mathcal{P}(A)$ l'insieme di tutti i suoi sottinsiemi, allora $|A| < |\mathcal{P}(A)|$.

DIM.: La dimostrazione usa un metodo diagonale che generalizza (in maniera un pochino più astratta) quello che abbiamo usato per mostrare che $\mathbf{R}$ non è numerabile.

Innanzitutto, notiamo che la funzione $f : A \to \mathcal{P}(A)$ definita da $a \mapsto \{a\}$ è totale e iniettiva, per cui $|A| \leq |\mathcal{P}(A)|$.

Mostriamo che una qualsiasi data funzione $g : A \to \mathcal{P}(A)$ *non può essere suriettiva*, per cui vale la disuguaglianza stretta. È sufficiente considerare il seguente insieme $B \in \mathcal{P}(A)$:

$$B = \{a \in A : \ a \notin g(a)\}.$$

Dico che non esiste alcun $a \in A$ tale che $g(a) = B$. Infatti, o $a \in B$ o $a \notin B$. Se $a \in B$ allora per definizione $a \notin g(a)$ e quindi $B \neq g(a)$. Viceversa, se $a \notin B$ allora $a \in g(a)$ e abbiamo ancora $B \neq g(a)$. Q.E.D.

Come ultimo argomento del corso di Matematica di Base I, vi propongo una brevissima discussione del principio di induzione:

PROPOSIZIONE: Se un insieme di numeri naturali $A \subset \mathbf{N}$ gode delle seguenti due proprietà:

- $0 \in A$*;*
- *per ogni $n \in A$ si ha anche $(n+1) \in A$;*

allora $A = \mathbf{N}$.

DIM.: Dimostriamo il principio di induzione a partire dalla proprietà di *buon ordinamento* di $\mathbf{N}$: ogni sottinsieme non vuoto di $\mathbf{N}$ ammette un elemento minimo[3]. Dimostriamo per assurdo che l'insieme complementare di A, cioè $B = N \setminus A$, è vuoto. Se non lo fosse, avrebbe un elemento minimo n_0. Questo elemento minimo non può essere 0 per la prima delle nostre ipotesi. Ma allora $n_0 - 1 \in A$ e, per la seconda delle nostre ipotesi $n_0 \in A$, il che è assurdo! Q.E.D.

[3]In alternativa, è possibile accettare il principio di induzione come assioma e dimostrare che $\mathbf{N}$ è bene ordinato. È l'approccio che si segue nella definizione assiomatica di $\mathbf{N}$ data da Giuseppe Peano.

Può essere utile anche la seguente variante del principio di induzione (nota a volte col nome di principio di induzione forte), che è del tutto equivalente e si dimostra in modo molto simile:
PROPOSIZIONE: Se un insieme di numeri naturali $A \subset \mathbf{N}$ *gode delle seguenti due proprietà:*

- $0 \in A$*;*
- *se per* $n \in \mathbf{N}$ *vale che* $k \in A$ *per ogni* $k \leq n$*, allora* $(n+1) \in A$*;*

allora $A = \mathbf{N}$.

Spesso il principio di induzione, in una delle sue forme, si usa per dimostrare che una certa affermazione, dipendente da un numero naturale n, vale per ogni $n \in \mathbf{N}$: si applica una delle due proposizioni all'*insieme dei numeri naturali per cui l'affermazione è vera*!

Ad esempio, in questo modo possiamo dimostrare che per ogni $n \in \mathbf{N}$ si ha
$$\sum_{k=0}^{n} k = 0 + 1 + 2 + 3 + 4 + 5 + \ldots + n = \frac{n(n+1)}{2}.$$
Infatti, l'affermazione è chiaramente vera per $n = 0$ ("base dell'induzione"). Supponiamola poi vera per un certo n ("ipotesi induttiva"). Allora
$$\sum_{k=0}^{n+1} k = \sum_{k=0}^{n} k + (n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{(n+1)(n+2)}{2}$$
e la nostra affermazione è vera anche per $n + 2$: grazie al principio di induzione, essa è vera sempre.

Analogamente, possiamo dimostrare che per ogni $a \in \mathbf{R} \setminus \{1\}$ ed ogni $n \in \mathbf{N}$ vale
$$\sum_{k=0}^{n} a^k = 1 + a + a^2 + a^3 + \ldots + a^n = \frac{a^{n+1} - 1}{a - 1}.$$
Oppure che, per $x > -1$ e $n \in \mathbf{N}$ vale la disuguaglianza di Bernoulli:
$$(1+x)^n \geq 1 + nx.$$

È poi il principio di induzione a giustificare le definizioni date per ricorrenza: per esempio, se $a \in \mathbf{R}$, $a \neq 0$ e $n \in \mathbf{N}$, la potenza a^n si definisce in modo rigoroso ponendo $a^0 = 1$ e $a^{n+1} = a \cdot a^n$ per ogni $n \in \mathbf{N}$. Il principio di induzione assicura che in questo modo la potenza a esponente naturale è definita qualunque sia l'esponente.

Con la versione forte del principio di induzione si dimostra poi facilmente che la *suggessione di Fibonacci*, definita per ricorrenza nel modo seguente:

$$a_0 = 1,\ a_1 = 2,\ a_{n+2} = a_n + a_{n+1} \text{ per ogni } n \in \mathbf{N}$$

si può esplicitare come segue per ogni $n \in \mathbf{N}$:

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

La formula è molto graziosa perché mette in relazione la successione di Fibonacci ed il rapporto aureo...

Abbiamo visto anche che, *come sempre in matematica*, occorre prestare attenzione ai dettagli: mi sono divertito a "dimostrarvi" per induzione che tutte le persone sono bionde con gli occhi azzurri... affermazione della quale sono un controesempio (non ho mai avuto i capelli biondi, neanche prima che diventassero grigi: quanto poi agli occhi azzurri...).

Per concludere il corso, abbiamo svolto qualche esercizio tipo esame su relazioni (d'equivalenza e d'ordine largo e stretto) e su cardinalità.

4 Lezione del 4/4/2014

Buona parte delle costruzioni matematiche dalla geometria di Euclide in poi sono basate sui *numeri reali*... ma siamo sicuri di conoscere bene questo insieme numerico?

Consideriamo la seguente catena di insiemi numerici, sempre più grandi:

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

Ogni volta che passiamo da un insieme al successivo, *guadagnamo qualcosa*... Nell'insieme $\mathbf{N} = \{0, 1, 2, \ldots\}$ dei numeri naturali non è possibile trovare l'elemento inverso di un numero rispetto alla somma (l'opposto): perché questo si possa fare dobbiamo allargarci all'insieme $\mathbf{Z} = \{0, \pm 1, \pm 2, \ldots\}$ dei numeri interi. Analogamente, in $\mathbf{Z}$ non è possibile definire l'operazione inversa del prodotto: per questo, si introduce l'insieme $\mathbf{Q}$ dei numeri razionali ($\mathbf{Q} = \{m/n : \ m, n \in \mathbf{Z}, \ n \neq 0\}$).

Con l'insieme dei numeri razionali potremmo dirci soddisfatti, almeno dal punto di vista delle quattro operazioni! E allora, perché sentiamo il bisogno di allargare ulteriormente l'insieme dei "numeri"?

Questa necessità divenne evidente già agli albori della matematica greca (anche se i greci avevano una visione più "geometrica" che "algebrica" della matematica): i pitagorici si accorsero che *la lunghezza della diagonale di un quadrato di lato 1 non è un numero razionale*. In termini moderni (e grazie al teorema di Pitagora), questo equivale a dire che $\sqrt{2}$ non è un numero razionale... In seguito ci divertiremo a dimostrare questo fatto!

Per i pitagorici, la scoperta dell'irrazionalità di $\sqrt{2}$ ebbe sconvolgenti conseguenze filosofiche...per noi, significa solo che dobbiamo trovare un insieme *più ampio* di numeri, in modo che almeno uno di essi abbia quadrato uguale a 2 (e in cui magari sia possibile risolvere altri interessanti problemi!).

Una buona risposta a queste necessità è l'insieme $\mathbf{R}$ dei numeri reali, che noi ben conosciamo. Vero?

Cosa sono i numeri reali di cui abbiamo parlato (e che abbiamo usato) per buona parte della nostra carriera scolastica?

Una possibile risposta: sono tutti i numeri decimali, eventualmente con infinite cifre dopo la virgola. Questo è un buon modello dei numeri reali, che presenta però un piccolo problema: se *definiamo* i reali come numeri decimali infiniti, non è poi facilissimo definire le operazioni e la relazione d'ordine, e mostrare poi che esse godono di tutte le proprietà che ci aspettiamo... Comunque, questo è possibile senza eccessive difficoltà.

Un approccio alternativo (adottato da molti testi di analisi matematica) è quello assiomatico: i numeri reali sono *per definizione* un *campo ordinato e completo*.

Questo significa che i numeri reali sono un *insieme* su cui sono definite due *operazioni* (la somma e il prodotto), ed una *relazione d'ordine* che godono di tutta una serie di buone proprietà che elencheremo meglio tra poco. Un tale oggetto si dice *campo ordinato*...ma anche **Q** è un campo ordinato!

Quel che distingue **R** da **Q** è l'*assioma di completezza*... di cui ci occuperemo tra poco!

Riprendiamo i nostri discorsi sull'insieme dei numeri reali. Siccome abbiamo affermato che la radice quadrata di 2 non è un numero razionale, vediamo di dimostrarlo!

In termini precisi, dire che $\sqrt{2}$ è razionale equivale ad affermare che esiste un numero razionale $q = m/n$ tale che $q > 0$ e $q^2 = 2$. Riducendo la frazione ai minimi termini, non è restrittivo supporre che i numeri naturali m e n non abbiano fattori primi in comune.

Ora, la nostra supposizione equivale a $m^2 = 2n^2$, da cui segue che m^2 è un numero pari. Poiché ogni fattore primo di m^2 deve essere presente anche in m, ne deriva che m è pari.

Dunque, $m = 2r$ per qualche numero naturale r, e la nostra identità diventa $4r^2 = 2n^2$, da cui $2r^2 = n^2$. Ripetendo esattamente il ragionamento appena fatto, questo mostra che n è pari. Assurdo perché abbiamo supposto che m e n non abbiano fattori in comune, e quindi essi non possono essere entrambi pari!

In sostanza, il teorema appena dimostrato ci dice che esistono tutta una serie di problemi "naturali" per la matematica che *non si possono risolvere* nel campo **Q** dei razionali.

È per questo che risulta necessario introdurre l'insieme **R** dei numeri reali, del quale diamo la seguente definizione:

DEFINIZIONE: l'insieme **R** *dei numeri reali è un campo ordinato completo.*

Questo significa che i numeri reali sono un *insieme* su cui sono definite due *operazioni* (la somma e il prodotto), entrambe associative e commutative. Inoltre, entrambe le operazioni hanno un elemento neutro (0 e 1 rispettivamente) e sono "invertibili" (cioè per ogni $x \in \mathbf{R}$ esiste un altro elemento che denotiamo $(-x)$ tale che $x + (-x) = 0$; per ogni $x \in \mathbf{R}$, $x \neq 0$ esiste un altro elemento x^{-1} tale che $x \cdot x^{-1} = 1$). Vale inoltre la proprietà distributiva, che lega la somma al prodotto. Un insieme con due operazioni che godono di queste proprietà è detto *campo.*

C'è poi una *relazione d'ordine*, che dati due numeri reali ci consente di dire qual è il più grande. Questa realazione d'ordine è compatibile con le operazioni (nel senso che possiamo manipolare le disuguaglianze nel modo in cui siamo abituati: sommando uno stesso numero reale ad ambo i membri di una disuguaglianza essa rimane vera, così come se moltiplichiamo ambo i membri per uno stesso numero reale positivo). Un campo che gode di queste proprietà è un *campo ordinato*: ma anche $\mathbf{Q}$ è un campo ordinato!

Il campo ordinato $\mathbf{R}$ è poi *completo*:

ASSIOMA DI COMPLETEZZA: Ogni sottoinsieme non vuoto e superiormente limitato di $\mathbf{R}$ *ammette estremo superiore in* $\mathbf{R}$.

Per capire la formulazione dell'assioma di completezza che si rifà al concetto di estremo superiore, ricordiamo alcune definizioni già viste nel primo modulo del corso tandem (parte fatta a scuola sulle relazioni d'ordine!):
DEFINIZIONI: Se $A \subset \mathbf{R}$, un *maggiorante* di A è un numero reale M tale che $M \geq a$ per ogni $a \in A$.

Un sottinsieme di $\mathbf{R}$ si dice superiormente limitato se ammette almeno un maggiorante.

L'*estremo superiore* di un sottinsieme A di $\mathbf{R}$ è il *minimo* dei maggioranti di A, se questo minimo esiste.

Cerchiamo di comprendere il significato di questa definizione di estremo superiore (che per inciso si indica col simbolo $\sup A$...).

Se l'insieme A ammette massimo, allora l'estremo superiore coincide col massimo. Infatti il massimo dell'insieme è un maggiorante per definizione, e inoltre nessun numero più piccolo del massimo può essere un maggiorante (perché è superato dal massimo stesso, che è un elemento dell'insieme).

D'altra parte, un insieme infinito non è detto che possegga massimo anche se è superiormente limitato: per esempio, la semiretta $A = \{x \in \mathbf{R} : \ x < 2\} = (-\infty, 2)$ non possiede un elemento massimo. Infatti, dato un qualunque elemento $a \in A$, il numero $\frac{a+2}{2}$ è ancora minore di 2, ed è maggiore di a: *a non può essere dunque l'elemento massimo della semiretta*. Invece, è immediato verificare che $\sup A = 2$... L'estremo superiore è la *naturale generalizzazione del concetto di massimo* agli insiemi superiormente limitati che non hanno massimo!

L'assioma di completezza dice una cosa non ovvia: *qualunque* sottinsieme non vuoto e superiormente limitato $A \subset \mathbf{R}$ ammette il sup. La cosa *non* è immediata perché il sup è definito come minimo di un insieme *infinito* (l'insieme dei maggioranti di A), e non sempre un insieme infinito ammette minimo!

Evidentemente, l'assioma di completezza (in una qualunque delle sue formulazioni equivalenti) *non è vero* nel campo dei razionali: per esempio, l'insieme $A = \{q \in \mathbf{Q} : \ q \geq 0, \ q^2 \leq 2\}$ è un sottinsieme di $\mathbf{Q}$ non vuoto e superiormente limitato (per esempio, 2 è un maggiorante), ma esso non ha estremo superiore in $\mathbf{Q}$: il problema è che possiamo trovare maggioranti razionali *arbitrariamente vicini* a $\sqrt{2}$, che peró non appartiene ai razionali. Ovviamente, se vediamo questo insieme come *sottinsieme di* $\mathbf{R}$, l'estremo superiore c'è ed è uguale a $\sqrt{2}$.

L'affermazione appena fatta è una conseguenza praticamente immediata della seguente, importante proprietà di *densità* dei numeri razionali in $\mathbf{R}$:

PROPOSIZIONE (Proprietà archimedea e densità di $\mathbf{Q}$ in $\mathbf{R}$): Siano a, b numeri reali con $0 < a < b$. Allora esiste un numero naturale n tale che $na > b$ (proprietà archimedea).

Se poi $a, b \in \mathbf{R}$, $a < b$ (di segno qualunque), esiste un numero razionale q tale che $a < q < b$ (densità di $\mathbf{Q}$ in $\mathbf{R}$).

DIM. (Non vista in classe): Proviamo che valgono la proprietà archimedea e la densità dei razionali. Per provare la proprietà archimedea, basta osservare che esiste un numero naturale n tale che $n > \frac{b}{a}$: basta prendere $n = [\frac{b}{a}] + 1$[4] Moltiplicando per a si ha la tesi.

Proviamo poi la densità dei razionali nel caso $0 < a < b$. Intanto, per la proprietà archimedea esiste $n \in \mathbf{N}$ tale che $n(b-a) > 1$, da cui $nb > 1 + na$. Ne segue che c'è un numero naturale m compreso tra nb e na (possiamo prendere semplicemente $[nb]$, tranne nel caso in cui nb è intero: in quel caso prendiamo $nb-1$). Ma allora $na < m < nb$, da cui $a < \frac{m}{n} < b$, come volevasi. Se poi $a < 0 < b$, basta prendere $q = 0$. Se invece $a < b < 0$, basta cambiare di segno... Q.E.D.

In matematica, oltre all'estremo superiore si usa spesso l'estremo inferiore che è l'oggetto simmetrico:

DEFINIZIONE: Un minorante di un insieme $A \subset \mathbf{R}$ è un numero reale c tale che $c \leq a$ per ogni $a \in A$. A si dice inferiormente limitato se possiede un minorante. L'estremo inferiore di A (se esiste) è il massimo dei minoranti di A, e si indica con $\inf A$.

Ovviamente, l'estremo inferiore coincide con il minimo di A quando questo esiste.

Inoltre, un'ulteriore formulazione equivalente dell'assioma di completezza consiste nel chiedere che ogni insieme inferiormente limitato ammette estremo inferiore (esercizio)!

Per comodità, vale anche la pena di introdurre una notazione per indicare l'estremo superiore e l'estremo inferiore di insiemi illimitati, e dell'insieme vuoto:

DEFINIZIONE: Il fatto che un insieme A non sia superiormente limitato si esprime con la scrittura $\sup A = +\infty$. *Analogamente, per un insieme*

[4] Con $[x]$ indichiamo la *parte intera* di un numero reale x, ossia il più grande numero intero minore o uguale a x.

illimitato inferiormente scriveremo $\inf A = -\infty$. *Poniamo poi* $\sup \emptyset = -\infty$, $\inf \emptyset = +\infty$.

La definizione assiomatica di $\mathbf{R}$ è comoda (è sostanzialmente un menù delle proprietà che possiamo utilizzare quando manipoliamo i numeri reali), ma rimane il problema di mostrare che *esiste almeno un insieme, dotato di operazioni e relazione d'ordine, che soddisfa tutti gli assiomi*: abbiamo bisogno di un *modello* dei numeri reali.

Come accennavamo, un tale modello è costituito dai *numeri decimali infiniti*. Non è difficile convincersi che con tale modello si possono definire le operazioni (non banalissimo!) e la relazione d'ordine, che esse godono di tutte le proprietà che ci servono e che vale l'assioma di completezza...

L'esistenza della radice quadrata di un numero reale positivo (necessaria per ... accontentare Pitagora) può essere recuperata usando l'estremo superiore: se $a \in \mathbf{R}$, $a > 0$, definiamo $\sqrt{a} = \sup\{x \in \mathbf{R} : \ x^2 \leq a\}$. Questo definisce un numero reale positivo, il cui quadrato si può dimostrare che è uguale ad a.

Non dimostriamo questo fatto, perché esso è una conseguenza immediata del teorema di esistenza degli zeri per le funzioni continue, che vedremo tra poco.

È utile ricordare le seguenti *caratterizzazione* di estremo superiore ed estremo inferiore: siano $A \subset \mathbf{R}$, $M, m \in \mathbf{R}$. Allora si ha $\sup A = M$ se e solo se

- $a \leq M$ per ogni $a \in A$ (M è un maggiorante);
- per ogni $\varepsilon > 0$ esiste $\overline{a} \in A$ tale che $\overline{a} > M - \varepsilon$ (nessun numero strettamente minore di M è un maggiorante).

Analogamente si ha $m = \inf A$ se e solo se

- $a \geq m$ per ogni $a \in A$ (m è un minorante);
- per ogni $\varepsilon > 0$ esiste $\overline{a} \in A$ tale che $\overline{a} < m + \varepsilon$ (nessun numero strettamente maggiore di m è un minorante).

Ci siamo poi dedicati ad un ripasso critico sulle potenze: potenze ad esponente naturale, intero, razionale (la definizione è "obbligata" se vogliamo che valgano le proprietà delle potenze!).

Potenze ed esponente reale: se $a > 1$ e $x \in \mathbf{R}$, definiamo

$$a^x = \sup\{a^q : \ q \in \mathbf{Q}, \ q \leq x\}.$$

Si può dimostrare che questo oggetto ha tutte le proprietà della funzione esponenziale che ben conosciamo (omettiamo però la dimostrazione).

A questo punto, per andare avanti, abbiamo ricordato il concetto di limite e quello di funzione continua...

Usando il concetto di estremo superiore non è difficile dimostrare il seguente teorema:

TEOREMA: Sia $\{a_n\}$ una successione crescente (cioè $a_{n+1} \geq a_n$ per ogni n). Allora $\{a_n\}$ ammette limite per $n \to +\infty$, e si ha

$$\lim_{n\to+\infty} a_n = \sup\{a_n : \ n \in \mathbf{N}\}.$$

Analogo risultato vale per successioni decrescenti: in questo caso il limite coincide con l'estremo inferiore dell'immagine della successione.

DIM.: Sia $S = \sup\{a_n : \ n \in \mathbf{N}\}$. Supponiamo dapprima che sia $S \in \mathbf{R}$: vedremo poi il caso $S = +\infty$.

Fissiamo $\varepsilon > 0$. Siccome S è un maggiorante dei valori assunti dalla successione, abbiamo che $a_n \leq S < S + \varepsilon$ per ogni n. D'altra parte, $S - \varepsilon$ *non* è più un maggiorante (per definizione di sup), per cui esiste un elemento della successione, chiamiamolo $a_{\overline{n}}$, tale che $a_{\overline{n}} > S - \varepsilon$. Siccome la successione è crescente, se $n \geq \overline{n}$ si ha $a_n \geq a_{\overline{n}} > S - \varepsilon$.

Mettendo insieme le due disuguaglianze ottenute sopra, abbiamo che per $n > \overline{n}$ si ha $|a_n - S| < \varepsilon$.

Se poi $S = +\infty$, l'immagine della successione non ammette maggioranti. Dunque, per ogni fissato $M > 0$ esiste $\overline{n}$ tale che $a_{\overline{n}} > M$. Per la crescenza della successione abbiamo allora $a_n \geq a_{\overline{n}} > M$ per ogni $n \geq \overline{n}$ e dunque $\lim\limits_{n\to+\infty} a_n = +\infty$ Q.E.D.

Si noti che il risultato non è ovvio: esistono successioni che non ammettono limite, per esempio $a_n = (-1)^n$. C'è anche di peggio: per esempio, abbiamo visto nel primo modulo che esiste una funzione biiettiva $f : \mathbf{N} \to \mathbf{Q}$... provate a pensare un po' a quanto è "selvaggia" la successione $a_n = f(n)$...

Un risultato fondamentale sulle funzioni continue, che dipende in modo essenziale dall'assioma di completezza, è il

TEOREMA DI ESISTENZA DEGLI ZERI: Sia $f : [a, b] \to \mathbf{R}$ una funzione *continua* tale che $f(a) < 0$ e $f(b) > 0$. Allora esiste un punto $c \in (a, b)$ tale che $f(c) = 0$.

DIM.: Usiamo il cosiddetto *metodo di bisezione.* Sia $d = (b - a)/2$ il punto medio dell'intervallo $[a, b]$: se $f(d) = 0$ siamo felicissimi perché abbiamo trovato il punto voluto, in caso contrario avremo $f(d) < 0$ oppure $f(d) > 0$. In ogni caso, in uno dei due "mezzi intervalli" $[a, d]$ oppure $[d, b]$ si ripropone la situazione di partenza: f è negativa nell'estremo sinistro dell'intervallo, positiva nell'estremo destro. Chiamiamo $[a_1, b_1]$ il semiintervallo che gode di questa proprietà.

Ripetiamo poi la stessa costruzione: prendiamo il punto di mezzo dell'intervallo $[a_1, b_1]$ e osserviamo che se la funzione non si annulla nel punto di mezzo (ma se così fosse avremmo finito), in uno dei due mezzi intervalli che chiameremo $[a_2, b_2]$ si ripropone la situazione di partenza: $f(a_2) < 0$ e $f(b_2) > 0$.

Iteriamo questa costruzione: se il processo non si arresta perché troviamo un punto in cui la funzione si annulla, avremmo individuato una successione infinita di intervalli $[a_n, b_n]$, ciascuno contenuto nel precedente e tali che $f(a_n) < 0$, $f(b_n) > 0$. Per costruzione abbiamo che la successione degli estremi sinistri a_n è crescente, la successione degli estremi destri b_n è decrescente e inoltre $b_n - a_n = (b - a)/2^n$. Siccome una successione crescente e limitata ammette limite finito, esisterà il limite $\lim_{n\to+\infty} a_n = c$, ed evidentemente $c \in [a, b]$.

È evidente che si ha anche $\lim_{n\to+\infty} b_n = c$ per quanto osservato sopra sulla differenza tra a_n e b_n. Grazie alla continuità di f, si ha poi

$$\lim_{n\to+\infty} f(a_n) = f(c) \qquad \lim_{n\to+\infty} f(b_n) = f(c).$$

D'altra parte, il primo limite deve essere necessariamente ≤ 0 in quanto limite di una successione di numeri negativi, mentre il secondo deve essere ≥ 0 in quanto limite di una successione di numeri positivi: siccome i due limiti sono entrambi uguali a $f(c)$, ne deriva che $f(c) = 0$. Q.E.D.

DIMOSTRAZIONE ALTERNATIVA: Vogliamo proporre un'altra dimostrazione del teorema, leggermente più rapida.

Poniamo $c = \sup A$, dove $A = \{x \in [a, b] : f(x) < 0\}$. Questo è evidentemente un numero reale compreso tra a e b. Dico che $f(c) = 0$.

Infatti, se per assurdo avessimo $f(c) > 0$, per definizione di limite avremmo $f(x) > 0$ anche per tutti gli x in un certo intorno sinistro $[c - \delta, c]$ di c[5]. Quindi $c - \delta$ sarebbe un maggiorante di A più piccolo di c, contro la definizione di estremo superiore.

Se poi fosse $f(c) < 0$, dovrebbe essere $c < b$ (perché $f(b) > 0$). Per lo stesso motivo di prima, troveremmo $\delta > 0$ tale che $f(x) < 0$ per $x \in [c, c+\delta]$, e c non sarebbe più un maggiorante di A. Q.E.D.

[5] Questo semplice fatto è noto come *teorema della permanenza del segno*: se una funzione ha limite positivo in x_0, allora è positiva in un intorno di x_0 (con la possibile esclusione di x_0).

Un immediato corollario del teorema di esistenza degli zeri è il seguente

TEOREMA (dei valori intermedi): Se $f : [a, b] \to \mathbf{R}$ *è una funzione continua, essa assume tutti i valori compresi tra* $f(a)$ *e* $f(b)$.

Dimostrazione: Sia y_0 un valore compreso tra $f(a)$ e $f(b)$. Basta applicare il Teorema di esistenza degli zeri alla funzione $g(x) = f(x) - y_0$...
Q.E.D.

Questo teorema è particolarmente utile per dimostrare che certe funzioni continue sono *invertibili.* Dal teorema dei valori intermedi segue che se $f :$

$[a, b] \to \mathbf{R}$ è una funzione (totale) continua e strettamente crescente (che è evidentemente iniettiva!), essa è *suriettiva* sull'intervallo $[f(a), f(b)]$: in altre parole, essa è *invertibile.* Questo ci assicura l'esistenza di radici, logaritmi, funzioni inverse delle funzioni trigonometriche... Non è difficile dimostrare che *anche la funzione inversa* di una tale funzione è continua!

Al termine della lezione, ho cercato di convincervi che *adattando opportunamente il metodo di bisezione* possiamo abbastanza facilmente dimostrare il

TEOREMA (di Weierstrass): Una funzione (totale e) continua $f : [a, b] \to \mathbf{R}$ *ammette massimo e minimo. (Attenzione: è importante che il dominio della funzione sia un intervallo chiuso e limitato, e che la funzione sia continua. si vede facilmente con qualche controesempio che senza queste ipotesi la tesi può anche essere falsa!).*

5 Lezione dell'11/4/2014

In questa lezione voglio discutere alcuni esempi che evidenziano... diffusi pregiudizi su alcuni concetti cardine di analisi matematica.

Per prima cosa, voglio analizzare il concetto di continuità di una funzione: spero di avervi convinto di quanto questo sia importante nel nostro precedente incontro! Toccheremo anche i legami che vi sono tra continuità e derivabilità.

Ricordo la seguente

Definizione: Supponiamo di avere una funzione (totale) $f : A \to \mathbf{R}$, ove $A \subset \mathbf{R}$ e sia $x_0 \in A$. f si dice continua in x_0 se *per ogni intorno J di $f(x_0)$ esiste un intorno I di x_0 tale che $f(I \cap A) \subset J$.*

Ricordo che un *intorno* di un punto $x_0 \in \mathbf{R}$ è, per definizione, un qualunque insieme che contiene un intervallo non banale centrato in x_0.

Questa definizione, se x_0 è un *punto di accumulazione* di A, è del tutto equivalente all'usuale richiesta che valga l'uguaglianza

$$f(x_0) = \lim_{x \to x_0} f(x).$$

Ma che succede se x_0 *non è* un punto di accumulazione per A, ovvero se x_0 è un *punto isolato* dell'insieme di definizione A?

Ebbene, la nostra definizione implica che f è continua in x_0... in contrasto con alcuni diffusi libri di testo di analisi matematica per la Scuola Secondaria Superiore che, in varie combinazioni,

- si limitano a definire la continuità per funzioni definite su intervalli aperti... per poi enunciare poco oltre il teorema di Lagrange chiedendo (correttamente) la continuità agli estremi di un intervallo chiuso;
- definiscono la continuità unicamente per punti del dominio che siano di accumulazione;
- danno una definizione di "punto di discontinuità" secondo la quale una funzione è sempre discontinua nei punti isolati del suo dominio.

La continuità o meno di una funzione nei punti isolati del suo dominio può sembrare una questione di lana caprina. Non lo è del tutto se si vuole mantenere la "compatibilità" con nozioni più generali di continuità che stanno alla base della matematica "moderna" (dal ventesimo secolo in poi): si è infatti trovato utile estendere il concetto di continuità a funzioni definite su insiemi più generali dei numeri reali, ossia su insiemi dotati di una distanza (spazi metrici) o di una topologia (spazi topologici). Secondo queste nozioni più generali di continuità, una funzione è sempre continua nei punti isolati del suo dominio!

Del resto, questo è in perfetto accordo con l'idea intuitiva di continuità: f è continua in x_0 se *avvicinandosi abbastanza a* x_0 la funzione assume valori *arbitrariamente vicini a* $f(x_0)$. Se x_0 è un punto isolato dell'insieme di definizione, "avvicinarsi abbastanza a x_0" significa "stare in x_0"... e la continuità diventa autoevidente!

Altra cosa su cui i libri di testo non sembrano riuscire a mettersi d'accorso è il concetto di *punto di discontinuità* di una funzione: alcune definizioni sono davvero fantasiose e a volte autocontraddittorie (perchè si scopre che vi possono essere dei punti in cui la funzione è sia continua che discontinua)!

La cosa più sensata da fare è dire che un punto di discontinuità di una funzione (totale) $f : A \to \mathbf{R}$ è un punto x_0 *appartenente all'insieme di definizione* A in cui la funzione *non è continua.* Meglio invece non pronunciarsi sui punti che *non* fanno parte dell'insieme di definizione della funzione.

In questo modo, la funzione $f(x) = 1/x$ è continua in tutti i punti del suo insieme di definizione $A = \{x \in \mathbf{R} : \ x \neq 0\}$. Si noti che, anche se rinunciamo a dichiarare che 0 è un punto di discontinuità per f, non stiamo affatto sminuendo l'interesse di tale punto ai fini del disegno del grafico della funzione: un asintoto verticale rimane un asintoto verticale!

Analogamente, le funzioni $\tan x$, $\log x$, $\frac{\sin x}{x}$ sono continue in tutti i punti del loro insieme di definizione, come lo è la funzione $\sqrt{\sin(x) - 1}$.

Vediamo alcuni esempi che mostrano funzioni "malamente discontinue":

ESEMPIO: Si consideri la funzione (totale) $f : \mathbf{R} \to \mathbf{R}$ data da

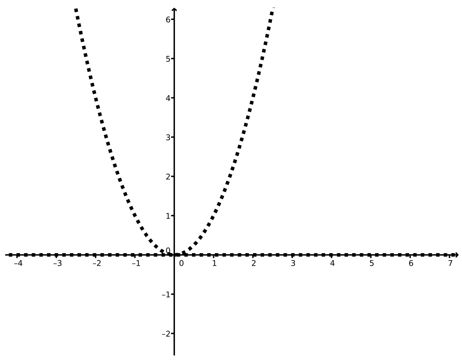
$$f(x) = \begin{cases} 1 & se\ x \in \mathbf{Q}, \\ 0 & se\ x \in \mathbf{R} \setminus \mathbf{Q}. \end{cases}$$

Questa funzione è ovunque discontinua: vicino a ogni punto razionale, esistono punti irrazionali arbitrariamente vicini ad esso. Viceversa, vicino ad ogni punto irrazionale esistono punti razionali ad esso arbitrariamente vicini.

ESEMPIO: Si consideri la funzione (totale) $f : \mathbf{R} \to \mathbf{R}$ data da

$$f(x) = \begin{cases} x^2 & se\ x \in \mathbf{Q}, \\ 0 & se\ x \in \mathbf{R} \setminus \mathbf{Q}. \end{cases}$$

Il grafico di questa funzione è più o meno il seguente:



Questa funzione è discontinua in tutti i punti diversi da 0, mentre è continua in 0 (si usi il *Teorema dei carabinieri*: abbiamo $0 \leq f(x) \leq x^2$ e per $x \to 0$... In 0, unico punto di continuità, f è anche derivabile, in quanto

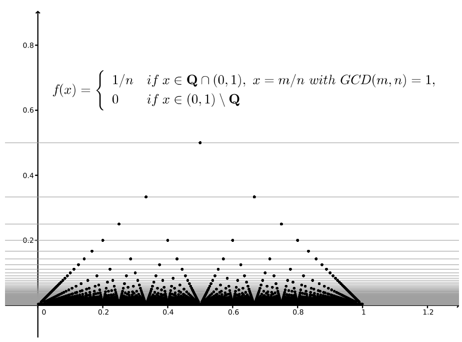
$$\frac{f(h) - f(0)}{h} = \begin{cases} h & se\ x \in \mathbf{Q}, \\ 0 & se\ x \in \mathbf{R} \setminus \mathbf{Q} \end{cases}$$

e il limite di questo rapporto incrementale per $h \to 0$ è 0: $f'(0) = 0$... Questo mostra che la derivabilità implica, è vero, la continuità... ma solo nel punto, non certo in un intorno!

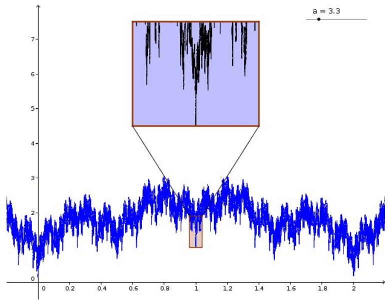
ESEMPIO: Si consideri la seguente funzione totale $f : (0, 1) \to \mathbf{R}$,

$$f(x) = \begin{cases} \frac{1}{q} & se\ x = \frac{p}{q}\ con\ p, q \in \mathbf{N}\ e\ \mathrm{MCD}(p, q) = 1, \\ 0 & se\ x \in (0, 1) \setminus \mathbf{Q}. \end{cases}$$

Questa strana funzione è discontinua in tutti i punti razionali, mentre è continua in tutti i punti irrazionali: il primo fatto è facile, il secondo un po' meno evidente... ma in classe ho cercato di spiegarvi perché è così a partire dal seguente grafico:



ESEMPIO (Weierstrass): Esistono funzioni che sono ovunque continue ma non sono derivabili in nessun punto: ecco un esempio (una variante di quello dovuto a Weierstrass)



Questa funzione è data da

$$f(x) = \sum_{k=0}^{\infty} \left(\frac{3}{4}\right)^k \phi(4^k x)$$

ove $\phi(x)$ è una funzione periodica di periodo 2 che su $[-1, 1]$ coincide con la funzione $|x|$ (un'"onda triangolare").

Analizziamo ora il legame che c'è tra la *monotonia* (crescenza/decrescenza) di una funzione derivabile ed il *segno della derivata*. Che un legame vi sia è intuitivamente evidente: in fondo, la derivata è la "pendenza" del grafico della funzione... per cui "derivata positiva" significa moralmente "grafico in salita"!

Osserviamo preliminarmente che *se una funzione (totale) $f : A \to \mathbf{R}$ è ovunque derivabile e crescente in A, allora $f'(x) \geq 0$ per ogni $x \in A$.* Infatti, il rapporto incrementale è sempre ≥ 0 grazie alla crescenza, e lo stesso rimane vero per il suo limite che è la derivata!

Il viceversa, però, non è sempre vero: ad esempio, la funzione $\tan x$ è ovunque derivabile nel suo insieme di definizione, ha derivata ovunque positiva ma *non* è globalmente crescente!

L'enunciato è però vero *su un intervallo*: se $I \subset \mathbf{R}$ è un intervallo (anche una semiretta o l'intera retta reale) e $f : I \to \mathbf{R}$ è una funzione totale e derivabile tale che $f'(x) \geq 0$ per ogni $x \in I$, allora f è crescente sull'intervallo I.

Questa è la più importante delle conseguenze del teorema di Lagrange: è proprio a causa di essa che lo si considera uno dei *più importanti teoremi* sul calcolo differenziale! Ci assicura infatti che *sugli intervalli* "avere il grafico ovunque in salita" (derivata ≥ 0) equivale ad "essere crescente".

Alcuni testi scolastici danno una definizione a mio avviso non tanto sensata: dicono che una funzione con derivata positiva in un punto è *ivi crescente*. La crescenza è un concetto globale, non puntuale... e la positività della derivata in un punto non implica la crescenza nemmeno in un suo intorno.

Un esempio di ciò lo possiamo ottenere modificando uno di quelli visti in precedenza: si consideri la funzione

$$f(x) = \begin{cases} x + x^2 & \textit{se } x \in \mathbf{Q} \\ x & \textit{se } x \in \mathbf{R} \setminus \mathbf{Q} \end{cases}$$

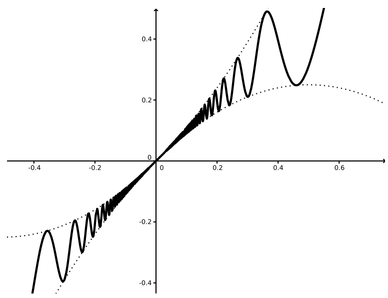
Grazie alla densità di razionali e di irrazionali in $\mathbf{R}$, si vede subito che questa funzione è discontinua per ogni $x \neq 0$. D'altra parte, essa è derivabile in 0 con derivata 1 (si usi la definizione di derivata).

Questa funzione ha ovviamente la proprietà di avere *derivata strettamente positiva in* 0, ma di *non essere crescente in alcun intorno di* 0.

Esempi di questo tipo non dipendono affatto dalla scelta di una funzione patologica come quella appena vista: ad esempio, potremmo scegliere

$$f(x) = \begin{cases} x + x^2 \sin(1/x^2) & \textit{se } x \neq 0 \\ 0 & \textit{se } x = 0 \end{cases}$$

Calcolando esplicitamente la derivata per $x \neq 0$, abbiamo verificato che questa cambia di segno infinite volte in un intorno di 0: non esiste alcun intorno di 0 in cui la funzione è crescente o decrescente. D'altra parte, un calcolo diretto usando la definizione di derivata mostra che f è derivabile in 0 con $f'(0) = 1$.



Questi due esempi ci mostrano tutta la potenza del teorema di Lagrange e della sua conseguenza!

Per concludere si noti che *se sapessimo che la derivata è continua nel punto*, un tale esempio non sarebbe possibile: la derivata dovrebbe essere positiva anche in un intorno ed il teorema di Lagrange ci permetterebbe di dedurre la crescenza almeno in un piccolo intervallo!

Passiamo ora a parlare di "punti di massimo e minimo relativo" ed "annullamento della derivata". Sappiamo che se una funzione, *definita in un intorno di* x_0, è derivabile in quel punto e vi assume massimo o minimo relativo, allora $f'(x_0) = 0$.

Un punto in cui la derivata si annulla è noto come *punto critico* o *punto stazionario*. Ovviamente, non è detto che un tale punto sia di massimo o minimo relativo...

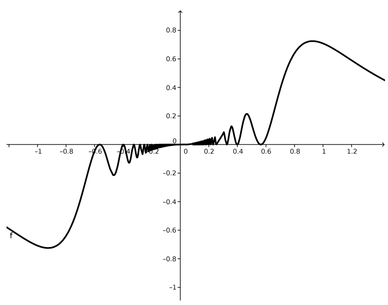
Vi è una diffusa convinzione secondo cui un punto critico è *di massimo relativo oppure di minimo relativo oppure di flesso con tangente orizzontale*: voglio provare a... sfatare questa leggenda!

Per farlo, occorrerebbe messersi d'accordo su cosa sia un flesso! Purtroppo, vi sono in commercio almeno 3 definizioni non equivalenti di punto di flesso. Sia f una funzione definita in un intorno di x_0 e derivabile in tale intorno. Allora x_0 si dice punto di flesso se

- *Prima versione:* esiste un intorno sinistro di x_0 in cui la funzione è concava ed un intorno destro in cui è convessa, o viceversa;
- *Seconda versione:* esistono un intorno sinistro ed un intorno destro di x_0 in cui il grafico della funzione giace da parti opposte rispetto alla retta tangente in x_0;
- *Terza versione:* non vi è nessun intorno (completo) di x_0 in cui la funzione sia concava o convessa;

È facile vedere che la prima definizione di flesso implica la seconda. Il viceversa però non è vero: si consideri infatti la funzione

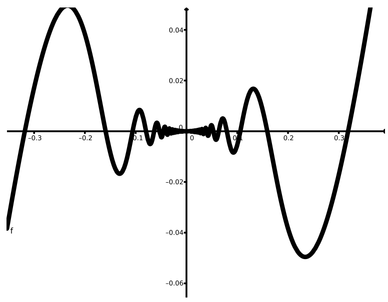
$$f(x) = \begin{cases} x^2 \sin^2(1/x^2) & se\ x > 0, \\ -x^2 \sin^2(1/x^2) & se\ x < 0, \\ 0 & se\ x = 0 \end{cases}$$



La seconda definizione non implica la terza: ad esempio, per la funzione $f(x) = x$ *tutti i punti* sono di flesso per la prima e per la seconda definizione ma non per la terza (una funzione affine è sia concava che convessa!).

Infine, la terza definizione non implica né la prima né la seconda: si consideri ad esempio la funzione

$$f(x) = \begin{cases} x^2 \sin(1/x) & se\ x \neq 0, \\ 0 & se\ x = 0 \end{cases}$$



Per le due funzioni oscillanti appena viste, 0 è un punto critico che non è né di massimo né di minimo relativo. Se però sia o meno un punto di flesso, come abbiamo visto è questione alquanto opinabile!

Un esempio simile consente anche di evidenziare tutto l'orrore di una definizione che ho trovato su un diffuso testo scolastico:
Una funzione $f : \mathbf{R} \to \mathbf{R}$*, derivabile due volte, si dice convessa in* x_0 *se* $f''(x_0) > 0$.

La convessità è un concetto *globale*: non si predica mai in un punto, ma in un *intervallo* non banale! Per capire perché, si consideri la funzione

$$f(x) = \begin{cases} x^2 + 5x^4 \sin(1/x) & \textit{se } x \neq 0, \\ 0 & \textit{se } x = 0 \end{cases}$$

Facendo un conticino, si verifica con relativa facilità che questa funzione è derivabile due volte, che $f''(0) = 2$, e che *in ogni intorno completo dell'origine vi sono infiniti intervalli di convessità e infiniti intervalli di concavità.*