Diario del Corso di Analisi - III Unità Didattica

Corsi di Laurea: Matematica, Fisica, Fisica Applicata
Docente: Sisto Baldo

ATTENZIONE: Il presente Diario del Corso vuole essere un riassunto abbastanza dettagliato di quello che è stato detto in aula, 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. D'altra parte, queste brevi paginette NON possono sostituire completamente un libro di testo, la lezione in aula o un'interazione diretta con il docente o l'esercitatrice: siete quindi invitati a servirvi ANCHE di queste altre opportunità per approfondire le vostre conoscenze!

Lezione del 16/4/2007 (2 ore): La nostra definizione di integrale di Riemann è abbastanza soddisfacente, ma nella pratica può essere utile definire l'integrale di una funzione su intervalli che siano *illimitati.* Precisamente, proponiamo in modo abbastanza naturale la seguente definizione:

DEFINIZIONE: Sia $f : [a, +\infty) \to \mathbf{R}$ una funzione continua. L'integrale (improprio o generalizzato) di f sulla semiretta $[a, +\infty)$ si definisce come

$$\int_a^{+\infty} f(x)\, dx = \lim_{M\to+\infty} \int_a^M f(x)\, dx$$

a patto che il limite esista.

A titolo di esempio, abbiamo verificato che $\int_1^{+\infty} \frac{1}{x^2}\, dx = 1$, mentre $\int_1^{+\infty} \frac{1}{x}\, dx = +\infty$ e infine $\int_0^{+\infty} \sin x\, dx$ non esiste.

In modo analogo possiamo definire l'integrale (generalizzato) di una funzione continua su un intervallo *aperto in uno dei suoi estremi*: se per esempio $f : (a, b] \to \mathbf{R}$ è continua, definiamo

$$\int_a^b f(x)\, dx = \lim_{\delta\to a^+} \int_\delta^b f(x)\, dx,$$

a patto che il limite esista.

Anche in questo caso, l'integrale può essere finito, infinito o non esistere.

Per esempio, un semplice conticino mostra che l'integrale $\int_0^1 \frac{1}{x^\alpha}\, dx$ è finito se e soltanto se $\alpha < 1$, mentre $\int_1^{+\infty} \frac{1}{x^\alpha}\, dx$ è finito se e solo se $\alpha > 1$ (e per $\alpha = 1$, *entrambi* gli integrali sono infiniti). Invece, l'integrale della funzione $f(x) = \frac{1}{x^2}\sin(1/x)$ non esiste su $(0, 1]$.

A questo proposito, vale la pena di notare che se $f \geq 0$, l'integrale improprio *esiste sempre* (sia nel caso delle semirette che nel caso degli intervalli semiaperti). Infatti, in quel caso il limite nella definizione di integrale improprio è il limite di una funzione monotona, e sappiamo bene che questo esiste sempre (finito o infinito). Dunque, nel caso delle funzioni non negative, l'integrale improprio può essere finito (e in quel caso diremo che *converge*), oppure può *divergere a* $+\infty$.

Un'altra semplicissima osservazione (che deriva dalla proprietà monotonia dell'integrale) è la seguente

PROPOSIZIONE (Principio del confronto per gli integrali impropri, prima versione): Siano $f, g : [a, +\infty) \to [0, +\infty)$ *due funzioni continue non negative, e supponiamo di sapere che* $f(x) \leq g(x)$ *per ogni* $x \in [a, +\infty)$. *Allora, se l'integrale di* g *converge, converge anche l'integrale di* f. *Se invece l'integrale di* f *diverge, diverge anche l'integrale di* g.

Un analogo principio di confronto vale anche per gli integrali impropri di funzioni continue su intervalli semiaperti.

Come conseguenza abbiamo la seguente Proposizione, che ha anch'essa un'ovvia estensione agli integrali impropri di funzioni continue su intervalli semiaperti:

PROPOSIZIONE (Principio dell'equivalenza asintotica): Siano $f, g : [a, +\infty) \to [0, +\infty)$ *due funzioni continue non negative tali che* $f(x) \sim g(x)$ *per* $x \to +\infty$ *(questo significa che* $\lim_{x \to +\infty} f(x)/g(x) = 1$, *e si legge "*$f$ *è asintoticamente equivalente a* g*"). Allora gli integrali impropri*

$$\int_a^{+\infty} f(x)\,dx, \qquad \int_a^{+\infty} g(x)\,dx$$

hanno lo stesso comportamento: sono entrambi convergenti, oppure entrambi divergenti a $+\infty$.

DIM.: Per definizione di limite all'infinito, visto che $f \sim g$ esisterà $\overline{a} \geq a$ tale che

$$\frac{1}{2}g(x) \leq f(x) \leq \frac{3}{2}g(x) \quad \forall x \geq \overline{a}.$$

La tesi segue allora dal principio del confronto applicato agli integrali impropri sulla semiretta $[\overline{a}, +\infty)$. Q.E.D.

Meno ovvia è la seguente versione del principio del confronto, valida per una funzione f *di segno qualunque*:

PROPOSIZIONE (Principio del confronto per gli integrali impropri, seconda versione): Siano $f, g : [a, +\infty) \to \mathbf{R}$ due funzioni continue, tali che $|f(x)| \leq g(x)$ per ogni $x \in [a, +\infty)$. Allora, se l'integrale improprio

$$\int_a^{+\infty} g(x)\, dx$$

converge, esiste finito anche l'integrale improprio di f. In particolare, se converge l'integrale improprio di $|f|$, esiste finito anche l'integrale improprio di f.

DIM.: Consideriamo la funzione $h(x) = f(x) + g(x)$.

Dalla disuguaglianza tra $|f|$ e g segue subito che $0 \leq h(x) \leq 2g(x)$, per cui l'integrale improprio di h è convergente. Possiamo scrivere allora:

$$\int_a^{+\infty} f(x)\, dx = \lim_{M\to+\infty} \int_a^M f(x)\, dx =$$
$$\lim_{M\to+\infty} \left[\int_a^M h(x)\, dx - \int_a^M g(x)\, dx \right],$$

e l'ultimo limite esiste finito perché i due integrali impropri coinvolti sono convergenti. Q.E.D.

Un modo particolarmente espressivo di leggere la tesi del teorema appena dimostrato è il seguente: se l'integrale improprio di una funzione di segno qualunque *converge assolutamente*, allora *converge*.

Il teorema del confronto si rivela spesso utilissimo per dimostrare la convergenza (o la divergenza, nel caso di funzioni non negative) dell'integrale improprio di una funzione di cui *non si sappia calcolare esplicitamente una primitiva*. Purtroppo, però, la convergenza assoluta è una condizione sufficiente, ma non necessaria, per la convergenza di un integrale improprio:

ESEMPIO: L'integrale improprio $\int_1^{+\infty} \frac{\sin x}{x}\, dx$ esiste finito, mentre $\int_1^{+\infty} \frac{|\sin x|}{x}\, dx = +\infty$.

Infatti, integrando per parti si ha

$$\int_1^M \frac{\sin x}{x}\, dx = \left[-\frac{\cos x}{x} \right]_1^M - \int_1^M \frac{\cos x}{x^2}\, dx \to \cos 1 - \int_1^{+\infty} \frac{\cos x}{x^2}\, dx.$$

L'integrale improprio $\int_1^{+\infty} \frac{\cos x}{x^2}\, dx$ converge per il principio del confronto, perché $\frac{|\cos x|}{x^2} \leq \frac{1}{x^2}$ e abbiamo visto che l'integrale dell'ultima funzione è finito. Dunque, $\int_1^{+\infty} \frac{\sin x}{x}\, dx$ esiste finito. La prossima volta mostreremo invece che l'integrale del modulo di questa funzione è infinito.

Lezione del 19/4/2007 (2 ore): Facciamo vedere che invece $\int_1^{+\infty} \frac{|\sin x|}{x}\, dx = +\infty$.

Infatti, per ogni $n \in \mathbf{N}$ ($n \geq 2$) avremo

$$\int_\pi^{n\pi} \frac{|\sin x|}{x}\, dx \geq \sum_{k=1}^{n-1} \int_{k\pi}^{(k+1)\pi} \frac{|\sin x|}{(k+1)\pi}\, dx = \frac{2}{\pi} \sum_{k=1}^{n-1} \frac{1}{(k+1)}.$$

Passando al limite per $n \to +\infty$ abbiamo quindi trovato la seguente disuguaglianza:

$$\int_\pi^{+\infty} \frac{|\sin x|}{x} \geq \frac{2}{\pi} \sum_{k=1}^{+\infty} \frac{1}{k+1} = \frac{2}{\pi}(\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots)$$

La serie nell'ultima disuguaglianza (cioè la serie dei reciproci dei numeri naturali) si chiama *serie armonica*: mostreremo ora che essa *diverge a* $+\infty$, concludendo così la dimostrazione della divergenza del nostro integrale.

Per farlo, useremo ancora una volta il principio del confronto per gli integrali impropri! Infatti, possiamo interpretare la serie armonica come l'integrale tra 2 e $+\infty$ della funzione a scala (con infiniti scalini) $\phi(x) = \frac{1}{[x]}$[1]. Visto che $\phi(x) \geq \frac{1}{x}$, e visto che$\int_2^{+\infty} \frac{1}{x}\, dx = +\infty$, il principio del confronto ci dice che l'integrale di ϕ (che è poi la serie armonica) diverge.

La discussione appena fatta per mostrare la divergenza dell'integrale improprio di $\frac{|\sin x|}{x}$, ci suggerisce un utilissimo *criterio di convergenza per le serie.* Ricordiamo a questo proposito, la fondamentale *definizione di serie* che abbiamo già visto parlando di polinomi di Taylor: se $\{a_n\}_{n\in\mathbf{N}}$ è una successione di numeri reali, poniamo *per definizione*

$$\sum_{k=1}^{+\infty} a_k = \lim_{n\to+\infty} \sum_{k=1}^{n} a_k,$$

a patto che il limite a secondo membro esista. La successione

$$s_n = \sum_{k=1}^{n} a_k$$

si chiama *successione delle somme parziali* della serie, per cui la somma della serie è semplicemente il limite delle somme parziali!

[1]Si noti che la definizione di integrale improprio si può estendere senza cambiare nulla a funzioni anche non necessariamente continue, che siano però integrabili secondo Riemann su tutti gli intervalli limitati. Quindi, è perfettamente lecito fare l'integrale improprio di ϕ.

PROPOSIZIONE (Criterio integrale di convergenza per le serie): Sia $f : [1, +\infty) \to [0, +\infty)$ una funzione non negativa e decrescente. Allora l'integrale improprio

$$\int_1^{+\infty} f(x)\, dx$$

ha lo stesso comportamento della serie

$$\sum_{k=1}^{\infty} f(k).$$

Con questo vogliamo dire che se l'integrale converge, converge anche la serie, mentre se l'integrale diverge a $+\infty$, diverge anche la serie.

DIM.: Basta osservare che $f([x+1]) \leq f(x) \leq f([x])$, e interpretare la serie come l'integrale della funzione costante a tratti $f([x])$. Q.E.D.

ESEMPIO: La *serie armonica generalizzata*

$$\sum_{k=1}^{+\infty} \frac{1}{k^\alpha}$$

converge se e solo se $\alpha > 1$: basta usare il criterio integrale con la funzione $f(x) = 1/x^\alpha$.

Da quanto abbiamo appena visto, non ci stupirà scoprire che esistono delle analogie tra integrali impropri e serie! Cominciamo infatti con un elenco di alcune proprietà delle serie che ricordano quanto abbiamo già visto per gli integrali:

- Se $a_n \geq 0$, la serie $\sum\limits_{k=1}^{+\infty} a_k$ *esiste sempre*, finita o infinita: infatti, le somme parziali costituiscono una successione monotona. In sostanza *una serie a termini positivi converge a una somma finita, oppure diverge a $+\infty$.*

- Vale il seguente *criterio del confronto per le serie a termini positivi*: se $0 \leq a_n \leq b_n$, allora la convergenza di $\sum\limits_{k=1}^{+\infty} b_k$ implica la convergenza di $\sum\limits_{k=1}^{+\infty} a_k$, mentre la divergenza di $\sum\limits_{k=1}^{+\infty} a_k$ implica la divergenza di $\sum\limits_{k=1}^{+\infty} b_k$.

- Vale anche un *criterio dell'equivalenza asintotica*: se $a_n, b_n > 0$ e

$$\lim_{n \to +\infty} \frac{a_n}{b_n} = 1,$$

allora le serie $\sum_{k=1}^{+\infty} a_k$ e $\sum_{k=1}^{+\infty} b_k$ hanno lo stesso comportamento. Basta infatti osservare che per n abbastanza grande $\frac{1}{2}b_n \leq a_n \leq \frac{3}{2}b_n$, e applicare il criterio del confronto.[2]

- Se una serie è *assolutamente convergente*, cioè se converge $\sum_{k=1}^{+\infty} |a_k|$, allora anche la serie $\sum_{k=1}^{+\infty} a_k$ esiste finita. La dimostrazione è analoga a quella del criterio del confronto per gli integrali (seconda versione): si considera la serie di termine generale $b_n = a_n + |a_n|$ e si nota che $0 \leq b_n \leq 2|a_n|$. Ne deriva che $\sum_{k=1}^{+\infty} b_k$ converge (principio del confronto), e la convergenza di $\sum_{k=1}^{+\infty} a_n$ si ricava subito scrivendo $a_n = b_n - |a_n|$.[3]

Un'altra proprietà interessante (che per gli integrali impropri *non* vale) è la seguente:

PROPOSIZIONE: Se la serie (a termini di segno qualunque) $\sum_{k=0}^{+\infty} a_k$ *converge, allora*

$$\lim_{k \to +\infty} a_k = 0$$

.

DIM.: Se s_n denota la somma parziale n-esima della serie e s la sua somma (cioè $s_n \to s$), si ha $a_n = s_n - s_{n-1}$, e passando al limite per $n \to +\infty$ si ha la tesi. Q.E.D.

Si noti che il viceversa *non è vero*: il termine generale della serie può essere infinitesimo senza che la serie converga. Come esempio, abbiamo visto il caso della serie armonica: $\sum_{k=1}^{+\infty} \frac{1}{k} = +\infty$.

ESEMPIO FONDAMENTALE (Serie geometrica): Utilizziamo ora la formula per la somma della progressione geometrica per studiare la *serie geometrica*

$$\sum_{k=0}^{+\infty} a^k,$$

[2]Questo criterio è falso per le serie a termini di segno qualunque. Infatti, le serie di termine generale $a_n = (-1)^n \frac{1}{\sqrt{n}} + \frac{1}{n}$ e $b_n = (-1)^n \frac{1}{\sqrt{n}}$ sono asintoticamente equivalenti. Grazie a un criterio di convergenza che vedremo la prossima volta (criterio di Leibniz), si verifica che la prima serie diverge a $+\infty$, mentre la seconda converge.

[3]Anche in questo caso, non vale il viceversa: vedremo che una serie può convergere anche se non converge assolutamente.

in cui a è un numero reale chiamato *ragione* della serie.

Se $s_n = \sum_{k=0}^{n} a^k$ denota la somma parziale n-esima della serie, abbiamo visto che $s_n = \frac{1-a^{n+1}}{1-a}$. Se studiamo il limite di questa espressione per $n \to +\infty$, vediamo che la serie converge a $1/(1-a)$ per $|a| < 1$, diverge a $+\infty$ per $a \geq 1$, non esiste per $a < -1$.

Lezione del 23/4/2007 (2 ore): Usando il nostro studio della serie geometrica e il criterio del confronto, otteniamo i seguenti semplici risultati.

PROPOSIZIONE (Criteri della radice e del rapporto): Sia a_n il termine generale di una serie, $\sum_{k=1}^{+\infty} a_k$.

- *Supponiamo che esista il limite*
$$\lim_{n\to+\infty} \sqrt[n]{|a_n|} = \ell_1.$$
Allora, se $\ell_1 > 1$ la serie non converge, se $\ell_1 < 1$ converge.

- *Sia $a_n \neq 0$, e supponiamo che esista il limite*
$$\lim_{n\to+\infty} \left|\frac{a_{n+1}}{a_n}\right| = \ell_2.$$
Allora, se $\ell_2 > 1$ la serie non converge, se $\ell_2 < 1$ converge[4].

DIM.: Cominciamo dal criterio della radice. Se $\ell_1 > 1$, scegliamo $\varepsilon > 0$ in modo che $\ell_1 - \varepsilon > 1$. Per la definizione di limite, sappiamo che per n abbastanza grande avremo $\sqrt[n]{|a_n|} > \ell_1 - \varepsilon$, ossia $|a_n| > (\ell_1 - \varepsilon)^n$. Passando al limite per $n \to +\infty$ otteniamo $|a_n| \to +\infty$, e la serie non può convergere perché il suo termine generale non tende a zero.

Se $\ell_1 < 1$, scegliamo $\varepsilon > 0$ in modo che $\ell_1 + \varepsilon < 1$. Per n abbastanza grande avremo $\sqrt[n]{|a_n|} < \ell_1 + \varepsilon$, da cui $|a_n| < (\ell_1 + \varepsilon)^n$. Poiché la serie geometrica di ragione $\ell_1 + \varepsilon$ converge, converge *assolutamente* anche la nostra serie (criterio del confronto), e quindi essa converge.

Dimostriamo il *criterio del rapporto*: supponiamo $\ell_2 > 1$, e scegliamo ε tanto piccolo che $\ell_2 - \varepsilon > 1$. Per definizione di limite, troviamo $\nu \in \mathbf{N}$ tale che $\left|\frac{a_{n+1}}{a_n}\right| > \ell_2 - \varepsilon$ per $n \geq \nu$. Allora, se $n > \nu$:
$$|a_n| = |a_\nu| \cdot \left|\frac{a_{\nu+1}}{a_\nu} \cdot \ldots \cdot \frac{a_{n-1}}{a_{n-2}} \cdot \frac{a_n}{a_{n-1}}\right| \geq |a_\nu| \cdot (\ell_2 - \varepsilon)^{n-\nu}.$$

[4]Si potrebbe far vedere che se esiste il limite del rapporto ℓ_2, allora esiste anche il limite della radice ℓ_1, e questi due limiti sono uguali: per questo, il criterio del rapporto è in realtà una conseguenza del criterio della radice.

Passando al limite vediamo che $|a_n| \to +\infty$, e la serie non converge di sicuro.

Se poi $\ell_2 < 1$, scegliamo $\varepsilon > 0$ in modo che $\ell_2 + \varepsilon < 1$. Troviamo $\nu \in \mathbf{N}$ tale che $\left|\frac{a_{n+1}}{a_n}\right| < \ell_2 + \varepsilon$ per $n \geq \nu$. Rifacendo il conto di prima abbiamo

$$|a_n| = |a_\nu| \cdot \left|\frac{a_{\nu+1}}{a_\nu} \cdot \ldots \cdot \frac{a_n}{a_{n-1}}\right| \leq |a_\nu| \cdot (\ell_2 + \varepsilon)^{n-\nu},$$

e la serie dei moduli risulta maggiorata da una serie geometrica convergente.[5]. Q.E.D.

Che altro dire delle serie a termini di segno qualunque? Se esse non convergono assolutamente, abbiamo ben pochi strumenti a nostra disposizione. Uno di questi è il seguente:

PROPOSIZIONE (Criterio di Leibniz): Sia $\{a_n\}$ una successione di numeri non negativi, e si consideri la serie a termini di segno alterno

$$\sum_{n=0}^{+\infty} (-1)^n a_n.$$

Se a_n è decrescente e tende a zero, allora la serie converge.

DIM.: Al solito, sia $s_n = \sum_{k=0}^{n} (-1)^n a_n$ la successione delle somme parziali. Si ha

$$s_{2n+2} = s_{2n} - a_{2n+1} + a_{2n+2} \leq s_{2n},$$

cioè la successione delle somme parziali *di indice pari* è decrescente.

Analogamente, $s_{2n+3} \geq s_{2n+1}$: la successione delle somme parziali di indice dispari è crescente. Inoltre, evidentemente $s_1 \geq 0$ e $s_{2n} \geq s_{2n-1}$: ne segue che la successione delle somme parziali pari è non negativa, e tenderà a un limite finito ℓ (uguale al suo inf). Anche le somme parziali dispari tenderanno allo stesso limite:

$$\lim_{k\to+\infty} s_{2k+1} = \lim_{k\to+\infty} (s_{2k} - a_{2k+1}) = \ell - 0.$$

[5]Mostriamo che se $|a_{n+1}/a_n| \to \ell$, allora si ha anche $\sqrt[n]{|a_n|} \to \ell$: questo fornisce evidentemente una dimostrazione alternativa del criterio del rapporto. Usando la definizione di limite e il ragionamento appena fatto, scopriamo che per qualunque $\varepsilon > 0$ esiste $\nu > 0$ tale che

$$|a_\nu|(\ell - \varepsilon)^{n-\nu} \leq |a_n| \leq |a_\nu|(\ell + \varepsilon)^{n-\nu}$$

da cui

$$\sqrt[n]{|a_\nu|}(\ell - \varepsilon)^{1-\nu/n} \leq \sqrt[n]{|a_n|} \leq \sqrt[n]{|a_\nu|}(\ell + \varepsilon)^{1-\nu/n}$$

. Il membro di sinistra e quello di destra tendono a $\ell - \varepsilon$ e a $\ell + \varepsilon$ rispettivamente, quindi per n abbastanza grande si avrà $\ell - 2\varepsilon \leq \sqrt[n]{|a_n|} \leq \ell + 2\varepsilon$. Dall'arbitrarieà di ε segue che $\sqrt[n]{|a_n|} \to \ell$.

Ne segue che ℓ è proprio la somma della serie. Q.E.D.

Il criterio di Leibniz ci dice ad esempio che la versione a segni alterni della serie armonica, $\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n}$, è convergente (in realtà, si può far vedere che converge a $\log 2$).

Si ricorderà che abbiamo incontrato per la prima volta le serie quando ci siamo accorti che alcune funzioni infinitamente derivabili possono essere sviluppate in *serie di Taylor*: la serie che si ottiene in quel caso è del tipo

$$\sum_{n=0}^{+\infty} a_n x^n,$$

e si chiama *serie di potenze.*

Abbiamo visto che se $f(x)$ è una funzione analitica e scegliamo $a_n = \frac{f^{(n)}(0)}{n!}$, allora la serie di potenze converge in un intorno di 0, e converge proprio a $f(x)$... a dire il vero, questa era proprio la definizione di funzione analitica!

È però interessante anche studiare il problema inverso: se ci viene data una serie di potenze $\sum_{n=0}^{+\infty} a_n x^n$, cosa possiamo dire del suo insieme di convergenza?

E se scopriamo che essa converge per gli x in un opportuno intorno di 0, sarà poi vero che la sua somma $f(x)$ è una funzione infinitamente derivabile, la cui serie di Taylor coincide con la serie di partenza?

Volendo rispondere a queste domande, cominciamo con un'osservazione semplice ma interessante:

LEMMA: Se la serie di potenze $\sum_{n=0}^{+\infty} a_n x^n$ converge per $x = x_0$, allora converge assolutamente per tutti gli x con $|x| < |x_0|$.

DIM.: Siccome la serie converge per $x = x_0$, abbiamo necessariamente $\lim_{n\to+\infty} a_n x_0^n = 0$. In particolare, per n abbastanza grande avremo $|a_n x_0^n| \leq 1$.

Se poi $|x| < |x_0|$ si ha (sempre per n abbastanza grande)

$$|a_n x^n| = |a_n x_0^n| \cdot \left|\frac{x}{x_0}\right|^n \leq \left|\frac{x}{x_0}\right|^n,$$

e la serie $\sum_{n=0}^{+\infty} |a_n x^n|$ converge perché è maggiorata da una serie geometrica convergente. Q.E.D.

Il lemma ci suggerisce di dare la seguente, fondamentale

DEFINIZIONE: Il *raggio di convergenza* della serie di potenze $\sum_{n=0}^{+\infty} a_n x^n$ è l'estremo superiore dei valori di x per cui la serie converge.

Sia r il raggio di convergenza della nostra serie di potenze. Grazie al lemma visto sopra, possiamo concludere che

- Se $r = +\infty$, la serie converge assolutamente per ogni $x \in \mathbf{R}$.
- Se $r > 0$, la serie converge assolutamente nell'intervallo aperto $(-r, r)$,mentre non converge per $|x| > r$.
- Se $r = 0$, la serie converge soltanto per $x = 0$.

Tutti e tre questi comportamenti sono possibili: la serie esponenziale $\sum_{n=0}^{+\infty} \frac{x^n}{n!}$ converge per ogni $x \in \mathbf{R}$, la serie geometrica $\sum_{n=0}^{+\infty} x^n$ ha raggio di convergenza 1, mentre la serie $\sum_{n=0}^{+\infty} n^n x^n$ ha raggio di convergenza 0 come si può facilmente verificare con il criterio della radice.

Si noti anche che il nostro lemma non dice nulla sul comportamento della serie per $x = \pm r$, cioè agli estremi dell'intervallo di convergenza: in effetti, in quei due punti può succedere qualunque cosa (la serie può convergere in tutti e due i punti, in uno solo di essi, oppure in nessuno dei due).

La convergenza di una serie di potenze agli estremi dell'intervallo di convergenza è spesso la cosa più difficile da valutare, e lo studio deve essere condotto caso per caso.

Vediamo ora come è possibile calcolare il raggio di convergenza di una serie di potenze: in *quasi tutti i casi* è possibile dare una risposta grazie al criterio della radice o del rapporto.

Supponiamo infatti di sapere che

$$\lim_{n\to+\infty} \sqrt[n]{|a_n|} = \ell$$

(oppure che $\lim_{n\to+\infty} \frac{|a_{n+1}|}{|a_n|} = \ell$). Usando il criterio della radice (del rapporto) vediamo subito che la serie converge per $|x| < \frac{1}{\ell}$, mentre non converge per $|x| > \frac{1}{\ell}$ (nei casi $\ell = 0$ e $\ell = +\infty$, il raggio di convergenza è rispettivamente $+\infty$ e 0...).

L'unica situazione in cui questo metodo non funziona, è quella in cui non esiste il limite della radice n-esima di $|a_n|$ (anche in questo caso, tuttavia, è possibile aggirare il problema: se volete sapere come fare, leggete la nota[6] qui sotto... è piuttosto complicato, non dite che non vi avevo avvertito!).

[6] *La semplice osservazione appena fatta può essere trasformata in una* ricetta

Lezione del 26/4/2007 (2 ore): Supponiamo ora di avere una serie di potenze con raggio di convergenza $r > 0$. Possiamo allora definire una funzione $f : (-r, r) \to \mathbf{R}$ nel modo seguente:

$$f(x) = \sum_{n=0}^{+\infty} a_n x^n.$$

Ci chiediamo quali siano le proprietà della funzione $f(x)$.

Osserviamo che essa si ottiene come limite di polinomi: $f(x)$ è il limite delle somme parziali $s_n(x) = \sum_{k=0}^{n} a_k x^k$ della serie.

D'altra parte, può accadere benissimo che una successione di funzioni molto regolari (derivabili infinite volte) converga ad una funzione discontinua: per esempio, la successione di funzioni $s_n(x) = \arctan(nx)$ ha per limite la funzione discontinua $\pi/2 \operatorname{sgn}(x)$.[7] Tuttavia, questa eventualità poco desiderabile non si verifica per le serie di potenze. Vale infatti il seguente

TEOREMA (Regolarità delle serie di potenze): Sia $f(x) = \sum_{n=0}^{+\infty} a_n x^n$ la somma di una serie di potenze con raggio di convergenza $r > 0$. Allora $f(x)$ è continua e derivabile in $(-r, r)$. Inoltre la serie delle derivate $\sum_{n=1} a_n \cdot nx^{n-1}$ ha ancora raggio di convergenza r, e converge in $(-r, r)$ proprio alla derivata f' di f.

---

universale *per trovare il raggio di convergenza: se è vero che non sempre esiste il limite della successione $\sqrt[n]{|a_n|}$, è però sempre possibile farne il* massimo limite.

Il massimo limite di una successione $\{b_n\}$ è una schifezza che si definisce come

$$\max \lim_{n\to+\infty} b_n = \inf_{n\in\mathbf{N}} \sup\{b_m : \ m \geq n\} = \lim_{n\to+\infty} \sup\{b_m : \ m \geq n\},$$

e analogamente si definisce il minimo limite

$$\min \lim_{n\to+\infty} b_n = \sup_{n\in\mathbf{N}} \inf\{b_m : \ m \geq n\} = \lim_{n\to+\infty} \inf\{b_m : \ m \geq n\}.$$

Non è troppo difficile far vedere che la successione ammette limite se e solo se il suo massimo limite e il suo minimo limite sono uguali. Il massimo e il minimo limite possono essere caratterizzati come il più grande e il più piccolo tra i limiti di tutte le sottosuccessioni di $\{b_n\}$ che ammettono limite.

Ora, il criterio della radice vale pari pari (e anche la dimostrazione non cambia granché) se si sostituisce il limite con il massimo limite: possiamo quindi affermare che il raggio di convergenza della nostra serie di potenze è dato dal reciproco di $\ell = \max \lim_{n\to+\infty} \sqrt[n]{|a_n|}$.

[7] Si noti che $f(x)$ può essere vista come la somma della serie di funzioni di termine generale $a_n(x) = s_n(x) - s_{n-1}(x)$.

Il teorema dice dunque che una serie di potenze si può derivare termine a termine. Inoltre, iterando il procedimento si ottiene che $f(x)$ è derivabile infinite volte, e che la serie di partenza non è altro che la serie di Taylor di f centrata in 0.

La dimostrazione del teorema sulla regolarità della somma delle serie di potenze è piuttosto complicata...ma siccome siete degli studenti di matematica ho deciso di non risparmiarvela! La vedremo dopo qualche esempio.

ESEMPIO: Come applicazione del teorema sulla somma delle serie di potenze, verifichiamo che

$$\log(1+x) = \sum_{n=1}^{+\infty} (-1)^{n+1} \frac{x^n}{n} \qquad 1 < x < 1;$$
$$\arctan(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{2n+1} \qquad 1 < x < 1.$$

Infatti, non è difficile vedere che entrambe le serie hanno raggio di convergenza 1. Se chiamiamo $f(x)$ la somma della prima e $g(x)$ la somma della seconda, derivando termine a termine si ottiene

$$f'(x) = \sum_{n=1}^{+\infty} (-1)^{n+1} x^{n-1} = \sum_{k=0}^{+\infty} (-x)^k = \frac{1}{1+x},$$
$$g'(x) = \sum_{n=0}^{+\infty} (-1)^n x^{2n} = \sum_{n=0}^{+\infty} (-x^2)^n = \frac{1}{1+x^2},$$

dove abbiamo usato la formula per la somma della serie geometrica. Integrando e tenendo conto del fatto che $f(0) = g(0) = 0$, si ottiene $f(x) = \log(1+x)$ e $g(x) = \arctan x$.

ESEMPIO: Consideriamo la *serie binomiale*

$$\sum_{n=0}^{+\infty} \binom{\alpha}{n} x^n,$$

in cui $\alpha \in \mathbf{R}$ e i coefficienti binomiali sono definiti da

$$\binom{\alpha}{0} = 1, \quad \binom{\alpha}{n} = \frac{\alpha \cdot (\alpha-1) \cdot (\alpha-2) \cdot \ldots \cdot (\alpha-n+1)}{n!} \quad (n \geq 1).$$

Usando il criterio del rapporto, si verifica subito che il questa serie ha raggio di convergenza 1[8]. Mostriamo che la somma $f(x)$ della serie è uguale a $(1+x)^\alpha$ per ogni $x \in (-1, 1)$.

[8] A meno che non si abbia $\alpha \in \mathbf{N}$: in tal caso solo i primi α coefficienti binomiali sono diversi da 0, e la serie si riduce a un polinomio.

Derivando termine a termine la serie si ottiene

$$f'(x) = \sum_{n=0}^{+\infty} \binom{\alpha}{n} n x^{n-1} = \sum_{n=0}^{+\infty} \binom{\alpha}{n} (\alpha - n) x^n$$

(la seconda espressione si ottiene osservando che $\binom{\alpha}{n} n = \binom{\alpha}{n-1} (\alpha - n + 1)$ e cambiando l'indice, $(n-1) \leftrightarrow n$). Utilizzando queste due scritture equivalenti di $f'(x)$ si ottiene subito

$$(1+x) f'(x) = \alpha f(x), \text{ ovvero } \frac{f'(x)}{f(x)} = \frac{\alpha}{1+x}.$$

Integrando, si ha allora $\log f(x) = \log(1+x)^\alpha + C$, da cui (osservando che $f(0) = 1$) $f(x) = (1+x)^\alpha$.

DIM. della regolarità della somma di una serie di potenze.

Per poter dimostrare il teorema di regolarità, abbiamo bisogno di alcuni lemmi. Il primo di essi riguarda la velocità con cui il resto N-esimo di una serie di potenze va a zero.

Se prendiamo una serie numerica convergente $\sum_{n=1}^{\infty} a_n$, il resto N-esimo è per definizione $R_N = \sum_{n=N}^{\infty} a_n$: è immediato verificare che $\lim_{N \to +\infty} R_N = 0$.

Se poi consideriamo il caso della nostra serie di potenze $\sum_{n=0}^{\infty} a_n x^n$ di raggio di convergenza $r > 0$, il resto N-esimo $R_N(x) = \sum_{n=N}^{\infty} a_n x^n$ tenderà a zero per ogni $x \in (-r, r)$. Il primo lemma dice che se $0 < \rho < r$ e $x \in [-\rho, \rho]$, allora possiamo stimare la velocità con cui $R_N(x)$ tende a zero in maniera *indipendente da* x: in matematichese, si dice che la serie di potenze converge *uniformemente* in $[-\rho, \rho]$.

LEMMA 1: Sia $\sum_{n=0}^{\infty} a_n x^n$ *una serie di potenze con raggio di convergenza* $r > 0$, $\rho \in (0, r)$. *Allora, per ogni* $x \in [-\rho, \rho]$ *si ha*

$$|R_N(x)| \leq \sum_{n=N}^{\infty} |a_n| \rho^n.$$

In altre parole, il resto N*-esimo della serie è maggiorato, per ogni* $x \in [-\rho, \rho]$, *dal resto* N*-esimo della serie numerica convergente (la serie di potenze converge assolutamente per* $x = \rho$*...)* $\sum_{n=N}^{\infty} |a_n| \rho^n$.

Dim.: Sia $x \in [-\rho, \rho]$, $M \geq N$. Usando la disuguaglianza triangolare otteniamo:

$$|\sum_{n=N}^{M} a_n x^n| \leq \sum_{n=N}^{M} |a_n||x|^n \leq \sum_{n=N}^{M} |a_n|\rho^n,$$

e passando al limite per $M \to +\infty$ si ha la tesi. Q.E.D.

Il prossimo lemma constituisce la prima parte del teorema di regolarità che vogliamo dimostrare: la somma di una serie di potenze è una funzione continua. Questo risultato è meno ovvio di quanto possa sembrare: se abbiamo una serie di funzioni continue, è in generale falso che la somma della serie sia continua:

ESEMPIO: Poniamo $g_n(x) = \arctan(nx)$, $f_n(x) = g_n(x) - g_n - 1(x)$. Allora la somma parziale N-esima della serie di funzioni continue

$$\sum_{n=1}^{\infty} f_n(x)$$

è $g_N(x)$, che tende chiaramente ad una funzione discontinua per $N \to +\infty$.

LEMMA 2: Sia $f(x) = \sum_{n=0}^{\infty} a_n x^n$ la somma di una serie di potenze con raggio di convergenza $r > 0$. Allora la funzione f è continua in $(-r, r)$.

Dim.: Sia $x_0 \in (-r, r)$: mostriamo che f è continua in x_0. A questo scopo, scegliamo $\rho > 0$ con $|x_0| < \rho < r$. Se $|x| \leq \rho$ e N è un qualunque numero naturale possiamo scrivere:

$$\begin{aligned}
|f(x) - f(x_0)| \quad = \quad & |\sum_{n=0}^{\infty} a_n(x^n - x_0^n)| \leq \\
& |\sum_{n=0}^{N} a_n(x^n - x_0^n)| + |\sum_{n=N}^{\infty} a_n x^n - \sum_{n=N}^{\infty} a_n x_0^n| \leq \\
& |\sum_{n=0}^{N} a_n(x^n - x_0^n)| + |R_N(x)| + |R_N(x_0)| \leq \\
& |\sum_{n=0}^{N} a_n(x^n - x_0^n)| + 2\sum_{n=N}^{\infty} |a_n|\rho^n,
\end{aligned}$$

dove nell'ultimo passaggio abbiamo fatto uso del Lemma 1. Fissiamo $\varepsilon > 0$ e osserviamo l'ultima riga della nostra stima: il secondo addendo può essere reso minore di $\varepsilon/2$ a patto di prendere N abbastanza grande (si tratta del resto N-esimo di una serie numerica convergente). A questo punto, prendendo x abbastanza vicino a x_0 il primo addendo può essere reso minore di $\varepsilon/2$

(perché il polinomio $\sum\limits_{n=0}^{N} a_n x^n$ è una funzione continua): in conclusione, se x è abbastanza vicino a x_0 si ha $|f(x) - f(x_0)| < \varepsilon$, Q.E.D.

Il terzo lemma dice che possiamo integrare termine a termine una serie di potenze sugli intervalli chiusi contenuti in $(-r, r)$.

LEMMA 3: Sia $f(x) = \sum\limits_{n=0}^{\infty} a_n x^n$ *la somma di una serie di potenze con raggio di convergenza* $r > 0$*,* $[x_1, x_2] \subset (-r, r)$*. Denotiamo con* $S_N(x) = \sum\limits_{n=0}^{N} a_n x^n$ *le somme parziali della serie. Allora*

$$\int_{x_1}^{x_2} f(x)\, dx = \lim_{N\to+\infty} \int_{x_1}^{x_2} S_N(x)\, dx.$$

In maniera più espressiva, questo è equivalente a scrivere

$$\int_{x_1}^{x_2} \sum_{n=0}^{\infty} a_n x^n\, dx = \sum_{n=0}^{\infty} \int_{x_1}^{x_2} a_n x^n\, dx,$$

cioè si poò scambiare il segno di serie con quello di integrale[9]*).*

Dim: La funzione $f(x)$ è integrabile perché abbiamo dimostrato nel Lemma 2 che è continua.

Sia $\rho = \max\{|x_1|, |x_2|\}$: grazie al Lemma 1 abbiamo

$$\begin{gathered}
|\int_{x_1}^{x_2} (f(x) - S_N(x))\, dx| = |\int_{x_1}^{x_2} R_N(x)\, dx| \leq \\
\int_{x_1}^{x_2} |R_N(x)|\, dx \leq \int_{x_1}^{x_2} \sum_{n=N}^{\infty} |a_n|\rho^n = (x_2 - x_1) \sum_{n=N}^{\infty} |a_n|\rho^n,
\end{gathered}$$

e l'ultima espressione tende a zero per $N \to +\infty$. Q.E.D.

Abbiamo così concluso il "lavoro sporco": la prossima volta, vedremo come la regolarità della somma di una serie di potenze segua facilmente dagli ultimi due lemmi!

Lezione del 3/5/2007 (2 ore): Concludiamo la dimostrazione del teorema sulla regolarità della somma di una serie di potenze. Abbiamo visto

[9] Infatti questo è chiaramente possibile per le somme parziali (additività dell'integrale): il lemma dice che è lecito passare al limite per $N \to +\infty$ ed estendere l'affermazione alla serie.

che la funzione somma $f(x)$ è continua, rimangono da dimostrare solo le affermazioni sulla derivabilità.

Innanzitutto, le serie derivata $\sum_{n=1}^{\infty} n a_n x^{n-1}$ ha ancora raggio di convergenza r. Infatti, il reciproco del raggio di convergenza della serie derivata è dato da $\max \lim_{n\to+\infty} \sqrt[n-1]{n|a_n|} = \max \lim_{n\to+\infty} \sqrt[n]{|a_n|}$, e l'ultima espressione sappiamo che vale $1/r$.

Grazie al Lemma 2, sappiamo dunque che le somme parziali $S'_N(x)$ della serie derivata convergono ad una certa funzione continua $g(x)$ (la somma della serie derivata).

Ora, siano $S_N(x)$ le somme parziali della serie di potenze non derivata: grazie al teorema fondamentale del calcolo integrale, se fissiamo $x_0 \in (-r, r)$ si ha $S_N(x) = S_N(x_0) + \int_{x_0}^{x} S'_N(t)\ dt$. Passiamo al limite per $N \to +\infty$: il primo mebro tende a $f(x)$, mentre il secondo membro tende a $f(x_0) + \int_{x_0}^{x} g(t)\ dt$ grazie al Lemma 3.

In conclusione, abbiamo mostrato che

$$f(x) = f(x_0) + \int_{x_0}^{x} g(t)\ dt,$$

per cui f è derivabile e la sua derivata è proprio g (teorema fondamentale del calcolo integrale). Q.E.D.

Passiamo ora ad un argomento completamente diverso: le funzioni di due (o più) variabili.

Nella seconda unità didattica abbiamo anticipato il concetto di *continuità per una funzione di n variabili*: una funzione $f : \mathbf{R}^n \to \mathbf{R}$ si dice continua in $x_0 \in \mathbf{R}^n$ se per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che

$$|f(x) - f(x_0)| < \varepsilon$$

per ogni $x \in \mathbf{R}^n$ tale che $|x - x_0| < \delta$. Ricordiamo che la *norma* di un vettore $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$ si definisce come $|x| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ e che geometricamente $|x - x_0|$ non è altro che la distanza euclidea tra x e x_0.

Analogamente si definisce il limite di una funzione di n variabili:

DEFINIZIONE: Scriveremo che

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

se per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che $0 < |x - x_0| < \delta$ implica $|f(x) - \ell| < \varepsilon$.

Esattamente come in una variabile, queste definizioni si estendono in modo ovvio a funzioni definite soltanto su un sottinsieme di $\mathbf{R}^n$: in questo caso,

x_0 dovrà essere un punto di accumulazione del dominio perché la definizione di limite abbia senso!

In buona sostanza, la definizione di continuità (e quella di limite) per una funzione di due variabili è molto simile a quella per una funzione di una variabile. E' anche facile vedere che somme, prodotti, quozienti e composizioni di funzioni continue sono ancora funzioni continue (purché non si annulli il denominatore, nel caso dei quozienti...): in particolare, sono continue le funzioni elementari.

E dunque, tutto bene?

Non proprio: con le funzioni di due variabili possono accadere fatti abbastanza sorprendenti! Per esempio, potremmo pensare che se

$$\lim_{x \to x_0} f(x, y_0) = f(x_0, y_0), \quad \lim_{y \to y_0} f(x_0, y) = f(x_0, y_0),$$

allora f sia continua in (x_0, y_0). Nulla di più sbagliato: la funzione

$$(A)\ f(x) = \begin{cases} \frac{xy}{x^2+y^2} & se\ (x,y) \neq (0,0) \\ 0 & se\ (x,y) = (0,0) \end{cases}$$

ha questa proprietà ma *non è continua nell'origine* $(0,0)$, perchè $f(x,x) = 1/2$ (e quindi f *non tende a* 0 se ci si avvicina all'origine lungo la bisettrice del primo e terzo quadrante). Siamo stati troppo ottimisti!

Ma certamente, se $g(x,y)$ tende a $g(x_0,y_0)$ quando ci si avvicina a (x_0,y_0) lungo *tutte le rette passanti per* (x_0,y_0), allora g sarà continua in quel punto?

Anche questo è clamorosamente falso: si consideri la funzione

$$(B)\ g(x,y) = \begin{cases} \frac{x^2y}{x^4+y^2} & se\ (x,y) \neq (0,0) \\ 0 & se\ (x,y) = (0,0) \end{cases}$$

La restrizione di g a qualunque retta passante per l'origine tende a 0 quando ci si avvicina all'origine. In compenso, $g(x,x^2) = 1/2$, e la funzione *non* tende a 0 se ci si avvicina all'origine lungo la parabola $y = x^2$...

Questo ci dice che non è possibile testare la continuità di una funzione di due variabili limitandosi a vedere come si comporta lungo le rette, o lungo una fissata famiglia di curve: è proprio necessario ricorrere alla definizione.

Se questi problemi si presentano per la continuità, possiamo immaginare che siano ancora più macroscopici quando andiamo a studiare il calcolo differenziale per funzioni di due variabili!

In particolare, ci poniamo il seguente problema. Il grafico di una funzione di due variabili rappresenta graficamente una superficie nello spazio $\mathbf{R}^3$: sotto quali condizioni tale grafico è dotato di piano tangente in un certo punto? In altre parole, sotto quali condizioni potremo "approssimare bene" il grafico di una funzione attorno al punto (x_0, y_0) con un piano?

Quanto visto sopra per quanto riguarda la continuità ci suggerisce che probabilmente la sola esistenza delle derivate parziali non sarà sufficiente. Ricordiamo che la derivata parziale di $f : \mathbf{R}^n \to \mathbf{R}$ rispetto a x_i si definisce nel modo seguente:

$$\frac{\partial f}{\partial x_i}(x_0) = \lim_{h \to 0} \frac{f(x_0 + he_i) - f(x_0)}{h},$$

dove e_i è l'i-esimo vettore della base canonica di $\mathbf{R}^n$ (e purché il limite esista finito). Si tratta cioè della derivata (in x_0) della restrizione di f ad una retta parallela all'asse x_i e passante per x_0.

Se riprendiamo la funzione f definita nell'equazione (A) sopra, vediamo che essa possiede derivate parziali nulle nell'origine (infatti la funzione si annulla identicamente su entrambi gli assi coordinati): questo ci dice che *la sola esistenza delle derivate parziali non garantisce nemmeno la continuità della funzione*, e tantomeno sarà sufficiente all'esistenza del piano tangente!

Le cose vanno ancor peggio con la funzione g definita sopra in (B): la funzione è dotata di *derivata direzionale* in ogni direzione, ma questo non le impedisce di essere discontinua!

Una *direzione* in $\mathbf{R}^n$ è semplicemente un vettore $v \in \mathbf{R}^n$ con $|v| = 1$. La *derivata direzionale* di $f : \mathbf{R}^n \to \mathbf{R}$ rispetto a v nel punto x_0 si definisce come

$$\frac{\partial f(x_0)}{\partial v} = \lim_{h \to 0} \frac{f(x_0 + hv) - f(x_0)}{v}.$$

E' la derivata della restrizione di f alla retta passante per x_0 in direzione v.

Facendo il conto per la funzione g scritta sopra, con $v = (v_1, v_2) \in \mathbf{R}^2$ direzione fissata, si ottiene

$$\frac{\partial f}{\partial v}(0,0) = v_1^2/v_2 \quad \textit{se } v_2 \neq 0,$$

mentre se $v_2 = 0$ la derivata vale evidentemente 0.

Per recuperare una decente regolarità, occorre una condizione più forte della derivabilità parziale o anche direzionale: conviene invece definire il piano tangente al grafico di una funzione di più variabili come "piano di migliore approssimazione", nella maniera che segue. Se ci restringiamo al caso di 2 variabili abbiamo

DEFINIZIONE: Il piano di equazione $z = f(x_0, y_0) + a\ (x - x_0) + b\ (y - y_0)$ si dice *piano tangente* al grafico di f nel punto (x_0, y_0) se e soltanto se

$$f(x, y) = f(x_0, y_0) + a\ (x - x_0) + b\ (y - y_0) + o(|(x - x_0, y - y_0)|),$$

cioè se e soltanto se

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f(x,y)-f(x_0,y_0)-a\,(x-x_0)-b\,(y-y_0)}{\sqrt{(x-x_0)^2+(y-y_0)^2}} = 0.$$

Se il grafico una funzione f ammette piano tangente nel punto (x_0, y_0), si dice che *f è differenziabile in* (x_0, y_0).

Lezione del 7/5/2007 (2 ore): A riprova che la nostra definizione di differenziabilità è quella "giusta", verifichiamo che una funzione differenziabile è continua ed anche derivabile parzialmente:

TEOREMA: Se f è differenziabile in (x_0, y_0), essa è anche continua nello stesso punto. f è poi derivabile parzialmente in (x_0, y_0), e l'equazione del piano tangente è data da

$$z = f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)(y-y_0).$$

Se definiamo il vettore gradiente

$$\nabla f(x_0,y_0) := (\frac{\partial f}{\partial x}(x_0,y_0), \frac{\partial f}{\partial y}(x_0,y_0)),$$

l'equazione del piano tangente può anche essere scritta

$$z = f(x_0,y_0) + \nabla f(x_0,y_0)\cdot(x-x_0, y-y_0).$$

DIM.: Si ha

$$\lim_{(x,y)\to(x_0,y_0)}[f(x,y)-f(x_0,y_0)] =$$
$$\lim_{(x,y)\to(x_0,y_0)}[a\,(x-x_0)+b\,(y-y_0)+o(|(x-x_0,y-y_0)|)] = 0,$$

per cui f è continua in (x_0, y_0). Inoltre, ponendo $x = x_0 + h$, $y = y_0$ nella definizione di piano tangente:

$$\frac{f(x_0+h,y_0)-f(x_0,y_0)}{h} = \frac{a\,h+o(h)}{h},$$

e passando al limite per $h \to 0$ si trova che f è derivabile parzialmente rispetto a x e $\frac{\partial f}{\partial x}(x_0, y_0) = a$.

In maniera del tutto analoga si mostra che $\frac{\partial f}{\partial y}(x_0, y_0) = b$. Q.E.D.

Evidentemente, la definizione di differenziabilità appena proposta non è molto agevole da verificare, in quanto occorre fare un limite in due variabili, e gli esempi visti sopra dovrebbero essere sufficienti a convincerci che questo non è sempre facile.

Per fare un esempio, in classe ci siamo divertiti (si fa per dire...) a verificare che la funzione $f(x, y) = x^2 - 3y^2$ è differenziabile in tutti i punti (x_0, y_0).... Questo esempio ci preoccupa un po': se è necessario riempire una lavagnata di conti per dimostrare che un polinomio di secondo grado è differenziabile, chissà cosa succederà con funzioni più complicate! Fortunatamente, vedremo che esistono condizioni sufficienti per la differenziabilità un po' più agevoli da verificare: in particolare vedremo che se una funzione è derivabile parzialmente in un intorno di (x_0, y_0) con *derivate parziali continue*, essa è anche differenziabile in (x_0, y_0).

Per avere un'idea "grafica" delle questioni trattate in questa lezione, potete consultare una mia pagina web sulle funzioni di più variabili[10] con animazioni interattive.

Il seguente teorema permette spesso di stabilire che una funzione è differenziabile:

TEOREMA (Del differenziale totale): Sia $f : \mathbf{R}^2 \to \mathbf{R}$ una funzione derivabile parzialmente in un intorno del punto (x_0, y_0) (cioè in un rettangolo che contiene (x_0, y_0) al suo interno). Se le derivate parziali $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ sono entrambe continue nel punto (x_0, y_0), allora la funzione f è differenziabile nel punto (x_0, y_0).

ESEMPIO: Riprendiamo la funzione $f(x, y) = x^2 - 3y^2$, di cui avevamo verificato "a mano" la differenziabilità. Si ha

$$\frac{\partial f}{\partial x}(x, y) = 2x, \quad \frac{\partial f}{\partial y}(x, y) = -6y.$$

Le derivate parziali esistono ovunque, e sono funzioni continue di (x, y), per cui la funzione è ovunque differenziabile.

In maniera analoga, un qualunque *polinomio di due variabili* è ovunque differenziabile, perché le sue derivate parziali sono ancora polinomi e sappiamo che i polinomi sono funzioni continue... Utilizzando il teorema del differenziale totale, si può così mostrare facilmente che un gran numero di funzioni elementari sono differenziabili.

DIMOSTRAZIONE del Teorema del Differenziale Totale: Sia (x, y) un punto sufficientemente vicino a (x_0, y_0), in modo che tutto il rettangolo di vertici

[10] http://www.science.unitn.it/baldo/divulgazione/Graficiescript/Grafici3D.html

(x_0, y_0), (x_0, y), (x, y) e (x, y_0) sia contenuto nell'insieme in cui esistono le derivate parziali di f.

Per mostrare che f è differenziabile in (x_0, y_0), dobbiamo far vedere che quando $(x, y) \to (x_0, y_0)$ il seguente rapporto tende a 0:

$$G(x,y) = \frac{f(x,y) - f(x_0,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) - \frac{\partial f}{\partial y}(x_0,y_0)(y-y_0)}{\sqrt{(x-x_0)^2+(y-y_0)^2}}$$

Il teorema di Lagrange per funzioni di una variabile mi assicura che esistono dei punti ξ (con ξ compreso tra x_0 ed x) ed η (con η compreso tra y_0 ed y) tali che

$$\begin{aligned} f(x,y_0) - f(x_0,y_0) &= \frac{\partial f}{\partial x}(\xi,y_0)(x-x_0), \\ f(x,y) - f(x,y_0) &= \frac{\partial f}{\partial y}(x,\eta)(y-y_0). \end{aligned}$$

Nel rapporto $G(x, y)$, aggiungiamo e togliamo la quantità $f(x, y_0)$ a numeratore, ed applichiamo le due relazioni appena trovate. Possiamo così scrivere

$$\begin{aligned} G(x,y) = &\frac{x-x_0}{\sqrt{(x-x_0)^2+(y-y_0)^2}}\left[\frac{\partial f}{\partial x}(\xi,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)\right] + \\ &\frac{y-y_0}{\sqrt{(x-x_0)^2+(y-y_0)^2}}\left[\frac{\partial f}{\partial y}(x,\eta) - \frac{\partial f}{\partial y}(x_0,y_0)\right]. \end{aligned}$$

Ora, quando $(x, y) \to (x_0, y_0)$ avremo che $\xi \to x_0$ e $\eta \to y_0$, e grazie alla continuità delle derivate parziali in (x_0, y_0) le due espressioni tra parentesi quadre tenderanno a 0. Invece, i rapporti

$$\frac{x-x_0}{\sqrt{(x-x_0)^2+(y-y_0)^2}}, \quad \frac{y-y_0}{\sqrt{(x-x_0)^2+(y-y_0)^2}}$$

sono limitati: precisamente, il modulo di entrambi è minore o uguale a 1. Se ne conclude che

$$\lim_{(x,y)\to(x_0,y_0)} G(x,y) = 0.$$

Q.E.D.

OSSERVAZIONE: Il teorema del differenziale totale ci fornisce solo una condizione *sufficiente* di differenziabilità: una funzione può benissimo essere differenziabile anche se non sono soddisfatte le ipotesi. Ad esempio, la funzione

$$f(x,y)=\begin{cases} 0 & \textit{se } x\in\mathbf{Q} \textit{ e } y\in\mathbf{Q} \\ x^2+y^2 & \textit{altrimenti} \end{cases}$$

è differenziabile nell'origine (con piano tangente $z=0$), ma non è neppure derivabile parzialmente in tutti gli altri punti.

Esattamente come accadeva in una variabile, l'analisi delle derivate parziali è un metodo potentissimo per cercare i *punti di massimo e minimo relativo per una funzione differenziabile di due variabili.* Una semplicissima osservazione ci assicura che in un punto di massimo o minimo relativo il piano tangente *deve essere orizzontale* (lo vedremo la prossima volta).

I punti in cui il piano tangente è orizzontale si chiamano *punti critici.* Evidentemente, questi non sono tutti punti di massimo o minimo relativo: ci possono essere per esempio *punti di sella* (punti corrispondenti a un "passo di montagna", se interpretiamo il grafico della funzione come rappresentazione grafica di un territorio montuoso...).

Così come facevamo in una variabile, ci piacerebbe classificare i punti critici in base al segno della derivata seconda. Un piccolo problemino è dato però dal fatto che in un punto (x_0,y_0) di possibili derivate seconde ce ne sono ben quattro:

$$\frac{\partial^2 f}{\partial x^2}(x_0,y_0),\ \frac{\partial^2 f}{\partial y^2}(x_0,y_0),\ \frac{\partial^2 f}{\partial x\partial y}(x_0,y_0),\ \frac{\partial^2 f}{\partial y\partial x}(x_0,y_0).$$

Per comodità, esse si raggruppano di solito nella *matrice hessiana*

$$Hf(x_0,y_0)=\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y\partial x} \\ \frac{\partial^2 f}{\partial x\partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

(in cui, ovviamente, tutte le derivate parziali sono calcolate in (x_0,y_0)).

Per nostra fortuna, se la funzione f è abbastanza buona le due derivate parziali miste *sono uguali* e la matrice hessiana è una matrice simmetrica: è il contenuto del seguente

TEOREMA (di Schwartz): Se f possiede entrambe le derivate seconde miste

$$\frac{\partial^2 f}{\partial x\partial y}(x,y),\ \frac{\partial^2 f}{\partial y\partial x}(x,y)$$

in un intorno del punto (x_0,y_0), e queste due derivate sono continue in (x_0,y_0), allora

$$\frac{\partial^2 f}{\partial x\partial y}(x_0,y_0)=\frac{\partial^2 f}{\partial y\partial x}(x_0,y_0).$$

Dimostreremo questo teorema la prossima volta. Per ora, diamo un esempio di funzione che non soddifa le ipotesi del teorema, per cui le due derivate parziali miste esistono ma sono diverse: la funzione

$$f(x,y) = \begin{cases} \frac{x^3 y}{x^2+y^2} & se\ (x,y) \neq (0,0), \\ 0 & se\ (x,y) = (0,0) \end{cases}$$

ha entrambe le derivate seconde miste nell'origine, ma esse sono *diverse*.

Lezione del 10/5/2007 (2 ore): Dimostriamo il teorema di Schwarz. Siano h, k sufficientemente piccoli, in modo che il rettangolo di diagonale delimitata da (x_0, y_0) e (x_0+h, y_0+k) sia tutto contenuto nell'insieme in cui esistono le derivate seconde miste. Consideriamo poi la funzione ausiliaria

$$g(h,k) = \frac{f(x_0+h, y_0+k) - f(x_0+h, y_0) - f(x_0, y_0+k) + f(x_0, y_0)}{hk}.$$

Dimostreremo che il limite di questa funzione per $(h,k) \to (0,0)$ esiste, ed è uguale *sia a* $f_{xy}(x_0, y_0)$ *che a* $f_{yx}(x_0, y_0)$, che implica la tesi del teorema.

Consideriamo infatti la funzione di una variabile $\phi(y) = f(x_0+h, y) - f(x_0, y)$: possiamo scrivere la funzione g in termini di ϕ ed applicare il teorema di Lagrange, ottenendo

$$g(h,k) = \frac{\phi(y_0+k) - \phi(y_0)}{hk} = \frac{\phi'(y_0+pk)}{h} = \frac{f_y(x_0+h, y_0+qk) - f_y(x_0, y_0+qk)}{h}$$

per un opportuno $q \in (0,1)$. Riapplicando il teorema di Lagrange (nella variabile x) all'ultimo rapporto si ottiene

$$g(h,k) = f_{xy}(x_0+ph, y_0+qk)$$

per opportuni $p, q \in (0,1)$. Passando al limite per $(h,k) \to (0,0)$ si ottiene, grazie alla continuità delle derivate parziali miste, $f_{xy}(x_0, y_0)$.

D'altra parte, se rifacciamo lo stesso giochetto usando la funzione $\psi(x) = f(x, y_0+k) - f(x, y_0)$, troviamo $p', q' \in (0,1)$ tali che $g(h,k) = f_{yx}(x_0 + p'h, y_0+q'k)$. Ques'ultima espressione tende a $f_{yx}(x_0, y_0)$ per $(h,k) \to (0,0)$. Q.E.D.

Il teorema di Schwartz, come il teorema del differenziale totale, è banalmente estendibile a più di due variabili.

L'ultimo teorema ci sarà utilissimo quando useremo le derivate seconde per caratterizzare i punti di massimo e minimo relativo di una funzione di più variabili. La simmetria della matrice hessiana non è infatti una questione di lana caprina: sarà invece una proprietà molto importante nel nostro studio dei punti di estremo[11].

Ovviamente, prima di studiare le derivate seconde è necessario studiare le derivate prime. Vale la seguente, ovvia osservazione.

OSSERVAZIONE: Se (x_0, y_0) è un punto di massimo o minimo relativo per la funzione differenziabile $f(x, y)$, allora

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0, \quad \frac{\partial f}{\partial y}(x_0, y_0) = 0.$$

Infatti, il punto (x_0, y_0) sarà a maggior ragione un punto di massimo o minimo relativo per la restrizione di f alle rette $y = y_0$ e $x = x_0$, per cui le derivate di tali restrizioni (che sono poi le derivate parziali) si devono annullare nel punto stesso.

Come per le funzioni di una variabile, lo studio dei punti critici mediante le derivate seconde è strettamente legato ad una *formula di Taylor di ordine 2 (con resto di Peano).* Per poterla dimostrare, ci serve una semplice regola di derivazione delle funzioni composte che è molto interessante anche per se stessa:

PROPOSIZIONE (Regola di derivazione delle funzioni composte): Sia $f(x, y)$ una funzione ovunque differenziabile, e siano $x(t)$, $y(t)$ due funzioni derivabili di una variabile. Allora la funzione composta $F(t) = f(x(t), y(t))$ è derivabile, e si ha

$$F'(t) = \frac{\partial f}{\partial x}(x(t), y(t))\, x'(t) + \frac{\partial f}{\partial y}(x(t), y(t))\, y'(t).$$

DIM: Per dimostrare che f è derivabile in t_0 e che vale la nostra espressione per la derivata, è sufficiente verificare che vale la formula di Taylor del primo ordine:

$$(A) \qquad f(x(t), y(t)) = f(x(t_0), y(t_0)) +$$

[11] In effetti, abbiamo imparato nel corso di geometria che le matrici reali simmetriche sono diagonalizzabili, e in ultima analisi questo è esattamente quel che serve... Siccome però abbiamo solo due variabili, riusciremo a dimostrare il nostro risultato principale in modo elementare, senza ricorrere ai teoremi di algebra lineare che ci serviranno in dimensione superiore.

$$\left(\frac{\partial f}{\partial x}(x(t_0), y(t_0))\; x'(t_0) + frac\partial f\partial y(x(t_0), y(t_0))\; y'(t_0)\right)(t-t_0) + o(t-t_0).$$

A questo scopo, notiamo che la differenziabilità di f implica

$$(B) \qquad f(x(t), y(t)) = f(x(t_0), y(t_0)) + \frac{\partial f}{\partial x}(x(t_0), y(t_0))\; (x(t) - x(t_0)) + \frac{\partial f}{\partial y}(x(t_0), y(t_0))\; (y(t) - y(t_0)) + o(|(x(t) - x(t_0), y(t) - y(t_0))|),$$

mentre per le funzioni derivabili $x(t)$ e $y(t)$ abbiamo

$$x(t) - x(t_0) = x'(t_0)(t-t_0) + o(t-t_0), \quad y(t) - y(t_0) = y'(t_0)(t-t_0) + o(t-t_0).$$

Sostituendo queste due ultime relazioni in (B) si ottiene (A), a patto di osservare che $o(|(x(t) - x(t_0), y(t) - y(t_0))|) = o(t - t_0)$.[12] Q.E.D.

OSSERVAZIONE: In n variabili il teorema diventa il seguente. Se $f : \mathbf{R}^n \to \mathbf{R}$ è differenziabile e $g : [a, b] \to \mathbf{R}^n$ è derivabile, allora posto $h(t) = f(g(t))$ la funzione di una variabile h è derivabile e si ha

$$h'(t) = \nabla f(g(t)) \cdot g'(t),$$

dove $v \cdot w$ indica il prodotto scalare standard dei due vettori v e w.

Vediamo ora che aspetto assume il Teorema di Taylor per le funzioni di due variabili. Siccome questo sarà sufficiente per i nostri scopi, ci fermiamo al secondo ordine:

TEOREMA (di Taylor al II ordine): Sia $f(x, y)$ una funzione di due variabili, derivabile due volte con derivate prime e seconde continue in un intorno di (x_0, y_0). Allora, per tutti i punti (x, y) sufficientemente vicini a (x_0, y_0) si ha

$$f(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + \frac{1}{2}[f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + f_{yy}(x_0, y_0)(y - y_0)^2] + o((x - x_0)^2 + (y - y_0)^2).$$

[12]Infatti, si verifica facilmente che $|(x(t) - x(t_0), y(t) - y(t_0))| \leq C|t - t_0|$, per cui la distanza tende a 0 con la stessa velocità di $t - t_0$, e la nostra affermazione segue facilmente...

DIM.: Sia (x, y) sufficientemente vicino a (x_0, y_0), in modo che tutto il segmento che congiunge i due punti sia contenuto nell'insieme in cui esistono le derivate prime e seconde (e sono continue).

Consideriamo la funzione (di una variabile) $F(t) = f(x_0 + t(x - x_0), y_0 + t(y - y_0))$: evidentemente, $F(0) = f(x_0, y_0)$ e $F(1) = f(x, y)$, mentre per valori intermedi di t la funzione f viene calcolata nei punti intermedi del segmento suddetto.

Applicando il teorema di derivazione della funzione composta si trova

$$F'(t) = f_x(\ldots)(x - x_0) + f_y(\ldots)(y - y_0),$$

dove l'argomento delle derivate (indicato con i puntini) è ovviamente $(x_0 + t(x - x_0), y_0 + t(y - y_0))$. Derivando ancora:

$$F''(t) = f_{xx}(\ldots)(x - x_0)^2 + 2f_{xy}(\ldots)(x - x_0)(y - y_0) + f_{yy}(\ldots)(y - y_0)^2$$

(in cui i puntini hanno ancora lo stesso significato...).

La formula di Taylor per funzioni di una variabile, con resto di Lagrange, dice che è possibile trovare un punto $c \in [0, 1]$ tale che

$$F(1) = F(0) + F'(0) \cdot 1 + \frac{1}{2}F''(c) \cdot 1,$$

e sostituendo le espressioni che abbiamo trovato per le derivate otteniamo

$$\begin{aligned}&f(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + \\ &\frac{1}{2}[f_{xx}(a, b)(x - x_0)^2 + 2f_{xy}(a, b)(x - x_0)(y - y_0) + f_{yy}(a, b)(y - y_0)^2],\end{aligned}$$

dove $a = x_0 + c(x - x_0)$, $b = x_0 + c(x - x_0)$. Nell'espressione sopra, aggiungiamo e togliamo il polinomio di secondo grado

$$\frac{1}{2}[f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + f_{yy}(x_0, y_0)(y - y_0)^2],$$

in modo da ottenere

$$\begin{aligned}&f(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) + \\ &\frac{1}{2}[f_{xx}(x_0, y_0)(x - x_0)^2 + 2f_{xy}(x_0, y_0)(x - x_0)(y - y_0) + f_{yy}(x_0, y_0)(y - y_0)^2] + \\ &R(x, y)\end{aligned}$$

dove

$$\begin{aligned}
R(x,y) &= \frac{1}{2}(x-x_0)^2[f_{xx}(a,b)-f_{xx}(x_0,y_0)]+\\
&\quad (x-x_0)(y-y_0)[f_{xy}(a,b)-f_{xy}(x_0,y_0)]+\\
&\quad \frac{1}{2}(y-y_0)^2[f_{yy}(a,b)-f_{yy}(x_0,y_0)].
\end{aligned}$$

Ora, è immediato verificare che

$$\lim_{(x,y)\to(x_0,y_0)} \frac{R(x,y)}{(x-x_0)^2+(y-y_0)^2} = 0.$$

Infatti, se $(x,y) \to (x_0,y_0)$ si ha anche $(a,b) \to (x_0,y_0)$ per cui le quantità tra parentesi quadre tendono a 0. Invece, i rapporti

$$\frac{(x-x_0)^2}{(x-x_0)^2+(y-y_0)^2},\ \frac{2(x-x_0)(y-y_0)}{(x-x_0)^2+(y-y_0)^2},\ \frac{(y-y_0)^2}{(x-x_0)^2+(y-y_0)^2}$$

sono tutti minori o uguali ad 1 in modulo. Q.E.D.

Lezione del 14/5/2007 (2 ore): Se abbiamo una funzione di n variabili $f(\mathbf{x}) = f(x_1,x_2,\ldots,x_n)$, le nozioni di derivabilità parziale e di differenziabilità si introducono in maniera del tutto identica al caso delle due variabili. Anche risultati fondamentali come il teorema del differenziale totale ed il teorema di Schwartz, continuano a valere esattamente allo stesso modo.

Per ragioni di "compattezza notazionale", conviene usare una notazione vettoriale: scriveremo

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$$

per indicare un vettore di $\mathbf{R}^n$. Data una funzione $f : \mathbf{R}^n \to \mathbf{R}$ derivabile parzialmente, introduciamo il vettore *gradiente*

$$\mathrm{grad}\ f(x) = \nabla f(x) = \begin{pmatrix} f_{x_1}(x) \\ f_{x_2}(x) \\ \cdots \\ f_{x_n}(x) \end{pmatrix}$$

Esattamente come in due variabili, l'equazione del piano tangente al grafico di f in $x_0 \in \mathbf{R}^n$, se questo c'è, avrà equazione $z = f(x_0)+\nabla f(x_0)\cdot(x-x_0)$, dove $v \cdot w$ indica il prodotto scalare standard in $\mathbf{R}^n$. Dunque, diremo che f è differenziabile in x_0 se è derivabile parzialmente in quel punto e si ha $f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + o(|x - x_0|)$ (dove al solito la norma di un vettore $v \in \mathbf{R}^n$ è data da $|v| = \sqrt{v_1^2 + \ldots + v_n^2}$).

Vale anche il teorema del differenziale totale: se f è derivabile parzialmente in un intorno di x_0, con derivate parziali continue in x_0, allora f è differenziabile in x_0.

Il vettore gradiente ha un'interessante interpretazione geometrica: la sua direzione è quella di *massima pendenza* per il grafico di f, come si vede nella seguente

OSSERVAZIONE: Se $v \in \mathbf{R}^n$ è una *direzione*, cioè se $|v| = 1$, abbiamo definito la derivata direzionale di f in direzione v in un punto x_0 come

$$\frac{\partial f(x_0)}{\partial v} = \lim_{h \to 0} \frac{f(x_0 + hv) - f(x_0)}{h}.$$

Si tratta della derivata della restrizione di f alla retta passante per x_0 in direzione v...

Se f è differenziabile in x_0, è immediato verificare che si ha $\frac{\partial f(x_0)}{\partial v} = \nabla f(x_0) \cdot v$. Evidentemente quest'ultimo prodotto scalare sarà massimo quando due vettori hanno la stessa direzione e lo stesso verso: la derivata direzionale è massima nella direzione del vettore gradiente!

Vediamo come si studiano i massimi e i minimi di una funzione di n variabili tramite la matrice hessiana. Il criterio che troveremo coinvolge il segno degli *autovalori* della matrice hessiana.

Ovviamente, continua a essere vero che in un punto di massimo e minimo relativo si annulla il gradiente della funzione, cioè il vettore

$$\operatorname{grad} f(x_1, \ldots, x_n) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

Nel caso di una funzione di n variabili, derivabile 2 volte con continuità, l'hessiana sarà la matrice quadrata $n \times n$, simmetrica perché il teorema di Schwarz vale come in 2 variabili:

$$Hf = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Per decidere se un punto critico è di massimo o minimo relativo, occorre andare a studiare il segno degli *autovalori della matrice hessiana*. Infatti, sappiamo dal corso di geometria che una matrice reale simmetrica ha autovalori reali. Ebbene, se la matrice hessiana ha gli autovalori *tutti positivi* in un punto critico, il punto è di minimo relativo, se gli autovalori sono tutti *negativi* il punto è di massimo, se infine esistono *sia autovalori positivi che negativi*, il punto è di sella:

TEOREMA: Sia f una funzione di n variabili derivabile due volte con continuità, e supponimo che x_0 sia un punto in cui si annulla il gradiente di f. Allora

1. *Se la matrice hessiana $Hf(x_0)$ è definita positiva, cioè se tutti i suoi autovalori sono strettamente positivi, allora x_0 è un punto di minimo relativo. Se invece $Hf(x_0)$ è definita negativa (tutti gli autovalori sono negativi), avremo un punto di massimo relativo.*

2. *Se x_0 è un punto di minimo relativo (risp. massimo relativo), allora la matrice $Hf(x_0)$ è semidefinita positiva (risp. semidefinita negativa).*

3. *Se la matrice $Hf(x_0)$ ha sia autovalori positivi che autovalori negativi, allora x_0 è un punto di sella.*

Dal punto di vista operativo, quindi, si tratta di studiare il *segno* degli autovalori della matrice hessiana. Dal corso di geometria, sappiamo che λ è un autovalore di H se e soltanto se

$$\det(\lambda I - H) = 0$$

(questo viene dal fatto che vogliamo avere soluzioni non nulle dell'equazione $Hv = \lambda v$, per cui la matrice $(\lambda I - H)$ deve essere singolare).

Si tratta quindi semplicemente di scrivere l'equazione degli autovalori (un'equazione polinomiale di grado n), e di studiare il segno delle sue n radici reali (sappiamo che sono tutte reali perché la matrice è simmetrica)[13]...

Si noti una cosa importante: se la matrice hessiana è semidefinita positiva (o semidefinita negativa) e però c'è qualche autovalore nullo, il teorema non è decisivo sulla natura del punto critico, che potrebbe essere di minimo o di sella (risp., di massimo o di sella).

[13]Ci sono dei trucchi che consentono di vedere il segno delle radici anche se non si riesce a risolvere esplicitamente l'equazione: questo è estremamente utile perchè non sempre siamo in grado risolvere un'equazione polinomiale di grado elevato! Questi trucchi però esulano dallo scopo di questo corso

Questo fatto diventa però interessante se lo leggiamo nell'altra direzione: in un punto di minimo relativo, la matrice hessiana deve essere *semidefinita positiva*, in un punto di massimo *semidefinita negativa* (se ci fossero degli autovalori col segno "sbagliato", dal teorema seguirebbe che...).

DIM.: La formula di Taylor del secondo ordine in n variabili si può scrivere

$$\begin{aligned}&f(x) = f(x_0) + \\ &\text{grad } f(x_0) \cdot (x - x_0) + \frac{1}{2}(Hf(x_0)(x - x_0)) \cdot (x - x_0) + o(|x - x_0|^2),\end{aligned}$$

dove

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad x_0 = \begin{pmatrix} x_0^1 \\ \vdots \\ x_0^n \end{pmatrix},$$

$a{\cdot}b$ denota l'ordinario prodotto scalare in $\mathbf{R}^n$, e con la scrittura $Hf(x_0)(x-x_0)$ indichiamo il prodotto della matrice hessiana con il vettore $(x - x_0)$.

Scriviamo $x = x_0 + rv$ con $v = (x - x_0)\|x - x_0|$, $r = |x - x_0|$. Il vettore v appartiene allora alla sfera unitaria $S = \{v : \ v_1^2 + v_2^2 + \dots v_n^2 = 1\}$.

In un *punto critico*, il termine della formula di Taylor che contiene il gradiente non c'è, e ci rimane soltanto

$$\begin{aligned}&f(x) - f(x_0) = \\ &\frac{1}{2}(Hf(x_0)(x - x_0)) \cdot (x - x_0) + o(|x - x_0|^2) = \\ &\frac{1}{2}r^2 \left[(Hf(x_0)v) \cdot v + o(r^2)\right].\end{aligned}$$

È ora abbastanza chiaro che il segno del membro di destra di questa equazione è strettamente legato al segno del *valore massimo* e del *valore minimo* del polinomio di secondo grado

$$Q(v) = (Hf(x_0)v) \cdot v = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) v_i v_j$$

sulla *sfera unitaria*. Innanzitutto, questo massimo e questo minimo esistono: anche in più variabili vale il teorema di Weierstrass, che assicura che una funzione continua ammette massimo e minimo su un insieme chiuso e limitato.

Il nostro enunciato si può allora ricavare facilmente grazie al risultato seguente:

PROPOSIZIONE: Il massimo e il minimo della funzione $Q(v)$ sulla sfera unitaria S non sono altro che il massimo ed il minimo degli autovalori della matrice simmetrica $Hf(x_0)$.

DIM.: Indichiamo per brevità $H := Hf(x_0)$, per cui $Q(v) = (Hv) \cdot v$. La matrice H è una matrice reale simmetrica $n \times n$: sappiamo dal corso di geometria che una tale matrice si può diagonalizzare. Precisamente, esiste una trasformazione ortogonale di coordinate $T : \mathbf{R}^n \to \mathbf{R}^n$ tale che nelle nuove coordinate la matrice è diagonale:

$$\widetilde{H} = \begin{pmatrix} \lambda_1 & 0 & \dots & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

(in altre parole, vale l'identità $H = {}^tT\widetilde{H}T$). I coefficienti λ_i sono gli autovalori della matrice H, e possiamo supporre senza perdita di generalità che essi siano ordinati in senso decrescente: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

La trasformazione ortogonale T è un'isometria, per cui manda la sfera unitaria in se: di conseguenza massimizzare e minimizzare $Q(v)$ al variare di v sulla sfera unitaria S di $\mathbf{R}^n$ è la stessa cosa che massimizzare e minimizzare la forma quadratica diagonalizzata $\widetilde{Q} : w \mapsto (\widetilde{H}w) \cdot w$ su S stessa. D'altra parte,

$$\widetilde{Q}(w) = \sum_{i=1}^{n} \lambda_i w_i,$$

e se $w \in S$ ne deduciamo $\widetilde{Q}(w) \geq \lambda_n \sum w_i^2 = \lambda_n$: sulla sfera S la forma quadratica è maggiore o uguale al minimo autovalore della matrice H. D'altra parte, esiste un elemento della sfera unitaria su cui la forma quadratica assume proprio il valore λ_n: basta prendere $w = e_n$ (l'n-esimo vettore della base canonica). Dunque λ_n è proprio il valore minimo di $\widetilde{Q}$ (e quindi di Q) sulla sfera unitaria.

Con un ragionamento del tutto analogo, il massimo autovalore λ_1 è proprio il massimo di Q su S.

Dimostrazione alternativa: Indichiamo con h_{ij} i coefficienti della matrice H. Si ha $Q(v) = \sum_{i,j=1}^{n} h_{ij} v_i v_j$, da cui (usando la simmetria di H) si ricava facilmente $\nabla Q(v) = 2Hv$.

Ora, la funzione Q è evidentemente continua (polinomio di secondo grado) su S che è un sottinsieme chiuso e limitato di $\mathbf{R}^n$: per il teorema di Weierstrass essa assumerà minimo assoluto m e massimo assoluto M rispettivamente in due punti v e w di S.

Consideriamo la funzione $F(v) = Q(v)/|v|^2$, definita per ogni $v \neq 0$: si vede subito che $F(tv) = F(v)$ per ogni $t > 0$. Ne deriva che F è costante sulle semirette uscenti dall'origine, ed in particolare assume tutti e soli i valori assunti da Q sulla sfera S. Inoltre, un punto di massimo o di minimo assoluto per Q sulla sfera sarà un punto di massimo e minimo assoluto per F su $\mathbf{R}^n \setminus \{0\}$. Dunque, se v è il punto di minimo di Q su S, si dovrà avere $\nabla F(v) = 0$. Dunque:

$$0 = \nabla F(v) = 2\frac{|v|^2 Hv - Q(v)v}{|v|^2}$$

da cui $Hv = \frac{Q(v)}{|v|^2}v = Q(v)v$, e $Q(v)$ è un autovalore di H con autovettore v. È necessariamente il *minimo autovalore*: se λ è un qualunque altro autovalore e $v_\lambda \in S$ un suo autovettore, si ha $Q(v_\lambda) = (Hv_\lambda) \cdot v_\lambda = \lambda v_\lambda \cdot v_\lambda = \lambda$, da cui $m \leq \lambda$. Analogamente, il punto di massimo w di Q su S sarà un autovettore con autovalore $Q(w)$ della matrice H, e quest'ultimo sarà il più grande degli autovalori. Q.E.D.

Concludiamo la dimostrazione del teorema: denotiamo $r = |x - x_0|$, $v = (x - x_0)/|x - x_0| \in S$. La formula di Taylor del secondo ordine diventa:

$$f(x) - f(x_0) = \frac{1}{2}r^2 \left[(Hf(x_0)v) \cdot v + \frac{o(r^2)}{r^2}\right].$$

Supponiamo che la matrice hessiana sia definita positiva, e che m sia il minimo dei suoi autovalori. Grazie alla proposizione precedente, l'ultima espressione è maggiore o uguale di

$$\frac{1}{2}r^2 \left[m + \frac{o(r^2)}{r^2}\right],$$

e quest'espressione è strettamente positiva a patto di prendere r sufficientemente piccolo (cioè x sufficientemente vicino a x_0): x_0 è un punto di minimo relativo (stretto) per f.

In maniera del tutto analoga, se $Hf(x_0)$ ha tutti gli autovalori negativi abbiamo un punto di massimo relativo (stretto).

Se infine la matrice ha sia autovalori positivi che autovalori negativi, allora esistono due vettori v, w su S tali che $Q(v) = m < 0$ e $Q(w) = M > 0$ (per esempio, i punti di minimo e di massimo di Q sulla sfera). Se prendiamo $x = x_0 + rv$ e sostituiamo nella formula di Taylor si ha

$$f(x) - f(x_0) = \frac{1}{2}r^2 \left[m + \frac{o(r^2)}{r^2}\right],$$

espressione negativa per r abbastanza piccolo. Analogamente, se prendiamo $x = x_0 + rw$ otteniamo

$$f(x) - f(x_0) = \frac{1}{2} r^2 \left[M + \frac{o(r^2)}{r^2} \right],$$

espressione positiva per r piccolo: in alcune direzioni x_0 è punto di massimo relativo, in altre di minimo relativo. x_0 è allora un punto di sella. Q.E.D.

OSSERVAZIONE: In due variabili, lo studio del segno della forma quadratica $Q(v)$ sulla sfera unitaria è più facile. Supponiamo infatti di voler verificare se $Q(v)$ è strettamente positivo su $\mathbf{R}^2 \setminus \{0\}$: questo è equivalente alla disequazione $Q(v) = f_{xx}(x_0)v_1^2 + 2f_{xy}(x_0)v_1v_2 + f_{yy}v_2^2 > 0$, che posso vedere come una disequazione di secondo grado nell'incognita v_1. Perché questa cosa non cambi di segno, occorre che $\det Hf(x_0) > 0$ (tale determinante è esattamente il discriminante della disequazione, cambiato di segno...). Perchè sia poi positiva serve che $f_{xx}(x_0) > 0$: sotto queste condizioni abbiamo un punto di minimo relativo.

Analogamente, se $\det Hf(x_0) > 0$ e $f_{xx}(x_0) > 0$ avremo un massimo relativo, se $\det Hf(x_0) < 0$ un punto di sella.

In questo caso particolare l'enunciato può essere riscritto come segue:

TEOREMA: Sia $f(x, y)$ una funzione derivabile due volte in un intorno di (x_0, y_0), con derivate prime e seconde continue. Supponiamo anche che $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ (cioè che il punto (x_0, y_0) sia un punto critico). Se il determinante della matrice hessiana è positivo,

$$\det Hf(x_0, y_0) = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - (f_{xy}(x_0, y_0))^2 > 0,$$

allora (x_0, y_0) è un punto di minimo relativo o di massimo relativo a seconda che sia $f_{xx}(x_0, y_0) > 0$ o$f_{xx}(x_0, y_0) < 0$ rispettivamente.

Se viceversa (x_0, y_0) è un punto di minimo (risp. massimo) relativo per f, allora $\det Hf(x_0, y_0) \geq 0$ e $f_{xx}(x_0, y_0) \geq 0$ (risp. $f_{xx}(x_0, y_0) \leq 0$).

In particolare, un punto critico in cui $\det Hf(x_0, y_0) < 0$ non è né di massimo né di minimo relativo (punto di sella).

Lezione del 17/5/2005 (2 ore): Per vari motivi, tra cui lo studio dei massimi e dei minimi di una funzione differenziabile sulla *frontiera di un insieme regolare*, è utile studiare come sono fatte le *curve di livello* di una funzione di due variabili $f(x, y)$, cioè gli insiemi

$$S_C = \{(x, y) \in \mathbf{R}^2 : f(x, y) = C\},$$

dove C è una fissata costante.

Se siamo degli appassionati escursionisti, questi oggetti dovrebbero esserci familiari: immaginiamo che il grafico $z = f(x, y)$ della nostra funzione rappresenti il "plastico" di un territorio montuoso. Supponiamo anche che la funzione f sia differenziabile: abbiamo dei pendii, delle colline e degli avvallamenti, ma non sono presenti pareti rocciose verticali a strapiombo. Se un cartografo vuole disegnare una mappa che renda efficacemente l'idea del territiorio e della sua orografia, uno degli espedienti più utili cui può ricorrere consiste nel disegnare le curve che *uniscono punti ad uguale altitudine*, cioè appunto gli *insiemi di livello*.

Ora, la nostra esperienza di lettori di carte topografiche ci suggerisce che, tranne in pochi casi eccezionali, gli insiemi di livello siano *curve regolari*, ossia che essi *localmente coincidano con il grafico di una funzione derivabile (di x o di y)*.

Se ci pensiamo un po' vedremo che i casi eccezionali si possono avere quando il piano tangente al grafico di f è orizzontale in alcuni punti dell'insieme di livello: per esempio, la cima di una collina può essere un punto isolato del suo insieme di livello, oppure si può avere una grossa zona piatta, che certo non è una curva. Inoltre, in un punto di sella l'insieme di livello può essere costituito da *due curve che si incrociano*, ed anche in questo caso non è molto regolare.

Vedremo invece che un insieme di livello sul quale non ci siano punti critici è effettivamente una curva regolare: precisamente, l'insieme di livello stesso può essere localmente descritto come grafico di una funzione derivabile di una variabile.

DEFINIZIONE: Un insieme $A \subset \mathbf{R}^2$ è una curva regolare se esso può essere descritto localmente come grafico di una funzione regolare di x o y. Precisamente, vogliamo che ogni punto $(x_0, y_0) \in A$ abbia un intorno rettangolare $[x_0 - \delta, x_0 + \delta] \times [y_0 - r, y_0 + r]$ tale che la parte di A contenuta nel rettangolo coincida con il grafico di una funzione derivabile di x oppure di y.

Il seguente teorema dice che un insieme di livello di una funzione regolare f su cui il gradiente di f *non si annulla mai* è una curva regolare. Nell'enunciato, ci occupiamo dell'insieme di livello zero di f, cioè dell'insieme dei punti in cui la funzione *si annulla*. Evidentemente, questo non è restrittivo: l'insieme di livello C della funzione f non è altro che l'insieme degli zeri della funzione $g(x, y) = f(x, y) - C$.

TEOREMA (Delle funzioni implicite o del Dini): Sia $f(x, y)$ una funzione derivabile in un intorno di (x_0, y_0), con derivate continue in tale intorno. Supponiamo anche di sapere che $f(x_0, y_0) = 0$, e che $f_y(x_0, y_0) \neq 0$. Allora esistono $\delta > 0$ e $r > 0$ tali che *(i) Per ogni $x \in [x_0 - \delta, x_0 + \delta]$ esiste un unico $y \in [y_0 - r, y_0 + r]$ tale che $f(x, y) = 0$. Grazie all'unicità, possiamo anche scrivere $y = g(x)$.*

(ii) La funzione $g : [x_0 - \delta, x_0 + \delta] \to \mathbf{R}$ è continua e derivabile, e si ha

$$g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))}.$$

OSSERVAZIONE: Vediamo di tradurre l'enunciato. Sia

$$Z = \{(x, y) : \ f(x, y) = 0\}$$

l'insieme degli zeri di f. Il teorema dice che *dentro un rettangolo sufficientemente piccolo* $[x_0 - \delta, x_0 + \delta] \times [y_0 - r, y_0 + r]$, centrato nel punto (x_0, y_0) (che appartiene a Z per ipotesi), l'insieme degli zeri coincide con il grafico di una funzione derivabile g. Abbiamo inoltre una simpatica formuletta per calcolare la derivata di g. Si noti che se invece di avere la condizione $f_y(x_0, y_0) \neq 0$, sapessimo che $f_x(x_0, y_0) \neq 0$, scambiando i ruoli delle due variabili potremmo dire che Z, in un rettangolino attorno a (x_0, y_0), coincide con il grafico di *una funzione della variabile y*... L'unico caso in cui il teorema non ci dice assolutamente nulla, è quello in cui *entrambe* le derivate si annullano: come abbiamo visto prima, questo è appunto il caso in cui l'insieme di livello può avere un aspetto orribile!

DIM. DEL TEOREMA: Supponiamo, per fissare le idee, che si abbia $f_y(x_0, y_0) > 0$. Siccome le due derivate parziali sono continue, per il teorema della permanenza del segno possiamo dire che f_y sarà strettamente positiva in un intero rettangolino centrato in (x_0, y_0): esiste $r > 0$ tale che $f_y(x, y) > 0$ per ogni $(x, y) \in [x_0 - r, x_0 + r] \times [y_0 - r, y_0 + r]$.

In particolare, la funzione di una variabile $y \mapsto f(x_0, y)$ è strettamente crescente sull'intervallo $[y_0 - r, y_0 + r]$. Siccome $f(x_0, y_0) = 0$, possiamo dedurne che

$$f(x_0, y_0 - r) < 0, \quad f(x_0, y_0 + r) > 0.$$

Ancora per il teorema della permanenza del segno, la funzione sarà negativa in un intorno di $(x_0, y_0 - r)$, e positiva in un intorno di $(x_0, y_0 + r)$: potremmo così trovare un numero $\delta > 0$ (minore di r) tale che

$$f(x, y_0 - r) < 0, \ f(x, y_0 + r) > 0 \quad \forall x \in [x_0 - \delta, x_0 + \delta].$$

Dunque, per ogni fissato $x \in [x_0 - \delta, x_0 + \delta]$ la funzione $y \mapsto f(x, y)$ è strettamente crescente su $[y_0 - r, y_0 + r]$ (per quanto detto sopra su f_y), è negativa nell'estremo sinistro dell'intervallo e positiva nell'estremo destro.

Essa si annullerà dunque in un uno ed un sol punto compreso tra $y_0 - r$ e $y_0 + r$, che è esattamente la parte (i) dell'enunciato.

Dimostriamo il punto (ii): per ogni $\overline{x} \in [x_0 - \delta, x_0 + \delta]$ abbiamo definito $g(\overline{x})$ come l'unico numero tra $y_0 - r$ e $y_0 + r$ tale che $f(\overline{x}, g(\overline{x})) = 0$. Se anche $x \in [x_0 - \delta, x_0 + \delta]$ abbiamo

$$(B) \qquad \begin{aligned} &0 = f(x, g(x)) - f(\overline{x}, g(\overline{x})) = \\ &f(x, g(x)) - f(\overline{x}, g(x)) + f(\overline{x}, g(x)) - f(\overline{x}, g(\overline{x})) = \\ &f_x(\xi, g(x))(x - \overline{x}) + f_y(\overline{x}, \eta)(g(x) - g(\overline{x})), \end{aligned}$$

dove ξ è compreso tra x e $\overline{x}$, mentre η è compreso tra $g(x)$ e $g(\overline{x})$: l'esistenza di valori di ξ ed η che rendano vera l'uguaglianza è garantita dal teorema di Lagrange (in una variabile).

Dunque,

$$(C) \quad g(x) - g(\overline{x}) = -\frac{f_x(\xi, g(x))}{f_y(\overline{x}, \eta)}(x - \overline{x}).$$

La frazione nel membro di destra si mantiene limitata nel nostro rettangolino (possiamo maggiorarla con il massimo di $|f_x|$ diviso il minimo di $|f_y|$: quest'ultimo sarà strettamente positivo perché f_y non si annulla nel rettangolo). Dunque, facendo tendere $x \to \overline{x}$ si ha $g(x) \to g(\overline{x})$, e la funzione g è continua.

A questo punto, riprendiamo ancora una volta (C), e scriviamola nella forma

$$\frac{g(x) - g(\overline{x})}{x - \overline{x}} = -\frac{f_x(\xi, g(x))}{f_y(\overline{x}, \eta)}.$$

Facendo tendere $x \to \overline{x}$ si ottiene la formula per la derivata di g. Q.E.D.

OSSERVAZIONE: Può valer la pena di osservare il seguente fatto geometrico: se f e (x_0, y_0) sono come nel teorema, allora grad $f(x_0, y_0)$ è un vettore *ortogonale* all'insieme Z degli zeri di f (mostrarlo per esercizio). Siccome il vettore gradiente di una funzione punta sempre nella *direzione di massima crescita* della funzione stessa, questo fatto non è sorprendente: infatti, sull'insieme Z la funzione non cresce affatto (è costante).

Lezione del 21/5/2005 (2 ore): La volta scorsa abbiamo accennato al fatto che la discussione sulle curve di livello di una funzione regolare ci sarà utile nella soluzione di certi problemi di minimo.

In effetti, tutto l'armamentario dell'Hessiano non è veramente necessario se siamo interessati a trovare i massimi e minimi *assoluti* di una funzione di due variabili $F(x, y)$ in un insieme chiuso e limitato $\overline{A}$ sufficientemente "buono" nel piano $\mathbf{R}^2$.

Tipicamente, $\overline{A}$ sarà una regione del piano delimitata da una curva regolare, che spesso sarà data in forma implicita (ossia il bordo di $\overline{A}$ sarà l'insieme degli zeri di una funzione $f(x, y)$). Per fare un esempio concreto, potremmo avere

$$\overline{A} = \{(x, y) \in \mathbf{R}^2 : \ x^2/4 + y^2 \leq 1\}.$$

In questo caso, l'insieme $\overline{A}$ è la regione di piano delimitata dall'ellisse $x^2/4 + y^2 - 1 = 0$.

In classe, abbiamo colto l'occasione della discussione sull'insieme $\overline{A}$ su cui vogliamo trovare i massimi e i minimi di F, per dare alcune importanti definizioni "topologiche": in $\mathbf{R}^2$, oppure in $\mathbf{R}^n$, un punto (x_0, y_0) si dice *interno* per un insieme A, se esiste un intorno del punto tutto contenuto in A. Analogamente si definisce un punto *esterno* ad A: è un punto che ha un intorno tutto contenuto nel complementare di A.

L'insieme A si dice poi *aperto* se e solo se tutti i suoi punti sono interni. La *frontiera* di A, indicata con ∂A, è l'insieme dei punti che non sono né interni né esterni ad A. La *chiusura di* A è poi l'insieme $\overline{A} = A \cup \partial A$: non è difficile vedere che si tratta del più piccolo chiuso che contiene A.

Infine, l'aperto A si dice *regolare* se ogni punto della sua frontiera ha un intorno rettangolare in cui A coincide con il sottografico di una funzione di una variabile (con opportuna scelta degli assi).

Sia A un aperto limitato regolare: per il teorema di Weierstrass, sappiamo che F ha sia massimo che minimo in $\overline{A}$: si tratta solo di individuarli!

In una sola variabile, è piuttosto semplice trovare i punti di massimo e minimo assoluto di una funzione derivabile su un intervallo chiuso e limitato: basta trovare i punti in cui si annulla la derivata, e confrontare poi i valori assunti dalla funzione in tali punti *e agli estremi dell'intervallo*.

Anche in due variabili, se un punto di massimo o minimo assoluto cade *all'interno di* A, il vettore gradiente si deve annullare (cioè un punto di massimo o minimo relativo interno è necessariamente un punto critico per la funzione F). Purtroppo, però, in questo caso abbiamo *infiniti* punti sulla frontiera ∂A di A, e non è così chiaro come possiamo confrontare il valore di F su *tutti questi punti*!

Se siamo in grado di parametrizzare esplicitamente il bordo di A (cioè se il bordo di A è l'immagine di una funzione vettoriale di due variabili, $t \mapsto (x(t), y(t))$), trovare il massimo e il minimo di F sul bordo di A corrisponde a trovare il massimo e il minimo della funzione di una variabile $F(x(t), y(t))$, cosa che sappiamo fare! Nell'esempio sopra, una parametrizzazione del bordo è per esempio $(2\cos t, \sin t)$ (con $t \in [0, 2\pi]$), per cui dovremmo massimizzare e minimizzare la funzione composta $F(2\cos t, \sin t)$ al variare di t tra 0 e 2π.

Che fare se invece la frontiera di A è l'insieme degli zeri di una funzione regolare di due variabili?

Il teorema delle funzioni implicite ha come conseguenza un'importante ricetta per *trovare i punti di massimo e minimo di una funzione regolare $F(x, y)$ sull'insieme Z degli zeri di un'altra funzione regolare $f(x, y)$* (tipicamente, Z potrà essere la frontiera di un insieme del piano su cui vogliamo trovare i massimi e i minimi assoluti di f).

TEOREMA (Dei moltiplicatori di Lagrange): Sia $f(x, y)$ una funzione derivabile con derivate parziali continue, e supponiamo che il suo insieme degli zeri

$$Z = \{(x, y) : \ f(x, y) = 0\}$$

abbia la proprietà che grad $f(x, y) \neq (0, 0)$ per ogni $(x, y) \in Z$. Sia $F(x, y)$ un'altra funzione di due variabili, derivabile con derivate continue, di cui siamo interessati a trovare i massimi e i minimi su Z.

Allora, se (x_0, y_0) è un punto di massimo o minimo relativo per la restrizione di F all'insieme Z, esiste un numero reale λ tale che

$$\begin{cases} F_x(x_0, y_0) = \lambda f_x(x_0, y_0) \\ F_y(x_0, y_0) = \lambda f_y(x_0, y_0) \end{cases}$$

OSSERVAZIONE: Dal punto di vista operativo, i massimi e i minimi di $F(x, y)$ sull'insieme Z sono da ricercare tra i punti (x, y) soluzioni del seguente sistema di *tre* equazioni nelle *tre incognite* x, y, λ:

$$(A) \begin{cases} F_x(x_0, y_0) = \lambda f_x(x_0, y_0) \\ F_y(x_0, y_0) = \lambda f_y(x_0, y_0) \\ f(x, y) = 0. \end{cases}$$

Il numero λ si chiama *moltiplicatore di Lagrange.*

Geometricamente, risolvere il sistema (A) consiste nel cercare punti sulla curva Z in cui i vettori grad F e grad f *abbiano la stessa direzione.* Poiché sappiamo che il vettore grad f è sempre ortogonale a Z, stiamo dunque cercando i punti della curva in cui la funzione da minimizzare *cresce in direzione normale alla curva stessa* (che è come dire che, nella direzione tangente alla curva, la funzione F "ha derivata 0").

DIM. TEOREMA: Supponiamo per fissare le idee che sia $f_y(x_0, y_0) \neq 0$ (se a non annullarsi è l'altra derivata, basta scambiare il ruolo delle due variabili...).

Grazie al teorema delle funzioni implicite, esiste un intorno di (x_0, y_0) entro il quale l'insieme Z coincide con il grafico di una funzione derivabile $g(x)$. Allora, la funzione di 1 variabile $\phi : x \mapsto F(x, g(x))$ è definita in un intorno di x_0, ed ha un punto di massimo o minimo relativo in x_0, per cui $\phi'(x_0) = 0$.

D'altra parte, ricordando la formula per la derivata di g si ottiene subito:

$$0 = \phi'(x_0) = F_x(x_0, y_0) - F_y(x_0, y_0)\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)}.$$

Basta allora porre

$$\lambda = \frac{F_y(x_0, y_0)}{f_y(x_0, y_0)},$$

e si ottiene la tesi. Q.E.D.

OSSERVAZIONE: L'ipotesi che il gradiente di f non si annulli mai sull'insieme Z è fondamentale. Si prenda per esempio $f(x, y) = (x-1)^3 - y^2$ e $F(x, y) = x^2 + y^2$. Vogliamo cioè trovare il minimo della funzione F (che rappresenta il quadrato della distanza dall'origine) sulla curva Z di equazione $f(x, y) = 0$... Tale minimo dovrà pur esistere!

Il metodo dei moltiplicatori di Lagrange ci suggerisce (suggerirebbe...) di cercare una soluzione del sistema

$$\begin{cases} 2x = 3\lambda(x-1)^2 \\ 2y = -2\lambda y \\ (x-1)^3 - y^2 = 0 \end{cases}$$

Dalla seconda equazione otteniamo $y = 0$ oppure $\lambda = -1$. Nel primo caso la terza equazione implica $x = 1$, ma questo è incompatibile con la prima equazione che diventa $2 = 0$... Se invece $\lambda = -1$, la prima equazione diventa un'equazione di secondo grado senza radici reali: il sistema dato non ha dunque alcuna soluzione.

D'altra parte, facendo un disegnino ci si rende subito conto che il minimo esiste, ed è raggiunto nel punto $(1, 0)$. Perché il metodo non ha funzionato? Semplicemente perchè grad $f(1, 0) = (0, 0)$, per cui il teorema non è affatto utilizzabile! D'altra parte, il nostro disegnino ci avrà mostrato che nel punto $(1, 0)$ la curva Z ha una cuspide...Si tratta di un punto in cui il luogo degli zeri non è una curva regolare!

Lezione del 24/5/2005 (2 ore): In questa lezione vogliamo introdurre il concetto di *successione di Cauchy*, sia in $\mathbf{R}$ che in $\mathbf{R}^n$, e vedere come da questo si ottenga una formulazione dell'assioma di completezza facilmente esportabile a spazi dove non c'è un ordinamento (come, appunto, $\mathbf{R}^n$).

Infatti, per le successioni *reali* esiste un semplice "test di convergenza" che ci permette di dire se una successione possiede un limite finito: precisamente, il limite finito c'è se e soltanto se la successione è di Cauchy.

DEFINIZIONE: Una successione $\{a_n\}_{n\in\mathbf{N}}$ si dice *di Cauchy* se per ogni $\varepsilon > 0$ esiste un indice $\overline{n} \in \mathbf{N}$ tale che

$$|a_n - a_m| < \varepsilon \qquad \forall n \geq \overline{n},\ \forall m \geq \overline{n}.$$

TEOREMA: Una successione $\{a_n\}$ ammette limite finito se e soltanto se essa è di Cauchy.

DIM.: Mostriamo la prima implicazione: supponiamo che esista $\lim\limits_{n\to+\infty} a_n = \ell \in \mathbf{R}$, e mostriamo che la nostra successione è di Cauchy.

Per definizione di limite, esiste $\overline{n} \in \mathbf{N}$ tale che $|a_n - \ell| < \frac{\varepsilon}{2}$ per ogni $n \geq \overline{n}$.

Se ora $m, n \geq \overline{n}$, grazie alla disuguaglianza triangolare avremo

$$|a_n - a_m| \leq |a_n - \ell| + |a_m - \ell| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

e la nostra successione è di Cauchy.

Il viceversa è decisamente più complicato da dimostrare: facciamo vedere che se $\{a_n\}$ è una successione di Cauchy, allora essa ammette un limite finito ℓ.

Cominciamo con l'osservare che una successione di Cauchy è *limitata*: usiamo la definizione di successione di Cauchy con $\varepsilon = 1$. Allora, per m e n maggiori o uguali a un certo indice ν avremo $|a_m - a_n| < 1$. In particolare, se prendiamo $m = \nu$ otteniamo

$$a_\nu - 1 < a_n < a_\nu + 1 \qquad \forall n \geq \nu,$$

e la nostra successione è limitata (a rigore, dalla nostra disuguaglianza rimangono fuori i termini della successione di indice minore di ν, ma essi sono in numero finito e non possono certo renderla illimitata!).

Dunque, $\{a_n\}$ è limitata e possiamo applicare il teorema di Bolzano-Weierstrass: esiste una sottosuccessione $\{a_{n_k}\}$ tale che $\lim\limits_{k\to+\infty} a_{n_k} = \ell \in \mathbf{R}$.

Dico che in realtà *tutta la successione* a_n, e non solo la sua sottosuccessione a_{n_k}, tende a ℓ. Prendiamo $\varepsilon > 0$. Per definizione di successione di Cauchy, troviamo $\overline{n}$ tale che per $m, n \geq \overline{n}$ si abbia $|a_n - a_m| < \frac{\varepsilon}{2}$. D'altra parte, per definizione di limite troviamo $\overline{k}$ tale che $|a_{n_k} - \ell| < \frac{\varepsilon}{2}$ per ogni $k \geq \overline{k}$.

Evidentemente, non è restrittivo supporre che $\overline{k}$ sia tanto grande che $n_{\overline{k}}$ sia maggiore o uguale a $\overline{n}$ (se così non fosse, basterà sostituire $\overline{k}$ con un valore opportunamente più grande). Allora, se $n > n_{\overline{k}}$ avremo

$$|a_n - \ell| \leq |a_n - a_{n_{\overline{k}}}| + |a_{n_{\overline{k}}} - \ell| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Q.E.D.

OSSERVAZIONE: Il teorema appena dimostrato può essere preso come *formulazione equivalente dell'assioma di completezza.*

Ci si convince subito che in **Q** esistono delle successioni di Cauchy che non convergono ad alcun numero razionale (perché esse, se le si guarda come successioni reali, convergono ad un numero irrazionale!). Si noti che l'altra metà del teorema è invece vera anche in **Q**: ogni successione convergente è di Cauchy.

Nel dimostrare che ogni successione di Cauchy in **R** converge ad un limite finito, abbiamo fatto uso dell'assioma di completezza di **R** (dove?). Viceversa, se assumiamo *come assioma* che ogni successione di Cauchy converge, possiamo mostrare che vale l'assioma di completezza in qualche altra sua formulazione: per esempio, possiamo mostrare che ogni insieme non vuoto ed inferiormente limitato ammette estremo inferiore in **R** (Questo è relativamente facile. Nella seconda lezione del corso (quella del 23 settembre), abbiamo escogitato un algoritmo per identificare l'inf come numero decimale infinito: non è difficile verificare che la successione di decimali finiti trovata con quell'algoritmo è una successione di Cauchy!).

La formulazione dell'assioma di completezza tramite le successioni di Cauchy, non è particolarmente "conveniente" per quanto riguarda i numeri reali (anche se ci tornerà utile sapere che le successioni di Cauchy convergono).

Essa si presta però particolarmente bene ad essere generalizzata a situazioni più complicate. Il vantaggio principale è che nella definizione di successione di Cauchy non c'è bisogno di sapere che $\{a_n\}$ vive *in un insieme ordinato.* Per questo motivo, l'assioma di completezza viene formulato in questo modo negli spazi funzionali, o negli spazi metrici in genere.

DEFINIZIONE: Sia X un insieme. Una *distanza* su X è una funzione $d : X \times X \to [0, +\infty)$ tale che

(i) $d(x, y) = 0$ se e solo se $x = y$;

(ii) $d(x, y) = d(y, x)$ per ogni $x, y \in X$ (simmetria);

(iii) $d(x, y) \leq d(x, z) + d(z, y)$ per ogni $x, y, z \in X$ (disuguaglianza triangolare).

Un insieme X dotato di distanza d si dice *spazio metrico.*

Data una successione $\{x_n\} \subset X$ e $\overline{x} \in X$, si dice che $x_n \to \overline{x}$ (rispetto alla distanza d) se e soltanto se $d(x_n, \overline{x}) \to 0$.

Infine, uno spazio metrico (X, d) si dice *completo* se e solo se ogni successione di Cauchy $\{x_n\}$ converge ad un elemento di X. Ovviamente, $\{x_n\}$ si dice di Cauchy se per ogni $\varepsilon > 0$ esiste $\nu \in \mathbf{N}$ tale che per ogni $m, n \geq \nu$ vale $d(x_n, x_m) < \varepsilon$.

ESEMPI: Ecco qualche esempio di spazio metrico.

(i) $\mathbf{R}^n$ con la distanza euclidea $d(x, y) = |x - y|$;

(ii) $\mathbf{R}^n$ con la distanza $d_\infty(x, y) = \max\{|x_i - y_i| : \ i = 1, 2, \ldots, n\}$;

(ii) $\mathbf{R}^n$ con la distanza $d_1(x, y) = \sum_{i=1}^{n} |x_i - y_i|$ (city-block metric);

(iii) Un insieme qualunque X con la *metrica discreta*

$$d(x, y) = \begin{cases} 0 & \textit{se } x = y; \\ 1 & \textit{altrimenti}; \end{cases}$$

.

Ci siamo divertiti a verificare che $\mathbf{R}^n$ dotato di ciascuna delle tre metriche (i), (ii) e (iii) ha le stesse successioni convergenti. In particolare, una successione di vettori di $\mathbf{R}^n$ converge ad un certo vettore se e soltanto se le n successioni reali formate dalla componenti convergono alla rispettiva componente del vettore limite.

La metrica discreta ha invece pochissime successioni convergenti: una successione converge in quella metrica se e solo se è definitivamente costante. In compenso, si tratta di uno spazio metrico completo (assai poco interessante, in verità...).

Verifichiamo ora che gli spazi euclidei $\mathbf{R}^n$ sono completi:

COROLLARIO: Se $\{x_k\}$ è una successione di Cauchy in $\mathbf{R}^n$, allora esiste $\overline{x} \in \mathbf{R}^n$ tale che $x_k \to \overline{x}$.

DIM.: La definizione di successione di Cauchy in $\mathbf{R}^n$ è la solita: per ogni $\varepsilon > 0$ esiste $\nu \in \mathbf{N}$ tale che per $h, k \geq \nu$ vale $|x_k - x_h| < \varepsilon$ (con l'unica differenza che al posto del modulo c'è la norma in $\mathbf{R}^n$...).

In particolare, le successioni reali corrispondenti alle *componenti* degli x_k (cioè, se $x_k = (x_k^1, x_k^2, \ldots, x_k^n)$, allora per $i = 1, \ldots, n$ le successioni reali $\{x_k^i\}_k$ sono di Cauchy in $\mathbf{R}$. Per la completezza di $\mathbf{R}$ dimostrata prima, esisterà

$x^i \in \mathbf{R}$ tale che $x_k^i \to x^i$ per $k \to +\infty$. Allora la successione vettoriale x_k converge al vettore $x = (x^1, x^2, \dots, x^n)$[14]. Q.E.D.

In $\mathbf{R}^n$ valgono anche il teorema di Bolzano-Weierstrass e il teorema di Weierstrass: in effetti, quest'ultimo è stato utilizzato piuttosto spesso in questo corso, per cui è decisamente venuta l'ora di dimostrarlo!

TEOREMA (Bolzano-Weierstrass): Una successione limitata in $\mathbf{R}^n$ ammette sempre una sottosuccessione convergente.

DIM.: Sia $\{x_k\}_k$ la nostra successione limitata, con $x_k = (x_k^1, \dots, x_k^n)$. La successione (reale) $\{x_k^1\}_k$ delle prime componenti è limitata: per il teorema di Bolzano Weiestrass in $\mathbf{R}$ possiamo estrarne una sottosuccessione $\{x_{k_h}^1\}$ convergente ad un numero reale x^1. Questo è come dire che la sottosuccessione (vettoriale) $\{x_{k_h}\}_h$ ha la proprietà che la successione delle *prime componenti* converge... Consideriamo la successione delle seconde componenti di questa, $\{x_{k_h}^2\}_h$: essa è una successione reale limitata, per cui possiamo estrarre un'ulteriore sottosuccessione convergente, $\{x_{k_{h_l}}^2\}$...

Proseguendo in questo modo, in n passi estraiamo una sottosuccessione per cui tutte le n componenti convergono. Q.E.D.

Vedremo la prossima volta come dal teorema di Bolzano Weierstrass segua immediatamente che una funzione continua ammette massimo e minimo assoluto su un insieme chiuso e limitato di $\mathbf{R}^n$.

Lezione del 28/5/2005 (2 ore): Come promesso, eccoci al teorema di Weierstrass per funzioni di più variabili:

TEOREMA (Weierstrass): Una funzione continua $f : C \to \mathbf{R}$, con C chiuso e limitato in $\mathbf{R}^n$, ammette massimo e minimo assoluti in C.

DIM.: Sia $M = \sup\{f(x) : \ x \in C\}$. Per definizione di estremo superiore, esiste una successione $y_k \subset f(C)$ tale che $y_k \to M$.

Ora, si ha $y_k = f(x_k)$ per appropriati punti $x_k \in C$: consideriamo la successione $\{x_k\} \subset C$. Essa è limitata perché lo è C: per il teorema di Bolzano-Weierstrass esiste una sottosuccessione $\{x_{k_h}\}$ e $\overline{x} \in \mathbf{R}^n$ tali che $x_{k_h} \to \overline{x}$. Siccome C è chiuso (e quindi contiene tutti i suoi punti di accumulazione), avremo che $\overline{x} \in C$.

[14]È infatti facile verificare che una successione in $\mathbf{R}^n$ converge se e solo se le n successioni delle sue componenti convergono in $\mathbf{R}$. Questo discende dalla seguente coppia di disuguaglianze. Se $x = (x^1, \dots, x^n) \in \mathbf{R}^n$, allora si ha

$$\max\{|x^1|, |x^2|, \dots, |x^n|\} \le |x| \le \sqrt{n} \max\{|x^1|, |x^2|, \dots, |x^n|\}.$$

Allora, usando la continuità di f in $\overline{x}$ si ha:

$$M = \lim_{h\to+\infty} f(x_{k_h}) = f(\overline{x}),$$

e $\overline{x}$ è un punto di massimo per f in C.

In maniera del tutto analoga si vede che esiste il minimo. Q.E.D.

Veniamo ad un risultato importante, valido in $\mathbf{R}^n$ ma anche in qualunque spazio metrico completo:

TEOREMA (delle contrazioni): Sia (X, d) uno spazio metrico completo, $T : X \to X$ una contrazione, cioè una funzione tale che esista $\delta \in (0, 1)$ tale che $d(T(x), T(y)) \leq \delta d(x, y)$ per ogni $x, y \in X$.[15] Allora esiste ed è unico un punto fisso della funzione T, cioè esiste un unico punto $x^* \in X$ tale che $x^* = T(x^*)$.

DIM.: Fissiamo $x_0 \in X$ e costruiamo per ricorrenza la successione $\{x_n\}$ tale che $x_{n+1} = T(x_n)$ per ogni $n \geq 0$.

Mostriamo che la successione $\{x_n\}$ converge, e converge proprio ad un punto fisso per T: lo facciamo provando che $\{x_n\}$ è una successione di Cauchy.

Intanto, usando la definizione di contrazione si ottiene per ogni n:

$$\begin{aligned} d(x_{n+1}, x_n) &= d(T(x_n), T(x_{n-1}) \leq \delta d(x_n, x_{n-1}) = \\ &\quad \delta d(T(x_{n-1}), T(x_{n-2})) \leq \delta^2 d(x_{n-1}, x_{n-2}) = \ldots \leq \delta^n d(x_1, x_0). \end{aligned}$$

Se adesso h è un qualunque numero naturale ≥ 1 avremo:

$$d(x_{n+h}, x_n) \leq d(x_{n+h}, x_{n+h-1}) + d(x_{n+h-1}, x_{n+h-2}) + \ldots + d(x_{n+1}, x_n) \leq$$
$$d(x_1, x_0)(\delta^{n+h-1} + \delta^{n+h-2} + \ldots + \delta^n) = d(x_1, x_0)\delta^n \frac{1-\delta^h}{1-\delta} \leq d(x_1, x_0)\frac{\delta^n}{1-\delta}$$

L'ultima espressione non dipende da h e tende a zero per $n \to +\infty$: possiamo dunque trovare $\overline{n}$ tale che per ogni $n \geq \overline{n}$ e per ogni h sia $d(x_{n+h}, x_n) < \varepsilon$, e la successione $\{x_n\}$ è di Cauchy. Sia x^* il suo limite: passando al limite per $n \to +\infty$ nella relazione $x_{n+1} = T(x_n)$ otteniamo $x^* = T(x^*)$, per cui x^* è il punto fisso cercato.

Mostriamone l'unicità: se $\tilde{x}$ è un altro punto fisso (cioè $\tilde{x} = T(\tilde{x})$) avremo

$$d(x^*, \tilde{x}) = d(T(x^*), T(\tilde{x})) \leq \delta d(x^*, \tilde{x}),$$

[15] Una contrazione riduce le distanze di un fattore fisso $\delta < 1$.

da cui $(1-\delta)d(x^*,\tilde{x}) \leq 0$, quindi $d(x^*,\tilde{x}) = 0$ e $x^* = \tilde{x}$. Q.E.D.

OSSERVAZIONE: Come già accennato, il Teorema delle contrazioni è applicabile ad una contrazione $T : \mathbf{R}^n \to \mathbf{R}^n$, perché $\mathbf{R}^n$ è uno spazio metrico completo. Ma la cosa funziona anche se $T : C \to C$, con C sottinsieme *chiuso* di $\mathbf{R}^n$. Infatti, C è uno spazio metrico (con la metrica euclidea ereditata da $\mathbf{R}^n$). È poi completo: una successione di Cauchy in C converge a qualche punto di $\mathbf{R}^n$ (grazie alla completezza di $\mathbf{R}^n$). Questo limite appartiene poi a C grazie alla chiusura dell'insieme.

Vedremo l'anno prossimo come il teorema delle contrazioni ci consentirà di generalizzare il Teorema del Dini a funzioni da $\mathbf{R}^n$ in $\mathbf{R}^k$ (caso generale del teorema delle funzioni implicite).

Per ora, vogliamo vederne un'altra importantissima applicazione: lo useremo per dimostrare il teorema di Cauchy-Lipschitz di esistenza e unicità locale per un'equazione differenziale ordinaria del I ordine.

A questo scopo, è necessario introdurre una metrica sullo spazio delle funzioni continue che le renda uno spazio metrico completo:

DEFINIZIONE: Denotiamo con $\mathcal{C}^0([a,b])$ lo spazio vettoriale delle funzioni continue sull'intervallo $[a,b]$. Se $f, g \in \mathcal{C}^0([a,b])$, definiamo la loro distanza nel modo seguente:

$$d_\infty(f,g) = \max\{|f(x) - g(x)| : \ x \in [a,b]\}.$$

La distanza d_∞ è nota come *distanza della convergenza uniforme*, perchè se $\{f_n\}$, f sono tali che $d_\infty(f_n, f) \to 0$, allora si dice che $f_n \to f$ *uniformemente in* $[a,b]$.

TEOREMA: Lo spazio $(\mathcal{C}^0([a,b]), d_\infty)$ *è uno spazio metrico completo.*

DIM.: Bisogna innanzitutto verificare che d_∞ è una distanza. Intanto è ben definita: il massimo esiste grazie al teorema di Weierstrass. E' poi evidentemente non negativa e simmetrica. Inoltre $d_\infty(f,g) = 0$ se e solo se $|f(x) - g(x)| = 0$ per ogni $x \in [a,b]$, cioè se e solo se $f = g$. Anche la disuguaglianza triangolare segue banalmente dalla disuguaglianzatriangolare in $\mathbf{R}$ e dal fatto che il massimo della somma di due funzioni è minore o uguale alla somma dei massimi. Abbiamo uno spazio metrico!

Sia poi $\{f_n\}$ una successione di Cauchy in $\mathcal{C}^0([a,b])$: questo significa che per ogni $\varepsilon > 0$ esiste $\nu \in \mathbf{N}$ tale che se $m, n \geq \nu$ allora $d_\infty(f_n, f_m) < \varepsilon$. In particolare, questo implica che

$$(*)|f_n(x) - f_m(x)| < \varepsilon \quad \forall x \in [a,b],\ \forall m,n \geq \nu.$$

Abbiamo così scoperto che per ogni fissato $x \in [a,b]$ la successione reale $\{f_n(x)\}$ è di Cauchy: essa convergerà dunque ad un numero reale che chiameremo $f(x)$ (grazie alla completezza di $\mathbf{R}$). Per concludere la dimostrazione, basta far vedere che la funzione f è continua e che $f_n \to f$ rispetto alla distanza d_∞. Il secondo fatto è il più semplice da dimostrare: da (*), passando al limite per $m \to +\infty$ otteniamo

$$(**) \ |f_n(x) - f(x)| \le \varepsilon \quad \forall x \in [a,b], \ \forall n \ge \nu,$$

da cui $d_\infty(f_n, f) \le \varepsilon$ per ogni $n \ge \nu$, come volevasi.

Mostriamo che f è continua: partiamo ancora da $(**)$ con $m = \nu$. La funzione f_ν è (uniformemente) continua: esiste $\delta > 0$ tale che se $|x-y| < \delta$ allora $|f_\nu(x) - f_\nu(y)| < \varepsilon$. Allora, se $|x-u| < \delta$ da quest'ultima relazione e da $(**)$ deduciamo:

$$|f(x) - f(y)| \le |f(x) - f_\nu(x)| + |f_\nu(x) - f_\nu(y)| + |f_\nu(y) - f(y)| < 3\varepsilon,$$

che è l'uniforme continuità di f. Q.E.D.

OSSERVAZIONE: La convergenza uniforme implica la *convergenza puntuale*: se $d_\infty(f_n, f) \to 0$, allora per ogni fissato $x \in [a,b]$ si ha $\lim_{n\to+\infty} f_n(x) = f(x)$.

Il viceversa non è vero: esistono successioni che convergono puntualmente ma non uniformemente. In particolare, abbiamo visto che esistono successioni di funzioni continue che convergono puntualmente ad una funzione discontinua (cosa che *non può succedere* se c'è convergenza uniforme: l'ultimo teorema mostra che il limite uniforme di una successione di funzioni continue è una funzione continua.

L'esempio che abbiamo visto è la successione $f_n : [0,1] \to \mathbf{R}$, $f_n(x) = x^n$.

Lezione del 31/5/2005 (2 ore): Siamo finalmente in grado di dimostrare il Teorema di Cauchy-Lipschitz sulle equazioni differenziali ordinarie:

TEOREMA (Di esistenza e unicità locale): Sia $f : [a,b] \times [c,d] \to \mathbf{R}$ una funzione di due variabili continua, per cui esiste una costante $L > 0$ tale che

$$|f(t,w) - f(t,z)| \le L|w-z| \quad \forall t \in [a,b], \ \forall w, z \in [c,d]$$

(si dice in tal caso che f è lipschitziana nella seconda variabile, con una costante che non dipende dalla prima).

Siano poi $t_0 \in (a,b)$, $y_0 \in (c,d)$. Allora esiste $\delta > 0$ ed una funzione derivabile $y(t) : [t_0 - \delta, t_0 + \delta] \to \mathbf{R}$ che risolve il problema di Cauchy

$$\begin{cases} y'(t) = f(t, y(t)) \\ y(t_0) = y_0. \end{cases}$$

Questa soluzione è unica, nel senso che ogni altra soluzione del problema coincide con $y(t)$ nell'intervallo in cui sono definite entrambe.

DIM.: Supponiamo dapprima, per semplicità, che $f(t, y)$ sia definita (e lipschitziana) per ogni y (cioè che $f : [a, b] \times \mathbf{R} \to \mathbf{R}$). Vedremo poi come fare per togliere quest'ipotesi in più.

Fissato $\delta > 0$ in modo che $[t_0 - \delta, t_0 + \delta] \in [a, b]$, definiamo un'applicazione $T : \mathcal{C}^0([a, b]) \to \mathcal{C}^0([a, b])$ nel modo seguente: se $y(t) \in \mathcal{C}^0([a, b])$ e $g = T(y)$, allora

$$(T(y))(t) = g(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds.$$

Chiaramente, T manda lo spazio $\mathcal{C}^0([a, b])$ in se stesso (grazie al teorema fondamentale del calcolo integrale, la funzione g è addirittura derivabile!). Diciamo che, se abbiamo cura di scegliere δ abbastanza piccolo, T è anche una contrazione.

Siano infatti $y_1(t), y_2(t)$ funzioni continue, $g_1 = T(y_1)$, $g_2 = T(y_2)$: per ogni $t \in [t_0 - \delta, t_0 + \delta]$ si ha

$$\begin{aligned}
&|g_1(t) - g_2(t)| = |\int_{t_0}^{t} [f(s, y_1(s)) - f(s, y_2(s))]\, ds| \leq \\
&|\int_{t_0}^{t} |f(s, y_1(s)) - f(s, y_2(s))|\, ds| \leq L|\int_{t_0}^{t} |y_1(s) - y_2(s)|\, ds| \leq \\
&L|t - t_0|\, d_\infty(y_1, y_2) \leq L\delta\, d_\infty(y_1, y_2)
\end{aligned}$$

Prendendo il massimo per $t \in [t_0 - \delta, t_0 + \delta]$ otteniamo

$$d_\infty(T(y_1), T(y_2)) \leq L\delta\, d_\infty(y_1, y_2)$$

e T è una contrazione se $L\delta < 1$, cosa vera per δ abbastanza piccolo.

Per il teorema delle contrazioni, esiste allora un'unica funzione $\overline{y}(t)$ tale che $T(\overline{y}) = \overline{y}$, cioè tale che

$$(**)\ \overline{y}(t) = y_0 + \int_{t_0}^{t} f(s, \overline{y}(s))\, ds \quad \forall t \in [t_0 - \delta, t_0 + \delta].$$

La funzione $\overline{y}$ soddisfa la condizione iniziale (basta prendere $t = t_0$ nell'identità integrale). Inoltre, applicando il teorema fondamentale del calcolo al secondo membro di $(**)$, otteniamo che $\overline{y}$ è derivabile e $\overline{y}'(t) = f(t, \overline{y}(t))$: abbiamo trovato una soluzione del nostro problema di Cauchy.

L'unicità del punto fisso ci regala anche l'unicità della soluzione: integrando l'equazione differenziale tra t_0 e t, si vede subito che ogni soluzione del problema di Cauchy soddifa l'equazione integrale $(**)$, ed è cioè un punto fisso di T.

Per concludere la dimostrazione, dobbiamo però affrontare il caso generale in cui $f : [a, b] \times [c, d] \to \mathbf{R}$. In questo caso, la difficoltà sta nel fatto che la funzione T non si può più definire per tutte le $y \in \mathcal{C}^0([t_0 - \delta, t_0 + \delta])$, ma solo per le funzioni a valori in $[c, d]$.

Fissiamo allora $r > 0$ tale che $[y_0 - r, y_0 + r] \subset [c, d]$ e definiamo

$$\mathcal{C} = \{y \in \mathcal{C}^0([t_0 - \delta, t_0 + \delta]) : \ y(t) \in [y_0 - r, y_0 + r] \ \forall t \in [t_0 - \delta, t_0 + \delta]\}.$$

Notiamo che $\mathcal{C}$ è uno spazio metrico completo con la metrica d_∞: una succesione di Cauchy in $\mathcal{C}$ converge infatti ad una funzione continua grazie alla completezza di $(\mathcal{C}^0, d_\infty)$. Tale funzione avrà poi immagine contenuta in $[y_0 - r, y_0 + r]$ perchè tutte le funzioni della successione sono di questo tipo, e l'intervallo è chiuso.

Possiamo allora ripetere pari pari il ragionamento fatto sopra, purché facciamo vedere che per δ abbastanza piccolo T manda $\mathcal{C}$ in se stesso. A questo scopo, poniamo $M = \max\{|f(t, y)| : \ (t, y) \in [a, b] \times [c, d]\}$. Sia dunque $y \in \mathcal{C}$, $g = T(y)$. Allora

$$|g(t) - y_0| = |\int_{t_0}^{t} f(s, y(s)) \, ds| \leq |\int_{t_0}^{t} |f(s, y(s)) \, ds| \leq M\delta.$$

Chiaramente, per $\delta \leq r/M$ vale $M\delta \leq r$, per cui $g \in \mathcal{C}$. Q.E.D.

OSSERVAZIONI: Il teorema si dimostra allo stesso modo (sostituendo un po' di moduli con delle norme in $\mathbf{R}^n$) se f e l'incognita $y(t)$ sono funzioni vettoriali. Dunque il teorema di Cauchy-Lipschitz risulta dimostrato anche per sistemi di n equazioni in n incognite.

Vale anche la pena di osservare una cosa sulla nostra dimostrazione del "caso semplificato" in cui $f(t, y)$ è definita su $[a, b] \times \mathbf{R}$ ed è lipschitziana nella seconda variabile, uniformemente rispetto alla prima. In quel caso il nostro ragionamento ci fornisce in realtà *esistenza globale* della soluzione in tutto l'intervallo $[a, b]$: infatti, abbiamo ottenuto una soluzione su un intervallo del tipo $[t_0 - \delta, t_0 + \delta]$ con l'unica condizione che $\delta < 1/L$. L'intervallo $[a, b]$ può essere ricoperto con un numero finito di intervallini di questo tipo: prendendo come condizione iniziale $(t_0 + \delta, y(t_0 + \delta))$ possiamo estendere la nostra soluzione su $[t_0 + \delta, t_0 + 2\delta]$ e così via... In un numero finito di passi troviamo la soluzione globale!

Questo ragionamento si applica in particolare ad equazioni (e sistemi!) lineari: è per questo che il problema di Cauchy per un'equazione lineare di ordine n ammmette una soluzione globale, definita su tutto l'intervallo in cui sono definiti i coefficienti.

Alla fine di questa carrellata sullo spazio $\mathcal{C}^0$ delle funzioni continue, abbiamo voluto fare un breve discorso anche sullo spazio $\mathcal{C}^1([a,b])$ delle funzioni continue con derivate continue. Su questo spazio la distanza si definisce come

$$d_{\mathcal{C}^1}(f,g) := d_\infty(f,g) + d_\infty(f',g').$$

Vogliamo mostrare che anche lo spazio metrico $(\mathcal{C}^1, d_{\mathcal{C}^1})$ è completo. Sia dunque $\{f_n\}$ una successione di Cauchy in $\mathcal{C}^1$: a maggior ragione, le successioni di funzioni continue $\{f_n\}$ e $\{f'_n\}$ sono di Cauchy in $(\mathcal{C}^0, d_\infty)$ (che è completo!) per cui esistono due funzioni continue f, g tali che $f_n \to f$, $f'_n \to g$ uniformemente in $[a,b]$. Per concludere la dimostrazione, resta solo da far vedere che f è derivabile e g è la sua derivata (perchè allora, chiaramente, $f_n \to f$ nella metrica $\mathcal{C}^1$).

Questo è conseguenza del seguente lemma, assai importante anche per meriti propri:

LEMMA: Se $\{f_n\}$ è una successione di funzioni continue che converge uniformemente a f in $[a,b]$, allora

$$\lim_{n\to+\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.$$

DIM.: Stimiamo la differenza degli integrali:

$$|\int_a^b (f_n(x) - f(x))\,dx \leq \int_a^b |f_n(x) - f(x)|\,dx \leq (b-a)d_\infty(f_n, f) \to 0.$$

Q.E.D.

Concludiamo la dimostrazione della completezza di $\mathcal{C}^1$: per ogni fissato x e per ogni n si ha

$$f_n(x) = f_n(a) + \int_a^x f'_n(t)\,dt.$$

Passando al limite per $n \to +\infty$ e usando il lemma si ottiene

$$f(x) = f(a) + \int_a^x g(t)\,dt,$$

per cui f è derivabile con derivata g, come volevasi.

OSSERVAZIONE: Il lemma è falso se si ha solo la convergenza puntuale di f_n a f. Per esempio, le funzioni $f_n : [0,1] \to \mathbf{R}$ definite come

$$f_n(x) = \begin{cases} n^2 x & \textit{se } 0 \leq x \leq 1/n, \\ -n^2 x + 2n & \textit{se } 1/n < x \leq 2/n, \\ 0 & \textit{altrimenti} \end{cases}$$

convergono puntualmente a 0 per ogni $x \in [0, 1]$, ma il loro integrale vale 1.

OSSERVAZIONE: Alla luce della definizione di convergenza uniforme, siamo in grado di reinterpretare la nostra dimostrazione della regolarità della somma di una serie di potenze (si veda la lezione del 26/4). Il Lemma 1 dice che le somme parziali di una serie di potenze *convergono uniformemente* sui sottointervalli chiusi e limitati dell'insieme di convergenza. Il Lemma 2, che ci dava la continuità della somma, dipende sostanzialmente dal fatto che il limite uniforme di una successione di funzioni continue è una funzione continua. Infine, il Lemma 3 non è altro che il lemma di convergenza integrale appena visto, applicato alla successione delle somme parziali della serie di potenze.