Funzioni di n variabili a valori scalari

Ultimo aggiornamento: 3 febbraio 2018

Cominciamo a studiare le funzioni di più variabili. Nella maggior parte dei casi studieremo funzioni a valori scalari, ma alcune cose le vedremo per funzioni a valori vettoriali.
Un utile esercizio è cercare di abituarsi ad immaginare grafici di funzioni da $\mathbf{R}^2$ in $\mathbf{R}$. A titolo di esempio proviamo con la funzione $f(x,y) = x^2 + y^2$. Per cercare di capire come è fatto un grafico può essere utile restringere la funzione a rette. Cominciamo considerando rette parallele agli assi: se consideriamo per esempio la retta $(x,0)$ la funzione ristretta a tale retta è

$$f(x,0) = x^2 \,.$$

Se consideriamo la retta $(x,1)$ la funzione ristretta a tale retta è $f(x,1) = x^2 + 1$. Se fissiamo $\bar{y}$ e restringiamo f alla retta $(x,\bar{y})$ abbiamo la parabola (si veda la Figura 1.a per alcuni esempi)

$$f(x,\bar{y}) = x^2 + \bar{y}^2 \,.$$

Analogamente si ottengono parabole fissando $\bar{x}$ e considerando $f(\bar{x},y) = \bar{x}^2 + y^2$ oppure fissando rette per l'origine. Per considerare tutte le rette per l'origine, compresi gli assi cartesiani che abbiamo già considerato, è sufficiente fissare un vettore $v = (v_1, v_2)$ che per semplicità considereremo di norma 1, cioè soddisfacente $v_1^2 + v_2^2 = 1$, e considerare la curva $r(t) := tv$ con $t \in \mathbf{R}$.
In questo caso si ottiene $f \circ r(t) = f(r(t)) = f(tv_1, tv_2) = t^2 v_1^2 + t^2 v_2^2 = t^2$. Si osservi che $f(r(t)) = t^2$ per ogni vettore $v \in \mathbf{R}^2$ di norma 1. Questo significa che considerando la restrizione di f ad una qualunque retta per l'origine il grafico è sempre lo stessa parabola. Questa serie di informazioni può farci intuire che il grafico è quello indicato in Figura 1.c o in Figura 1.d. La differenza tra i due è che in Figura 1.c il dominio è un quadrato, in Figura 1.d il dominio è un cerchio. Si osservi come il grafico in Figura 1.d può essere ottenuto ruotando una qualunque delle parabole in Figura 1.b.

Si noti che questo procedimento equivale a sezionare il grafico di f

con i piani verticali $\{(x,y,z) \in \mathbf{R}^3 \,|\, y = \bar{y}\}$, che brevemente scriveremo $y = \bar{y}$, nel primo caso (o analogamente con i piani $x = \bar{x}$); con i piani verticali $\{(x,y,z) \in \mathbf{R}^3 \,|\, x = tv_1\,, y = tv_2\, t \in \mathbf{R}\}$ dove v è un vettore di norma 1.
Il procedimento di sezionare il grafico, abbinato alla ricerca degli insiemi di livello, non è una tecnica generale, ma può essere qualche volta utile ad avere informazioni, anche se parziali, sul grafico della funzione che si sta considerando.

Figura 1.a

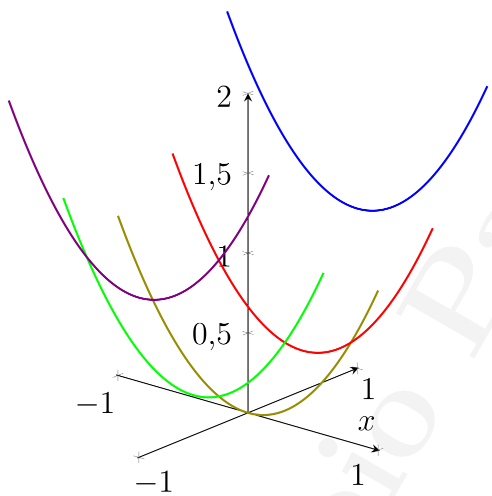


Figura 1.b

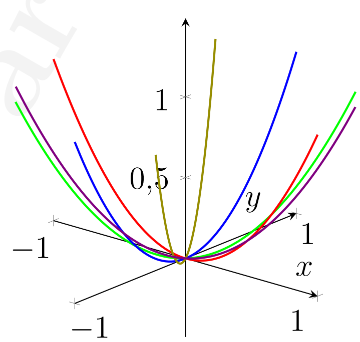


Figura 1.c

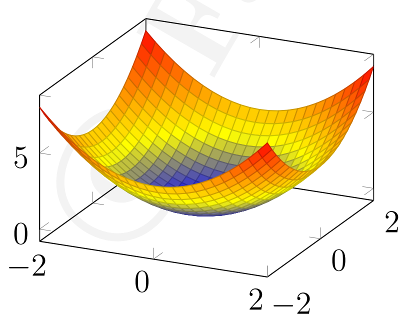
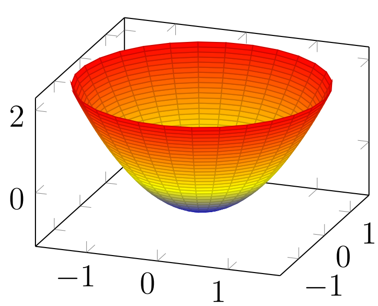


Figura 1.d



Il caso più semplice è quello in cui una funzione f di due variabili x, y è il prodotto di due funzioni g, h di una variabile definita come segue

$$f(x,y) = g(x)h(y)\,.$$

In tal caso, fissando prima una variabile e poi l'altra è ancora più semplice immaginare il grafico. Ad esempio, immaginare il grafico della funzione (il cui grafico è parzialmente mostrato in Figura 1.e limitatamente al rettangolo $x \in [-1, 3]$ e $y \in [0, \pi]$)

$$f(x, y) = e^x \operatorname{sen} y$$

è relativamente semplice. Fissando prima un valore per y, sia $\bar{y}$, ci si rende conto che sopra la retta $y = \bar{y}$ la restrizione del grafico di f è un'esponenziale moltiplicata per il fattore $\operatorname{sen} \bar{y}$. In particolare per $y = 0$ e $y = \pi$ tale restrizione è nulla.
Viceversa, fissando $x = \bar{x}$, si ha che la restrizione del grafico di f alla retta $x = \bar{x}$ è il grafico della funzione seno moltiplicata per $e^{\bar{x}}$.

Figura 1.e

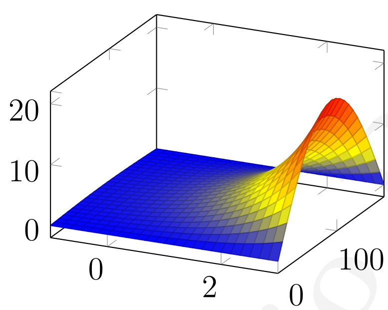
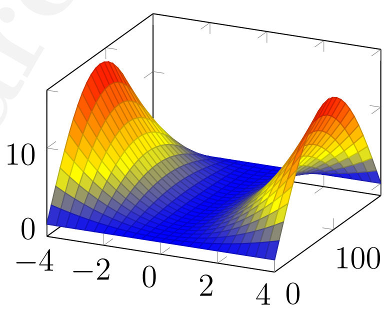


Figura 1.f



Analogamente si può fare per la funzione $g(x, y) = x^2 \operatorname{sen} y$, il cui grafico, limitatamente al rettangolo $[-4, 4] \times [0, \pi]$, è mostrato in Figura 1.f.

Altro esempio molto semplice, ma allo stesso tempo non banale, è il seguente: si consideri la funzione

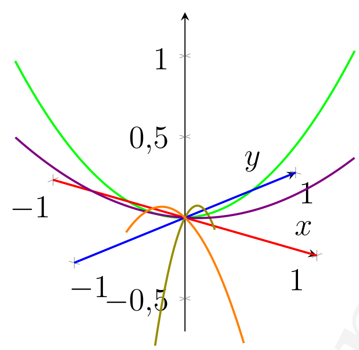
$$f(x, y) = xy\,.$$

Anche questa funzione, come il primo esempio visto, è un polinomio di secondo grado. Se si considerano le restrizioni di tale funzione alle rette per l'origine del tipo

$$y = kx \quad \text{oppure} \quad x = ky \qquad \text{con} \qquad k \neq 0$$

ci si accorge subito che tali grafici sono parabole, che degenerano ad una retta nel caso in cui $k = 0$, come evidenziato nella figura che segue

Figura 1.g



1. Insiemi di livello

Data una funzione $f : A \to \mathbf{R}$ ed una costante $c \in \mathbf{R}$ ci si può chiedere qual è l'insieme

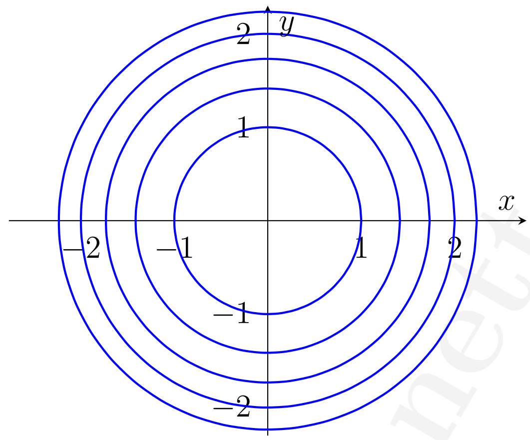
$$\Gamma_c = \left\{ x \in A \,\middle|\, f(x) = c \right\}.$$

Procediamo sempre con l'esempio di $f(x, y) = x^2 + y^2$ definita in $\mathbf{R}^2$ e chiediamoci quali sono gli insiemi di livello. Si osserva subito che

$$\Gamma_c = \begin{cases} \emptyset & \text{se } c < 0\,, \\ \{(0,0)\} & \text{se } c = 0\,, \\ \{(x,y) \in \mathbf{R}^2 \mid x^2 + y^2 = c\} & \text{se } c > 0\,, \end{cases}$$

cioè, se $c > 0$, è una circonferenza di raggio $\sqrt{c}$, se $c = 0$ abbiamo un solo punto, se $c < 0$ abbiamo l'insieme vuoto.

Figura 2



In Figura 2 sono riportati gli insiemi di livello corrispondenti alle costanti $1, 2, 3, 4, 5$. Si noti come dall'origine, l'insieme Γ_0 di livello 0, alla prima curva, l'insieme Γ_1 di livello 1, ci sia molto più spazio che tra Γ_1 e Γ_2, e così via. Questi curve ci forniscono una doppia informazione: che la forma dell'insieme sul quale la funzione assume uno stesso valore è una circonferenza e che all'aumentare della costante c che indica il livello la funzione cresce sempre più velocemente poiché gli insiemi invece si allontanano tra loro sempre più lentamente. Se idealmente ci muovessimo lungo una semiretta uscente dall'origine per salire dal livello c al livello $c+1$ dovrei percorrere un segmento lungo $\sqrt{c+1} - \sqrt{c} = \dfrac{1}{\sqrt{c+1} + \sqrt{c}}$, che evidentemente decresce all'aumentare di c.

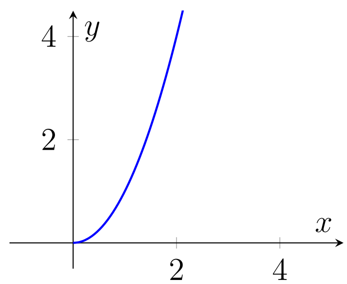
Spesso per funzioni da $\mathbf{R}^2$ in $\mathbf{R}$ si parla di *curve di livello*, anche se gli insiemi non sono sempre delle curve.

La funzione appena vista è detta *radiale*. Una funzione $f : \mathbf{R}^n \to \mathbf{R}$ è detta radiale se esiste una funzione $\tilde{f} : [0, +\infty) \to \mathbf{R}$ tale che

$$f(x_1, \ldots x_n) = \tilde{f}(|x|) \, ,$$

cioè il valore della funzione f nel generico punto $(x_1, \ldots x_n)$ dipende di fatto solo dalla distanza dall'origine di tale punto. La funzione $\tilde{f}$ nel caso di $f(x, y) = x^2 + y^2$ è semplicemente $\tilde{f}(\rho) = \rho^2$ definita per $\rho \geqslant 0$ e rappresentata in figura

Figura 3



Come conseguenza si ha il fatto che, se il dominio di f è $\mathbf{R}^2$, gli insiemi di livello possono essere solo dei seguenti tipi: l'insieme vuoto, un punto, una circonferenza, una corona circolare, oppure unione di questi (ad esempio più circonferenze). Per cui se le curve di livello di una funzione da $\mathbf{R}^2$ in $\mathbf{R}$ sono circonferenze la funzione in questione sarà radiale.
In generale, se $f : \mathbf{R}^n \to \mathbf{R}$, gli insiemi di livello potranno essere sempre l'insieme vuoto o un punto, ma anziché circonferenze si avranno sfere se $n = 3$, ipersfere in dimensione più alta (cioè insiemi che soddisfano $|x|$ costante con $x \in \mathbf{R}^n$).

2. Continuità

Cominciamo con il dare la definizione di funzione continua.

Definizione 2.1 (continuità)**.** *Una funzione $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, è continua nel punto $x_o \in A$ se per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che*

$$\|x - x_o\| < \delta\,,\ x \in A\,, \quad \Longrightarrow \quad \|f(x) - f(x_o)\| < \varepsilon;$$

equivalentemente se per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che

$$x \in A \cap B_\delta(x_o) \quad \Longrightarrow \quad f(x) \in B_\varepsilon(f(x_o))\,;$$

equivalentemente se $\displaystyle\lim_{\substack{x \to x_o \\ x \in A}} f(x) = f(x_o)$.

Si considerino ora una funzione $f : A \to \mathbf{R}$ e una curva continua $\gamma : [a, b] \to A$, A aperto di $\mathbf{R}^n$, in modo tale che il punto x_o appartenga al sostegno della curva. Per semplicità (a meno di riparametrizzazioni è sempre vero) supponiamo che γ sia definita in $[-\varepsilon, \varepsilon]$ per qualche $\varepsilon > 0$ e che $\gamma(0) = x_o$. Se f è una funzione continua in $x_o \in A$ si ha che per ogni $\delta_2 > 0$ esiste $\delta_1 > 0$ tale che

$$x \in A \cap B_{\delta_1}(x_o) \quad \Longrightarrow \quad f(x) \in B_{\delta_2}(f(x_o))\,.$$

Se ora ci limitiamo a considerare alcuni valori per x, quelli che stanno sul sostegno della curva che indicheremo con Γ, chiaramente si avrà che

$$x \in A \cap B_{\delta_1}(x_o) \cap \Gamma \quad \Longrightarrow \quad f(x) \in B_{\delta_2}(f(x_o))\,.$$

Questo significa in particolare che la funzione $f \circ \gamma : [-\varepsilon, \varepsilon] \to \mathbf{R}$ è continua. Possiamo facilmente concludere quindi che

$$f \text{ continua in } x_o \in A \quad \Longrightarrow \quad f \circ \gamma \;\text{ continua in } 0$$

per ogni curva continua $\gamma : [-\varepsilon, \varepsilon] \to A$ con $\gamma(0) = x_o$ (si veda anche la Figura 2).

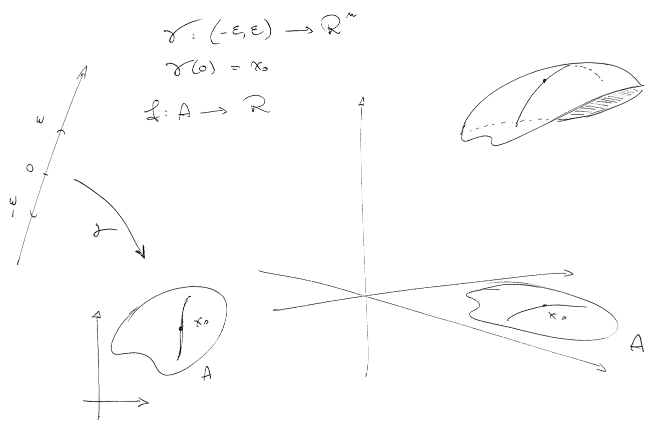


Figura 4

Ricordo - Dati due numeri $a, b \in \mathbf{R}$ vale la disuguaglianza $2ab \leqslant a^2+b^2$ e si ha $2ab = a^2 + b^2$ solo se $a = b$. (Dimostrazione: $(a-b)^2 \geqslant 0$.)

Esempio 2.2. - Vediamo ora un esempio che farà capire che la continuità in 0 di $f \circ \gamma : [-\varepsilon, \varepsilon] \to \mathbf{R}$ per qualche curva γ o anche per una ampia classe di curve γ tali che $\gamma(0) = x_o$ non garantisce la continuità in x_o per f.
Si consideri la funzione $f : \mathbf{R}^2 \to \mathbf{R}$ definita da

$$f(x,y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2} & \text{se } (x,y) \neq (0,0) \\[2ex] 0 & \text{se } (x,y) = (0,0). \end{cases}$$

Ovviamente la funzione è continua in tutti punti diversi da $(0,0)$. Cerchiamo di capire se lo è anche in $(0,0)$. Consideriamo tutte le rette per l'origine e restringiamo f a tali rette. Per fare ciò si fissi un vettore $v = (v_1, v_2)$ di norma 1 e si consideri la retta $t \mapsto tv$ con $t \in \mathbf{R}$. Studiamo la funzione ristretta a tali rette con $t \neq 0$:

$$f(tv_1, tv_2) = \frac{t^3 v_1^2 v_2}{t^4 v_1^4 + t^2 v_2^2}\,.$$

Se $v_1 = 0$ oppure $v_2 = 0$ la funzione è identicamente nulla; diversamente studiando il limite

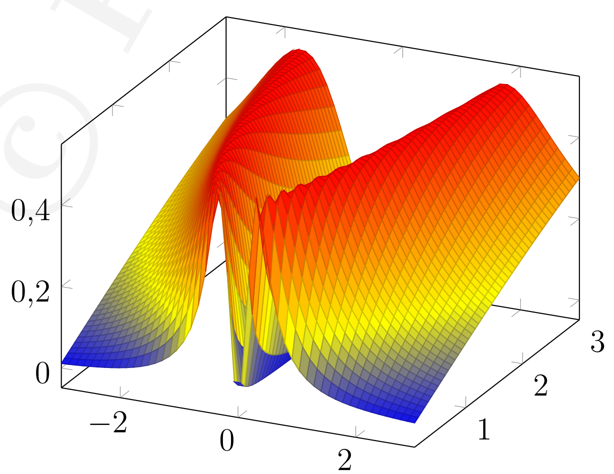
$$\lim_{t\to 0} f(tv_1, tv_2) = \lim_{t\to 0} \frac{t^3 v_1^2 v_2}{t^4 v_1^4 + t^2 v_2^2} = 0$$

si ottiene nuovamente zero. Si conclude che la restrizione di f lungo tutte le rette per l'origine ha limite 0 nel punto $(0,0)$, valore che coincide con il valore di f nell'origine. Quindi le restrizioni lungo tutte le rette per l'origine sono continue.
Può bastare a dire che f è continua in $(0,0)$?
Si consideri la curva $y = x^2$ e si restringa la f a tale curva, per esempio parametrizzandola con $t \mapsto (t, t^2)$. Si ottiene

$$f(t, t^2) = \frac{t^4}{t^4 + t^4} \qquad \text{per cui} \qquad \lim_{t\to 0} f(t, t^2) = \frac{1}{2}\,.$$

Di conseguenza concludiamo che f non è continua in $(0,0)$. Si veda la Figura 3 nella quale è rappresentato il grafico di f con $y > 1/10$.

Figura 3



EX - Si osservi che seguendo l'esempio precedente è facile produrre funzioni le cui restrizioni lungo le rette e le parabole per l'origine sono continue e non lo sono lungo le cubiche. Basta considerare

$$f(x,y) = \begin{cases} \dfrac{x^3 y}{x^6 + y^2} & \text{se } (x,y) \neq (0,0) \\ \\ 0 & \text{se } (x,y) = (0,0). \end{cases}$$

Verificarlo! Si scriva un esempio di funzione la cui restrizione lungo tutte le curve del tipo (at, bt^k) e (at^k, bt) risulta continua nell'origine per $k = 1, 2, \dots n-1$ ($a, b \in \mathbf{R}$), ma invece non lo è lungo le curve (t, t^n).

EX - Si verifichi che la seguente funzione definite in $\mathbf{R}^2$ e a valori reali non è continua in $(0,0)$:

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

Dati $f : A \to \mathbf{R}$, $x_o \in A$, $\gamma : [-\varepsilon, \varepsilon] \to A$ ($\varepsilon > 0$) con $\gamma(0) = x_o$ diremo *restrizione di* f a γ e la denoteremo con $f|_\gamma$ la funzione $f \circ \gamma$.
Dato $x_o \in A$, per semplicità diremo che la restrizione $f|_\gamma$ è continua in x_o anziché dire che $f \circ \gamma$ è continua in 0.

Esempio 2.3. - Un esempio molto più semplice rispetto a 2.2, ma analogo, che mostra che lungo tutte le rette per un punto le restrizioni di una certa funzione sono continue nonostante la funzione non sia continua è il seguente.
Si consideri la funzione $f : \mathbf{R}^2 \to \mathbf{R}$ definita da

$$f(x,y) = \begin{cases} 1 & \text{se } (x,y) \in A\,, \\ 0 & \text{se } (x,y) \notin A. \end{cases}$$

dove $A = \{(x,y) \in \mathbf{R}^2 \,|\, x^2 < y < 2x^2\}$.

Figura 4.a

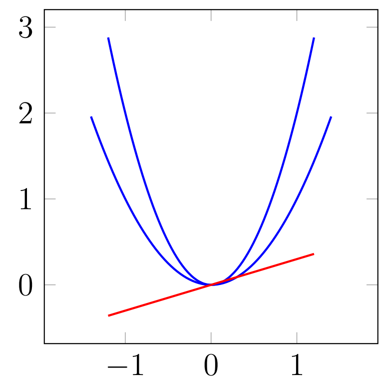
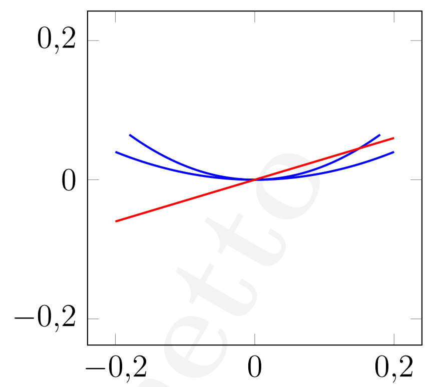


Figura 4.b



Chiaramente f non è continua nell'origine, dove f assume il valore 0, perché in qualunque intorno $B_\delta((0,0))$ dell'origine f assume sia il valore 0 che il valore 1. Valutando però il limite in $(0,0)$ della restrizione di f lungo una qualunque retta per l'origine ($\lim_{t\to 0} f(tv_1, tv_2)$ con $|(v_1, v_2)| = 1$) si ha che tale limite è zero. Infatti se si considera la retta tv con $v = (1,0)$ si ha che la restrizione di f a tale retta è identicamente nulla; per ogni altro vettore v si ha che, definitivamente per $t \to 0$, la funzione assume il valore zero (verificarlo! fare l'esercizio che segue). A tal proposito si veda la Figura 4. Nella seconda, la 4.b, c'è un ingrandimento vicino all'origine della Figura 4.a, che evidenzia (ma non dimostra) come la retta (disegnata in rosso) sufficientemente vicino all'origine passi solamente nella zona in cui f assume identicamente il valore zero.

EX - Date le due funzioni $r(x) = mx$ e $p(x) = ax^2$ ($m \neq 0$ e $a > 0$) mostrare che esiste un intorno I dello 0 tale che per ogni $x \in I \setminus \{0\}$ si ha che $r(x) < 0$ oppure $r(x) > p(x)$.

Valgono però i seguenti risultati, le cui dimostrazioni sono molto simili tra loro.

Teorema 2.4. *Siano $f : A \to \mathbf{R}$ con A aperto di $\mathbf{R}^N$ e $x_o \in A$. Allora f è continua in x_o se e solo se per ogni successione $(x_n)_n \subset A$ convergente ad x_o si ha che*

$$\lim_{n\to+\infty} f(x_n) = f(x_o).$$

Dimostrazione - È immediato vedere che se f è continua in x_o allora per ogni successione $(x_n)_{n\in\mathbf{N}}$ convergente ad x_o si ha

$$\lim_{n\to+\infty} f(x_n) = f(x_o).$$

Vediamo l'implicazione inversa. Se, per assurdo, non fosse vera si avrebbe l'esistenza di $\varepsilon > 0$ tale che per ogni $n \in \mathbf{N}$ esisterebbe

$$x_n \in B_{1/n}(x_o) \qquad \text{tale che} \qquad |f(x_n) - f(x_o)| \geqslant \varepsilon.$$

Ma ciò contraddirebbe il fatto che f è continua in x_o. $\square$

Il Teorema 2.4 precedente può essere enunciato in maniera analoga per i limiti, visto che la continuità di f in x_o è equivalente a $\lim\limits_{\substack{x \to x_o \\ x \in A}} f(x) = f(x_o)$.

Di conseguenza si avrà, per f definita in A, che ($l \in \mathbf{R} \cup \{-\infty, +\infty\}$)

$$\lim_{\substack{x \to x_o \\ x \in A}} f(x) = l \qquad \Longleftrightarrow \qquad \lim_{n \to +\infty} f\big(x_n\big) = l$$

per **ogni** successione $(x_n)_n \subset A$ convergente ad x_o.

Osservazione 2.5. - È chiaramente impossibile usare questa equivalenza per provare che un certo limite è l, ma è molto utile usata in negativo, come vedremo con qualche esercizio: se si trovano due successioni $(x_n)_n$ e $(y_n)_n$ con $\lim_{n\to+\infty} f(x_n) \neq \lim_{n\to+\infty} f(y_n)$ allora non esiste $\lim_{x\to x_o} f(x) = l$. Si può concludere la stessa cosa se si trovano due curve γ_1 e γ_2 tali che $\lim_{t\to 0} f(\gamma_1(t))$ e $\lim_{t\to 0} f(\gamma_2(t))$ sono diversi.

Enunciamo ora un risultato, senza mostrarlo, già visto per funzioni di una variabile e che vale anche per funzioni di più variabili.

Teorema 2.6 (permanenza del segno)**.** *Siano $f : A \to \mathbf{R}$ con A aperto di $\mathbf{R}^N$ e x_o di accumulazione per A. Se $\lim\limits_{\substack{x \to x_o \\ x \in A}} f(x) = l \neq 0$ allora esiste un intorno di x_o $B_r(x_o)$ contenuto in A tale che $f(x)$ assume lo stesso segno di l per ogni $x \in \big(B_r(x_o) \setminus \{x_o\}\big) \cap A$.*

Concludiamo con il seguente risultato, già visto per funzioni di una variabile e la cui dimostrazione è uguale al caso di funzioni di una variabile.

Teorema 2.7 (Weierstrass)**.** *Se $f : K \to$ è una funzione continua su K, K insieme chiuso e limitato di $\mathbf{R}^n$, allora f ammette massimo e minimo su K.*

Dimostrazione - Definiamo $\mu := \inf_K f$. Tale quantità potrà essere finita o infinita ($-\infty$ per la precisione). Il problema è mostrare che $\mu \neq -\infty$ e che esiste un qualche punto $\bar{x} \in K$ nel quale f assume il valore μ: allora avremo mostrato che f assume il suo valore minimo in

$\bar{x} \in K$. Analogamente si farà per il valore massimo.
Se K è compatto, come abbiamo già visto, è compatto per successioni (analogamente al caso unidimensionale).
Si consideri ora una successione $\{x_n\}_n \subset K$ tale che

$$\lim_{n\to+\infty} f(x_n) = \mu.$$

Dalla compattezza di K possiamo trovare una sottosuccessione $\{x_{n_k}\}_{k\in\mathbf{N}}$ convergente ad un certo $\bar{x} \in K$. Si noti che ovviamente $\lim_{k\to+\infty} f(x_{n_k}) = \mu$. Usando la continuità di f possiamo concludere

$$\mu = \lim_{k\to+\infty} f(x_{n_k}) = f(\bar{x}). \qquad \square$$

In generale, se f è una funzione definita in A sottoinsieme generico di $\mathbf{R}^n$, l'insieme immagine $f(A)$ ammette sempre estremo superiore ed estremo inferiore, denotati al solito con $\sup_A f$ e $\inf_A f$. Può capitare che $\sup_A f$ non appartenga a $f(A)$, cioè che non esista alcun elemento di A la cui immagine tramite f sia l'estremo superiore di $f(A)$. Quando ciò accade, quando cioè $f(A)$ ammette massimo, esiste (almeno) un punto $\bar{x} \in A$ tale che

$$\sup_A f = f(\bar{x})\,.$$

Tale punto è detto **punto di massimo**, f ammette massimo che si denota con $\max_A f$ e vale $\sup_A f = \max_A f$ e inoltre ovviamente

$$f(x) \leqslant f(\bar{x}) \qquad \text{per ogni } x \in A\,.$$

Si dice che un **punto** $\bar{x}$ è **di massimo stretto** se vale la disuguaglianza

$$f(x) < f(\bar{x}) \qquad \text{per ogni } x \in A,\ x \neq \bar{x}\,.$$

Analoghe considerazioni si fanno per estremo inferiore e minimo.
Il teorema di Weiestrass ci assicura, sotto opportune ipotesi, l'esistenza dei valori minimo e massimo per f e di almeno un punto di minimo e di almeno un punto di massimo.
Un punto $\bar{x} \in A$ sarà detto invece di **massimo locale** per f se esiste $\delta > 0$ tale che $\bar{x}$ risulta di massimo per f ristretta all'insieme $A \cap B_\delta(\bar{x})$, cioè

$$f(x) \leqslant f(\bar{x}) \qquad \text{per ogni } x \in A \cap B_\delta(\bar{x})$$

e di **massimo locale stretto** se esiste $\delta > 0$ tale che $\bar{x}$ risulta di massimo stretto per f ristretta all'insieme $A \cap B_\delta(\bar{x})$, cioè

$$f(x) < f(\bar{x}) \qquad \text{per ogni } x \in A \cap B_\delta(\bar{x}),\ x \neq \bar{x}\,.$$

Analogamente un punto sarà di minimo locale se esiste $\delta > 0$ tale che $\bar{x}$ risulta di minimo per f ristretta all'insieme $A \cap B_\delta(\bar{x})$.

Esempio - Si consideri $f : (0,1) \to \mathbf{R}$ definita da $f(x) = x + 1$. Tale funzione, se pur continua, non ammette né massimo, né minimo e si ha $\sup_{(0,1)} f = 2$ e $\inf_{(0,1)} f = 1$. Considerando $g : [0,1] \to \mathbf{R}$ definita da $g(x) = x + 1$ si ha invece che g ammette sia massimo che minimo e $\max_{[0,1]} g = 2$ e 1 è punto di massimo, $\min_{[0,1]} g = 1$ e 0 è punto di minimo.

3. Derivate parziali e derivabilità

Sia $f : A \to \mathbf{R}$ una funzione di n variabili, A aperto di $\mathbf{R}^n$ e $\bar{x} \in A$. Se fissiamo $n-1$ componenti di $\bar{x}$ e consideriamo la funzione di una variabile

$$f_i(y) := f(\bar{x}_1, \ldots, \bar{x}_{i-1}, y, \bar{x}_{i+1}, \ldots \bar{x}_n)$$

possiamo considerare la sua derivata nel punto $\bar{x}_i$

$$\frac{df_i}{dy}(\bar{x}_i) = \lim_{t \to 0} \frac{f_i(\bar{x}_i + t) - f_i(\bar{x}_i)}{t}$$

ammesso che tale limite esista e sia finito, dove

$$\frac{f_i(\bar{x}_i + t) - f_i(\bar{x}_i)}{t} =$$

(1)

$$= \frac{f(\bar{x}_1, \ldots, \bar{x}_{i-1}, \bar{x}_i + t, \bar{x}_{i+1}, \ldots \bar{x}_n) - f(\bar{x}_1, \ldots, \bar{x}_{i-1}, \bar{x}_i, \bar{x}_{i+1}, \ldots \bar{x}_n)}{t}.$$

Questo corrisponde a restringere f alla retta

$$(\bar{x}_1, \ldots \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_n)$$

e a fare la derivata ordinaria della funzione di una variabile così ottenuta. Se, per esempio, ci limitiamo a $n = 2$ e se denotiamo con $(\bar{x}, \bar{y})$ il punto nel quale si vogliono definire le derivate si avrebbero le due funzioni, che chiamiamo g_1 e g_2,

$$g_1(x) = f(x, \bar{y}), \qquad g_2(y) = f(\bar{x}, y).$$

La procedura definita sopra si riduce in questo caso a calcolare le due derivate, se esistono,

$$\frac{dg_1}{dx}(\bar{x}) \quad \text{e} \quad \frac{dg_2}{dy}(\bar{y}),$$

cioè rispettivamente i due limiti, se esistono finiti,

$$\begin{aligned}
\lim_{t\to 0} \frac{g_1(\bar{x}+t)-g_1(\bar{x})}{t} &= \lim_{h\to 0} \frac{f(\bar{x}+t,\bar{y})-f(\bar{x},\bar{y})}{t}\,,\\
\lim_{t\to 0} \frac{g_2(\bar{y}+t)-g_2(\bar{y})}{t} &= \lim_{h\to 0} \frac{f(\bar{x},\bar{y}+t)-f(\bar{x},\bar{y})}{t}\,.
\end{aligned}$$

I rapporti incrementali appena scritti possono equivalentemente essere riscritti come segue

$$\begin{aligned}
\frac{f(\bar{x}+t,\bar{y})-f(\bar{x},\bar{y})}{t} &= \frac{f\big((\bar{x},\bar{y})+t(1,0)\big)-f(\bar{x},\bar{y})}{t}\,,\\
\frac{f(\bar{x},\bar{y}+t)-f(\bar{x},\bar{y})}{t} &= \frac{f\big((\bar{x},\bar{y})+t(0,1)\big)-f(\bar{x},\bar{y})}{t}\,,
\end{aligned}$$

e in generale in (1)

$$\frac{f(\bar{x}+t\,e_i)-f(\bar{x})}{t}\,.$$

Allo stesso modo è possibile considerare un generico vettore v di modulo 1 e considerare la funzione f ristretta alla retta $t \mapsto \bar{x}+tv$ e valutarne, se possibile, il limite dei rapporti incrementali

$$\lim_{t\to 0} \frac{f(\bar{x}+t\,v)-f(\bar{x})}{t}\,.$$

Se tale limite esiste, finito, lo chiameremo derivata direzionale di f nella direzione v nel punto $\bar{x}$ e lo denoteremo con

$$\frac{\partial f}{\partial v}(\bar{x})\,.$$

Le derivate lungo i particolari vettori e_i si denotano semplicemente con $D_i f(\bar{x})$, con $f_{x_i}(\bar{x})$ o più spesso con

$$\frac{\partial f}{\partial x_i}(\bar{x})$$

e sono dette *derivate parziali* di f nel punto $\bar{x}$.

Esempio 3.1. - Calcoliamo le derivate parziali della funzione $g(x,y) = x^2 \operatorname{sen} y$. Se vogliamo calcolare la derivata rispetto alla variabile x, poiché ciò equivale a fissare un valore di y e muoversi lungo una retta parallela all'asse x (e analogamente per la derivata rispetto alla variabile y), si avrà semplicemente

$$\frac{\partial g}{\partial x}(x,y) = 2\,x \operatorname{sen} y \qquad \text{e} \qquad \frac{\partial g}{\partial y}(x,y) = x^2 \cos y\,.$$

Se la funzione è $f(x, y, z) = x^2 z^3 \mathrm{sen}\, y + z$ si fisseranno due valori per due delle variabili per fare la derivata rispetto alla terza variabile. Di conseguenza, come se le altre due fossero costanti, si ha

$$\begin{aligned}
\frac{\partial f}{\partial x}(x, y, z) &= 2\, x\, z^3 \,\mathrm{sen}\, y\,, \\
\frac{\partial f}{\partial y}(x, y, z) &= x^2\, z^3 \cos y\,, \\
\frac{\partial f}{\partial z}(x, y, z) &= 3\, x^2 z^2 \mathrm{sen}\, y + 1\,.
\end{aligned}$$

Proviamo a calcolare le derivate direzionali della funzione $f(x, y) = x^2 + y^2$. Fintanto che ci limitiamo alle direzioni principali otteniamo

$$\frac{\partial f}{\partial x}(x_o, y_o) = 2\, x_o\,, \qquad \frac{\partial f}{\partial x}(x_o, y_o) = 2\, y_o\,.$$

In un'altra direzione $v = (v_1, v_2)$ si ottiene

$$\begin{aligned}
&\lim_{t\to 0} \frac{f((x_o, y_o) + t(v_1, v_2)) - f(x_o, y_o)}{t} = \\
&\quad = \lim_{t\to 0} \frac{x_o^2 + y_o^2 + t^2 + 2t(x_o v_1 + y_o v_2) - x_o^2 - y_o^2}{t} = \langle (2x_o, 2y_o), (v_1, v_2) \rangle\,.
\end{aligned}$$

Si intuisce che le derivate direzionali sono meno immediate da calcolare rispetto a quelle parziali, ma vedremo che nei casi "buoni" anche queste sono immediate da calcolare (si veda (9)).

Esercizio 3.2. - Il vettore v, come già detto, avrà sempre lunghezza 1. Cosa succede se il vettore v non ha lunghezza 1 (e ovviamente non zero)? Proviamo a riducri al caso di un vettore di modulo 1 riparametrizzando la retta $t \mapsto \bar{x} + tv$. Scriviamo la retta rispetto ad un nuovo parametro come segue

$$\bar{x} + tv = \bar{x} + t\frac{v}{|v|}|v| = \bar{x} + s\frac{v}{|v|}$$

dove $s = t|v|$. Si osservi che $v/|v|$ è un vettore con la stessa direzione e lo stesso verso di v e di modulo 1. In tal modo, detto w il vettore $v/|v|$, si ha

$$\begin{aligned}
\frac{\partial f}{\partial v}(\bar{x}) &= \lim_{t\to 0} \frac{f(\bar{x} + t\, v) - f(\bar{x})}{t} = \lim_{\frac{s}{|v|}\to 0} \frac{f\left(\bar{x} + s\, \dfrac{v}{|v|}\right) - f(\bar{x})}{\dfrac{s}{|v|}} = \\
&= \lim_{s\to 0} |v|\, \frac{f\,(\bar{x} + s\, w) - f(\bar{x})}{s} = |v| \frac{\partial f}{\partial w}(\bar{x}) = |v| \frac{\partial f}{\partial \frac{v}{|v|}}(\bar{x})\,.
\end{aligned}$$

Per cui derivare lungo la direzione individuata da v o da $v/|v|$ fornisce lo stesso risultato semplicemente riscalato di un fattore $|v|$.
Di conseguenza si osservi che, dato un generico vettore v non nullo e f funzione differenziabile in $\bar{x}$, si ha

$$\frac{\partial f}{\partial v}(\bar{x}) = |v| \frac{\partial f}{\partial \frac{v}{|v|}}(\bar{x}) = |v| \left\langle \nabla f(\bar{x}), \frac{v}{|v|} \right\rangle = \langle \nabla f(\bar{x}), v \rangle. \tag{2}$$

Nel caso dell'esempio precedente, se derivassimo $g(x,y) = x^2 \operatorname{sen} y$ rispetto al vettore $v = (2,0) = 2(1,0) = 2e_1$ in un qualunque punto (x,y) si avrebbe

$$\frac{\partial g}{\partial v}(x,y) = 2\, \frac{\partial g}{\partial x}(x,y)\,.$$

Esempio 3.3. - Calcoliamo le derivate direzionali della funzione dell'Esempio 2.3. Si noti che $f(0,0) = 0$, quindi

$$\frac{\partial f}{\partial v}(0,0) = \lim_{t \to 0} \frac{f((0,0) + t\,(v_1, v_2)) - f(0,0)}{t} = 0$$

qualunque sia il vettore v. Quindi la funzione f è derivabile in ogni direzione nel punto $(0,0)$ e tali derivate direzionali valgono tutte 0. Nonostante ciò, la funzione f *non* è continua nel punto $(0,0)$.

Esempio 3.4. - Calcoliamo le derivate direzionali della funzione dell'Esempio 2.2. Considerando un vettore $v = (v_1, v_2)$ di modulo 1 si ha che, se esiste, la derivata direzionale nell'origine è

$$\begin{aligned}\frac{\partial f}{\partial v}(0,0) &= \lim_{t \to 0} \frac{f((0,0) + t\,(v_1, v_2)) - f(0,0)}{t} = \\ &= \lim_{t \to 0} \frac{t^3 v_1^2 v_2}{t^4 v_1^4 + t^2 v_2^2}\, \frac{1}{t} = \lim_{t \to 0} \frac{v_1^2 v_2}{t^2 v_1^4 + v_2^2} = \begin{cases} \dfrac{v_1^2}{v_2} & \text{se } v_2 \neq 0 \\ 0 & \text{se } v_2 = 0\,. \end{cases}\end{aligned}$$

EX - Si trovino gli insiemi massimali di definizione e si scrivano le derivate parziali delle seguenti funzioni

$$f_1(x,y) = x\, \log(xy^2)\,, \qquad f_2(x,y) = \cos\frac{x}{y} + \cos\frac{y}{x}\,,$$

$$f_3(x,y) = x\, \log(x \log y)\,, \qquad f_4(x,y,z) = \sqrt{x^2 - y^2 + z^2}\,,$$

dopodiché si verifichino le uguaglianze

$$\frac{\partial}{\partial x}\left(\frac{\partial f_i}{\partial y}\right) = \frac{\partial}{\partial y}\left(\frac{\partial f_i}{\partial x}\right) \qquad i = 1, 2, 3, 4\,,$$
$$\frac{\partial}{\partial x}\left(\frac{\partial f_4}{\partial z}\right) = \frac{\partial}{\partial z}\left(\frac{\partial f_4}{\partial x}\right)\,,$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f_4}{\partial z}\right) = \frac{\partial}{\partial z}\left(\frac{\partial f_4}{\partial y}\right)\,.$$

EX - Data $g(x, y) = y\sqrt{|x|}$ si calcoli, se esiste, la derivata $\dfrac{\partial g}{\partial x}(0, 0)$.
Si dica quali derivate direzionali esistono nell'origine.
Si dica se è differenziabile nell'origine.

Definizione 3.5 (funzione derivabile e gradiente)**.** *Diremo che una funzione $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, è derivabile in $x_o \in A$ se esistono tutte le derivate parziali $\frac{\partial f}{\partial x_i}(x_o)$. Si dice che f è derivabile in A se lo è in ogni punto di A. Diremo che f è di classe C^1 in A, e scriveremo $f \in C^1(A)$, se f ammette tutte le derivate parziali in ogni punto di A e tali derivate sono continue in A.*
Se f è derivabile in x_o si chiama gradiente *di f in x_o il vettore*

$$\nabla f(x_o) = \left(\frac{\partial f}{\partial x_1}(x_o), \frac{\partial f}{\partial x_2}(x_o), \dots \frac{\partial f}{\partial x_n}(x_o)\right)\,.$$

4. Differenziale e differenziabilità

Definizione 4.1 (funzione differenziabile)**.** *Diremo che una funzione $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, è differenziabile in $x_o \in A$ se esiste un'applicazione lineare $L : \mathbf{R}^n \to \mathbf{R}$ tale che ($h \in \mathbf{R}^n$)*

$$\lim_{|h| \to 0} \frac{f(x_o + h) - f(x_o) - L(h)}{|h|} = 0\,. \tag{3}$$

Ricordo: tutte e sole le applicazioni lineari $L : \mathbf{R}^n \to \mathbf{R}$ sono funzioni del tipo $L(x) = a_1x_1 + a_2x_2 + \dots a_nx_n$. Per cui un'applicazione lineare è univocamente determinata da un vettore $a = (a_1, \dots a_n) \in \mathbf{R}^n$.
Possiamo allora dire in maniera equivalente a (3) che $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, è differenziabile in $x_o \in A$ se esistono n numeri reali

$a_1, \dots, a_n$ tali che

$$\lim_{|h|\to 0} \frac{f(x_o+h) - f(x_o) - a_1h_1 - a_2h_2 - \dots - a_nh_n}{|h|} =$$
$$= \lim_{|h|\to 0} \frac{f(x_o+h) - f(x_o) - \langle a, h\rangle}{|h|} = 0\,.$$

Vediamo, prima di procedere, cosa significa essere differenziabile in dimensione 1. Affinché $f : \mathbf{R} \to \mathbf{R}$ sia differenziabile dovrà esistere una applicazione lineare da $\mathbf{R}$ a $\mathbf{R}$, cioè una funzione del tipo $x \mapsto ax$ con $a \in \mathbf{R}$, che sia tale che

$$\lim_{|h|\to 0} \frac{f(x_o+h) - f(x_o) - ah}{|h|} = 0\,.$$

Ci siamo ridotti quindi a dire che f è differenziabile in x_o se esiste un numero a tale che l'uguaglianza appena scritta è vera.
Si osservi come tale limite è zero se e solo se $\lim_{|h|\to 0} \frac{|f(x_o+h)-f(x_o)-ah|}{|h|} = 0$ e infine se e sole se

$$\lim_{|h|\to 0} \frac{f(x_o+h) - f(x_o) - ah}{h} = 0\,. \tag{4}$$

Ma nel caso in cui f sia derivabile in x_o esiste un numero a per il quale (4) sia vera ed è $f'(x_o)$.
Quindi la differenziabilità in dimensione 1 non è altro che la derivabilità.
Si noti inoltre che (4) è equivalente alla formula di Taylor arrestata al primo ordine per f nel punto x_o (dove il resto è espresso nella forma di Peano), cioè

$$f(x_o+h) = f(x_o) + ah + o(h)$$

o equivalentemente (scrivendo x al posto di $x_o + h$)

$$f(x) = f(x_o) + a(x - x_o) + o(x - x_o),$$

dove $a = f'(x_o)$ e la funzione $p(x) = f(x_o) + f'(x_o)(x - x_o)$ è il polinomio di Taylor di primo grado il cui grafico rappresenta la retta tangente al grafico di f nel punto $(x_o, f(x_o))$.

Fatto importante equivalente alla differenziabilità - In maniera analoga a quanto appena detto per una funzione di una variabile si ha che per una funzione $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, la differenziabilità di f in un punto $\bar{x} \in A$ è equivalente alla scrittura

$$\begin{aligned} f(\bar{x}+h) &= f(\bar{x}) + a_1h_1 + \dots + a_nh_n + o(|h|) = \\ &= f(\bar{x}) + \langle a, h\rangle + o(|h|) \end{aligned} \tag{5}$$

o equivalentemente $f(x) = f(\bar{x}) + \langle a, x - \bar{x} \rangle + o(|x - \bar{x}|)$. Questa è anche l'espressione della formula di Taylor al primo ordine con resto di Peano, formula che più avanti riprenderemo e vedremo meglio fino al secondo ordine. La funzione

$$(6)\quad p(x) = f(\bar{x}) + a_1(x_1 - \bar{x}_1) + \cdots + a_n(x_n - \bar{x}_n) = f(\bar{x}) + \langle a, x - \bar{x} \rangle$$

ha come grafico un piano che è detto piano tangente al grafico di f nel punto $(\bar{x}, f(\bar{x}))$.

Conseguenze della differenziabilità - Vediamo ora due conseguenze della differenziabilità in un punto per una funzione f.

1° **fatto** - Se $f : A \to \mathbf{R}$ è differenziabile in x_o allora è continua in x_o. Per dimostrarlo sarà sufficiente verificare che $\lim_{x \to x_o} f(x) = f(x_o)$. Per fare ciò verifichiamo che $\lim_{x \to x_o}[f(x) - f(x_o)] = 0$ o equivalentemente $\lim_{|h| \to 0}[f(x_o + h) - f(x_o)] = 0$. Poiché f è differenziabile in x_o esiste L lineare tale che vale (3) e si osservi che una funzione lineare è continua, per cui $\lim_{|h| \to 0} L(h) = 0$. Di conseguenza si ha

$$\begin{aligned} \lim_{|h| \to 0}[f(x_o + h) - f(x_o)] &= \lim_{|h| \to 0}[f(x_o + h) - f(x_o) - L(h)] = \\ &= \lim_{|h| \to 0} \frac{f(x_o + h) - f(x_o) - L(h)}{|h|}\, |h| \end{aligned}$$

e

$$\lim_{|h| \to 0} \frac{f(x_o + h) - f(x_o) - L(h)}{|h|}\, |h| = 0$$

perché prodotto di due quantità infinitesime.

2° **fatto** - Se $f : A \to \mathbf{R}$ è differenziabile in x_o allora è derivabile in tutte le direzioni x_o. Per dimostrarlo si consideri l'incremento $h = tv$ dove v è un generico vettore v di modulo 1. Usando il fatto che $|h|$ tende zero se e solo se $|t|$ tende a zero e quindi se e solo se t tende a zero e la linearità di L, per cui $L(tv) = tL(v)$, dalla differenziabilità di f in x_o si ricava

$$\begin{aligned} 0 = \lim_{t \to 0} \frac{f(x_o + tv) - f(x_o) - L(tv)}{|t|} &= \lim_{t \to 0} \left| \frac{f(x_o + tv) - f(x_o) - tL(v)}{t} \right| = \\ = \lim_{t \to 0} \frac{f(x_o + tv) - f(x_o) - tL(v)}{t} & \end{aligned}$$

per cui

$$\lim_{t \to 0} \frac{f(x_o + tv) - f(x_o)}{t} = L(v)\,.$$

Deduciamo due cose: primo che la derivata direzionale esiste, seconda cosa che è Lv, cioè

$$\frac{\partial f}{\partial v}(x_o) = L(v)\,. \tag{7}$$

Poiché L è lineare sarà del tipo $L(v) = a_1 v_1 + \cdots + a_n v_n$ per qualche $a_1, \ldots a_n \in \mathbf{R}$; di conseguenza, se f è differenziabile in x_o, si dovrà avere

$$\frac{\partial f}{\partial v}(x_o) = a_1 v_1 + \cdots + a_n v_n\,. \tag{8}$$

In particolare (si consideri $v = e_i$) si ha $\frac{\partial f}{\partial x_i}(x_o) = a_i$ da cui, se f è differenziabile in x_o,

$$\frac{\partial f}{\partial v}(x_o) = \frac{\partial f}{\partial x_1}(x_o) v_1 + \cdots + \frac{\partial f}{\partial x_n}(x_o) v_n = \big\langle \nabla f(x_o), v \big\rangle\,. \tag{9}$$

Osservazione 4.2. - Questi due fatti possono essere usati in negativo per verificare che una funzione f non è differenziabile in un punto x_o:

i) se f in tale punto non è continua non sarà nemmeno differenziabile;

ii) se f in tale punto non è derivabile in tutte le direzioni allora in tale punto non sarà nemmeno differenziabile;

iii) se f in tale punto è derivabile in tutte le direzioni, ma la funzione

$$v \mapsto \frac{\partial f}{\partial v}(x_o) \tag{10}$$

non è lineare allora in tale punto f non sarà nemmeno differenziabile, poiché se lo fosse non solo sarebbe derivabile in tutte le direzioni, ma la funzione (10) dovrebbe essere del tipo $a_1 v_1 + \cdots + a_n v_n$, come ottenuto in (8).

Differenziale e sua rappresentazione - Il *differenziale* di f si denota con df e si ha che, se $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, f differenziabile in ogni punto di A,

$$df : A \to \mathcal{L}(\mathbf{R}^n; \mathbf{R})$$

dove $\mathcal{L}(\mathbf{R}^n; \mathbf{R})$ denota l'insieme delle applicazioni lineari da $\mathbf{R}^n$ in $\mathbf{R}$. L'applicazione L della Definizione 4.1 sarà detta *differenziale* di f in x_o e si denoterà con

$$df_{x_o} \quad \text{o con} \quad df(x_o)\,.$$

In dimensione 1 si avrà $df_{x_o} = f'(x_o)$. In dimensione più alta invece, utilizzando (7) e (8), si ha

$$df_{x_o}(v) = a_1 v_1 + \cdots + a_n v_n = \frac{\partial f}{\partial v}(x_o)$$

e si noti che in particolare (e_i, $i = 1, \dots n$, rappresenta la base canonica di $\mathbf{R}^n$)

$$df_{x_o}(e_i) = a_i = \frac{\partial f}{\partial x_i}(x_o)\,.$$

Dalla linearità di df_{x_o} si ha, supposto che $v = \sum_{i=1}^n v_i e_i$,

$$\begin{aligned} df_{x_o}(v) &= df_{x_o}\left(\sum_{i=1}^n v_i e_i\right) = \sum_{i=1}^n v_i\, df_{x_o}(e_i) = \\ &= \sum_{i=1}^n v_i\, \frac{\partial f}{\partial x_i}(x_o) = \langle \nabla f(x_o), v\rangle\,. \end{aligned}$$

Per cui il gradiente rappresenta il differenziale, nel senso che

$$df_{x_o}(v) = \langle \nabla f(x_o), v\rangle \qquad \text{per ogni } v \in \mathbf{R}^n\,. \tag{11}$$

Piano tangente e formula di Taylor al primo ordine - Come già visto il grafico di una funzione del tipo $p(x) = a_0 + a_1 x_1 + \dots a_n x_n$ è un piano (meglio: un iperpiano in $\mathbf{R}^n$, anche se spesso si dice semplicemente piano anche in dimensione più alta di 2). Mettendo insieme le informazioni che abbiamo, ricaviamo da (6) e da (8) e (11) che il piano tangente al grafico di f nel punto $(\bar{x}, f(\bar{x}))$ è il grafico della funzione

$$\begin{aligned} p(x) &= f(\bar{x}) + \frac{\partial f}{\partial x_1}(x_1 - \bar{x}_1) + \dots + \frac{\partial f}{\partial x_n}(x_n - \bar{x}_n) = \\ &= f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle \end{aligned} \tag{12}$$

e ricaviamo inoltre che la formula di Taylor al primo ordine, già vista in (5), diventa

$$f(\bar{x} + h) = f(\bar{x}) + \langle \nabla f(\bar{x}), h\rangle + o(|h|). \tag{13}$$

o equivalentemente

$$f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x}\rangle + o(|x - \bar{x}|). \tag{14}$$

Nel caso più semplice di $n = 2$ e considerata una funzione $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^2$, differenziabile in un punto $(x_o, y_o) \in A$, il piano tangente al grafico di f nel punto $(x_o, y_o, f(x_o, y_o))$ sarà il grafico della funzione

$$p(x, y) = f(x_o, y_o) + \frac{\partial f}{\partial x}(x_o, y_o)\big(x - x_o\big) + \frac{\partial f}{\partial y}(x_o, y_o)\big(y - y_o\big)$$

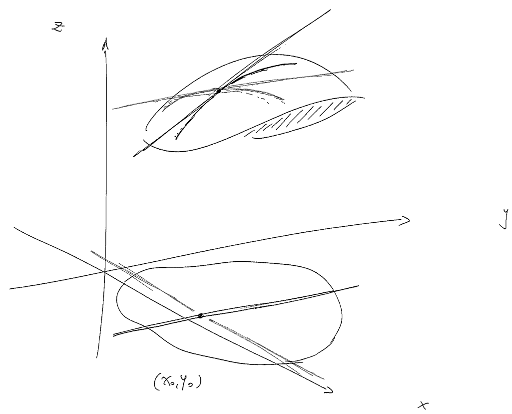


Figura 5

In maniera implicita tale piano può essere scritto come

$$z - z_o = \frac{\partial f}{\partial x}(x_o, y_o)\big(x - x_o\big) + \frac{\partial f}{\partial y}(x_o, y_o)\big(y - y_o\big) \tag{15}$$

dove $z_o = f(x_o, y_o)$. Tale piano può essere visto come un fascio di rette ottenute come rette tangenti in questo modo: dato un vettore v, $|v| = 1$ e la restrizione di f alla retta $(x_o, y_o) + t(v_1, v_2)$ si consideri la funzione di una variabile

$$g_v(t) := f\big((x_o, y_o) + t(v_1, v_2)\big)\,.$$

Allora la retta tangente a g_v nel punto 0 è data da

$$r_v(t) = g_v(0) + t g_v'(0) = f(x_o, y_o) + t\,\frac{\partial f}{\partial v}(x_o, y_o)\,.$$

Poiché $\frac{\partial f}{\partial v}(x_o, y_o) = \langle \nabla f(x_o, y_o), v\rangle$ si ha che la retta è

$$r_v(t) = f(x_o, y_o) + t v_1\,\frac{\partial f}{\partial x}(x_o, y_o) + t v_2\,\frac{\partial f}{\partial y}(x_o, y_o)\,.$$

Il grafico del differenziale di f nel punto (x_o, y_o) è un piano parallelo al piano tangente al grafico di f nel punto (x_o, y_o). Infatti il differenziale di f nel punto (x_o, y_o) è la funzione $df_{x_o} : \mathbf{R}^2 \longrightarrow \mathbf{R}$ definita da

$$df_{x_o}(x, y) = \frac{\partial f}{\partial x}(x_o, y_o)x + \frac{\partial f}{\partial y}(x_o, y_o)y\,.$$

Tale piano, opportunamente traslato in modo da passare dal punto $(x_o, y_o, f(x_o, y_o))$, è il piano tangente al grafico di f.

In dimensione più alta si parla sempre di piano tangente (anche se in realtà è un iperpiano tangente) al grafico di una funzione

$$f : \mathbf{R}^n \to \mathbf{R}$$

nel punto $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$ considerando, analogamente a (15), l'insieme dei punti $(x_1, \dots, x_n, x_{n+1}) \in \mathbf{R}^{n+1}$ che soddisfano l'equazione

$$x_{n+1} - f(\bar{x}) = \frac{\partial f}{\partial x_1}(\bar{x})(x_1 - \bar{x}_1) + \dots \frac{\partial f}{\partial x_n}(\bar{x})(x_n - \bar{x}_n)\,.$$

Esempio 4.3. - Si studi la continuità, derivabilità, differenziabilità in $(0,0)$ della funzione

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

Per vedere se è continua valutiamo il limite nell'origine di f. Usando la solita stima $2ab \leqslant a^2 + b^2$ si ricava

$$0 \leqslant \left| \frac{x^2 y}{x^2+y^2} \right| \leqslant \frac{1}{2}|x|$$

per cui il limite per $(x,y) \to (0,0)$ è zero, che è $f(0,0)$. Alternativamente usando le coordinate polari possiamo scrivere

$$f(\rho\cos\vartheta, \rho\,\mathrm{sen}\,\vartheta) = \frac{\rho^3 \cos^2\!\vartheta\,\mathrm{sen}\,\vartheta}{\rho^2}$$

e

$$0 \leqslant |f(\rho\cos\vartheta, \rho\,\mathrm{sen}\,\vartheta)| \leqslant \rho$$

che per il teorema dei due carabinieri tende a zero al tendere di ρ a zero, quindi $f(\rho\cos\vartheta, \rho\,\mathrm{sen}\,\vartheta)$ tende uniformemente rispetto a ϑ a zero quando $\rho \to 0$.
Per vedere se è derivabile nella direzione $v = (v_1, v_2)$ (con $v_1^2 + v_2^2 = 1$) calcoliamo

$$\lim_{t\to 0} \frac{f((0,0) + t(v_1,v_2)) - f(0,0)}{t} = \lim_{t\to 0} \frac{t^3 v_1^2 v_2}{t^2 v_1^2 + t^2 v_2^2}\,\frac{1}{t} = v_1^2 v_2$$

per cui $\dfrac{\partial f}{\partial v}(0,0) = v_1^2 v_2$. Poiché questa espressione non è lineare in v sicuramente la funzione non è differenziabile in $(0,0)$.

Se non si calcolassero tutte le derivate direzionali, ma solo le due derivate parziali, si avrebbe

$$\frac{\partial f}{\partial x}(0,0) = 0 \qquad \text{poiché } f(x,0) = 0\,,$$
$$\frac{\partial f}{\partial y}(0,0) = 0 \qquad \text{poiché } f(0,y) = 0\,.$$

In questo modo per verificare se la funzione è differenziabile in $(0,0)$ dobbiamo, usando la definizione, verificare se vale (3) dove l'applicazione L, se c'è, deve essere necessariamente (si veda (11))

$$L(h) = \frac{\partial f}{\partial x}(0,0)h_1 + \frac{\partial f}{\partial y}(0,0)h_2$$

che nel nostro caso è l'applicazione nulla visto che $\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0$. Valutiamo

$$\lim_{|h|\to 0} \frac{f((0,0)+(h_1,h_2)) - f(0,0)}{|h|} = \lim_{|h|\to 0} \frac{h_1^2 h_2}{(h_1^2+h_2^2)^{3/2}}\,.$$

La quantità $\frac{h_1^2 h_2}{(h_1^2+h_2^2)^{3/2}}$ espressa in coordinate polari diventa

$$\frac{\rho^3 \cos^2\vartheta \operatorname{sen}\vartheta}{\rho^3} = \cos^2\vartheta \operatorname{sen}\vartheta$$

che chiaramente rimane costante al tendere di ρ a zero e, poiché dipende da ϑ, possiamo concludere che il limite non esiste. In particolare non è zero, per cui f non è differenziabile nel punto $(0,0)$.

Direzione di massima pendenza e legame con il gradiente - Vediamo ora qualche caratteristica del gradiente.

Proposizione 4.4. *Siano $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, $x_o \in A$ e f differenziabile in x_o. Si consideri il differenziale di f in x_o ristretto all'insieme $\mathbf{S}^{n-1}$ denotato con $df_{x_o}|_{\mathbf{S}^{n-1}}$, cioè la funzione*

$$df_{x_o} : \mathbf{S}^{n-1} \to \mathbf{R} \qquad \textit{dove } \mathbf{S}^{n-1} = \big\{ v \in \mathbf{R}^n \,\big|\, |v| = 1 \big\}\,.$$

Allora la funzione $df_{x_o}|_{\mathbf{S}^{n-1}}$ assume il suo massimo e il suo minimo rispettivamente nei vettori

$$\frac{\nabla f(x_o)}{|\nabla f(x_o)|} \qquad e \qquad -\frac{\nabla f(x_o)}{|\nabla f(x_o)|}\,.$$

Osservazione 4.5. - Il risultato appena enunciato può essere interpretato come segue: poiché (si veda (7))

$$df_{x_o}(v) = \frac{\partial f}{\partial v}(x_o),$$

definito il vettore
$$\bar{v} = \frac{\nabla f(x_o)}{|\nabla f(x_o)|},$$
si ha che
$$\frac{\partial f}{\partial \bar{v}}(x_o) \geqslant \frac{\partial f}{\partial v}(x_o) \qquad \text{per ogni } v \text{ di modulo 1.}$$
Poiché la derivata rispetto ad una certa direzione w fornisce un'informazione sulla "pendenza" del grafico di f nella direzione w possiamo dire un po' impropriamente che la direzione del vettore gradiente indica la direzione lungo la quale la funzione f ha la massima pendenza.
Con "pendenza" si intende una quantità che può essere anche negativa. Di conseguenza il vettore $-\bar{v}$ indicherà la direzione e il verso lungo i quali si ha la minima pendenza.

Osservazione 4.6. - Si supponga di avere una funzione $g : A \to \mathbf{R}$ e di sapere riguardo g le due seguenti cose:
$$f(x) \leqslant M \qquad \text{per ogni } x \in A, \quad f(\bar{x}) = M \qquad \text{per un qualche } \bar{x} \in A.$$
È chiaro allora che si può concludere che f ammette massimo in A e che il punto $\bar{x}$ è un punto di massimo.
Analoga considerazione vale per il minimo, ammesso che si sappia che $f(x) \geqslant m$ per ogni $x \in A$ e che esista un punto $\bar{\bar{x}} \in A$ tale che $f(\bar{\bar{x}}) = m$.
Si osservi che non si è richiesto che f sia continua e che A sia compatto.

Dimostrazione della Proposizione 4.4 - Prima di tutto si osservi che l'insieme $\mathbf{S}^{n-1}$ è chiuso e limitato e che $df_{x_o}|_{\mathbf{S}^{n-1}}$ è continua, per cui ammette sia massimo che minimo in $\mathbf{S}^{n-1}$ per il teorema di Weierstrass.
Possiamo riscrivere la quantità $df_{x_o}(v)$ con $v \in \mathbf{S}^{n-1}$ come
$$\frac{\partial f}{\partial v}(x_o) = \langle \nabla f(x_o), v \rangle .$$
Analizzando quindi l'ultima espressione dobbiamo trovare il massimo e il minimo di
$$v \mapsto \langle \nabla f(x_o), v \rangle \qquad \text{con } v \in \mathbf{S}^{n-1} .$$
Utilizzando la disuguaglianza di Schwarz
$$|\langle \nabla f(x_o), v \rangle| \leqslant |\nabla f(x_o)||v| = |\nabla f(x_o)|$$
per ogni $v \in \mathbf{S}^{n-1}$ si deduce che tale funzione assume tutti i valori tra
$$-|\nabla f(x_o)| \quad \text{ e } \quad |\nabla f(x_o)| .$$
Poiché per
$$v = \frac{\nabla f(x_o)}{|\nabla f(x_o)|} \qquad \text{e} \qquad w = -\frac{\nabla f(x_o)}{|\nabla f(x_o)|}$$

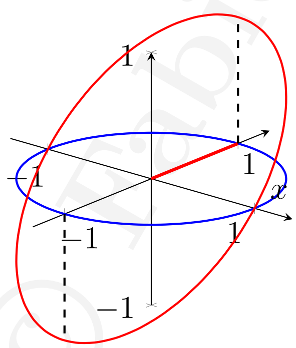
si ha $|\langle \nabla f(x_o), v\rangle| = |\nabla f(x_o)|$ e $|\langle \nabla f(x_o), w\rangle| = -|\nabla f(x_o)|$ si sono trovati i punti di massimo e minimo. $\square$

Per fare un semplice esempio si consideri la funzione $f(x,y) = y$ definita in $\mathbf{R}^2$. Chiaramente la massima pendenza, in qualunque punto di $\mathbf{R}^2$, si ha nella direzione $(0,1)$, quella dell'asse delle y. Calcolando le derivate direzionali in un qualunque punto $(\bar{x},\bar{y})$ nella direzione $v = (v_1, v_2) \in \mathbf{S}^1$ si ha

$$\lim_{t\to 0} \frac{f\big((\bar{x},\bar{y}) + t(v_1,v_2)\big) - f(\bar{x},\bar{y})}{t} = \lim_{t\to 0} \frac{\bar{x} + tv_1 - \bar{x}}{t} = v_2\,.$$

Chiaramente tale quantità è massima rispetto a v che soddisfa $v_1^2 + v_2^2 = 1$ quando $v_2 = 1$, è minima quando $v_2 = -1$.
In Figura 6 è rappresentato graficamente: in blu l'insieme $\mathbf{S}^1$ dei vettori di $\mathbf{R}^2$ di modulo 1 (**attenzione!** l'insieme blu giace sul piano $\{(x,y,z) \in \mathbf{R}^3 \,|\, z = 0\}$) sui quali andiamo a valutare il differenziale; in rosso il grafico del differenziale ristretto a $\mathbf{S}^1$; in rosso grassetto il vettore gradiente in $(0,0)$ (ma nel nostro caso, solo perché f è lineare, anche in qualunque altro punto). I punti di massimo e di minimo del differenziale valutato su $\mathbf{S}^1$ sono in corrispondenza dei tratteggi.

Figura 6



Vediamo ora un risultato di derivazione di funzioni composte che ci sarà immediatamente utile.

Teorema 4.7. *Siano $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, e $\gamma : I \to A$, I intervallo di $\mathbf{R}$, una curva regolare a tratti. Sia t_o un punto interno ad I e si supponga che*

i) *γ sia derivabile in t_o,*
ii) *f sia differenziabile in $x_o = \gamma(t_o)$.*

Allora $f \circ \gamma : I \to \mathbf{R}$ *è derivabile in* t_o *e vale la formula*

$$\frac{d}{dt}\big(f \circ \gamma\big)(t_o) = \langle \nabla f(x_o), \gamma'(t_o) \rangle \,. \tag{16}$$

Dimostrazione - Se f è differenziabile in x_o si ha per definizione che

$$f(x) - f(x_o) = \langle \nabla f(x_o), x - x_o \rangle + o(|x - x_o|)\,.$$

Se γ è derivabile in t_o si ha per definizione che

$$\gamma(t) - \gamma(t_o) = \gamma'(t_o)(t - t_o) + o(|t - t_o|)\,.$$

Valutiamo il rapporto incrementale di $f \circ \gamma$ in t_o usando entrambe le uguaglianze appena scritte:

$$\begin{aligned}
\frac{f\big(\gamma(t)\big) - f\big(\gamma(t_o)\big)}{t - t_o} &= \frac{\langle \nabla f(x_o), \gamma(t) - \gamma(t_o) \rangle + o(|\gamma(t) - \gamma(t_o)|)}{t - t_o} = \\
&= \frac{\langle \nabla f(x_o), \gamma'(t_o) + o(|t - t_o|) \rangle + o(|\gamma(t) - \gamma(t_o)|)}{t - t_o} = \\
&= \langle \nabla f(x_o), \gamma'(t_o) \rangle + \frac{\langle \nabla f(x_o), o(|t - t_o|) \rangle}{t - t_o} + \frac{o(|\gamma(t) - \gamma(t_o)|)}{t - t_o}\,.
\end{aligned}$$

Ora passando al limite si ha

$$\left| \frac{\langle \nabla f(x_o), o(|t - t_o|) \rangle}{t - t_o} \right| \leqslant |\nabla f(x_o)| \, \frac{|o(|t - t_o|)|}{|t - t_o|} \longrightarrow_{t \to 0} 0$$

e poiché $\lim_{t \to t_o} \frac{|\gamma(t) - \gamma(t_o)|}{|t - t_o|} = |\gamma'(t_o)|$ si ha

$$\left| \frac{o(|\gamma(t) - \gamma(t_o)|)}{t - t_o} \right| = \left| \frac{o(|\gamma(t) - \gamma(t_o)|)}{|\gamma(t) - \gamma(t_o)|} \right| \frac{|\gamma(t) - \gamma(t_o)|}{|t - t_o|} \longrightarrow_{t \to 0} 0$$

da cui si conclude. Nell'ultima uguaglianza abbiamo supposto che $|\gamma(t) - \gamma(t_o)| \neq 0$: possiamo farlo, almeno per t in un intorno sufficientemente piccolo di t_o, se $\gamma(t)$ non è costante per un tratto, perciò si è supposto γ regolare a tratti. □

Osservazione 4.8. - Si osservi che se in t_o la curva γ ha derivata uguale ad un certo vettore v di modulo 1 si ottiene che

$$\frac{d}{dt}\big(f \circ \gamma\big)(t_o) = \langle \nabla f(x_o), \gamma'(t_o) \rangle = \langle \nabla f(x_o), v \rangle = \frac{\partial f}{\partial v}(x_o)\,.$$

Il gradiente è ortogonale alle curve di livello - Si consideri ora una funzione $f : A \subset \mathbf{R}^2 \to \mathbf{R}$ e si supponga che sia differenziabile in A. Si supponga anche che l'insieme di livello $c \in \mathbf{R}$ sia una curva che denotiamo con Γ_c e si supponga ancora che esista una parametrizzazione derivabile di Γ_c, che denoteremo con $\gamma : I \to A$, I intervallo di $\mathbf{R}$. Per la formula appena mostrata si ha

$$(17) \qquad \frac{d}{dt}\big(f \circ \gamma\big)(t) = \langle \nabla f(\gamma(t)), \gamma'(t) \rangle \qquad \text{per ogni } t \in I\,.$$

D'altra parte, poiché la funzione su tale curva assume il valore costante c si ha anche

$$(18) \qquad \frac{d}{dt}\big(f \circ \gamma\big)(t) = 0 \qquad \text{per ogni } t \in I\,.$$

Mettendo assieme le cose si ottiene quindi

$$\langle \nabla f(\gamma(t)), \gamma'(t) \rangle = 0 \qquad \text{per ogni } t \in I\,,$$

cioè il gradiente di f in ogni punto della curva è ortogonale alla curva stessa.

Il gradiente è ortogonale alle (iper-)superfici di livello - La cosa non cambia (ma questo sarà visto bene quando si vedranno le superfici) se anziché una funzione definita in $A \subset \mathbf{R}^2$ si considera una funzione definita in $A \subset \mathbf{R}^3$ (o $\mathbf{R}^n$). In questo caso supponiamo che l'insieme di livello c sia una superficie che avrà equazione

$$f(x, y, z) = c\,.$$

Se si considera una qualunque curva sopra la superficie, cioè $\gamma : I \to A$, I intervallo di $\mathbf{R}$, tale che il sostegno sia contenuto in

$$\Gamma_c = \{(x, y, z) \in \mathbf{R}^3 \,|\, f(x, y, z) = c\}\,,$$

si avrà che valgono ancora (17) e (18), e questo per ogni curva il cui sostegno sia contenuto in Γ_c. Poiché ciò corrisponde a dire che il gradiente in un dato punto è ortogonale al vettore tangente alla curva γ per qualunque curva γ in Γ_c si può concludere che il gradiente è ortogonale alla superficie Γ_c.
Vediamo ora come una funzione che ammette derivate parziali continue sia anche differenziabile.

Teorema 4.9 (differenziale totale)**.** *Si considerino A aperto di $\mathbf{R}^n$, $x_o \in A$ e $f : A \to \mathbf{R}$. Si supponga che f abbia derivate parziali in un intorno $B_r(x_o)$ del punto x_o e si supponga che queste siano continue nel punto x_o. Allora f è differenziabile in x_o.*

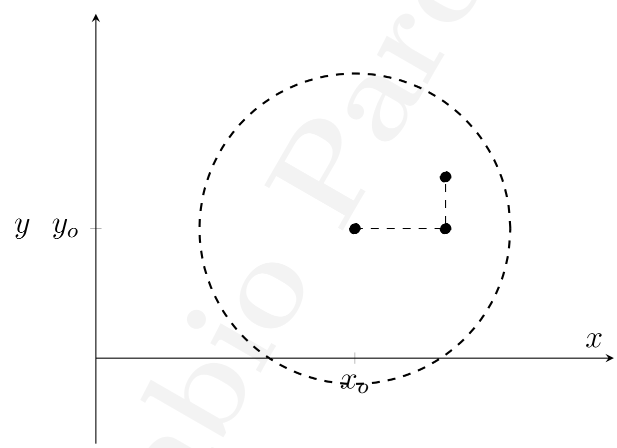
Dimostrazione - Vediamo la dimostrazione per $n = 2$. La funzione f risulta derivabile in un intorno del punto (x_o, y_o). Per mostrare che è

differenziabile in (x_o, y_o) bisogna verificare (si veda (11))
(19)
$$\lim_{|h|\to 0} \frac{f((x_o, y_o) + (h_1, h_2)) - f(x_o, y_o) - \langle \nabla f(x_o, y_o), (h_1, h_2)\rangle}{|h|} = 0\,.$$

L'idea è quella di muoversi lungo segmenti paralleli agli assi principali (si veda la Figura 7) prima dal punto (x_o, y_o) al punto $(x_o + h_1, y_o)$, poi dal punto $(x_o + h_1, y_o)$ al punto $(x_o + h_1, y_o + h_2)$, dopodiché usare la derivabilità per applicare il teorema di Lagrange alle due funzioni di una variabile

$$g_1(x) := f(x, y_o) \qquad \text{e} \qquad g_2^{h_1}(y) := f(x_o + h_1, y)\,.$$

Figura 7



La funzione g_2 in realtà dipende anche da h_1: mentre x_o e y_o sono fissati, h_1 varia. Otteniamo l'esistenza di due punti ξ_1, compreso tra x_o e $x_o + h_1$, e $\xi_2 = \xi_2(h_1)$ anch'esso dipendente da h_1, compreso tra y_o e $y_o + h_2$, per i quali

$$\begin{aligned} g_1(x_o + h_1) - g_1(x_o) &= g_1'(\xi_1) h_1\,, \\ g_2^{h_1}(y_o + h_2) - g_2^{h_1}(y_o) &= (g_2^{h_1})'(\xi_2) h_2\,, \end{aligned}$$

cioè

$$\begin{aligned} f(x_o + h_1, y_o) - f(x_o, y_o) &= \frac{\partial f}{\partial x}(\xi_1, y_o) h_1\,, \\ f(x_o + h_1, y_o + h_2) - f(x_o + h_1, y_o) &= \frac{\partial f}{\partial y}(x_o + h_1, \xi_2) h_2\,. \end{aligned}$$

Per cui

$$
\begin{aligned}
& f(x_o + h_1, y_o + h_2) - f(x_o, y_o) = \\
& = f(x_o + h_1, y_o + h_2) - f(x_o + h_1, y_o) + f(x_o + h_1, y_o) - f(x_o, y_o) = \\
& = \frac{\partial f}{\partial x}(\xi_1, y_o) h_1 + \frac{\partial f}{\partial y}(x_o + h_1, \xi_2) h_2 \,.
\end{aligned}
$$

Il limite (19) allora diventa

$$
\begin{aligned}
& \lim_{|h| \to 0} \frac{\left| \left[\frac{\partial f}{\partial x}(\xi_1, y_o) - \frac{\partial f}{\partial x}(x_o, y_o) \right] h_1 + \left[\frac{\partial f}{\partial y}(x_o + h_1, \xi_2) - \frac{\partial f}{\partial y}(x_o, y_o) \right] h_2 \right|}{|h|} \leqslant \\
& \qquad \leqslant \lim_{|h| \to 0} \frac{\left| \left[\frac{\partial f}{\partial x}(\xi_1, y_o) - \frac{\partial f}{\partial x}(x_o, y_o) \right] h_1 \right|}{|h|} + \\
& \qquad\qquad + \lim_{|h| \to 0} \frac{\left| \left[\frac{\partial f}{\partial y}(x_o + h_1, \xi_2) - \frac{\partial f}{\partial y}(x_o, y_o) \right] h_2 \right|}{|h|} \,.
\end{aligned}
$$

Possiamo stimare l'argomento del primo di questi due limiti (e analogamente il secondo) come segue:

$$
\frac{\left| \left[\frac{\partial f}{\partial x}(\xi_1, y_o) - \frac{\partial f}{\partial x}(x_o, y_o) \right] h_1 \right|}{|h|} \leqslant \left| \frac{\partial f}{\partial x}(\xi_1, y_o) - \frac{\partial f}{\partial x}(x_o, y_o) \right| \,.
$$

Dalla continuità delle derivate parziali in (x_o, y_o) si conclude che tali limiti, e quindi anche (19), sono zero. □

Si osservi che abbiamo utilizzato l'esistenza delle derivate parziali in un intorno del punto x_o, ma è sufficiente tali derivate siano continue solo nel punto x_o. Se tale regolarità per f si ha in tutto un insieme A, cioè $f \in C^1(A)$, allora f risulterà differenziabile in tutto l'insieme A.

EX - Se $f \in C^1(A)$ sono continue anche tutte le derivate direzionali?

Derivazione della composizione di due funzioni - Si considerino due funzioni

$$
f : A \to I \,, \qquad g : I \to \mathbf{R} \,,
$$

A aperto di $\mathbf{R}^n$ e I intervallo aperto di $\mathbf{R}$. **Se f è differenziabile in $x_o \in A$ e g è derivabile in $y_o = f(x_o) \in I$ allora**

$$
g \circ f \qquad \text{è differenziabile in } x_o
$$

e vale

$$\frac{\partial}{\partial x_j}(g \circ f)(x_o) = g'\big(f(x_o)\big)\frac{\partial f}{\partial x_j}(x_o)\,. \tag{20}$$

Teorema 4.10 (del valor medio). *Sia* $f : A \to \mathbf{R}$ *differenziabile in* A, *A aperto di* $\mathbf{R}^n$. *Allora per ogni coppia di punti* $x_o, y_o \in A$ *con il segmento* $[x_o, y_o] \subset A$ *esiste* $\xi \in [x_o, y_o]$ *tale che*

$$f(y_o) - f(x_o) = \langle \nabla f(\xi), y_o - x_o \rangle\,.$$

Dimostrazione - Si consideri la curva $r(t) = (1-t)x_o + ty_o$, $t \in [0,1]$, che unisce i due punti x_o, y_o. Per ipotesi il sostegno di r è contenuto in A. È sufficiente allora applicare il teorema di Lagrange alla funzione

$$h(t) := f\big(r(t)\big)\,.$$

Esiste quindi $\bar{t} \in (0,1)$ tale che $h(1) - h(0) = h'(\bar{t})$. Poiché $h(0) = f(x_o)$, $h(1) = f(y_o)$ e, usando (16), $h'(\bar{t}) = \langle \nabla f(r(\bar{t})), r'(\bar{t}) \rangle = \langle \nabla f(r(\bar{t})), y_o - x_o \rangle$. Si conclude scegliendo ξ il punto $r(\bar{t})$. □

Teorema 4.11. *Siano A un insieme aperto connesso per poligonali e* $f : A \to \mathbf{R}$ *differenziabile. Se* $\nabla f(x) = (0, \dots 0)$ *per ogni* $x \in A$ *allora* f *è costante.*

Dimostrazione - Per ipotesi si ha che, per ogni coppia di punti arbitraria di A, esiste una poligonale che unisce i due punti. Siano x, y due punti arbitrari: vogliamo mostrare che $f(x) = f(y)$ il che implica che f è costante. Sia P una poligonale, contenuta in A, che unisce i punti x, y. Denotiamo con $x_0, x_1, x_2, \dots x_n$ i vertici di tale poligonale, con $x_0 = x$, $x_n = y$. Applicando il teorema del valor medio sul segmento $[x_0, x_1]$ si ha che esiste un punto ξ_1 su tale segmento tale che

$$f(x_1) - f(x_0) = \langle \nabla f(\xi_1), x_1 - x_0 \rangle\,.$$

Per ipotesi $\nabla f(\xi_1)$ è il vettore nullo, per cui $f(x_1) - f(x_0) = 0$, cioè $f(x_1) = f(x_0)$. Ripetendo il ragionamento sul segmento $[x_1, x_2]$ si ottiene che $f(x_2) = f(x_1)$. Iterando su tutti i segmenti si conclude che

$$f(x_0) = f(x_1) = f(x_2) = \dots = f(x_n)$$

il che ci fa concludere. □

È immediato vedere che l'ipotesi che A sia connesso per poligonali è ottimale. È sufficiente prendere un aperto fatto di due componenti connesse per poligonali che siano disgiunte e fissare una funzione che vale una costante c_1 nel primo aperto e una costante $c_2 \neq c_1$ nel secondo.

Concludiamo con un'osservazione semplice, ma molto utile quando si devono cercare massimo e minimo di una funzione regolare.
Sia $x_o \in E$ un punto di massimo locale (oppure di minimo locale) per una funzione $f : E \to \mathbf{R}$, E aperto di $\mathbf{R}^n$. Allora esiste $\rho > 0$ tale che x_o è di massimo per f ristretta a $B_\rho(x_o)$. Supponiamo inoltre che f ammetta tutte le derivate direzionali in x_o.
Poiché x_o è un punto di massimo locale, in particolare la funzione

$$g_v(t) := f(x_o + tv)$$

avrà un massimo locale in corrispondenza di $t = 0$. Allora la sua derivata sarà nulla, ma poiché

$$g_v'(0) = \frac{\partial f}{\partial v}(x_o)$$

concludiamo che $\frac{\partial f}{\partial v}(x_o) = 0$, e ciò per ogni vettore v. In particolare si ha

$$\frac{\partial f}{\partial x_i}(x_o) = 0 \qquad \text{per ogni } i = 1, \dots n\,.$$

Concludiamo che se x_o è un punto di massimo (locale) o di minimo (locale) interno allora

$$\nabla f(x_o) = (0, 0, \dots 0)\,.$$

Dovendo cercare i punti di massimo e minimo per una funzione tra i candidati ci saranno allora i punti nei quali f è derivabile e ha gradiente nullo.
I punti che soddisfano l'equazione $\nabla f(x) = (0, 0, \dots 0)$ sono detti *stazionari* o *critici* per f.

5. Derivate di ordine successivo

Data una funzione $f : A \subset \mathbf{R}^n \to \mathbf{R}$ derivabile rispetto a x_j ci si può chiedere se la funzione $f_{x_j} = \frac{\partial f}{\partial x_j}$ sia a sua volta derivabile, ad esempio rispetto a x_i. Se lo è denoteremo con

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

tale funzione e la chiameremo derivata seconda mista.

Diremo che una funzione $f : A \subset \mathbf{R}^n \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, è di classe C^2 in A se f è continua in A, esistono le derivate prime

$$\frac{\partial f}{\partial x_j}(x), \quad j = 1, \dots n,$$

in A e sono continue in A, esistono le derivate seconde

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(x), \quad i,j = 1,\dots n,$$

in ogni punto $x \in A$ e sono continue in A.

Teorema 5.1 (Schwarz)**.** *Si considerino $f : A \subset \mathbf{R}^n \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, e $x_o \in A$. Si supponga che per i, j interi fissati tra 1 ed n le derivate seconde miste*

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right), \qquad \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

esistano in un intorno di x_o e siano entrambe continue in x_o. Allora

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(x_o) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x_o)\,.$$

Dimostrazione - Vediamola nel caso di $\mathbf{R}^2$. Si fissi un punto (x_o, y_o) nel dominio di f e si supponga che le derivate miste

$$\frac{\partial^2 f}{\partial y \partial x} \quad \text{e} \quad \frac{\partial^2 f}{\partial x \partial y} \qquad \begin{array}{l} \text{esistano in un intorno } I \times J \text{ di } (x_o, y_o) \\ \text{e siano entrambe continue in } (x_o, y_o). \end{array}$$

Si considerino ora h, k in modo tale che $h + x_o \in I$, $k + y_o \in J$. Definiamo ora le due funzioni

$$F^{(k)}(h) := f(x_o + h, y_o + k) - f(x_o + h, y_o),$$
$$G^{(h)}(k) := f(x_o + h, y_o + k) - f(x_o, y_o + k).$$

Si osservi ora che

$$F^{(k)}(h) - F^{(k)}(0) = G^{(h)}(k) - G^{(h)}(0). \tag{21}$$

Applicando il teorema di Lagrange a $F^{(k)}$ si ottiene l'esistenza di ξ_1 compreso tra x_o e $x_o + h$ tale che

$$F^{(k)}(h) - F^{(k)}(0) = h\,\left[\frac{\partial f}{\partial x}(\xi_1, y_o + k) - \frac{\partial f}{\partial x}(\xi_1, y_o)\right]$$

Poiché $\frac{\partial f}{\partial x}$ ammette derivata rispetto a y applichiamo nuovamente il teorema di Lagrange e otteniamo l'esistenza di η_1 compreso tra y_o e $y_o + k$ tale che

$$\begin{aligned} F^{(k)}(h) - F^{(k)}(0) &= h\,\left[\frac{\partial f}{\partial x}(\xi_1, y_o + k) - \frac{\partial f}{\partial x}(\xi_1, y_o)\right] = \\ &= h\,k\,\frac{\partial}{\partial y}\frac{\partial f}{\partial x}(\xi_1, \eta_1). \end{aligned}$$

In maniera analoga si ottengono ξ_2, η_2 tali che

$$G^{(h)}(k) - G^{(h)}(0) = h\,k\,\frac{\partial}{\partial x}\frac{\partial f}{\partial y}(\xi_2, \eta_2).$$

Da (21) si ottiene

$$\frac{\partial}{\partial y}\frac{\partial f}{\partial x}(\xi_1, \eta_1) = \frac{\partial}{\partial x}\frac{\partial f}{\partial y}(\xi_2, \eta_2)$$

e passando al limite per $h, k \to 0$ si ottiene la tesi. $\square$

Osservazione 5.2. - In particolare se $f \in C^2(A)$ allora $\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(x) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x)$ per ogni $x \in A$.

Osservazione 5.3. - Se f soddisfa le ipotesi del teorema di Schwarz in x_o non è difficile verificare che anche tutte le derivate direzionali soddisfano la tesi del teorema, cioè per ogni $v, w \in \mathbf{R}^n$ con $|v| = |w| = 1$ vale

$$\frac{\partial}{\partial v}\left(\frac{\partial f}{\partial w}\right)(x_o) = \frac{\partial}{\partial w}\left(\frac{\partial f}{\partial v}\right)(x_o)\,.$$

Osservazione 5.4. - Usando il precedente teorema si può mostrare che se $f \in C^k(A)$ allora le derivate miste di ordine minore o uguale a k non dipendono dall'ordine nel quale si eseguono le derivate.

EX - Si può verificare che la seguente funzione

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

è $C^1(\mathbf{R}^2) \cap C^2(\mathbf{R}^2 \setminus \{(0,0)\}$, e ammette le derivate seconde anche in $(0,0)$, ma

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)(0,0) = -1, \qquad \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)(0,0) = 1\,.$$

6. Forme quadratiche

Ricordiamo qui di seguito qualcosa che dovrebbe essere già noto dai corsi di algebra lineare.

Una forma bilineare è un'applicazione B definita in $\mathbf{R}^n \times \mathbf{R}^n$ e a valori in $\mathbf{R}$ che sia lineare sulle singole componenti di $\mathbf{R}^n$, cioè ogni volta che si fissa $y \in \mathbf{R}^n$ la mappa $x \mapsto B(x,y)$ è lineare in x e ogni volta che si

fissa $x \in \mathbf{R}^n$ la mappa $y \mapsto B(x,y)$ è lineare in y.
In sostanza un'applicazione bilineare è rappresentata nel modo seguente

$$(x,y) \mapsto \sum_{i,j=1}^{n} a_{ij} x_j y_i \,, \qquad\qquad x, y \in \mathbf{R}^n \,,$$

cioè esiste una matrice quadrata $n \times n$ $A = [a_{ij}]_{i,j=1}^{n}$ tale che

$$B(x,y) = \langle A \cdot x, y \rangle$$

dove $A \cdot x$ indica il prodotto matrice-vettore. Ricordiamo che una matrice $A = [a_{ij}]_{i,j=1}^{n}$ si dice simmetrica se $a_{ij} = a_{ji}$. Una forma quadratica Q altro non è che una mappa da $\mathbf{R}^n$ in $\mathbf{R}$ tale che

$$Q(x) = B(x,x) \qquad\qquad \text{con } B \text{ bilineare}$$

cioè un polinomio omogeneo di secondo grado del tipo

$$Q(x) = \langle A \cdot x, x \rangle = \sum_{i,j=1}^{n} a_{ij} x_i x_j \,, \qquad x \in \mathbf{R}^n \,.$$

Si noti come matrici diverse possono definire la stessa forma quadratica. Ad esempio le tre matrici (ma ce ne sono infinite)

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 3 \\ -1 & 1 \end{pmatrix},$$

danno luogo alla stessa forma quadratica $\mathbf{R}^2 \ni (x,y) \mapsto x^2 + 2xy + y^2$.
Ci si può sempre ridurre al caso di matrici simmetriche sostituendo ad a_{ij} la quantità $\tilde{a}_{ij} := \dfrac{a_{ij} + a_{ji}}{2}$. La forma quadratica così ottenuta coincide con la prima e ha i coefficienti simmetrici. Infatti $\tilde{a}_{ij} x_i x_j + \tilde{a}_{ji} x_j x_i = a_{ij} x_i x_j + a_{ji} x_j x_i$.
Quindi, anche se ci sono infinite matrici che definiscono la medesima forma quadratica, una sola fra queste è simmetrica (ed è l'unica matrice simmetrica che la definisce). Nell'esempio precedente *la* matrice simmetrica è ovviamente

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Nel seguito considereremo per lo più matrici simmetriche quindi qualche volta confonderemo la forma quadratica con la matrice, come ad esempio nella definizione che segue, che potrebbe essere data per le forme quadratiche anziché per le matrici.

Definizione 6.1. *Diremo che una matrice A, quadrata di ordine n e a coefficienti reali, è*

i) *definita positiva* (*rispettivamente negativa*) *se* $\langle A \cdot x, x \rangle > 0$ (*rispettivamente* $\langle A \cdot x, x \rangle < 0$) *per ogni* $x \neq (0, \dots 0)$*;*

ii) *diremo che una matrice A è semi-definita positiva (negativa) se $\langle A\cdot x, x\rangle \geqslant 0$ ($\langle A\cdot x, x\rangle \leqslant 0$) per ogni $x \in \mathbf{R}^n$ ed esiste un punto $\bar{x} \neq (0, \dots 0)$ tale che $\langle A\cdot\bar{x}, \bar{x}\rangle = 0$;*

iii) *diremo che una matrice è indefinita se esistono $\bar{x}, \bar{y}$ tali che $\langle A\cdot\bar{x}, \bar{x}\rangle > 0$ e $\langle A\cdot\bar{y}, \bar{y}\rangle < 0$.*

Ricordiamo una disuguaglianza utile (vedi EX): data una matrice $n\times n$ reale e simmetrica A (che quindi ha autovalori reali) vale che

$$\lambda_{\min}|x|^2 \leqslant \langle A\cdot x, x\rangle \leqslant \lambda_{\max}|x|^2 \tag{22}$$

dove $\lambda_{\min}$ indica il minore e $\lambda_{\max}$ il maggiore degli autovalori di A.
Di conseguenza possiamo affermare che una matrice A quadrata di ordine n, a coefficienti reali e simmetrica è

i) definita positiva (rispettivamente negativa) se tutti gli autovalori sono positivi (rispettivamente negativi) o equivalentemente se $\lambda_{\min} > 0$ (rispettivamente $\lambda_{\max} < 0$);

ii) diremo che una matrice A è semi-definita positiva (negativa) se $\lambda_{\min} = 0$ ($\lambda_{\max} = 0$);

iii) diremo che una matrice è indefinita se $\lambda_{\min} < 0$ e $\lambda_{\max} > 0$.

Da ciò si ricava come per capire quale sia la natura di una forma quadratica (o della matrice simmetrica che la definisce) sia sufficiente capire qual è il segno degli autovalori.

Data una matrice $A = [a_{ij}]_{i,j=1}^n$ $n\times n$ denotiamo con A_1, A_2, $\dots A_n$ i cosiddetti minori principali di *nord-ovest*, cioè se

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

i minori principali sono

$$A_1 = (a_{11}), A_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad A_3 = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \dots A_n = A.$$

I minori principali sono di più e quelli che noi abbiamo appena considerato. Noi però ci limiteremo a considerare solo questi e perciò qualche volta li chiameremo semplicemente minori principali.

Una matrice A a coefficienti reali e simmetrica può essere sempre diagonalizzata, cioè è possibile trovare Q altra matrice tale che

$$A = Q^T \Lambda Q \tag{23}$$

dove Q^T denota la trasposta di Q, Q ha la proprietà che $Q^T = Q^{-1}$ e

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

dove $\lambda_1, \dots \lambda_n$ sono gli autovalori, che sono tutti reali. Chiaramente la matrice Λ è:

definita positiva	se e solo se	$\lambda_i > 0$ per ogni i
semi-definita positiva	se e solo se	$\lambda_i \geqslant 0$ per ogni i ed esiste j tale che $\lambda_j = 0$,
definita negativa	se e solo se	$\lambda_i < 0$ per ogni i,
semi-definita negativa	se e solo se	$\lambda_i \leqslant 0$ per ogni i ed esiste j tale che $\lambda_j = 0$,
indefinita	se e solo se	esistono i, j tali che $\lambda_i > 0, \lambda_j < 0$.

Si osservi che un modo diverso per dire che la matrice Λ è definita positiva è

$$\lambda_1 \lambda_2 \cdot \ldots \cdot \lambda_k > 0 \qquad \text{per ogni } k \text{ tra } 1 \text{ e } n,$$

cioè se il prodotto dei primi k autovalori è sempre positivo. Un modo per dire che la matrice Λ è definita negativa è

$$(-1)^k \lambda_1 \cdot \ldots \cdot \lambda_k > 0 \qquad \text{per ogni } k \text{ tra } 1 \text{ e } n,$$

cioè se il prodotto dei primi k autovalori è a segni alterni, negativo quando k è dispari, positivo quando k è pari.

Si ricava (va mostrato, e noi non lo faremo) che più in generale, ma in maniera in qualche modo analoga, una matrice A simmetrica

i) è definita positiva se e solo se $|A_k| > 0$ per ogni $k = 1, \ldots n$, oppure (equivalentemente) se tutti gli autovalori sono positivi;

ii) è definita negativa se e solo se $(-1)^k |A_k| > 0$ per ogni $k = 1, \ldots n$, oppure (equivalentemente) se tutti gli autovalori sono negativi,

dove con $|M|$ denotiamo il determinante di M, M matrice quadrata.

Se quello che interessa è capire il segno degli autovalori si comprende come possa essere utile conoscere il segno dei determinanti dei minori

principali perlomeno, e di fatto solamente, nei casi i) e ii) appena visti. A tal proposito si vedano i due esempi più in basso.
Un modo utile e alternativo al calcolo del determinante dei minori è calcolare direttamente gli autovalori trovando gli zeri del polinomio caratteristico; ma, primo, non è necessario trovarli se siamo interessati solo al loro segno, secondo, può essere complicato o impossibile (a parte se il grado del polinomio è due). Qualche volta può essere utile per capire il segno degli autovalori considerare non solo il determinante, che è il prodotto degli autovalori, ma anche la traccia, che ne è la somma, cioè

$$\mathrm{tr}(A) = \sum_{j=1}^{n} \lambda_j, \qquad \det(A) = \prod_{j=1}^{n} \lambda_j, \qquad \lambda_i \text{ autovalori.}$$

Se la matrice in questione è una 2×2 questo risolve completamente il problema perché se conosciamo il segno di $\lambda_1\lambda_2$ e di $\lambda_1 + \lambda_2$ ricaviamo anche il segno di λ_1 e λ_2, ma queste informazioni possono essere utili anche per matrici più grandi, anche se non bastano a determinare il segno di tutti i λ_i.

Attenzione! Per quanto sia utile studiare il segno dei determinanti dei minori principali, questi non ci forniscono il segno degli autovalori, o perlomeno lo fanno solo in casi particolari (come i casi i) e ii) della pagina precedente). Si consideri ad esempio la matrice

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Si ha $|A_1| = 0$, $|A_2| = -1$, $|A_3| = |A| = -1$. Si può concludere che un autovalore è nullo e due negativi? Ovviamente no, visto che il determinanate di A, $|A|$, è -1 ed è anche il prodotto dei tre autovalori. Poiché il prodotto dei tre autovalori è negativo o uno o tutti e tre sono negativi e in particolare nessuno dei tre è zero; poiché la traccia della matrice, cioè la somma degli elementi sulla diagonale che coincide con la somma degli autovalori, è 1, cioè è positiva, almeno un autovalore è positivo. In effetti gli autovalori sono 1 (con molteplicità 2) e -1.
Anche se **tutti** i determinanti dei minori principali sono diversi da zero non possiamo concludere, in generale, che tali determinanti siano gli autovalori e nemmeno che i segni dei determinanti dei minori principali forniscano i segni degli autovalori. Ad esempio, i determinanti dei

minori principali della matrice

$$A = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

sono tutti e tre uguali a -1, ma gli autovalori non sono tutti e tre negativi. Infatti si vede immediatamente che la traccia è nulla, per cui la somma degli autovalori è zero, ma essendo il determinante non nullo sicuramente gli autovalori sono tutti diversi da zero e la loro somma si compensa. C'è quindi un autovalore negativo e almeno uno *positivo*, nonostante tutti e tre i minori principali siano negativi.
Volendo calcolarli si ottengono i tre valori $1, (\sqrt{6}-1)/2, -(\sqrt{6}+1)/2$.

7. Differenziale secondo, matrice hessiana e formula di Taylor al 2° ordine

Dati una funzione $f : A \subset \mathbf{R}^n \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, e un punto $x_o \in A$ si supponga che f sia derivabile due volte in un intorno di x_o e tali derivate siano continue in x_o (questo solo per far valere il teorema di Schwarz e avere le derivate seconde miste uguali). Consideriamo la funzione (che risulta derivabile due volte)

$$F_v(t) := f(x_o + tv) \tag{24}$$

con v vettore di modulo 1 e valutiamone lo sviluppo di Taylor al secondo ordine. Cominciamo con il calcolarne le derivate:

$$F_v'(t) = \langle \nabla f(x_o + tv), v \rangle = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_o + tv) v_i,$$

$$F_v''(t) = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(x_o + tv) v_i v_j.$$

Avremo allora

$$F_v(t) = F_v(0) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_o) v_i t + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(x_o) v_i v_j t^2 + o(t^2)\,,$$

cioè

$$f(x_o + tv) = f(x_o) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_o) v_i t + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(x_o) v_i v_j t^2 + o(t^2)\,,$$

Questo fa intuire che, volendo definire un differenziale secondo, tale differenziale dovrà essere (perlomeno se l'incremento rispetto ad x_o è

lineare come tv) una forma quadratica rappresentata dalla matrice delle derivate seconde nel punto x_o.
Definiremo le derivate direzionali in direzione v del gradiente di f come

$$\lim_{t\to 0} \frac{\nabla f(x_o+tv)-\nabla f(x_o)}{t}$$

dove il limite sarà un vettore dato da

$$\left(\lim_{t\to 0}\frac{\frac{\partial f}{\partial x_1}(x_o+tv)-\frac{\partial f}{\partial x_1}(x_o)}{t},\dots,\lim_{t\to 0}\frac{\frac{\partial f}{\partial x_n}(x_o+tv)-\frac{\partial f}{\partial x_n}(x_o)}{t}\right)=$$
$$=\left(\left\langle\nabla\frac{\partial f}{\partial x_1}(x_o),v\right\rangle,\dots,\left\langle\nabla\frac{\partial f}{\partial x_n}(x_o),v\right\rangle\right)=H_f(x_o)\cdot v$$

dove

$$H_f(x_o) = \begin{pmatrix} \nabla\frac{\partial f}{\partial x_1}(x_o) \\ \\ \nabla\frac{\partial f}{\partial x_n}(x_o)\end{pmatrix} =$$

$$= \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_o) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_o) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x_o) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_o) & \frac{\partial^2 f}{\partial x_2^2}(x_o) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x_o) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x_o) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x_o) & \cdots & \frac{\partial^2 f}{\partial^2 x_n}(x_o) \end{pmatrix}$$

è chiamata matrice *hessiana*.

Di conseguenza la matrice hessiana di f nel punto x_o può essere associata alla forma bilineare

$$B(v,w)=\langle H_f(x_o)\cdot v,w\rangle=\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_j\partial x_i}(x_o)v_jw_i.$$

Definizione 7.1. *Diremo che una funzione $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, è differenziabile due volte in $x_o \in A$ se $df : A \to \mathcal{L}(\mathbf{R}^n;\mathbf{R})$ è differenziabile in x_o o, equivalentemente, se $\nabla f : A \to \mathbf{R}^n$ è differenziabile in x_o, il che equivale a dire che $\frac{\partial f}{\partial x_i}$ è differenziabile in x_o per ogni $i \in \{1,\dots n\}$.*

Commento alla definizione - Dire che il differenziale primo è differenziabile in x_o significa dire che esiste un'applicazione L tale che

$$\lim_{|h|\to 0} \frac{df_{x_o+h} - df_{x_o} - Lh}{|h|} = 0. \tag{25}$$

L'applicazione lineare L apparterrà allo spazio $\mathcal{L}\big(\mathbf{R}^n; \mathcal{L}(\mathbf{R}^n; \mathbf{R})\big)$ che è isomorfo allo spazio delle applicazioni bilineari da $\mathbf{R}^n \times \mathbf{R}^n$ in $\mathbf{R}$.
Poiché il differenziale primo è rappresentato dal gradiente di f in x_o (25) è equivalente a dire che

$$\lim_{|h|\to 0} \frac{\nabla f(x_o + h) - \nabla f(x_o) - Lh}{|h|} = 0.$$

Questo limite è vettoriale e, componente per componente, possiamo scrivere

$$\lim_{|h|\to 0} \frac{\frac{\partial f}{\partial x_i}(x_o + h) - \frac{\partial f}{\partial x_i}(x_o) - L_i h}{|h|} = 0$$

dove L_i è la i-esima componente dell'applicazione L. Sappiamo che in questo caso

$$L_i \quad \text{è rappresentata da} \quad \nabla \frac{\partial f}{\partial x_i}(x_o) = \left(\frac{\partial^2 f}{\partial x_1 \partial x_i}(x_o), \ldots, \frac{\partial^2 f}{\partial x_n \partial x_i}(x_o) \right)$$

e ritroviamo quanto visto prima se l'incremento h è uguale a tv con $t \in \mathbf{R}$ e v vettore di modulo 1.
Per cui il differenziale secondo è una funzione

$$d^2 f : A \to \mathcal{L}\big(\mathbf{R}^n; \mathcal{L}(\mathbf{R}^n; \mathbf{R})\big) \simeq \mathcal{BL}\big(\mathbf{R}^n \times \mathbf{R}^n; \mathbf{R}\big)$$

rappresentato dalla matrice hessiana definita precedentemente, dove con $\mathcal{BL}$ denotiamo l'insieme delle forme bilineari e con $\simeq$ intendiamo uguaglianza a meno di un isomorfismo.
Nel caso in cui si valuti la forma bilineare $d^2 f_{x_o}$ nella coppia (v, v) scriveremo semplicemente

$$d^2 f_{x_o}(v)$$

per denotare la forma quadratica. Tale forma quadratica altro non è che

$$v \mapsto \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x_o) v_i v_j$$

e ha la proprietà che

$$\lim_{|h|\to 0} \frac{f(x_o + h) - f(x_o) - df_{x_o}(h) - \frac{1}{2} d^2 f_{x_o}(h)}{|h|^2} = 0.$$

Come già detto chiameremo matrice hessiana la matrice delle derivate seconde

$$\big(H_f(x_o)\big)_{ij} = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_o)$$

che risulta simmetrica se f ha le derivate seconde continue in x_o e in generale risulta simmetrica in A se $f \in C^2(A)$.
In realtà vale il seguente risultato, più generale rispetto al Teorema 5.1, la cui dimostrazione non riportiamo anche se non molto diversa da quella del Teorema 5.1.

Teorema 7.2. *Sia $f : A \subset \mathbf{R}^n \to \mathbf{R}$, A aperto di $\mathbf{R}^n$ una funzione due volte differenziabile in $x_o \in A$. Allora la forma bilineare $d^2 f_{x_o}$ è simmetrica.*

Dimostrazione - Senza dimostrazione. □

Ovviamente, nelle ipotesi del teorema appena enunciato, la matrice $H_f(x_o)$ che rappresenta il differenziale secondo in x_o è simmetrica.
D'ora in poi, per semplicità, anziché considerare una funzione differenziabile due volte in un punto oppure di classe C^1 le cui derivate siano differenziabili in un punto considereremo funzioni di classe C^2.

Dopo lo sviluppo di Taylor al primo ordine visto in (13), possiamo scrivere lo sviluppo di Taylor in x_o al secondo ordine per $f \in C^2(A)$

$$f(x_o + h) = f(x_o) + \langle \nabla f(x_o), h \rangle + \frac{1}{2} \langle H_f(x_o) \cdot h, h \rangle + o(|h|^2) \tag{26}$$

per $x_o \in A$ e il segmento $[x_o, x_o + h] \subset A$.
Vediamo ora come la matrice hessiana, o la forma quadratica ad essa legata, possa essere utile nello studio della natura dei punti critici di una funzione di classe C^2.

Teorema 7.3. *Siano $f : A \to \mathbf{R}$, A aperto di $\mathbf{R}^n$, $x_o \in A$, $f \in C^2(A)$. Se x_o è un punto stazionario per f allora*

- *i) se x_o è un punto di minimo locale per f allora $\langle H_f(x_o) \cdot v, v \rangle \geqslant 0$ per ogni $v \in \mathbf{R}^n$;*
- *ii) se la matrice hessiana $H_f(x_o)$ è definita positiva allora il punto x_o è un punto di minimo stretto;*
- *iii) se x_o è un punto di massimo locale per f allora $\langle H_f(x_o) \cdot v, v \rangle \leqslant 0$ per ogni $v \in \mathbf{R}^n$;*
- *iv) se la matrice hessiana $H_f(x_o)$ è definita negativa allora il punto x_o è un punto di massimo stretto.*

Dimostrazione - Per ipotesi ricordiamo che in tutti i casi si ha che $\nabla f(x_o) = (0, \dots 0)$.

Punto i) - Se x_o è di minimo locale per f, in particolare $t = 0$ è punto di minimo locale per la funzione F_v definita in (24). Allora $F_v''(0) \geqslant 0$ e come già visto

$$F_v''(0) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x_o) v_i v_j \, .$$

Poiché ciò è vero per ogni vettore v la tesi è dimostrata.
Punto ii) - Per definizione per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che $o(|h|^2) < \varepsilon |h|^2$ per ogni $h \in B_\delta(0)$. Poiché la matrice hessiana è definita positiva si ha $\langle H_f(x_o) \cdot h, h \rangle \geqslant \lambda |h|^2$ con $\lambda > 0$ (dove λ è il minimo autovalore di $H_f(x_o)$, si veda (22)). Di conseguenza utilizzando (26) con $x = x_o + h$ e scegliendo $\varepsilon < \lambda$ si ha

$$f(x) - f(x_o) = \frac{1}{2} \langle H_f(x_o) \cdot (x - x_o), x - x_o \rangle + o(|x - x_o|^2) \geqslant (\lambda - \varepsilon)|x - x_o|^2$$

il che conclude la dimostrazione. I punti *iii)* e *iv)* si mostrano in maniera analoga. □

Nel teorema appena visto non sono contemplati tutti i casi possibili, solamente perché negli altri casi nulla si può dire, in generale. Vediamo alcuni esempi, prima riguardanti le condizioni analizzate nel teorema, poi riguardanti i casi non contemplati. Si considerino le funzioni

$$\begin{aligned} f_1(x,y) &= x^2, \\ f_2(x,y) &= x^2 + y^2, \\ f_3(x,y) &= -x^2 - y^2. \end{aligned}$$

Il punto $(0,0)$ è stazionario per tutte e tre le funzioni. Nel primo caso, in cui $(0,0)$ è di minimo, ma non di minimo stretto, la matrice hessiana è data da

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \tag{27}$$

che risulta solamente semidefinita positiva e non definita positiva (punto i) del teorema). Nel secondo caso la matrice hessiana è data da

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

ed è definita positiva e il punto risulta di minimo stretto. Analoga considerazione per f_3, nel qual caso il punto $(0,0)$ risulta di massimo stretto. Se la matrice è indefinita il punto, che non è né di minimo né di massimo locale, si dice di sella, ed è il caso per esempio della funzione

$$f_4(x,y) = x^2 - y^2.$$

Se però la matrice è semidefinita positiva (o negativa) non possiamo concludere nulla. Ad esempio la matrice in (27) potrebbe essere la matrice hessiana di entrambe le funzioni

$$f_5(x,y) = x^2 + y^4,$$
$$f_6(x,y) = x^2 - y^4.$$

Quello che si può dire è solamente che il punto non sarà di massimo, ma non si sa se è di sella o di minimo. Nel caso in cui si avesse la matrice nulla, cioè

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

tutti i casi sono possibili. Tale matrice infatti potrebbe essere l'hessiana di

$$f_7(x,y) = x^4 + y^4,$$
$$f_8(x,y) = x^4 - y^4,$$
$$f_9(x,y) = -x^4 - y^4.$$

In questi casi $(0,0)$ risulta di minimo stretto per f_7, di sella per f_8, di massimo stretto per f_9.
Più complicate sono le considerazioni se la dimensione è più alta di 2. L'idea è quella di cercare di conoscere il segno degli autovalori, come analizzato alla fine del paragrafo precedente.
Qualora un autovalore fosse nullo un tentativo che può essere fatto è quello di cercare di analizzare il segno della funzione.

8. Funzioni convesse

Definizione 8.1. *Una funzione $f : A \to \mathbf{R}$, A insieme convesso di $\mathbf{R}^n$, si dice* convessa *se per ogni $x, y \in A$ e per ogni $t \in [0,1]$ vale*

$$f\big((1-t)x + ty\big) \leqslant (1-t)f(x) + tf(y).$$

La funzione f è detta strettamente convessa *se la disuguaglianza sopra è stretta per ogni $t \in (0,1)$.*

Osservazione 8.2. - Si osservi che f è convessa se e solo se la sua restrizione ad ogni segmento contenuto in A è una funzione convessa di una variabile, cioè se e solo se per ogni $x, y \in A$ la funzione

$$F(t) := f\big(r(t)\big) \qquad \text{dove } r(t) = (1-t)x + ty \tag{28}$$

è convessa.
La cosa è ovvia, ma verifichiamolo. A tal proposito si definisca, per $x, y \in A$ fissati, la funzione

$$F(t) := f\big(r(t)\big) \qquad \text{dove } r(t) = (1-t)x + ty. \tag{29}$$

Per $s \in [0,1]$ e $t_1, t_2 \in [0,1]$ si ha che se f è convessa

$$\begin{aligned} F\big((1-s)t_1 + st_2\big) &= f\big((1-s)[(1-t_1)x + t_1 y] + s[(1-t_2)x + t_2 y]\big) \leqslant \\ &\leqslant (1-s) f\big((1-t_1)x + t_1 y\big) + s f\big((1-t_2)x + t_2 y\big) = \\ &= (1-s)F(t_1) + sF(t_2). \end{aligned}$$

Viceversa, si supponga, per ogni $x, y \in A$ e per ogni curva r, che F sia convessa. Allora

$$f\big((1-t)x + ty\big) = F(t) = F\big((1-t)0 + t1\big) \leqslant (1-t)F(0) + tF(1) = (1-t)f(x) + tf(y).$$

Ricordo - In dimensione 1 una funzione $f \in C^1(I)$ è convessa se e solo se

$$f(x) \geqslant f(x_o) + f'(x_o)\big(x - x_o\big) \tag{30}$$

per ogni $x, x_o \in I$; inoltre è convessa se e solo se

$$x \mapsto f'(x) \qquad \text{è crescente.} \tag{31}$$

Teorema 8.3. *Sia $f : A \to \mathbf{R}$, A insieme convesso di $\mathbf{R}^n$, f differenziabile in A. Le seguenti affermazioni sono equivalenti:*

a) f è convessa;
b) $f(x) \geqslant f(x_o) + \langle \nabla f(x_o), x - x_o \rangle$ per ogni $x, x_o \in A$;
c) $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geqslant 0$ per ogni $x, y \in A$.

Dimostrazione - Mostriamo le tre implicazioni $a\,) \Rightarrow b\,)$, $b\,) \Rightarrow c\,)$, $c\,) \Rightarrow a\,)$, da cui avremo le tre equivalenze.

$a\,) \Rightarrow b\,)$ - Dall'ipotesi si ha

$$f\big((1-t)x + t\,x_o\big) \leqslant (1-t)f(x) + tf(x_o) = f(x) + t\big(f(x_o) - f(x)\big)$$

da cui, per $t \in (0,1)$,

$$f(x) - f(x_o) \geqslant \frac{f\big(x + t(x_o - x)\big) - f(x)}{t}.$$

Passando al limite per $t \to 0$ e utilizzando la formula (16) si ottiene la tesi.

$b\,) \Rightarrow c\,)$ - Per ipotesi valgono

$$\begin{aligned} &f(x) - f(y) \geqslant \langle \nabla f(y), x - y \rangle, \\ &f(y) - f(x) \geqslant \langle \nabla f(x), y - x \rangle = -\langle \nabla f(x), x - y \rangle \end{aligned}$$

per ogni $x, y \in A$. Sommando si ottiene la tesi.

$c\,) \Rightarrow a\,)$ - Dati $x, y \in A$ si consideri la funzione

$$F(t) = f\big(r(t)\big) = f\big((1-t)x + ty\big), \qquad r(t) = (1-t)x + ty.$$

Si osservi che

$$F'(t) = \big\langle \nabla f\big(r(t)\big), r'(t) \big\rangle, \tag{32}$$

e che, per $t, s \in [0.1]$,

$$r(t) - r(s) = (1-t)x + ty - (1-s)x + sy = (t-s)(y-x),$$

da cui, per $t > s$,

$$\begin{aligned} F'(t) - F'(s) &= \big\langle \nabla f\big(r(t)\big), r'(t)\big\rangle - \big\langle \nabla f\big(r(s)\big), r'(s)\big\rangle = \\ &= \big\langle \nabla f\big(r(t)\big) - \nabla f\big(r(s)\big), y - x\big\rangle = \\ &= \frac{1}{t-s}\big\langle \nabla f\big(r(t)\big) - \nabla f\big(r(s)\big), r(t) - r(s)\big\rangle. \end{aligned}$$

Per ipotesi si deduce quindi che, per $t > s$,

$$\Big(F'(t) - F'(s)\Big)(t-s) \geqslant 0,$$

cioè F' crescente. Da (31) e dall'Osservazione 8.2 si conclude. $\square$

Ricordo - In dimensione 1 una funzione $f \in C^1(I)$ con f' derivabile in I è convessa se e solo se

$$f''(x) \geqslant 0. \tag{33}$$

In particolare la funzione F definita in (29) ha derivata seconda non negativa.

Teorema 8.4. *Sia $f : A \to \mathbf{R}$, A insieme convesso di $\mathbf{R}^n$, $f \in C^2(A)$. Le seguenti affermazioni sono equivalenti:*

a) f è convessa;
b) $H_f(x) \geqslant 0$ per ogni $x \in A$.

Dimostrazione - $a) \Rightarrow b)$ - Da quanto visto la funzione F definita in (29) ha derivata seconda non negativa. Chiamando v il vettore $y - x$ e calcolando F'' si ottiene

$$F''(t) = \big\langle H_f\big(r(t)\big) \cdot v, v\big\rangle \geqslant 0.$$

In particolare per $t = 0$ si ha che $\big\langle H_f(x) \cdot v, v\big\rangle \geqslant 0$ per ogni v.

$b) \Rightarrow a)$ - Viceversa, dal teorema del valor medio (per funzioni di una variabile) applicato a F' si ha che per $t, s \in [0.1]$ dati esiste ξ compreso tra t ed s tale che

$$F'(t) - F'(s) = F''(\xi)\big(t - s\big).$$

Dal calcolo di F'' e da (32) quest'ultima uguaglianza, valutata per $t = 1$ e $s = 0$, diventa

$$\big\langle \nabla f(y), y - x\big\rangle - \big\langle \nabla f(x), y - x\big\rangle = \big\langle H_f\big(r(\xi)\big) \cdot (y-x), y - x\big\rangle \geqslant 0.$$

Dal punto $c)$ del del teorema precedente si conclude. $\square$

Osservazione 8.5. - Sia $f : A \to \mathbf{R}$ una funzione convessa di classe C^1 e A è un insieme convesso. Se un punto x_o è critico per f allora è di minimo globale.
Infatti, dal punto b) del Teorema 8.3 si ha che $f(x_o) \leqslant f(x)$ per ogni $x \in A$.
Inoltre se f è strettamente convessa e x_o è critico allora questo è l'unico punto di minimo.
Infatti se ce ne fossero due, x_o e y_o, si avrebbe che

$$f\left(\frac{x_o + y_o}{2}\right) < \frac{1}{2} f(x_o) + \frac{1}{2} f(y_o) = \min f$$

e questo non è possibile.

9. Il teorema delle funzioni implicite: caso di funzioni scalari di due variabili

Abbiamo già affrontato il problema di studiare gli insiemi di livello e visto che possono essere di diverso tipo. Nel caso, ad esempio, di una funzione $f : A \subset \mathbf{R}^2 \to \mathbf{R}$, A aperto, spesso vengono chiamati *curve di livello*, anche se sempre curve non sono, come ad esempio se si considerano gli insiemi di livello 0 e -1 della funzione $f(x, y) = x^2 + y^2$: il primo è l'insieme $\{(0, 0)\}$, il secondo è il vuoto.
In altri casi si possono avere unione di curve incidenti o anche tangenti tra loro, comunque più curve con un punto in comune. Altre volte, anche se l'insieme di livello è una curva, questa non sempre è regolare.
Si vedano gli esempi in Figura 8: nella prima è indicato l'insieme degli zeri di $f(x, y) = x^2 - y^2$, nella seconda di $f(x, y) = (y - x^2)(y - 2x^2)$, nella terza di $f(x, y) = x^3 - y^2$.

Figura 8.a

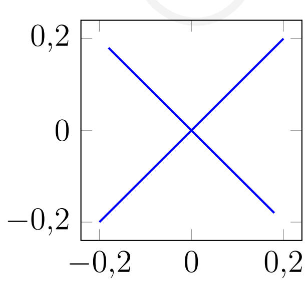


Figura 8.b

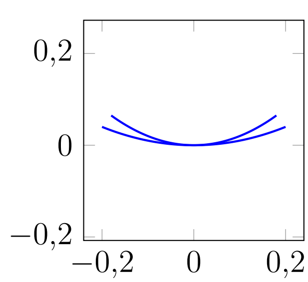
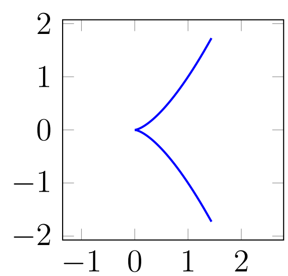


Figura 8.c



Affrontiamo ora il seguente problema: si supponga di avere una funzione $f : A \subset \mathbf{R}^n \to \mathbf{R}$, A aperto. Ci si chiede quando, data una costante $c \in \mathbf{R}$, l'insieme di livello c di f è localmente grafico di una funzione ed eventualmente quale regolarità ha tale funzione.
Per semplicità si considerano gli insiemi di livello zero, cioè i luoghi degli zeri, di f; se si vuole studiare l'insieme di livello c è sufficiente infatti considerare il luogo degli zeri della funzione $f - c$.

Caso di funzioni scalari di 2 variabili - Tornando per semplicità al caso $n = 2$, ci chiediamo quando il luogo degli zeri di una certa funzione è una curva che inoltre è localmente un grafico, cioè quando il luogo degli zeri è una curva cartesiana.
Il teorema delle funzioni implicite, noto anche come teorema del Dini, fornisce condizioni sufficienti a garantire, per funzioni definite in $\mathbf{R}^2$, che il luogo degli zeri sia una curva cartesiana.

Prima di vedere il teorema facciamo qualche considerazione euristica: si supponga che il luogo degli zeri di f definita in $\mathbf{R}^2$ sia una curva cartesiana regolare in un intorno di un punto (x_o, y_o), diciamo almeno derivabile, in modo da avere retta tangente. Tale retta sarà esprimibile a sua volta come luogo di zeri della funzione $p(x, y) = ax + by + c$ per qualche $a, b, c \in \mathbf{R}$.
Abbiamo due possibilità: la curva cercata può essere del tipo $(x, g(x))$ o $(h(y), y)$ con g e h funzioni di una variabile. Si supponga il primo caso, cioè che la curva sia grafico di una funzione della variabile x. Allora nel punto x_o tale curva avrà la derivata uguale a quella della sua retta tangente. Poiché la curva è cartesiana rispetto alla variabile x anche la retta tangente lo sarà, cioè la retta $ax + by + c$ potrà essere espressa anch'essa come funzione di x come segue

$$y = -\frac{a}{b}x - \frac{c}{b}.$$

! Attenzione! Ciò è vero solamente se $b \neq 0$, e si avrà $y'(x_o) = -a/b$. Questa condizione può essere vista anche come

$$\frac{\partial p}{\partial y}(x, y) \neq 0 \qquad \text{giacché} \qquad p(x, y) = ax + by + c$$

Il caso $b = 0$ corrisponde a rette parallele all'asse y, che non possono essere viste come grafici. Se una di tali rette fosse tangente alla curva che stiamo cercando potrebbe essere che questa non sia localmente un grafico (potrebbe anche esserlo, ma non ne abbiamo la certezza, come evidenziato nella figura che segue dove abbiamo disegnato un possibile luogo di zeri e la retta tangente).

Figura 9.a

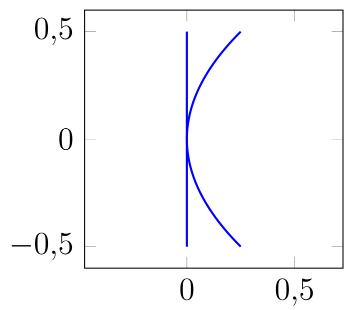
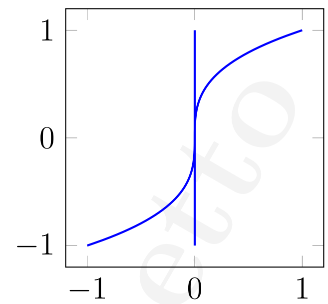


Figura 9.b



Un altro modo per vedere le cose - Si considerino le curve in Figura 9.c e in Figura 9.d e si supponga che tali curve siano una parte del luogo degli zeri di una certa funzione $f(x, y)$. Le due curve in figura non possono essere entrambe grafici di una qualche funzione g rispetto alla variabile x in un intorno del punto segnato in figura, punto che chiameremo (x_o, y_o), cioè immagini di curve cartesiane del tipo $(x, g(x))$: la seconda lo è, ma la prima no. Se f è sufficientemente regolare sappiamo che il suo gradiente è un vettore ortogonale alle curve di livello, ed in particolare nel primo caso il gradiente di f valutato nel punto (x_o, y_o), punto marcato in figura, sarà del tipo

$$(a, 0)\,,$$

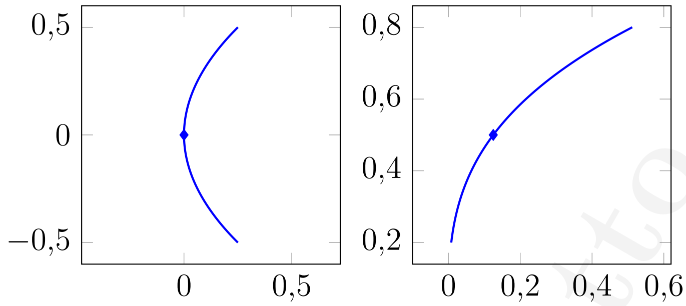
sarà cioè un vettore con seconda componente nulla. Precisamente

$$\nabla f(x_o, y_o) = \left(\frac{\partial f}{\partial x}(x_o, y_o), 0\right)\,.$$

Concludiamo che il fatto che tale curva non sia grafico rispetto alla variabile x nel punto (x_o, y_o) implica la condizione

$$\frac{\partial f}{\partial y}(x_o, y_o) = 0\,.$$

Figura 9.c Figura 9.c



Teorema 9.1 (del Dini)**.** *Siano A aperto di $\mathbf{R}^2$, $f : A \to \mathbf{R}$ una funzione continua che ammetta derivata parziale rispetto a y continua in A. Si considerino l'insieme*

$$Z = \{(x, y) \in A \,|\, f(x, y) = 0\}$$

e un punto $(x_o, y_o) \in Z$ e si supponga che

$$f_y(x_o, y_o) \neq 0\,.$$

Allora esistono un intorno I di x_o, un intorno J di y_o e una funzione $g : I \to J$ tali che

$$f(x, g(x)) = 0 \qquad \textit{per ogni } x \in I$$

e inoltre l'insieme

$$Z \cap (I \times J)$$

è il grafico della funzione g.

Dimostrazione - Supponiamo dapprima che $f_y(x_o, y_o) > 0$. Dalla continuità di f_y e dal teorema della permanenza del segno si ha che esiste un rettangolo

$$R = I \times J = [x_o - a, x_o + a] \times [y_o - b, y_o + b] \subset A$$

nel quale f_y ha lo stesso segno di $f_y(x_o, y_o)$ e inoltre soddisfa

$$m := \frac{1}{2}\, f_y(x_o, y_o) \leqslant f_y(x, y) \leqslant \frac{3}{2}\, f_y(x_o, y_o) =: M. \tag{34}$$

Si consideri ora lo spazio metrico $(X, d) = \big(C^0(I), d_\infty\big)$ dove d_∞ è la norma indotta dalla norma infinito. A questo punto si consideri il seguente operatore

$$P : (X, d) \longrightarrow (X, d)$$

definito da (si confronti con l'Esempio 5.10.5 del capitolo **Topologia in $\mathbf{R}^n$ - Cenni agli spazi metrici**)

$$Pg(x) = g(x) - \frac{1}{f_y(x_o, y_o)}\, f\big(x, g(x)\big), \qquad x \in [x_o - a, x_o + a].$$

Ora vediamo che P è una contrazione. Prima si osservi che per il teorema di Lagrange, dati due punti (x, y) e $(x, \tilde{y})$, si ha che esiste un valore ξ compreso tra y e $\tilde{y}$, che a priori dipenderà anche da x, tale che

$$f(x, y) - f(x, \tilde{y}) = f_y(x, \xi)\big(y - \tilde{y}\big)\,.$$

Analogamente, date due funzioni $g, \tilde{g} \in C^0\big([a, b]\big)$ esiste una funzione $\xi : [a, b] \to \mathbf{R}$ tale che per ogni $x \in I$ il valore $\xi(x)$ è compreso tra $g(x)$ e $\tilde{g}(x)$ e inoltre

$$\begin{aligned}
\big|Pg(x) - P\tilde{g}(x)\big| &= \left| g(x) - \tilde{g}(x) - \frac{1}{f_y(x_o, y_o)}\,\Big(f\big(x, g(x)\big) - \Big(f\big(x, \tilde{g}(x)\big)\right| = \\
&= \left| g(x) - \tilde{g}(x) - \frac{1}{f_y(x_o, y_o)}\, f_y(x, \xi(x))\big(g(x) - \tilde{g}(x)\big)\right| = \\
&= \left| 1 - \frac{1}{f_y(x_o, y_o)}\, f_y\big(x, \xi(x)\big)\right|\, |g(x) - \tilde{g}(x)|\,.
\end{aligned} \tag{35}$$

Si osservi che per la scelta di m ed M fatta in (34) si deduce che

$$f_y(x_o, y_o) = \frac{M + m}{2}.$$

Ora, poiché $-f_y(x, y) \leqslant -m$ e $f_y(x, y) \leqslant M$ per $(x, y) \in I \times J$, da (35) si ottiene che

$$\begin{aligned}
1 - \frac{2}{M + m}\, f_y\big(x, \xi(x)\big) &\leqslant \frac{M - m}{M + m}, \\
\frac{2}{M + m}\, f_y\big(x, \xi(x)\big) - 1 &\leqslant \frac{M - m}{M + m}.
\end{aligned}$$

da cui

$$\|Pg - P\tilde{g}\|_{\infty,[a,b]} \leqslant \alpha\, \|g - \tilde{g}\|_{\infty,[a,b]}, \qquad \alpha = \frac{M - m}{M + m}.$$

Dal teorema delle contrazioni si deduce che esiste un unico punto fisso $g^* \in X$.

Nel caso in cui $f_y(x_o, y_o)$ fosse negativo si può supporre che f_y sia negativa in R e

$$m := \frac{1}{2}\,|f_y(x_o, y_o)| \leqslant |f_y(x, y)| \leqslant \frac{3}{2}\,|f_y(x_o, y_o)| =: M$$

da cui $-M \leqslant f_y(x,y) \leqslant -m$ in R. A questo punto la dimostrazione è analoga al caso precedente. $\square$

Teorema 9.2. *Con A, f e g come nel Teorema* 9.1*, se $f \in C^1(A)$ allora la funzione $g \in C^1(I)$ e*

$$g'(x) = -\frac{f_x(x,g(x))}{f_y(x,g(x))} \qquad \textit{per ogni } x \in I\,. \tag{36}$$

Dimostrazione - Siano R, I, J come nella dimostrazione del teorema precedente. Fissiamo $x_1, x_2 \in I$ e chiamiamo $y_1 = g(x_1)$, $y_2 = g(x_2)$. Dal teorema del valor medio esiste (ξ_1, ξ_2) appartenente al segmento $[(x_1,y_1),(x_2,y_2)]$ tale che

$$f(x_2,y_2) - f(x_1,y_1) = \langle \nabla f(\xi_1,\xi_2), (x_2 - x_1, y_2 - y_1)\rangle$$

e poiché $f(x_j,y_j) = 0$, $j = 1,2$ si ha

$$(x_2 - x_1) f_x(\xi_1,\xi_2) + (y_2 - y_1) f_y(\xi_1,\xi_2) = 0.$$

Si osservi che $f_y(\xi_1,\xi_2) \neq 0$ perché $(\xi_1,\xi_2) \in R$. Da questa uguaglianza si ricava

$$\frac{g(x_2) - g(x_1)}{x_2 - x_1} = -\frac{f_x(\xi_1,\xi_2)}{f_y(\xi_1,\xi_2)}.$$

Se chiamiamo $\bar{x}$ il punto x_1, quindi $x_2 = \bar{x} + h$, passando al limite per $h \to 0$ si ricava che g è derivabile in $\bar{x}$ e

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

Forniamo di seguito, per chi fosse curioso, una dimostrazione che non fa uso del teorema delle contrazioni, ma che usa semplicemente la monotonia derivante dal fatto che una derivata di f non è nulla.

Teorema 9.3 (Dini)**.** *Siano A aperto di $\mathbf{R}^2$, $f : A \to \mathbf{R}$ di classe $C^1(A)$ e si considerino l'insieme*

$$Z = \{(x,y) \in A \,|\, f(x,y) = 0\}$$

e un punto $(x_o, y_o) \in Z$. Si supponga che

$$f_y(x_o, y_o) \neq 0\,.$$

Allora esistono un intorno I di x_o, un intorno J di y_o e una funzione $g : I \to J$ tali che

$$f(x, g(x)) = 0 \qquad \textit{per ogni } x \in I$$

e inoltre l'insieme

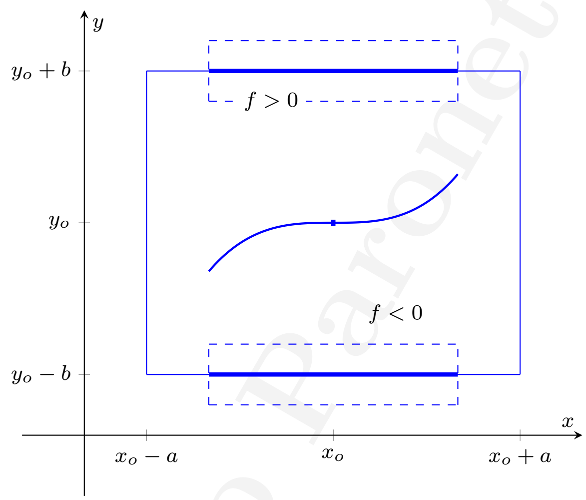
$$Z \cap (I \times J)$$

è il grafico della funzione g. Inoltre g risulta di classe C^1 in I e

$$g'(x) = -\frac{f_x(x,g(x))}{f_y(x,g(x))} \qquad \textit{per ogni } x \in I\,. \tag{37}$$

Dimostrazione - Si supponga, senza perdita di generalità, che il valore $f_y(x_o, y_o)$ sia positivo (un'analoga dimostrazione si potrà fare nel caso in cui $f_y(x_o, y_o)$ sia negativo).
1. Come prima cosa vediamo che la cosiddetta funzione implicita esiste. Per ipotesi si ha che f_y è continua. Grazie al Teorema 2.6 esiste un rettangolo (lo prendiamo chiuso per comodità) $R = I_o \times J_o = [x_o - a, x_o + a] \times [y_o - b, y_o + b]$ nel quale $f_y > 0$. Questo significa che, per ogni $x \in I_o$, la funzione definita in J_o da $y \mapsto f(x, y)$ è strettamente crescente. In particolare per $x = x_o$, poiché $f(x_o, y_o) = 0$, si avrà che

$$f(x_o, y_o - b) < 0 \qquad \text{e} \qquad f(x_o, y_o + b) > 0\,.$$



Ancora dal teorema della permanenza del segno esiste un intorno $I = (x_o - h, x_o + h)$, con $0 < h < a$, tale che

$$f(x, y_o - b) < 0 \quad \text{e} \quad f(x, y_o + b) > 0 \qquad \text{per ogni } x \in (x_o - h, x_o + h)\,.$$

Dalla stretta monotonia di $y \mapsto f(x, y)$ per ogni $x \in I$ si ha dal teorema degli zeri

$$\forall\, x \in (x_o - h, x_o + h)\ \exists!\ y \in (y_o - b, y_o + b) \text{ t.c. } f(x, y) = 0.$$

Per $x \in I$ chiamiamo $g(x)$ il valore di y che soddisfa $f(x, y) = 0$ e abbiamo definito la funzione cercata.
2. Vediamo ora la continuità e derivabilità di g. Fissiamo $x_1, x_2 \in I$ e chiamiamo $y_1 = g(x_1)$, $y_2 = g(x_2)$. Dal teorema del valor medio esiste (ξ_1, ξ_2) appartenente al segmento $[(x_1, y_1), (x_2, y_2)]$ tale che

$$f(x_2, y_2) - f(x_1, y_1) = \langle \nabla f(\xi_1, \xi_2), (x_2 - x_1, y_2 - y_1) \rangle$$

e poiché $f(x_j, y_j) = 0$, $j = 1, 2$ si ha

$$(x_2 - x_1) f_x(\xi_1, \xi_2) + (y_2 - y_1) f_y(\xi_1, \xi_2) = 0 \tag{38}$$

e quindi

$$|g(x_2) - g(x_1)| \leqslant \frac{\max_R |f_x(x, y)|}{\min_R |f_y(x, y)|}\, |x_2 - x_1|\,. \tag{39}$$

Il massimo e il minimo rispettivamente di f_x e f_y esistono dalla continuità delle funzioni e dalla compattezza di R, inoltre R può essere scelto in modo tale che il minimo di f_y su R sia positivo. Di conseguenza da (39) la funzione g risulta continua. Dalla disuguaglianza (38) si ricava

$$\frac{g(x_2) - g(x_1)}{x_2 - x_1} = -\frac{f_x(\xi_1, \xi_2)}{f_y(\xi_1, \xi_2)}.$$

Se chiamiamo $\bar{x}$ il punto x_1, quindi $x_2 = \bar{x} + h$, passando al limite per $h \to 0$ si ricava che g è derivabile e

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

Osservazione 9.4. - Ovviamente se in un punto $f(x_o, y_o) = 0$ e $f_y(x_o, y_o) = 0$, ma $f_x(x_o, y_o) \neq 0$ il teorema vale ugualmente. Esisterà una funzione h tale che $f(h(y), y) = 0$.
Il teorema in una forma più generale potrebbe essere enunciato sostituendo l'ipotesi $f_y(x_o, y_o) \neq 0$ con $|\nabla f(x_o, y_o)| \neq 0$, cioè almeno una delle due derivate è non nulla in (x_o, y_o).

Osservazione 9.5. - Il teorema, come già osservato, vale in un qualunque punto del dominio di f, anche se f in tale punto non si annulla. Se il valore di f in un certo punto (x_o, y_o) è c è sufficiente considerare la funzione

$$\tilde{f}(x, y) := f(x, y) - c.$$

Per tale funzione sono verificate le ipotesi del teorema del Dini per cui esisterà una funzione g il cui grafico sarà, localmente, l'insieme di livello 0 di $\tilde{f}$ e quindi l'insieme di livello c di f.

Osservazione 9.6. - Si osservi che, una volta dimostrata la (36), la si può ricavare semplicemente derivando l'espressione

$$f(x, g(x)) = 0.$$

Poiché $f(x, g(x))$ è costante (non importa che sia zero, si veda l'osservazione precedente) la sua derivata sarà nulla. Per cui

$$0 = \frac{d}{dx}\big(f(x, g(x))\big) = f_x(x, g(x)) + f_y(x, g(x))g'(x).$$

Poiché il luogo di zeri è una curva cartesiana $(x, g(x))$ il suo vettore tangente sarà $(1, g'(x))$. Abbiamo quindi ritrovato un fatto già visto: il gradiente è ortognale alle curve di livello, infatti

$$0 = \frac{d}{dx}\big(f(x, g(x))\big) = \langle \nabla f(x, g(x)), \gamma'(x) \rangle$$

dove $\gamma(x) = (x, g(x))$. Più precisamente, il gradiente è ortogonale alla retta individuata dal vettore $(1, g'(x))$ (che passa per l'origine) e che è parallela alla retta tangente alla curva γ, come mostrato in Figura 11.
Ricordiamo che la retta tangente passante per il punto $\gamma(x_o)$ è

$$\gamma(x_o) + \gamma'(x_o)(x - x_o).$$

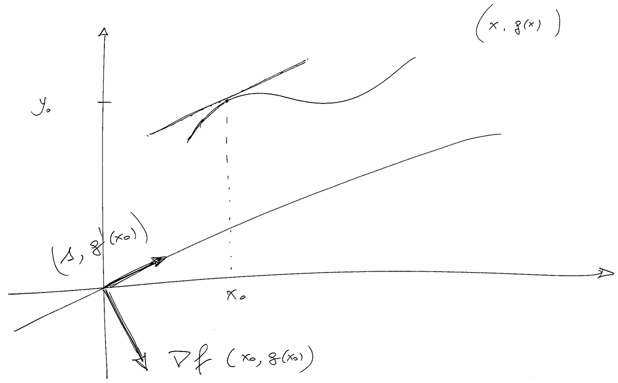


Figura 11

Nel nostro caso, con $\gamma(x) = (x, g(x))$, si avrà che la retta tangente in un generico punto $(x_o, g(x_o))$ sarà

$$(x_o, g(x_o)) + (1, g'(x))(x - x_o).$$

Osservazione 9.7. - Se $f \in C^2(A)$, conoscendo l'espressione della derivata prima di g si può ricavare anche quella della sua derivata seconda.
Infatti se nell'espressione (36) il termine a destra è derivabile lo è anche quello a sinistra. Poiché, come visto nell'osservazione precedente

$$f_x(x, g(x)) + f_y(x, g(x))g'(x) = 0,$$

la sua derivata (se esiste) sarà nulla. Per cui se $f \in C^2(A)$ si ha

$$\begin{aligned} f_{xx}(x, g(x)) + & f_{xy}(x, g(x))g'(x) + f_{xy}(x, g(x))g'(x) + \\ & + f_{yy}(x, g(x))(g'(x))^2 + f_y(x, g(x))g''(x) = 0 \end{aligned}$$

da cui si ricava $g''(x)$. Oppure derivando l'espressione (36) si ottiene (diversa, ma uguale, a quella di sopra)

$$g'' = \frac{-f_y(f_{xx} + f_{xy}g') + f_x(f_{xy} + f_{yy}g')}{f_y^2}. \tag{40}$$

Se $f \in C^k(A)$ possiamo ricavare le prime k derivate di g.

Esempio 9.8. - La funzione $f(x, y) = y^3 - x^3$ si annulla in $(0, 0)$. Le derivate parziali si annullano entrambe in $(0, 0)$. Il teorema del Dini fornisce condizioni *sufficienti*, ma non necessarie. In effetti il luogo di zeri di f è una retta (già vista) che è un grafico sia rispetto alla variabile x che alla y.

La funzione $f(x,y) = y^2 - x^3$ soddisfa $f(0,0) = f_x(0,0) = f_y(0,0) = 0$ ed è grafico rispetto alla variabile y. Infatti si ha $x(y) = \sqrt[3]{y^2}$.

Esempio 9.9. - La funzione $f(x,y) = x^2 + y^2 - 1$ ha come luogo di zeri la circonferenza di raggio 1 centrata nell'origine. Si consideri il punto $(0,1)$. Si ha

$$f(0,1) = 0 \qquad \text{e} \qquad f_y(0,1) \neq 0\,.$$

Per il teorema appena visto esiste un intorno I di 0 ed una funzione $g : I \to \mathbf{R}$ tale che $f(x,g(x)) = 0$, cioè l'insieme di livello zero localmente (cioè in un intorno del punto $(0,1)$) è il grafico di una funzione g. In questo caso, ma non sempre è immediato, possiamo anche ricavare l'espressione della funzione che sarà

$$g(x) = \sqrt{1 - x^2}, \qquad x \in (-1,1).$$

Si osservi che se si valuta $f_x(0,1)$ si ottiene zero e che la circonferenza in $(0,1)$ non è localmente grafico di una funzione della variabile y, e analogamente se si considera il punto $(1,0)$ si ha che $f_y(1,0) = 0$ e in nessun intorno di $(1,0)$ la circonferenza è grafico di una funzione della variabile x.
Possiamo anche sviluppare g, usando solamente le informazioni che ci vengono dal teorema e non l'espressione esplicita di g, e ottenere lo sviluppo, per esempio al secondo ordine, di g

$$g(x) = 1 - \frac{1}{2}x^2 + o(x^2).$$

EX - Per esercizio si considerino le funzioni $f_\alpha(x,y) = (x^2+y^2)^\alpha$, $\alpha > 0$, e si verifichi che il luogo dei punti in cui f_α vale 1 è lo stesso insieme considerato nell'esempio precedente e che nel punto $(0,1)$ la funzione implicita che descrive tale insieme in un intorno del punto $(0,1)$ ha lo stesso sviluppo di Taylor della funzione g dell'esempio precedente, qualunque sia α.
Si faccia lo stesso esercizio con la funzione $(x,y) \mapsto \log(x^2 + y^2)$.

Esempio 9.10. - Sia $f(x,y) = xe^y - y$. Il punto $(0,0)$ appartiene al luogo degli zeri di f. Si ha che $f_y(0,0) = -1$, per cui è possibile vedere localmente tale luogo di zeri come il grafico di una funzione g in modo tale che $f(x,g(x)) = 0$. Si ha che

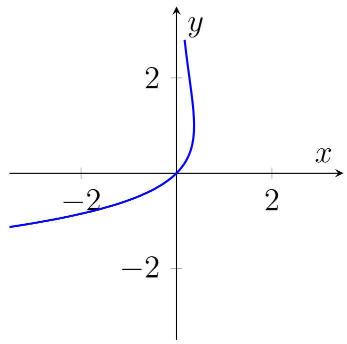
$$\begin{aligned} &f_x(x,y) = e^y, \quad f_y(x,y) = xe^y - 1, \\ &f_{xx}(x,y) = 0, \quad f_{xy}(x,y) = e^y, \quad f_{yy}(x,y) = xe^y. \end{aligned}$$

Utilizzando per g l'espressione per la derivata prima del teorema del Dini e l'espressione per la derivata seconda vista in (40) si ricava

$$g(0) = 0, \quad g'(0) = 1, \quad g''(0) = 2$$

da cui
$$g(x) = x + x^2 + o(x^2).$$
Analogamente, visto che si conosce f, per ricavare $g'(0)$ si può derivare l'espressione $x\,e^{g(x)} - g(x)$ e scrivere
$$\frac{d}{dx}\big(xe^{g(x)} - g(x)\big) = 0$$
da cui
$$g'(x) = \frac{e^{g(x)}}{1 - xe^{g(x)}}\,.$$
A questo punto per $g''(0)$ si può riderivare tale espressione (oppure l'espressione generale $-f_x(x, g(x))/f_y(x, g(x)))$ che in questo caso diventa
$$g''(x) = \frac{d}{dx}\frac{e^{g(x)}}{1 - xe^{g(x)}}\,.$$
Si osservi che anche $f_x(0, 0) \neq 0$, per cui il luogo di zeri può anche essere visto come grafico di una funzione h della variabile y e $f(h(y), y) = 0$. In questo caso si avrà
$$h'(y) = -\frac{f_y(h(y), y)}{f_x(h(y), y)}$$
per cui $h'(0) = 1$. Derivando l'espressione di sopra si trova
$$h''(y) = -2e^{-y} + ye^{-y}$$
da cui $h''(0) = -2$. Per cui lo sviluppo al secondo ordine di h è dato da
$$h(y) = y - y^2 + o(y^2).$$
Si osservi che se la funzione g e la funzione h sono una l'inversa dell'altra, per cui
$$g \circ h(y) = y \quad \text{e} \quad h \circ g(x) = x$$
sempre in un intorno, che a priori non è noto, del punto $y_o = 0$ nel primo caso e di $x_o = 0$ nel secondo.
In questo caso è facile esplicitare la funzione h, che è data da $h(y) = ye^{-y}$, ma è un caso fortunato. L'espressione di g, ad esempio, non si riesce a determinare esplicitamente. Nonostante ciò si noti come sia possibile trovarne uno sviluppo di Taylor fino all'ordine desiderato, visto che in questo casa f è infinitamente derivabile. Per completezza mostriamo il luogo degli zeri di f intorno all'origine nella figura che segue.

Figura 10



Osservazione 9.11. - Se esistono g ed h tali che $f(x, g(x)) = f(h(y), y) = 0$ in un intorno di un certo punto (x_o, y_o) allora valgono (localmente)

$$h = g^{-1} \quad \text{e} \quad g = h^{-1}.$$

Possiamo anche trovare un versore tangente ed uno normale al luogo di zeri visto che localmente è una curva (cartesiana). Se, ad esempio, il luogo degli zeri fosse in un intorno di un punto (x_o, y_o) il grafico di una funzione g e

$$f(x, g(x)) = 0$$

il luogo degli zeri è il sostegno della curva

$$I \ni x \mapsto (x, g(x)).$$

La derivata di tale curva è $(1, g'(x))$ per cui il versore tangente (nel verso di percorrenza della curva) è

$$\tau = \frac{\big(1, g'(x)\big)}{\sqrt{1 + (g'(x))^2}} = \left(\frac{1}{\sqrt{1 + (g'(x))^2}}, \frac{g'(x)}{\sqrt{1 + (g'(x))^2}} \right).$$

Volendo valutare i versori normali possiamo cercare un vettore di modulo 1 ortogonale al vettore tangente. Si hanno i due versori normali

$$\nu = \left(-\frac{g'(x)}{\sqrt{1 + (g'(x))^2}}, \frac{1}{\sqrt{1 + (g'(x))^2}} \right) \quad \text{e} \quad -\nu.$$

Usando l'espressione (36) per la derivata di g si ottiene che ν ha la seguente espressione

$$\nu = \left(\frac{f_x}{f_y\sqrt{1+f_x^2/f_y^2}}, \frac{1}{\sqrt{1+f_x^2/f_y^2}} \right) =$$

$$= \left(\frac{f_x|f_y|}{f_y\sqrt{f_x^2+f_y^2}}, \frac{|f_y|}{\sqrt{f_x^2+f_y^2}} \right) = \begin{cases} \dfrac{\nabla f(x,y)}{|\nabla f(x,y)|} & \text{se } f_y > 0, \\ -\dfrac{\nabla f(x,y)}{|\nabla f(x,y)|} & \text{se } f_y < 0. \end{cases}$$

In entrambi i casi si ha un vettore di modulo 1 ortogonale alla curva di livello (in questo caso è proprio una curva) $f(x,y) = 0$ nel punto (x_o, y_o) (e in un suo intorno).