

QUESTA PARTE E' LARGAMENTE INCOMPLETA. ALLA FINE HO RITENUTO COMUNQUE UTILE RENDERLA DISPONIBILE SOPRATTUTTO PER QUANTO RIGUARDA LA SEZIONE SUL CALCOLO DIFFERENZIALE IN PIU' VARIABILI. E' CONSIGLIABILE COMUNQUE CONFRONTARLA CON GLI APPUNTI PRESI A LEZIONE (TEMO CI SIANO PARECCHIE IMPRECISIONI).

1 Proprietà “topologiche” di \mathbb{R}^N

Ricordiamo che, se N è un numero intero, \mathbb{R}^N è lo spazio N -dimensionale i cui punti si possono rappresentare come N -uple di numeri reali: $\mathbf{x} \in \mathbb{R}^N \Leftrightarrow \mathbf{x} = (x_1, \dots, x_N)$. Se $N > 1$ useremo nel seguito la convenzione di indicare in grassetto (\mathbf{x}) i punti di \mathbb{R}^N e in carattere normale le sua coordinate x_1, \dots, x_n . Vogliamo introdurre la nozione di limite per funzioni di più variabili. Per questo premettiamo alcune definizioni riguardanti \mathbb{R}^N . Ricordiamo innanzitutto che, se $\mathbf{x} \in \mathbb{R}^N$ la sua *norma* è il numero positivo

$$\|\mathbf{x}\| := \sqrt{x_1^2 + \dots + x_N^2}$$

mentre se $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ la *distanza* tra di loro è data da

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{y} - \mathbf{x}\| = \sqrt{(y_1 - x_1)^2 + \dots + (y_N - x_N)^2}$$

Se $\mathbf{x}_0 \in \mathbb{R}^N$ e $r > 0$ chiamiamo *disco* di centro \mathbf{x}_0 e raggio r l'insieme

$$I(\mathbf{x}_0, r) := \{\mathbf{x} \in \mathbb{R}^N : d(\mathbf{x}, \mathbf{x}_0) < r\}$$

Siano ora A un sottoinsieme di \mathbb{R}^N : $A \subset \mathbb{R}^N$ e \mathbf{x}_0 un punto di \mathbb{R}^N : $\mathbf{x}_0 \in \mathbb{R}^N$. Diciamo che:

- \mathbf{x}_0 è interno ad A se esiste $r > 0$ tale che $I(\mathbf{x}_0, r) \subset A$;
- \mathbf{x}_0 è esterno ad A se esiste $r > 0$ tale che $I(\mathbf{x}_0, r) \cap A = \emptyset$;
- \mathbf{x}_0 è di frontiera per A se non è né interno né esterno ad A

Non è difficile vedere che le tre definizioni sopra si escludono mutualmente e viceversa ogni punto di \mathbb{R}^N ricade in una delle tre. Quindi, dato A , \mathbb{R}^N si spezza in tre insiemi disgiunti:

- la *parte interna* di A , denotata con $\text{int}(A)$, cioè l'insieme di tutti i punti interni ad A ;

- la *frontiera* di A , denotata con ∂A , cioè l'insieme di tutti i punti di frontiera per A ;
- la *parte esterna* di A , cioè l'insieme dei punti esterni ad A , che può risultare (come si vede facilmente) la parte interna del complementare di A , $\text{int}(\mathcal{C}A)$.

Notiamo che:

- \mathbf{x}_0 è di frontiera per A se e solo se ogni disco $I(\mathbf{x}_0, r)$ contiene almeno un punto di A (eventualmente lo stesso \mathbf{x}_0) e almeno un punto di $\mathcal{C}A$;
- dire che \mathbf{x}_0 è interno ad A è di più che $\mathbf{x}_0 \in A$.

Diciamo infine che

- A è aperto se tutti i suoi punti sono interni; questo è equivalente a dire che $A = \text{int}(A)$ o anche che A non contiene nessun punto di frontiera: $A \cap \partial A = \emptyset$;
- A è chiuso se A contiene tutta la sua frontiera: $\partial A \subset A$.

Notiamo che ci sono insiemi che non sono né aperti né chiusi. Per esempio, in \mathbb{R} , si può prendere $A =]0, 1]$ (intervallo “semiaperto”); in \mathbb{R}^N è chiaro che la situazione può essere assai più complicata.

Esempio 1.1. Consideriamo in \mathbb{R}^2 l'insieme $A := I(0, 1) = \{\mathbf{x} : \|\mathbf{x}\| < 1\}$. Allora

- A è aperto;
- $\partial A = S := \{\mathbf{x} : \|\mathbf{x}\| = 1\}$ (la sfera unitaria);
- S è chiuso.

Vediamo come si verificano le affermazioni sopra.

• Prendiamo un qualunque punto \mathbf{x} in A ; deve essere $r := \|\mathbf{x}\| < 1$. Allora se prendiamo $\epsilon = (1 - r)/2$ risulta $r + \epsilon < 1$. Sia \mathbf{y} tale che $\|\mathbf{y} - \mathbf{x}\| < \epsilon$, per le proprietà della distanza si ha

$$\|\mathbf{y}\| = \|\mathbf{x} + (\mathbf{y} - \mathbf{x})\| \leq \|\mathbf{x}\| + \|\mathbf{y} - \mathbf{x}\| = r + \|\mathbf{y} - \mathbf{x}\| < r + \epsilon < 1$$

Questo dimostra che tutto il disco $I(\mathbf{x}, \epsilon)$ è contenuto in A e cioè \mathbf{x} è interno ad A . Dato che \mathbf{x} è un generico punto di A se ne deduce che A è aperto.

· Sia ora \mathbf{x} un punto di S , cioè $\|\mathbf{x}\| = 1$, e sia $\epsilon > 0$. Prendiamo $\mathbf{y} := (1 - \epsilon/2)\mathbf{x}$: allora \mathbf{y} è in A (ci siamo avvicinati a zero) ma

$$\|\mathbf{y} - \mathbf{x}\| = \|(1 - \epsilon/2 - 1)\mathbf{x}\| = (\epsilon/2)\|\mathbf{x}\| = \epsilon/2 < \epsilon$$

Abbiamo quindi trovato un punto \mathbf{y} dentro A che dista da \mathbf{x} meno di ϵ . D'altra parte troviamo anche un punto fuori di A che dista da \mathbf{x} meno di ϵ (possiamo prendere lo stesso \mathbf{x}) e quindi ogni $I(\mathbf{x}, \epsilon)$ contiene (almeno) un punto di A e (almeno) un punto di $\mathcal{C}A$; quindi \mathbf{x} è di frontiera. In questo modo si vede che $S \subset \partial A$.

· Prendiamo infine \mathbf{x} con $r := \|\mathbf{x}\| > 1$. Imitando i discorsi del primo punto si vede facilmente che, scegliendo $\epsilon := (1 - r)/2$ tutto il disco $I(\mathbf{x}, \epsilon)$ è fatto di punti fuori di A . Quindi i punti che distano più di 1 da zero sono esterni ad A .

Tutto questo implica che la frontiera di A è fatta esattamente dai punti che distano 1 da zero, cioè $\partial A = S$. Con gli stessi argomenti si vede anche che $\partial S = S$ (che è automaticamente contenuto in S) e quindi S è chiuso. Per la verità si potrebbe dimostrare che la frontiera di un qualunque insieme è sempre un chiuso.

Prendiamo ora

$$A_1 := \left\{ \mathbf{x} = (x_1, x_2) : -1 \leq x_1 \leq 1, -\sqrt{1 - x_1^2} < x_2 \leq \sqrt{1 - x_1^2} \right\}$$

Si vede che $\text{int}(A_1) = A$, $\partial A_1 = S$ ma A_1 non coincide con A e non contiene tutto S (A_1 è "intermedio" tra A e $A \cup S$), dunque A_1 non è né aperto né chiuso.

Introduciamo ancora alcune definizioni.

Dato A sottoinsieme di \mathbb{R}^N chiamiamo *chiusura di A* l'insieme

$$\bar{A} := A \cup \partial A;$$

è chiaro che un insieme è chiuso se e solo se coincide con la sua chiusura.

Siano infine $A \subset \mathbb{R}^N$ e $\mathbf{x}_0 \in \mathbb{R}^N$. Diciamo che \mathbf{x}_0 è *di accumulazione* per A se per ogni $\epsilon > 0$ esiste un punto \mathbf{y} tale che

$$\mathbf{y} \neq \mathbf{x}_0 \quad , \quad \mathbf{y} \in A \quad , \quad \mathbf{y} \in I(\mathbf{x}_0, \epsilon)$$

In altre parole \mathbf{x}_0 è di accumulazione per A se e solo se si possono trovare punti di A diversi da \mathbf{x}_0 , vicini quanto si vuole a \mathbf{x}_0 .

2 Limiti di funzioni di più variabili

Consideriamo un sottoinsieme A di \mathbb{R}^N e una funzione $\mathbf{f} : A \rightarrow \mathbb{R}^M$.

Supponiamo inoltre che \mathbf{x}_0 in \mathbb{R}^N sia un punto di accumulazione per A .

Definizione 2.1 (di limite). *Un punto \mathbf{l} di \mathbb{R}^M si dice limite di \mathbf{f} per \mathbf{x} che tende a \mathbf{x}_0 , e si scrive*

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{l}$$

se per ogni $\epsilon > 0$ esiste $\rho > 0$ tale che

$$\mathbf{x} \neq \mathbf{x}_0, \mathbf{x} \in A, \|\mathbf{x} - \mathbf{x}_0\|_N < \rho \implies \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)\|_M < \epsilon$$

(si sono usati gli indici N ed M per indicare le norme relative a \mathbb{R}^N e \mathbb{R}^M). Spesso, quando il punto \mathbf{x}_0 è chiaro dal contesto useremo la scrittura più concisa $\mathbf{f}(x) \rightarrow \mathbf{l}$.

Elenchiamo ora le principali proprietà del limite.

Proposizione 2.2 (unicità). *Se il limite esiste, allora è unico.*

In altre parole se, per \mathbf{x} tendente a \mathbf{x}_0 , si ha $\mathbf{f}(\mathbf{x}) \rightarrow \mathbf{l}_1$ e $\mathbf{f}(\mathbf{x}) \rightarrow \mathbf{l}_2$, deve essere $\mathbf{l}_1 = \mathbf{l}_2$.

Proposizione 2.3 (linearità). *Se per $\mathbf{x} \rightarrow \mathbf{x}_0$ si ha $\mathbf{f}_1(\mathbf{x}) \rightarrow \mathbf{l}_1$ e $\mathbf{f}_2(\mathbf{x}) \rightarrow \mathbf{l}_2$, e se c_1, c_2 sono numeri reali, allora*

$$c_1 \mathbf{f}_1(\mathbf{x}) + c_2 \mathbf{f}_2(\mathbf{x}) \rightarrow c_1 \mathbf{l}_1 + c_2 \mathbf{l}_2$$

Proposizione 2.4 (prodotto). *Siano $\mathbf{f} : A \rightarrow \mathbb{R}^M$, $g : A \rightarrow \mathbb{R}$ e sia \mathbf{x}_0 un punto di accumulazione per A . Allora*

$$\mathbf{f}(\mathbf{x}) \rightarrow \mathbf{l}_1, g(\mathbf{x}) \rightarrow l_2 \implies g(\mathbf{x})\mathbf{f}(\mathbf{x}) \rightarrow l_2 \mathbf{l}_1$$

Proposizione 2.5 (reciproco). *Sia $f : A \rightarrow \mathbb{R}$ e sia \mathbf{x}_0 un punto di accumulazione per A . Allora*

$$f(\mathbf{x}) \rightarrow l \neq 0 \implies \frac{1}{f(\mathbf{x})} \rightarrow \frac{1}{l}$$

Proposizione 2.6 (composizione/cambio di variabile nei limiti). *Siano $A \subset \mathbb{R}^N$, $B \subset \mathbb{R}^M$, $\mathbf{f} : A \rightarrow B$, $\mathbf{g} : B \rightarrow \mathbb{R}^P$. Sia \mathbf{x}_0 un punto di accumulazione per A , \mathbf{y}_0 un punto di accumulazione per B . Supponiamo che*

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{y}_0 \quad , \quad \lim_{\mathbf{y} \rightarrow \mathbf{y}_0} \mathbf{g}(\mathbf{y}) = \mathbf{l}$$

per \mathbf{l} in \mathbb{R}^P . Supponiamo ancora che

$$\mathbf{f}(\mathbf{x}) \neq \mathbf{y}_0 \quad \text{se } \mathbf{x} \in A, \mathbf{x} \text{ è vicino a } \mathbf{x}_0.$$

Allora risulta

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{l}.$$

Notiamo ora che dire $\mathbf{f} : A \rightarrow \mathbb{R}^M$ significa che per ogni \mathbf{x} di A è definito $\mathbf{f}(\mathbf{x})$ in \mathbb{R}^M . Ma allora $\mathbf{f}(\mathbf{x})$ è una m -pla di numeri (y_1, \dots, y_M) . Possiamo allora definire $f_1(\mathbf{x}) := y_1, \dots, f_M(\mathbf{x}) := y_M$; in questo modo abbiamo definito M funzioni reali $f_1 : A \rightarrow \mathbb{R}, \dots, f_M : A \rightarrow \mathbb{R}$ che vengono dette *le componenti* di \mathbf{f} :

$$\mathbf{f}(x_1, \dots, x_N) = (f_1(x_1, \dots, x_N), \dots, f_M(x_1, \dots, x_N))$$

Proposizione 2.7. Se $\mathbf{l} = (l_1, \dots, l_M)$ si ha (per $\mathbf{x} \rightarrow \mathbf{x}_0$)

$$\mathbf{f}(\mathbf{x}) \rightarrow \mathbf{l} \Leftrightarrow f_1(\mathbf{x}) \rightarrow l_1, \dots, f_M(\mathbf{x}) \rightarrow l_M$$

L'ultima proposizione dice che calcolare il limite di una funzione vettoriale è come calcolare gli M limiti delle sue componenti. Quindi *in arrivo* si potrebbe sempre considerare di avere una sola variabile ($f : A \rightarrow \mathbb{R}$); ben diverso è il ruolo delle variabili *in partenza*.

Proposizione 2.8 (limiti di funzioni e limiti di successioni).

Siano $A \subset \mathbb{R}^N$, \mathbf{x}_0 un punto di accumulazione per A e \mathbf{l} in \mathbb{R}^M . Allora sono equivalenti i due fatti che seguono.

1. $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{l}$;
2. per ogni successione $(\mathbf{x}_n)_n$ di punti di A tale che $\mathbf{x}_n \neq \mathbf{x}_0 \forall n, \mathbf{x}_n \rightarrow \mathbf{x}_0$ si ha che $\mathbf{f}(\mathbf{x}_n) \rightarrow \mathbf{l}$.

Definizione 2.9 (limiti infiniti). Nei casi in cui c'è una sola variabile, in partenza o in arrivo, si possono dare le definizioni di limiti infiniti.

- Se $f : A \rightarrow \mathbb{R}$ e \mathbf{x}_0 è di accumulazione per A possiamo dire che il limite di $f(\mathbf{x})$ per \mathbf{x} tendente a \mathbf{x}_0 è più infinito (meno infinito), e scrivere

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = +\infty \quad (-\infty)$$

(o più brevemente $f(\mathbf{x}) \rightarrow +\infty / -\infty$) se per ogni M in \mathbb{R} esiste $\rho > 0$ tale che

$$\mathbf{x} \in A, 0 < \|\mathbf{x} - \mathbf{x}_0\| < \rho \Rightarrow f(\mathbf{x}) > M \quad (< M).$$

A volte si usa dire che f tende all'infinito ($f(\mathbf{x}) \rightarrow \infty$) se si ha che $|f(\mathbf{x})| \rightarrow +\infty$.

- Se $A \subset \mathbb{R}$, se A non è superiormente (inferiormente) limitato e se $\mathbf{f} : A \rightarrow \mathbb{R}^M$, allora possiamo dire che il limite di $\mathbf{f}(x)$ per x tendente a $+\infty$ ($-\infty$) è \mathbf{l} ($\in \mathbb{R}^M$) se per ogni $\epsilon > 0$ esiste c in \mathbb{R} tale che

$$x \in A, x > c \ (x < c) \Rightarrow \|\mathbf{f}(x) - \mathbf{l}\| < \epsilon$$

In questo caso scriviamo

$$\lim_{x \rightarrow +\infty / -\infty} \mathbf{f}(x) = \mathbf{l}$$

- Se $A \subset \mathbb{R}$, se A non è superiormente (inferiormente) limitato e se $f : A \rightarrow \mathbb{R}$, allora possiamo combinare le due definizioni sopra e dire che f tende a $+\infty / -\infty$ per x tendente a $+\infty$ ($-\infty$) se per ogni M in \mathbb{R} esiste c in \mathbb{R} tale che

$$x \in A, x > c \ (x < c) \Rightarrow f(x) > M / f(x) < M$$

- Se $\mathbf{f} : A \rightarrow \mathbb{R}^M$ e se A non è limitato, possiamo dire che il limite di $\mathbf{f}(\mathbf{x})$ per $\|\mathbf{x}\|$ tendente all'infinito è \mathbf{l} ($\in \mathbb{R}^M$) se per ogni $\epsilon > 0$ esiste $R > 0$ tale che

$$\mathbf{x} \in A, \|\mathbf{x}\| \geq R \Rightarrow \|\mathbf{f}(\mathbf{x}) - \mathbf{l}\| < \epsilon$$

In questo caso scriviamo

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} \mathbf{f}(\mathbf{x}) = \mathbf{l}$$

- Se $\mathbf{f} : A \rightarrow \mathbb{R}$ e se A non è limitato, possiamo dire che il limite di $\mathbf{f}(\mathbf{x})$ per $\|\mathbf{x}\|$ tendente all'infinito è più infinito (meno infinito) se per ogni $M \in \mathbb{R}$ esiste $R > 0$ tale che

$$\mathbf{x} \in A, \|\mathbf{x}\| \geq R \Rightarrow f(\mathbf{x}) > M \ (f(\mathbf{x}) < M)$$

In questo caso scriviamo

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} f(\mathbf{x}) = +\infty \ (-\infty)$$

Proposizione 2.10 (algebra dei limiti infiniti). Si ????????????

CONTINUITA' E PROPRIETA' DELLE FUNZIONI CONTINUE ???????
TEOREMA DI WEIERSTRASS ??????????

3 Derivate parziali, differenziale e gradiente

Vogliamo ora estendere le nozioni del calcolo differenziale per una funzione \mathbf{f} da \mathbb{R}^N in \mathbb{R}^M . Come prima cosa notiamo che, se $N = 1$ possiamo dare facilmente la nozione di derivata per \mathbf{f}

Definizione 3.1. *Sia I un intervallo di \mathbb{R} e sia $\mathbf{f} : I \rightarrow \mathbb{R}^M$. Sia x_0 un punto di I . Diciamo che \mathbf{f} è derivabile in x_0 se esiste finito il limite*

$$\lim_{x \rightarrow x_0} \frac{\mathbf{f}(x) - \mathbf{f}(x_0)}{x - x_0}$$

e in tal caso chiamiamo derivata di \mathbf{f} in x_0 tale limite, che indichiamo con $\mathbf{f}'(x_0)$. Se indichiamo con f_1, \dots, f_M le componenti di \mathbf{f} non è difficile vedere che

$$\mathbf{f} \text{ derivabile in } x_0 \Leftrightarrow f_1, \dots, f_M \text{ sono derivabili in } x_0$$

e $\mathbf{f}'(x_0) = (f_1'(x_0), \dots, f_M'(x_0))$.

Quindi la derivabilità di una funzione vettoriale di UNA variabile reale si riconduce facilmente alla derivabilità di M funzioni reali di una variabile reale

Osservazione 3.2. *Se $\mathbf{f} : [a, b] \rightarrow \mathbb{R}^M$, \mathbf{f} rappresenta una curva nello spazio \mathbb{R}^M . La sua derivata $\mathbf{f}'(x_0)$, se c'è, rappresenta il vettore tangente alla curva nel punto $\mathbf{f}(x_0)$. Se si interpreta x come un tempo allora \mathbf{f} si può interpretare come la traiettoria di un punto (che al tempo x sta nella posizione $\mathbf{f}(x)$) e in questo caso $\mathbf{f}'(x_0)$ rappresenta la velocità del punto all'istante x_0 .*

Ben più ricco si presenta il caso generale con N più grande di uno. Supponiamo d'ora in poi che A sia un aperto di \mathbb{R}^N , che $\mathbf{f} : A \rightarrow \mathbb{R}^M$ e che $x_0 \in A$.

Definizione 3.3 (derivate direzionali e derivate parziali). *Sia \mathbf{y} in \mathbb{R}^N . Chiamiamo derivata direzionale di \mathbf{f} nella direzione di \mathbf{y} (o lungo \mathbf{y}) il limite*

$$\mathbf{f}'(\mathbf{x})(\mathbf{y}) := \lim_{t \rightarrow 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{y}) - \mathbf{f}(\mathbf{x})}{t}$$

tale limite (che ha senso perché t è reale, e dunque si può dividere per t) rappresenta in sostanza la derivata della funzione $t \mapsto \mathbf{f}(\mathbf{x} + t\mathbf{y})$ nel punto $t = 0$, cioè della restrizione di \mathbf{f} sulla retta passante per \mathbf{x}_0 con direzione data da \mathbf{y} .

Un caso particolare è quello in cui \mathbf{y} è uno dei versori della base di \mathbb{R}^N , cioè

$$\mathbf{y} = \mathbf{e}_i := (0, \dots, 1, \dots, 0) \quad 1 \text{ al posto } i\text{-esimo}$$

In questo caso si vede che, scrivendo $\mathbf{x}_0 = (x_1^0, \dots, x_i^0, \dots, x_N^0)$, si ha

$$\mathbf{f}'(x_0)(\mathbf{e}_i) = \lim_{h \rightarrow 0} \frac{\mathbf{f}(x_1^0, \dots, x_i^0 + h, \dots, x_N^0)}{h}$$

cioè $\mathbf{f}'(x_0)(\mathbf{e}_i)$ è la derivata della funzione di una variabile $x \mapsto \mathbf{f}(x_1^0, \dots, x, \dots, x_N^0)$, ottenuta “congelando” le componenti di \mathbf{x} diverse dalla i -esima, nel punto $x = x_i^0$. Tale derivata si chiama derivata parziale i -esima di \mathbf{f} nel punto \mathbf{x}_0 e si indica con

$$\frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0) = \frac{\partial \mathbf{f}}{\partial x_i}(x_1^0, \dots, x_N^0) \quad \text{o anche } (D_i \mathbf{f})(\mathbf{x}_0) = (D_i \mathbf{f})(x_1^0, \dots, x_N^0)$$

Naturalmente le derivate direzionali e in particolare le derivate parziali sono dei vettori di \mathbb{R}^M , È chiaro d'altronde che

$$\frac{\partial \mathbf{f}}{\partial x_i}(x_0) = \left(\frac{\partial f_1}{\partial x_i}(x_0), \dots, \frac{\partial f_M}{\partial x_i}(x_0) \right)$$

Definizione 3.4 (matrice Jacobiana e gradiente). Su pponiamo che \mathbf{f} abbia derivate parziali in $\mathbf{x} : 0$. Chiamiamo Jacobiano di \mathbf{f} in \mathbf{x}_0 la matrice $M \times N$

$$J_{\mathbf{f}}(x_0) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \dots & \frac{\partial f_1}{\partial x_N}(x_0) \\ \vdots & & \vdots \\ \frac{\partial f_M}{\partial x_1}(x_0) & \dots & \frac{\partial f_M}{\partial x_N}(x_0) \end{pmatrix}$$

Nel caso particolare in cui $M = 1$ (funzione scalare) lo Jacobiano ha una sola riga di N elementi. Chiamiamo gradiente di f nel punto x_0 il trasposto dello Jacobiano:

$$\nabla f(x_0) := \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \vdots \\ \frac{\partial f}{\partial x_N}(x_0) \end{pmatrix}$$

Osservazione 3.5. Dobbiamo pensare al gradiente come al vettore avente come componenti le derivate parziali di f nel punto x_0 . Il motivo per cui si prende il trasposto è che, nelle convenzioni dell'algebra lineare, i vettori dello spazio sono rappresentati come colonne, in modo da far tornare le regole dei prodotti matriciali (per esempio se \mathbf{v} e \mathbf{w} sono vettori allora il prodotto scalare $\mathbf{v} \cdot \mathbf{w}$ si rappresenta come il prodotto matriciale tra \mathbf{v}^t e \mathbf{w}).

Avendo introdotto le definizioni precedenti, che sembrano naturali, ci si trova subito di fronte ad un fatto imprevisto: può succedere che una funzione abbia tutte le derivate parziali in un punto *senza essere neppure continua*.

Esempio 3.6. Sia $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definita da

$$f(x_1, x_2) := \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2} & \text{se } (x_1, x_2) \neq (0, 0) \\ 0 & \text{se } (x_1, x_2) = (0, 0) \end{cases}$$

Allora si vede subito che

$$f(x_1, 0) = 0 \quad , \quad f(0, x_2) = 0 \quad \forall x_1, x_2$$

e dunque $\frac{\partial f}{\partial x_1}(0, 0) = \frac{\partial f}{\partial x_2}(0, 0) = 0$. Se poi si prende un qualunque $\mathbf{y} = (y_1, y_2)$ si vede che

$$f(\mathbf{0} + t\mathbf{y}) = f(ty_1, ty_2) = \frac{t^2(y_1 y_2)}{t^2(y_1^2 + y_2^2)} = \frac{y_1 y_2}{y_1^2 + y_2^2}$$

e quindi, essendo $t \mapsto f(\mathbf{0} + t\mathbf{y})$ costante in t , risulta $f'(\mathbf{0})(\mathbf{y}) = 0$ per ogni \mathbf{y} . Ciò nonostante la funzione *NON* è continua in zero perché da quanto visto sopra segue

$$\lim_{t \rightarrow 0} f(\mathbf{0} + t\mathbf{y}) = \frac{y_1 y_2}{y_1^2 + y_2^2}$$

e dunque avvicinandosi a zero su rette diverse f tende a limiti diversi (e dunque *NON* ha limite in zero).

L'esempio precedente si giustifica pensando che la continuità di f richiede che $f(\mathbf{x})$ sia vicino a $f(\mathbf{x}_0)$ quando \mathbf{x} è vicino a \mathbf{x}_0 *in qualunque modo* \mathbf{x} si avvicini a \mathbf{x}_0 . Al contrario conoscere le derivate parziali o le derivate direzionali dice qualcosa sul comportamento di $f(\mathbf{x})$ quando \mathbf{x} si avvicina a \mathbf{x}_0 *sulle rette* e questo non basta.

per superare il problema precedente serve una definizione più forte.

Definizione 3.7 (differenziabilità). Diciamo che \mathbf{f} è differenziabile in \mathbf{x}_0 se esiste una applicazione lineare $L : \mathbb{R}^N \rightarrow \mathbb{R}^M$ tale che

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

La definizione sopra esprime il fatto che

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + L(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|)$$

(estendendo opportunamente la nozione di o piccolo), cioè che, per \mathbf{x} vicino a \mathbf{x}_0 la funzione “somiglia” alla funzione affine $\mathbf{x} \mapsto \mathbf{f}(\mathbf{x}_0) + L(\mathbf{x} - \mathbf{x}_0)$, il cui grafico rappresenta un “piano” passante per $(\mathbf{x}_0, \mathbf{f}(\mathbf{x}_0))$ (il termine piano è corretto se $M = 1$ se no è uno spazio lineare di dimensione N dentro \mathbb{R}^{N+M}). Quindi dire che \mathbf{f} è differenziabile “significa” dire che il grafico di \mathbf{f} ammette piano tangente in $(\mathbf{x}_0, \mathbf{f}(\mathbf{x}_0))$, dato dall’equazione affine sopra indicata.

Proposizione 3.8. *Supponiamo che \mathbf{f} sia differenziabile in \mathbf{x}_0 e sia L come nella definizione sopra. Allora*

1. \mathbf{f} è continua in \mathbf{x}_0 ;
2. per ogni \mathbf{y} in \mathbb{R}^N esiste la derivata direzionale di \mathbf{f} lungo \mathbf{y} e si ha

$$\mathbf{f}'(\mathbf{x}_0)(\mathbf{y}) = L(\mathbf{y});$$

3. come conseguenza del punto precedente L è unico con la proprietà suddetta ed è individuato da

$$L(\mathbf{y}) = J_{\mathbf{f}}(\mathbf{x}_0)\mathbf{y} \quad \forall \mathbf{y} \in \mathbb{R}^N$$

dove nella formula sopra, a destra dell’eguale, si intende il prodotto matriciale della matrice Jacobiana $J_{\mathbf{f}}(\mathbf{x}_0)$ per il vettore \mathbf{y} (che va messo in colonna!). Nel caso di una funzione scalare $f : A \rightarrow \mathbb{R}$, allora $L : \mathbb{R}^N \rightarrow \mathbb{R}$ deve verificare

$$L(\mathbf{y}) = \nabla f(\mathbf{x}_0) \cdot \mathbf{y} \quad \forall \mathbf{y} \in \mathbb{R}^N$$

denotando con \cdot il prodotto scalare tra i vettori di \mathbb{R}^N .

Formalizziamo anche la nozione di piano tangente - consideriamo solo il caso scalare, che è più semplice da visualizzare.

Definizione 3.9 (iperpiano tangente). *Se $f : A \rightarrow \mathbb{R}$ è differenziabile in \mathbf{x}_0 , allora il grafico dell’equazione*

$$z = f(\mathbf{x}_0) + J_{\mathbf{f}}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$$

si definisce iperpiano tangente (piano se $N = 2$) al grafico di f nel punto $(\mathbf{x}_0, f(\mathbf{x}_0)) = (x_1^0, \dots, x_N^0, f(\mathbf{x}_0))$ (tale iperpiano “vive” in \mathbb{R}^{N+1}), così come la retta tangente vive in \mathbb{R}^2).

Il teorema che segue mostra che, “nei casi buoni” trovare le derivate parziali è sufficiente per le buone proprietà di \mathbf{f} .

Teorema 3.10 (del differenziale totale). *Supponiamo che $\mathbf{f} : A \rightarrow \mathbb{R}^M$ abbia derivate parziali rispetto a x_i , per ogni $i = 1, \dots, N$ e in ogni \mathbf{x} di un opportuno $I(\mathbf{x}_0, r)$ con $r > 0$ e supponiamo che tali derivate parziali siano continue in \mathbf{x}_0 . Allora \mathbf{f} è differenziabile in \mathbf{x}_0 .*

Si noti che nel controesempio fatto in precedenza il problema che si verifica è proprio che le derivate parziali, pur esistendo per ogni (x_1, x_2) , non sono continue in $(0, 0)$.

Enunciamo ora vari teoremi sulle proprietà della differenziabilità e delle derivate parziali. In questi enunciati la parte veramente “nuova” è la differenziabilità, in quanto le formule si potrebbero ricavare dalle formule di derivazione per funzioni di una variabile (eccetto il caso della funzione composta).

Teorema 3.11 (linearità). *Se $\mathbf{f}_1, \mathbf{f}_2 : A \rightarrow \mathbb{R}^M$ sono differenziabili in \mathbf{x}_0 e se $c_1, c_2 \in \mathbb{R}$, allora $c_1\mathbf{f}_1 + c_2\mathbf{f}_2$ è differenziabile in \mathbf{x}_0 e si ha, per ogni $i = 1, \dots, N$ e $j = 1, \dots, M$*

$$\frac{\partial(c_1\mathbf{f}_1 + c_2\mathbf{f}_2)_j(\mathbf{x}_0)}{\partial x_i} = c_1 \frac{\partial(\mathbf{f}_1)_j(\mathbf{x}_0)}{\partial x_i} + c_2 \frac{\partial(\mathbf{f}_2)_j(\mathbf{x}_0)}{\partial x_i}$$

o intermini di matrici Jacobiane

$$J_{c_1\mathbf{f}_1 + c_2\mathbf{f}_2}(\mathbf{x}_0) = c_1 J_{\mathbf{f}_1}(\mathbf{x}_0) + c_2 J_{\mathbf{f}_2}(\mathbf{x}_0).$$

Teorema 3.12 (prodotto). *Se $\mathbf{f} : A \rightarrow \mathbb{R}^M$ e $g : A \rightarrow \mathbb{R}$ sono differenziabili in \mathbf{x}_0 , allora $g\mathbf{f} : A \rightarrow \mathbb{R}^M$ è differenziabile in \mathbf{x}_0 e si ha, per ogni $i = 1, \dots, N$ e $j = 1, \dots, M$*

$$\frac{\partial(g\mathbf{f})_j(\mathbf{x}_0)}{\partial x_i} = g(\mathbf{x}_0) \frac{\partial \mathbf{f}_j(\mathbf{x}_0)}{\partial x_i} + \mathbf{f}_j(\mathbf{x}_0) \frac{\partial g(\mathbf{x}_0)}{\partial x_i}$$

o intermini di matrici Jacobiane

$$J_{g\mathbf{f}}(\mathbf{x}_0) = g(\mathbf{x}_0) J_{\mathbf{f}}(\mathbf{x}_0) + \nabla g(\mathbf{x}_0) \otimes \mathbf{f}(\mathbf{x}_0).$$

(se si introduce il simbolo \otimes per il quale dati \mathbf{v} in \mathbb{R}^N di coordinate v_1, \dots, v_N e \mathbf{w} in \mathbb{R}^M di coordinate w_1, \dots, w_M , allora $\mathbf{v} \otimes \mathbf{w}$ è la matrice M per N di componenti $w_j v_i$????)

Teorema 3.13 (prodotto scalare). Se $\mathbf{f}_1, \mathbf{f}_2 : A \rightarrow \mathbb{R}^M$ sono differenziabili in \mathbf{x}_0 allora $\mathbf{f}_1 \cdot \mathbf{f}_2$ è differenziabile in \mathbf{x}_0 e

$$\frac{\partial(\mathbf{f}_1 \cdot \mathbf{f}_2)}{\partial x_i}(\mathbf{x}_0) = \sum_{j=1}^M \left(\frac{\partial f_{1,j}}{\partial x_i}(\mathbf{x}_0) f_{2,j}(\mathbf{x}_0) + f_{1,j}(\mathbf{x}_0) \frac{\partial f_{2,j}}{\partial x_i}(\mathbf{x}_0) \right)$$

che in termini di matrici Jacobiane e gradienti si può scrivere

$$\nabla(\mathbf{f}_1 \cdot \mathbf{f}_2)(\mathbf{x}_0) = J_{\mathbf{f}_1}(\mathbf{x}_0)\mathbf{f}_2(\mathbf{x}_0) + J_{\mathbf{f}_2}(\mathbf{x}_0)\mathbf{f}_1(\mathbf{x}_0)$$

Il seguente teorema, che non si riuscirebbe a dedurre con argomenti di una variabile, mostra come sia “giusta” l’impostazione introdotta sopra.

Teorema 3.14 (composizione). Siano $A \subset \mathbb{R}^N$, $B \subset \mathbb{R}^M$ aperti; siano $\mathbf{f} : A \rightarrow B$ e $\mathbf{g} : B \rightarrow \mathbb{R}^P$; siano infine \mathbf{x}_0 in A e \mathbf{y}_0 in B .

Se \mathbf{f} è differenziabile in \mathbf{x}_0 e \mathbf{g} è differenziabile in \mathbf{y}_0 allora la funzione composta $\mathbf{g} \circ \mathbf{f} : A \rightarrow \mathbb{R}^P$ è differenziabile in \mathbf{x}_0 . In questo caso si ha

$$J_{\mathbf{g} \circ \mathbf{f}}(\mathbf{x}_0) = J_{\mathbf{g}}(\mathbf{y}_0)J_{\mathbf{f}}(\mathbf{x}_0)$$

(prodotto di matrici) che in termini di derivate parziali significa

$$\frac{\partial(\mathbf{g} \circ \mathbf{f})_k}{\partial x_i} = \sum_{j=1}^M \frac{\partial g_k(\mathbf{y}_0)}{\partial y_j} \frac{\partial f_j(\mathbf{x}_0)}{\partial x_i} \quad i = 1, \dots, N \quad k = 1, \dots, P$$

Osservazione 3.15. Supponiamo che $f : A \rightarrow \mathbb{R}$ sia una funzione (scalare) differenziabile in A e che $\gamma : [0, T] \rightarrow A$ sia una curva definita in un intervallo $[0, T]$, derivabile in $[0, T]$. Allora $f \circ \gamma$ è una funzione reale definita su $[0, T]$ che ha come derivata (per la formula sulla composizione)

$$(f \circ \gamma)'(t) = \nabla f(\gamma(t)) \cdot \gamma'(t)$$

Supponiamo che γ descriva una linea di livello, cioè che $f(\gamma(t))$ sia costante; allora il termine di sinistra nella formula sopra è eguale a zero e cioè $\nabla f(\gamma(t))$ è ortogonale a $\gamma'(t)$. Dato che $\gamma'(t)$ è la direzione tangente alla curva, questo fatto si interpreta dicendo che il gradiente è in un punto $\mathbf{x}_0 (= \gamma(t))$ è ortogonale alla linea di livello passante per \mathbf{x}_0 (tutto questo è corretto se il gradiente è diverso da zero).

Facciamo un'altra considerazione. Sia \mathbf{y} un qualsiasi vettore di \mathbb{R}^N con $\|\mathbf{y}\| = 1$; allora la derivata di f nella direzione di \mathbf{y} è pari a

$$f'(\mathbf{x}_0)(\mathbf{y}) = \nabla f(\mathbf{x}_0) \cdot \mathbf{y}.$$

Non è difficile vedere che tale espressione ha valore massimo esattamente quando $\mathbf{y} = \nabla f(x_0)/\|\nabla f(\mathbf{x}_0)\|$ e il valore massimo è $\|\nabla f(\mathbf{x}_0)\|$; quindi il gradiente è un vettore diretto verso la “direzione di massima crescita” di f (a partire da \mathbf{x}_0), mentre il modulo del gradiente è la derivata di f lungo questa direzione.

Definizione 3.16 (massimi e minimi relativi). Sia A un sottoinsieme di \mathbb{R}^N e sia $f : A \rightarrow \mathbb{R}$ una funzione. Diciamo che un punto \mathbf{x}_0 di A è di massimo (minimo) relativo per f se esiste un disco $I(\mathbf{x}_0, r)$ con $r > 0$ tale che

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) \quad (f(\mathbf{x}) \leq f(\mathbf{x}_0)) \quad \text{per ogni } x \text{ in } I(\mathbf{x}_0, r) \cap A$$

Teorema 3.17 (massimi o minimi relativi e punti stazionari). Sia A un sottoinsieme di \mathbb{R}^N , $f : A \rightarrow \mathbb{R}$ una funzione e \mathbf{x}_0 un punto di A . Supponiamo che

- \mathbf{x}_0 sia punto di massimo (minimo) relativo per f ;
- \mathbf{x}_0 sia interno ad A ;
- f sia differenziabile in \mathbf{x}_0 .

Allora $\nabla f(x_0) = 0$; esprimeremo di solito quest’ultima proprietà dicendo che \mathbf{x}_0 è un punto stazionario (o critico) per f .

Osservazione 3.18. Come nel caso di una variabile non vale l’implicazione opposta, stazionario non è equivalente a massimo o minimo relativo. Nel caso di una variabile il controesempio era dato dai punti di flesso (come nel caso di $f(x) = x^3$, in cui 0 è stazionario ma f è strettamente crescente). Nel caso di funzioni di più variabili la situazione è più ricca: consideriamo la funzione

$$f(x_1, x_2) = x_1^2 - x_2^2$$

definita su tutto \mathbb{R}^2 . Si vede facilmente che

$$\nabla f(x_1, x_2) = (2x_1, -2x_2)$$

e quindi $(0, 0)$ è (l’unico) punto stazionario. Si vede anche, però, che

$$x \mapsto f(x, 0) \text{ ha minimo in } x = 0 \quad y \mapsto f(0, y) \text{ ha massimo in } y = 0$$

(massimi e minimi stretti) e quindi $(0, 0)$ non è né di massimo, né di minimo (e non possiamo neanche, ragionevolmente, chiamarlo un flesso). Un punto con tali caratteristiche si usa chiamare punto di sella per f .

Il teorema appena citato estende in maniera naturale quello già visto nel caso di funzioni di una variabile dicendo in sostanza che i massimi e i minimi relativi per f vanno cercati tra:

- i punti stazionari,
- i punti in cui f non è differenziabile,
- i punti di frontiera del dominio A .

Nel caso di una variabile il terzo punto era piuttosto facile in quanto tipicamente A era un intervallo e i suoi punti di frontiera gli estremi, che sono due punti. Nel caso di più variabili la ricerca dei punti di massimo o minimo relativo alla frontiera è ben più difficile. Ci sono dei teoremi che si occupano di studiare questo tipo di problemi in generale, ma noi vediamo come si può trattare un caso particolare.

Supponiamo che A sia un sottoinsieme di \mathbb{R}^2 tale che la sua frontiera ∂A sia descrivibile come il supporto di una curva chiusa e regolare γ . Con questo intendiamo che esiste una $\gamma : [0, T] \rightarrow \mathbb{R}^2$, iniettiva, derivabile, con $\gamma(0) = \gamma(T)$, $\gamma'(0) = \gamma'(T)$ e tale che $\partial A = \gamma([0, T])$. Quindi $\mathbf{x}_0 \in \partial A \Leftrightarrow \mathbf{x}_0 = \gamma(t_0)$ per un opportuno $t_0 \in [0, T]$. D'altra parte è chiaro che se \mathbf{x}_0 è punto di massimo (minimo) relativo per f , allora t_0 è punto di massimo (minimo) relativo per $f \circ \gamma$. Questo ci dice che gli eventuali punti di massimo o minimo relativo che si trovano sul bordo di A si possono cercare tra le immagini $\gamma(t)$ dei t in $[0, T]$ critici per $f \circ \gamma$.

Tutte le condizioni sopra, vale la pena ripeterlo, sono solo necessarie - per dimostrare che un punto stazionario è effettivamente un massimo o un minimo relativo bisogna trovare qualche altro argomento. Se però i punti stazionari che si trovano sono un numero finito e se si sa per altri motivi che f ha massimo (o minimo) su A (per esempio perché A è chiuso e limitato e dunque vale il teorema di Weierstrass), allora il punto stazionario su cui f ha il valore più alto (più basso), tra i punti stazionari, deve essere il punto di massimo (minimo) assoluto.

Esempio 3.19. *Consideriamo*

$$A := \{(x, y) : x^2 + 4y^2 \leq 1\} \quad f(x, y) = x^2 - y^2$$

Non è difficile verificare che

$$\partial A = \{(x, y) : x^2 + 4y^2 = 1\}$$

e dunque A è chiuso; inoltre è facile vedere che f è continua.

Per Weierstrass f DEVE avere massimo e minimo (assoluti) su A . Dato

che f è differenziabile su $\text{int}(A)$ il punto di massimo (minimo) o è interno ad A ed è stazionario per f oppure si trova su ∂A . Vediamo allora chi sono i punti stazionari per f . Si ha $\nabla f(x, y) = (2x, -2y)$ e dunque l'unico punto stazionario per f è il punto $(0, 0)$, in cui f vale zero. Abbiamo già visto nell'esempio precedente che questo punto è una sella, quindi il massimo e il minimo di f sono sul bordo di A .

Per studiare f su ∂A possiamo considerare la curva $\gamma : [0, 2\pi] \rightarrow \mathbb{R}^2$

$$\gamma(t) := \left(\cos(t), \frac{1}{2} \sin(t) \right)$$

che al variare di t dà tutti e soli i punti di ∂A (descrizione parametrica dell'ellisse). Cerchiamo quindi i punti stazionari di $f \circ \gamma$ cioè di

$$(f \circ \gamma)(t) = \cos^2(t) - \frac{1}{4} \sin^2(t) = 1 - \frac{5}{4} \sin^2(t)$$

su $[0, 2\pi]$. Derivando si trova $(f \circ \gamma)'(t) = \frac{5}{2} \sin(t) \cos(t)$ che ha le quattro radici

$$t_1 = 0, \quad t_2 = \frac{\pi}{2}, \quad t_3 = \pi, \quad t_4 = \frac{3\pi}{2}$$

(ci sarebbe anche $t_5 = 2\pi$ ma rappresenta lo stesso punto di t_1). Quindi i punti del bordo che dobbiamo considerare sono

$$\mathbf{x}_1 = (1, 0), \quad \mathbf{x}_2 = (0, 1/2), \quad \mathbf{x}_3 = (-1, 0), \quad \mathbf{x}_4 = (0, -1/2)$$

e $f(\mathbf{x}_1) = f(\mathbf{x}_3) = 1$ mentre $f(\mathbf{x}_2) = f(\mathbf{x}_4) = -1/4$; il massimo di f su A vale 1 ed è assunto nei punti $(\pm 1, 0)$, il minimo di f vale $1/4$ ed è assunto in $(0, \pm 1/2)$. Notiamo che anche se non avessimo saputo che $(0, 0)$ è punto di sella avremmo comunque potuto escludere che fosse massimo o minimo ASSOLUTO, perché $f(0, 0) = 0$.

Esempio 3.20. Consideriamo $f(x, y) := xy$ sul quadrato $Q := \{(x, y) : -1 \leq x, y \leq 1\}$ e cerchiamo, come prima, di trovare il massimo e il minimo di f su Q (che ci sono per Weierstrass). Si ha

$$\nabla f(x, y) = (y, x)$$

e dunque $\nabla f(x, y) = (0, 0)$ se e solo se $(x, y) = (0, 0)$. Dato che f è strettamente positiva sul primo e terzo quadrante, mentre è strettamente negativa sul secondo e sul quarto quadrante, tale punto è una sella. Il bordo ∂Q è composto dai quattro lati di Q

$$\begin{aligned} L_1 &= \{(1, y) : -1 \leq y \leq 1\}, & L_2 &= \{(x, 1) : -1 \leq x \leq 1\}, \\ L_3 &= \{(-1, y) : -1 \leq y \leq 1\}, & L_4 &= \{(x, -1) : -1 \leq x \leq 1\}. \end{aligned}$$

Per studiare f su L_1 possiamo studiare $y \mapsto f(1, y) = y$ per y in $[-1, 1]$ (che significa descrivere L_1 mediante la curva $\gamma_1(t) := (1, t)$, $\gamma : [-1, 1] \rightarrow \mathbb{R}^2$). Si vede allora che f ristretta a L_1 ha massimo 1, assunto nel vertice $(1, 1)$ e ha minimo -1 , assunto nel vertice $(1, -1)$. Ripetendo gli stessi ragionamenti su L_2, L_3 ed L_4 si vede che i punti $(1, 1)$ e $(-1, -1)$ sono di massimo, mentre i punti $(-1, 1)$ e $(1, -1)$ sono di minimo.

4 Derivate seconde

Consideriamo per semplicità in quest'ultimo paragrafo una funzione scalare: $f : A \rightarrow \mathbb{R}$.

Definizione 4.1. Se f è differenziabile in A allora assegnata una direzione \mathbf{y}_1 in \mathbb{R}^N è definita la derivata direzionale $f'(\mathbf{x})(\mathbf{y}_1)$, per ogni \mathbf{x} di A ; risulta così definita una nuova funzione che possiamo indicare con $f'_{\mathbf{y}_1} : A \rightarrow \mathbb{R}$. Prendiamo ora un punto \mathbf{x}_0 e un'altra direzione \mathbf{y}_2 . Può capitare che $f'_{\mathbf{y}_1}$ sia derivabile lungo \mathbf{y}_2 nel punto \mathbf{x}_0 : in questo caso diremo che \mathbf{f} ammette derivata direzionale seconda lungo $(\mathbf{y}_1, \mathbf{y}_2)$ e indicheremo con $f''(\mathbf{x}_0)(\mathbf{y}_1, \mathbf{y}_2)$ la derivata direzionale $(f'_{\mathbf{y}_1})'(\mathbf{x}_0)(\mathbf{y}_2)$. Se $\mathbf{y}_1 = \mathbf{y}_2 = \mathbf{y}$ è usuale scrivere $f''(\mathbf{x}_0)(\mathbf{y}^2)$ in luogo di $f''(\mathbf{x}_0)(\mathbf{y}, \mathbf{y})$.

Supponiamo che $\mathbf{y}_1 = \mathbf{e}_i$, $\mathbf{y}_2 = \mathbf{e}_j$, dove come prima \mathbf{e}_i denota il versore i -esimo della base canonica $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$, (1 al posto i -esimo). Allora se in \mathbf{x}_0 esiste la derivata seconda $f''(\mathbf{x}_0)(\mathbf{e}_i, \mathbf{e}_j)$ diremo che f ammette derivate parziali seconde rispetto a x_i e x_j , in \mathbf{x}_0 , e useremo la notazione

$$\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}_0) = f''(\mathbf{x}_0)(\mathbf{e}_i, \mathbf{e}_j);$$

nel caso $i = j$ scriveremo semplicemente $\frac{\partial^2}{\partial x_i^2} f(\mathbf{x}_0)$ in luogo di $\frac{\partial^2}{\partial x_i \partial x_i} f(\mathbf{x}_0)$

In generale NON è detto che $f''(\mathbf{x})(\mathbf{y}_1, \mathbf{y}_2) = f''(\mathbf{x})(\mathbf{y}_2, \mathbf{y}_1)$. Nei "casi normali" tutto funziona come ci si aspetta a causa del seguente teorema.

Teorema 4.2 (di Schwartz). Se le derivate parziali seconde $\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x})$ esistono per ogni $i, j = 1, \dots, N$ e se sono CONTINUE rispetto a \mathbf{x} , allora

$$\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) = \frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{x}) \quad \forall \mathbf{x}.$$

Notiamo che in tal caso, come conseguenza del teorema del differenziale totale (applicato alle varie $\frac{\partial}{\partial x_i} f(x)$), si ha

$$f''(\mathbf{x})(\mathbf{y}_1 \cdot \mathbf{y}_2) = \sum_{i,j=1}^N \frac{\partial^2}{\partial x_i \partial x_j} f(x) y_{1,i} y_{1,j}$$

(dove $\mathbf{y}_1 = (y_{1,1}, \dots, y_{1,N})$ e $\mathbf{y}_2 = (y_{2,1}, \dots, y_{2,N})$).

Definizione 4.3 (matrice Hessiana). Se f ha derivate parziali seconde, allora la matrice $N \times N$

$$H_f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} f(\mathbf{x}) & \frac{\partial^2}{\partial x_1 \partial x_2} f(\mathbf{x}) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_N} f(\mathbf{x}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(\mathbf{x}) & \frac{\partial^2}{\partial x_2^2} f(\mathbf{x}) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_N} f(\mathbf{x}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2}{\partial x_N \partial x_1} f(\mathbf{x}) & \frac{\partial^2}{\partial x_N \partial x_2} f(\mathbf{x}) & \cdots & \frac{\partial^2}{\partial x_N^2} f(\mathbf{x}) \end{pmatrix}$$

si chiama matrice Hessiana di f nel punto \mathbf{x} . Per il teorema di Schwartz, se le derivate parziali sono continue tale matrice risulta SIMMETRICA.

Osservazione 4.4. Nelle ipotesi del teorema di Schwartz ha

$$f''(\mathbf{x})(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{y}_1 \cdot H_f(\mathbf{x}) \mathbf{y}_2 = \mathbf{y}_2 \cdot H_f(\mathbf{x}) \mathbf{y}_1$$

Teorema 4.5 (formula di Taylor al secondo ordine). Supponiamo che f abbia derivate seconde continue e sia \mathbf{x}_0 un punto di A . Allora

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot H_f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|^2)$$

In altri termini se poniamo

$$P_2(\mathbf{y}) := f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \mathbf{y} + \mathbf{y} \cdot H_f(\mathbf{x}_0) \mathbf{y}$$

(P_2 è il polinomio di Taylor di ordine due relativo a \mathbf{x}_0), allora

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{f(\mathbf{x}) - P_2(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|^2} = 0$$

Tutto quanto fatto sopra per le derivate seconde si potrebbe iterare, dando delle nozioni di derivate parziali $\frac{\partial^K}{\partial x_1^{k_1} \cdots \partial x_N^{k_n}} f$ di ordine K , k_1 volte rispetto

a x_1, \dots, k_N volte rispetto a x_N ($k_1 + \dots + k_N = K$). In questo modo si potrebbe costruire il polinomio di Taylor di ordine K (con un'opportuna miscela delle derivate di ordine K e scrivere la corrispondente formula di approssimazione a meno di $o(\|\mathbf{x} - \mathbf{x}_0\|^K)$ - ma questa è un'altra storia, che non abbiamo tempo di raccontare. Mostriamo solamente come la nozione di matrice Hessiana, introdotta prima, possa essere utile per classificare i punti stazionari di una funzione, estendendo il semplice criterio per le funzioni di una variabile che dice $f''(x_0) > 0 / < 0 \Rightarrow x_0$ punto di minimo/massimo.

Osservazione 4.6. Ricordiamo che una matrice simmetrica $N \times N$ ammette sempre N autovalori $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$, anzi esiste una base ortonormale di \mathbb{R}^N in cui tale matrice assume la forma diagonale

$$\begin{pmatrix} \lambda_1 & , & 0 & \cdots & , & 0 \\ 0 & , & \lambda_2 & \cdots & , & 0 \\ \vdots & , & \vdots & \cdots & , & \vdots \\ 0 & , & 0, & \cdots & , & \lambda_N \end{pmatrix}$$

Teorema 4.7 (classificazione dei punti stazionari). Sia $f : A \rightarrow \mathbb{R}$ derivabile due volte in A con derivate seconde continue. Sia \mathbf{x}_0 in A un punto stazionario per f ($\nabla f(\mathbf{x}_0) = 0$). Siano

$$\lambda_1(\mathbf{x}_0) \leq \lambda_2(\mathbf{x}_0) \leq \dots \leq \lambda_N(\mathbf{x}_0)$$

gli autovalori della matrice Hessiana $H_f(\mathbf{x}_0)$. Allora

- se $\lambda_1 > 0$ (tutti gli autovalori positivi), allora \mathbf{x}_0 è un punto di minimo per f ;
- se $\lambda_N < 0$ (tutti gli autovalori negativi), allora \mathbf{x}_0 è un punto di massimo per f ;
- se $\lambda_1 < 0$ e $\lambda_N > 0$ il punto \mathbf{x}_0 è di sella per f : per la verità una vera sella dovrebbe avere tutti gli autovalori diversi da zero, in questo caso si potrebbe vedere che f scende (vicino a \mathbf{x}_0) nelle direzioni degli autovettori relativi agli autovalori negativi mentre sale nelle direzioni degli autovettori relativi agli autovalori positivi.

Se $\lambda_1 = 0$ o $\lambda_N = 0$ non si può dire nulla (porebbe esserci un analogo dei punti di flesso)