Analisi Matematica 2

Appunti delle lezioni tenute dal Prof. A. Fonda

Università di Trieste, CdL Matematica, a.a. 2017/2018

1 La derivata direzionale

In questa sezione, E sarà un sottoinsieme aperto di $\mathbb{R}^N$, $\boldsymbol{x}_0$ un punto di E e $f : E \to \mathbb{R}$ una funzione. Vogliamo estendere il concetto di derivata già introdotto nel caso $N = 1$. Iniziamo con il fissare una "direzione", ossia un vettore $\boldsymbol{v} \in \mathbb{R}^N$ tale che $\|\boldsymbol{v}\| = 1$ (detto anche "versore"). Chiamiamo, se esiste, "derivata direzionale" di f in $\boldsymbol{x}_0$ nella direzione $\boldsymbol{v}$ il seguente limite

$$\lim_{t\to 0} \frac{f(\boldsymbol{x}_0 + t\boldsymbol{v}) - f(\boldsymbol{x}_0)}{t},$$

che verrà indicato con il simbolo

$$\frac{\partial f}{\partial \boldsymbol{v}}(\boldsymbol{x}_0).$$

Se $\boldsymbol{v}$ coincide con un elemento $\boldsymbol{e}_k$ della base canonica $(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_N)$ di $\mathbb{R}^N$, la derivata direzionale si chiamerà "derivata parziale" $k-$esima di f in $\boldsymbol{x}_0$ e si indicherà con

$$\frac{\partial f}{\partial x_k}(\boldsymbol{x}_0).$$

Se $\boldsymbol{x}_0 = (x_1^0, x_2^0, \dots, x_N^0)$, si ha quindi:

$$\begin{aligned}\frac{\partial f}{\partial x_k}(\boldsymbol{x}_0) &= \lim_{t\to 0} \frac{f(\boldsymbol{x}_0 + t\boldsymbol{e}_k) - f(\boldsymbol{x}_0)}{t} \\ &= \lim_{t\to 0} \frac{f(x_1^0, x_2^0, \dots, x_k^0 + t, \dots, x_N^0) - f(x_1^0, x_2^0, \dots, x_k^0, \dots, x_N^0)}{t},\end{aligned}$$

per cui si usa parlare di "derivata rispetto alla $k-$esima variabile".

Esistono delle funzioni che, pur avendo derivate direzionali in tutte le possibili direzioni, non sono continue. Ad esempio, la funzione $f : \mathbb{R}^2 \to \mathbb{R}$ definita da

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

ha tutte le derivate direzionali nulle in $\boldsymbol{x}_0 = (0,0)$, ma non è continua in tale punto, come si vede considerando la restrizione alla parabola $\{(x,y) \in \mathbb{R}^2 : y = x^2\}$. Questo fatto ci porta a cercare una generalizzazione più appropriata del concetto di derivata.

2 Il differenziale di una funzione a valori scalari

Definizione. *Diremo che la funzione f è "differenziabile" in $\boldsymbol{x}_0$ se esiste una applicazione lineare $\ell : \mathbb{R}^N \to \mathbb{R}$ per cui si possa scrivere*

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \ell(\boldsymbol{x} - \boldsymbol{x}_0) + r(\boldsymbol{x}) \,,$$

dove r è una funzione tale che

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \frac{r(\boldsymbol{x})}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} = 0 \,.$$

Se f è differenziabile in $\boldsymbol{x}_0$, l'applicazione lineare ℓ si chiama "differenziale" di f in $\boldsymbol{x}_0$ e si indica con il simbolo

$$df(\boldsymbol{x}_0) \,.$$

Teorema. *Se f è differenziabile in $\boldsymbol{x}_0$, allora f è continua in $\boldsymbol{x}_0$.*

Dimostrazione. Sappiamo che l'applicazione $\ell = df(\boldsymbol{x}_0)$, essendo lineare, è continua e $\ell(\boldsymbol{0}) = 0$. Ne segue che

$$\begin{aligned}
\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} f(\boldsymbol{x}) &= \lim_{\boldsymbol{x} \to \boldsymbol{x}_0} [f(\boldsymbol{x}_0) + \ell(\boldsymbol{x} - \boldsymbol{x}_0) + r(\boldsymbol{x})] \\
&= f(\boldsymbol{x}_0) + \ell(\boldsymbol{0}) + \lim_{\boldsymbol{x} \to \boldsymbol{x}_0} r(\boldsymbol{x}) \\
&= f(\boldsymbol{x}_0) + \lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \frac{r(\boldsymbol{x})}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} \lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \|\boldsymbol{x} - \boldsymbol{x}_0\| \\
&= f(\boldsymbol{x}_0) \,.
\end{aligned}$$

■

Seguendo un'abitudine consolidata per le applicazioni lineari, si usa spesso scrivere $df(\boldsymbol{x}_0)\boldsymbol{h}$ invece di $df(\boldsymbol{x}_0)(\boldsymbol{h})$.

Teorema. *Se f è differenziabile in $\boldsymbol{x}_0$, allora esistono tutte le derivate direzionali di f in $\boldsymbol{x}_0$: per ogni direzione $\boldsymbol{v} \in \mathbb{R}^N$ si ha*

$$\frac{\partial f}{\partial \boldsymbol{v}}(\boldsymbol{x}_0) = df(\boldsymbol{x}_0)\boldsymbol{v} \,.$$

Dimostrazione. Usando la definizione di differenziale, abbiamo

$$\begin{aligned}\lim_{t\to0}\frac{f(\boldsymbol{x}_0+t\boldsymbol{v})-f(\boldsymbol{x}_0)}{t} &= \lim_{t\to0}\frac{df(\boldsymbol{x}_0)(t\boldsymbol{v})+r(\boldsymbol{x}_0+t\boldsymbol{v})}{t}\\ &= \lim_{t\to0}\frac{t\,df(\boldsymbol{x}_0)\boldsymbol{v}+r(\boldsymbol{x}_0+t\boldsymbol{v})}{t}\\ &= df(\boldsymbol{x}_0)\boldsymbol{v}+\lim_{t\to0}\frac{r(\boldsymbol{x}_0+t\boldsymbol{v})}{t}\,;\end{aligned}$$

d'altra parte, essendo $\|\boldsymbol{v}\|=1$, si ha

$$\lim_{t\to0}\left|\frac{r(\boldsymbol{x}_0+t\boldsymbol{v})}{t}\right| = \lim_{\boldsymbol{x}\to\boldsymbol{x}_0}\frac{|r(\boldsymbol{x})|}{\|\boldsymbol{x}-\boldsymbol{x}_0\|}=0\,,$$

da cui la tesi. ■

In particolare, se $\boldsymbol{v}$ coincide con un elemento $\boldsymbol{e}_k$ della base canonica $(\boldsymbol{e}_1,\boldsymbol{e}_2,\ldots,\boldsymbol{e}_N)$, si ha:

$$\frac{\partial f}{\partial x_k}(\boldsymbol{x}_0)=df(\boldsymbol{x}_0)\boldsymbol{e}_k\,.$$

Scrivendo il vettore $\boldsymbol{h}\in\mathbb{R}^N$ come $\boldsymbol{h}=h_1\boldsymbol{e}_1+h_2\boldsymbol{e}_2+\ldots+h_N\boldsymbol{e}_N$, abbiamo

$$\begin{aligned}df(\boldsymbol{x}_0)\boldsymbol{h} &= h_1df(\boldsymbol{x}_0)\boldsymbol{e}_1+h_2df(\boldsymbol{x}_0)\boldsymbol{e}_2+\ldots+h_Ndf(\boldsymbol{x}_0)\boldsymbol{e}_N\\ &= h_1\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0)+h_2\frac{\partial f}{\partial x_2}(\boldsymbol{x}_0)+\ldots+h_N\frac{\partial f}{\partial x_N}(\boldsymbol{x}_0)\,,\end{aligned}$$

ossia

$$df(\boldsymbol{x}_0)\boldsymbol{h}=\sum_{k=1}^N\frac{\partial f}{\partial x_k}(\boldsymbol{x}_0)h_k\,.$$

Introducendo il vettore "gradiente" di f in $\boldsymbol{x}_0$

$$\nabla f(\boldsymbol{x}_0)=\left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0),\frac{\partial f}{\partial x_2}(\boldsymbol{x}_0),\ldots,\frac{\partial f}{\partial x_N}(\boldsymbol{x}_0)\right)\,,$$

si può scrivere

$$df(\boldsymbol{x}_0)\boldsymbol{h}=\nabla f(\boldsymbol{x}_0)\cdot\boldsymbol{h}\,.$$

3 Funzioni di classe C^1

Il seguente risultato è noto come "teorema del differenziale totale".

Teorema. *Se f possiede le derivate parziali in un intorno di $\boldsymbol{x}_0$ ed esse sono continue in $\boldsymbol{x}_0$, allora f è differenziabile in $\boldsymbol{x}_0$.*

Dimostrazione. Supporremo per semplicità di notazioni $N=2$. Definiamo l'applicazione lineare $\ell:\mathbb{R}^2\to\mathbb{R}$ che ad ogni vettore $\boldsymbol{h}=(h_1,h_2)$ associa

$$\ell(\boldsymbol{h})=\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0)h_1+\frac{\partial f}{\partial x_2}(\boldsymbol{x}_0)h_2\,.$$

Vedremo che ℓ è proprio il differenziale di f in $\boldsymbol{x}_0$. Intanto, è lineare, come si vede immediatamente. Inoltre, scrivendo $\boldsymbol{x}_0 = (x_1^0, x_2^0)$ e $\boldsymbol{x} = (x_1, x_2)$, per il teorema di Lagrange si ha

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{x}_0) &= (f(x_1, x_2) - f(x_1^0, x_2)) + (f(x_1^0, x_2) - f(x_1^0, x_2^0)) \\ &= \frac{\partial f}{\partial x_1}(\xi_1, x_2)(x_1 - x_1^0) + \frac{\partial f}{\partial x_2}(x_1^0, \xi_2)(x_2 - x_2^0)\,, \end{aligned}$$

per un certo $\xi_1 \in [x_1^0, x_1]$ e un certo $\xi_2 \in [x_2^0, x_2]$. Quindi,

$$\begin{aligned} r(\boldsymbol{x}) &= f(\boldsymbol{x}) - f(\boldsymbol{x}_0) - \ell(\boldsymbol{x} - \boldsymbol{x}_0) \\ &= \left[\frac{\partial f}{\partial x_1}(\xi_1, x_2) - \frac{\partial f}{\partial x_1}(x_1^0, x_2^0)\right](x_1 - x_1^0) + \\ &\quad + \left[\frac{\partial f}{\partial x_2}(x_1^0, \xi_2) - \frac{\partial f}{\partial x_2}(x_1^0, x_2^0)\right](x_2 - x_2^0)\,, \end{aligned}$$

ed essendo $|x_1 - x_1^0| \leq \|\boldsymbol{x} - \boldsymbol{x}_0\|$ e $|x_2 - x_2^0| \leq \|\boldsymbol{x} - \boldsymbol{x}_0\|$,

$$\frac{|r(\boldsymbol{x})|}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} \leq \left|\frac{\partial f}{\partial x_1}(\xi_1, x_2) - \frac{\partial f}{\partial x_1}(x_1^0, x_2^0)\right| + \left|\frac{\partial f}{\partial x_2}(x_1^0, \xi_2) - \frac{\partial f}{\partial x_2}(x_1^0, x_2^0)\right|\,.$$

Facendo tendere $\boldsymbol{x}$ a $\boldsymbol{x}_0$, si ha che $(\xi_1, x_2) \to (x_1^0, x_2^0)$ e $(x_1^0, \xi_2) \to (x_1^0, x_2^0)$ per cui, essendo $\frac{\partial f}{\partial x_1}$ e $\frac{\partial f}{\partial x_2}$ continue in $\boldsymbol{x}_0 = (x_1^0, x_2^0)$, si ha

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \frac{|r(\boldsymbol{x})|}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} = 0\,,$$

da cui la tesi. ■

Diremo che la funzione f è di classe $\mathcal{C}^1$ su E se f possiede le derivate parziali ed esse sono continue su tutto E. Dal teorema precedente segue che una funzione di classe $\mathcal{C}^1$ è "differenziabile su E", ossia in ogni punto di E.

4 Derivate parziali successive

Supponiamo, per semplicità, $N = 2$. Consideriamo E, un insieme aperto di $\mathbb{R}^2$ e una funzione $f : E \to \mathbb{R}$ che abbia le derivate parziali $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$ in tutti i punti di E. Se esse posseggono a loro volta derivate parziali in un punto $\boldsymbol{x}_0$, queste si dicono "derivate parziali seconde" della f in $\boldsymbol{x}_0$ e si denotano con i simboli

$$\begin{aligned} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}_0) &= \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0)\,, \quad \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}_0) = \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0)\,, \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}_0) &= \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_2}(\boldsymbol{x}_0)\,, \quad \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}_0) = \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_2}(\boldsymbol{x}_0)\,. \end{aligned}$$

Teorema (di Schwarz). *Se esistono le derivate parziali seconde* $\frac{\partial^2 f}{\partial x_2 \partial x_1}$, $\frac{\partial^2 f}{\partial x_1 \partial x_2}$ *in un intorno di* $\boldsymbol{x}_0$ *ed esse sono continue in* $\boldsymbol{x}_0$, *allora*

$$\frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}_0) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}_0)\,.$$

Dimostrazione. Sia $\rho > 0$ tale che $B(\boldsymbol{x}_0, \rho) \subseteq E$. Scriviamo $\boldsymbol{x}_0 = (x_1^0, x_2^0)$ e prendiamo un $\boldsymbol{x} = (x_1, x_2) \in B(\boldsymbol{x}_0, \rho)$ tale che $x_1 \neq x_1^0$ e $x_2 \neq x_2^0$. Possiamo allora definire

$$g(x_1, x_2) = \frac{f(x_1, x_2) - f(x_1, x_2^0)}{x_2 - x_2^0}\,, \quad h(x_1, x_2) = \frac{f(x_1, x_2) - f(x_1^0, x_2)}{x_1 - x_1^0}\,.$$

Si verifica che vale l'uguaglianza

$$\frac{g(x_1, x_2) - g(x_1^0, x_2)}{x_1 - x_1^0} = \frac{h(x_1, x_2) - h(x_1, x_2^0)}{x_2 - x_2^0}\,.$$

Per il teorema di Lagrange, esiste un $\xi_1 \in]x_1^0, x_1[$ tale che

$$\frac{g(x_1, x_2) - g(x_1^0, x_2)}{x_1 - x_1^0} = \frac{\partial g}{\partial x_1}(\xi_1, x_2) = \frac{\frac{\partial f}{\partial x_1}(\xi_1, x_2) - \frac{\partial f}{\partial x_1}(\xi_1, x_2^0)}{x_2 - x_2^0}\,,$$

ed esiste un $\xi_2 \in]x_2^0, x_2[$ tale che

$$\frac{h(x_1, x_2) - h(x_1, x_2^0)}{x_2 - x_2^0} = \frac{\partial h}{\partial x_2}(x_1, \xi_2) = \frac{\frac{\partial f}{\partial x_2}(x_1, \xi_2) - \frac{\partial f}{\partial x_2}(x_1^0, \xi_2)}{x_1 - x_1^0}\,.$$

Di nuovo per il teorema di Lagrange, esiste un $\eta_2 \in]x_2^0, x_2[$ tale che

$$\frac{\frac{\partial f}{\partial x_1}(\xi_1, x_2) - \frac{\partial f}{\partial x_1}(\xi_1, x_2^0)}{x_2 - x_2^0} = \frac{\partial^2 f}{\partial x_2 \partial x_1}(\xi_1, \eta_2)\,,$$

ed esiste un $\eta_1 \in]x_1^0, x_1[$ tale che

$$\frac{\frac{\partial f}{\partial x_2}(x_1, \xi_2) - \frac{\partial f}{\partial x_2}(x_1^0, \xi_2)}{x_1 - x_1^0} = \frac{\partial^2 f}{\partial x_1 \partial x_2}(\eta_1, \xi_2)\,.$$

Quindi,

$$\frac{\partial^2 f}{\partial x_2 \partial x_1}(\xi_1, \eta_2) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(\eta_1, \xi_2)\,.$$

Facendo tendere $\boldsymbol{x} = (x_1, x_2)$ a $\boldsymbol{x}_0 = (x_1^0, x_2^0)$, si ha che sia (ξ_1, η_2) che (η_1, ξ_2) tendono a $\boldsymbol{x}_0$, e per la continuità delle derivate seconde miste si ha la tesi. ■

Diremo che la funzione f è di classe $\mathcal{C}^2$ su E se f possiede tutte le derivate parziali seconde ed esse sono continue su tutto E. Dal teorema precedente segue che se una funzione di classe $\mathcal{C}^2$, le derivate parziali "miste" sono uguali.

È utile definire la "matrice hessiana" di f nel punto $\boldsymbol{x}_0$:

$$Hf(\boldsymbol{x}_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}_0) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}_0) \\ \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}_0) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}_0) \end{pmatrix};$$

se f è di classe $\mathcal{C}^2$, si tratta di una matrice simmetrica.

Quanto sopra si può estendere senza difficoltà alle funzioni di N variabili, con N qualunque. Se f è di classe $\mathcal{C}^2$, la matrice hessiana risulta allora una matrice simmetrica del tipo $N \times N$.

Procedendo per induzione, si possono definire le derivate parziali $n-$esime di una funzione. Si dice che la funzione f è di classe $\mathcal{C}^n$ su E se f possiede tutte le derivate parziali $n-$esime ed esse sono continue su tutto E.

5 La formula di Taylor

Supponiamo ora che $f : E \to \mathbb{R}$ sia una funzione di classe $\mathcal{C}^{n+1}$, per un certo $n \geq 1$.

Consideriamo come sopra, per semplicità, il caso $N = 2$. Introduciamo le seguenti notazioni:

$$D_{x_1} = \frac{\partial}{\partial x_1}, \quad D_{x_2} = \frac{\partial}{\partial x_2},$$

$$D_{x_1}^2 = \frac{\partial^2}{\partial x_1^2}, \quad D_{x_1} D_{x_2} = \frac{\partial^2}{\partial x_1 \partial x_2}, \quad D_{x_2}^2 = \frac{\partial^2}{\partial x_2^2},$$

e così via, per le derivate parziali successive. Si noti che, per un vettore $\boldsymbol{h} = (h_1, h_2) \in \mathbb{R}^2$, si ha

$$df(\boldsymbol{x}_0)\boldsymbol{h} = h_1 D_{x_1} f(\boldsymbol{x}_0) + h_2 D_{x_2} f(\boldsymbol{x}_0),$$

che risulterà conveniente scrivere

$$df(\boldsymbol{x}_0)\boldsymbol{h} = [h_1 D_{x_1} + h_2 D_{x_2}] f(\boldsymbol{x}_0).$$

In questo modo, possiamo pensare che f viene trasformata dall'operatore $[h_1 D_{x_1} + h_2 D_{x_2}]$ nella nuova funzione $[h_1 D_{x_1} + h_2 D_{x_2}] f = h_1 D_{x_1} f + h_2 D_{x_2} f$.

Dati due punti $\boldsymbol{x}_0$ e $\boldsymbol{x}$ in $\mathbb{R}^N$, si definisce il "segmento" che li congiunge:

$$[\boldsymbol{x}_0, \boldsymbol{x}] = \{\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0) : t \in [0, 1]\};$$

analogamente, scriveremo

$$]\boldsymbol{x}_0, \boldsymbol{x}[= \{\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0) : t \in]0, 1[\}.$$

Supponiamo ora che $[\boldsymbol{x}_0, \boldsymbol{x}]$ sia un segmento contenuto in E e consideriamo la funzione $\phi : [0, 1] \to \mathbb{R}$ definita da

$$\phi(t) = f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0)).$$

Dimostriamo che ϕ è derivabile $n+1$ volte su $[0,1]$. Per $t \in [0,1]$, essendo f differenziabile in $\boldsymbol{u}_0 = \boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0)$, si ha

$$f(\boldsymbol{u}) = f(\boldsymbol{u}_0) + df(\boldsymbol{u}_0)(\boldsymbol{u} - \boldsymbol{u}_0) + r(\boldsymbol{u}),$$

con

$$\lim_{\boldsymbol{u} \to \boldsymbol{u}_0} \frac{r(\boldsymbol{u})}{\|\boldsymbol{u} - \boldsymbol{u}_0\|} = 0.$$

Quindi,

$$\begin{aligned}
\lim_{s \to t} \frac{\phi(s) - \phi(t)}{s - t} &= \lim_{s \to t} \frac{f(\boldsymbol{x}_0 + s(\boldsymbol{x} - \boldsymbol{x}_0)) - f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0))}{s - t} \\
&= \lim_{s \to t} \frac{df(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0))((s-t)(\boldsymbol{x} - \boldsymbol{x}_0)) + r(\boldsymbol{x}_0 + s(\boldsymbol{x} - \boldsymbol{x}_0))}{s - t} \\
&= df(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0))(\boldsymbol{x} - \boldsymbol{x}_0) + \lim_{s \to t} \frac{r(\boldsymbol{x}_0 + s(\boldsymbol{x} - \boldsymbol{x}_0))}{s - t},
\end{aligned}$$

ed essendo

$$\lim_{s \to t} \left| \frac{r(\boldsymbol{x}_0 + s(\boldsymbol{x} - \boldsymbol{x}_0))}{s - t} \right| = \lim_{\boldsymbol{u} \to \boldsymbol{u}_0} \frac{|r(\boldsymbol{u})|}{\|\boldsymbol{u} - \boldsymbol{u}_0\|} \|\boldsymbol{x} - \boldsymbol{x}_0\| = 0,$$

si ha

$$\phi'(t) = \lim_{s \to t} \frac{\phi(s) - \phi(t)}{s - t} = df(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0))(\boldsymbol{x} - \boldsymbol{x}_0).$$

Con le nuove notazioni, ponendo $\boldsymbol{x} - \boldsymbol{x}_0 = \boldsymbol{h} = (h_1, h_2)$, abbiamo

$$\phi'(t) = [h_1 D_{x_1} + h_2 D_{x_2}] f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0)) = g(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0)),$$

dove g è la nuova funzione $[h_1 D_{x_1} + h_2 D_{x_2}] f$. Possiamo allora iterare il procedimento, e calcolare la derivata seconda di ϕ:

$$\begin{aligned}
\phi''(t) &= [h_1 D_{x_1} + h_2 D_{x_2}] g(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0)) \\
&= [h_1 D_{x_1} + h_2 D_{x_2}][h_1 D_{x_1} + h_2 D_{x_2}] f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0)).
\end{aligned}$$

Per brevità, scriveremo

$$\phi''(t) = [h_1 D_{x_1} + h_2 D_{x_2}]^2 f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0)).$$

Notiamo che, usando la linearità delle derivate parziali e l'uguaglianza delle derivate miste (teorema di Schwarz), si ha

$$\begin{aligned}
[h_1 D_{x_1} + h_2 D_{x_2}]^2 f &= h_1^2 D_{x_1}^2 f + 2h_1 h_2 D_{x_1} D_{x_2} f + h_2^2 D_{x_2}^2 f \\
&= [h_1^2 D_{x_1}^2 + 2h_1 h_2 D_{x_1} D_{x_2} + h_2^2 D_{x_2}^2] f.
\end{aligned}$$

Osserviamo che l'espressione

$$[h_1 D_{x_1} + h_2 D_{x_2}]^2 = [h_1^2 D_{x_1}^2 + 2h_1 h_2 D_{x_1} D_{x_2} + h_2^2 D_{x_2}^2]$$

si ottiene formalmente come il quadrato di un binomio. Procedendo in questo modo, si può dimostrare per induzione che, per $k = 1, 2, \ldots, n+1$, la formula della derivata $k-$esima di ϕ è

$$\phi^{(k)}(t) = [h_1 D_{x_1} + h_2 D_{x_2}]^k f(\boldsymbol{x}_0 + t(\boldsymbol{x} - \boldsymbol{x}_0))\,,$$

e che, usando formalmente la formula del binomio di Newton

$$(a_1 + a_2)^k = \sum_{j=0}^{k} \binom{k}{j} a_1^{k-j} a_2^j,$$

si ha

$$[h_1 D_{x_1} + h_2 D_{x_2}]^k = \left[\sum_{j=0}^{k} \binom{k}{j} h_1^{k-j} h_2^j D_{x_1}^{k-j} D_{x_2}^j \right]$$

(in questa formula, i simboli $D^0_{x_1}$ e $D^0_{x_2}$ vanno interpretati come l'operatore identità).

Per poter scrivere agevolmente la formula di Taylor, introduciamo la notazione

$$d^k f(\boldsymbol{x}_0)\boldsymbol{h}^k = [h_1 D_{x_1} + h_2 D_{x_2}]^k f(\boldsymbol{x}_0)\,.$$

Teorema. *Sia $f : E \to \mathbb{R}$ di classe $\mathcal{C}^{n+1}$ e $[\boldsymbol{x}_0, \boldsymbol{x}]$ un segmento contenuto in E. Allora esiste un $\boldsymbol{\xi} \in]\boldsymbol{x}_0, \boldsymbol{x}[$ tale che*

$$f(\boldsymbol{x}) = p_n(\boldsymbol{x}) + r_n(\boldsymbol{x})\,,$$

dove

$$p_n(\boldsymbol{x}) = f(\boldsymbol{x}_0) + df(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2!} d^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)^2 + \ldots + \frac{1}{n!} d^n f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)^n$$

è il "polinomio di Taylor di grado n associato alla funzione f nel punto $\boldsymbol{x}_0$" e

$$r_n(\boldsymbol{x}) = \frac{1}{(n+1)!} d^{n+1} f(\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{x}_0)^{n+1}$$

è il "resto di Lagrange".

Dimostrazione. Per la formula di Taylor applicata alla funzione ϕ, si ha

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{1}{2!}\phi''(0)t^2 + \ldots + \frac{1}{n!}\phi^{(n)}(0)t^n + \frac{1}{(n+1)!}\phi^{(n+1)}(\xi)t^{n+1}\,,$$

per un certo $\xi \in]0, t[$. La formula cercata si ottiene prendendo $t = 1$ e sostituendo i valori delle derivate di ϕ trovati sopra. ■

Il polinomio di Taylor si può anche scrivere nella forma compatta

$$p_n(\boldsymbol{x}) = \sum_{k=0}^{n} \frac{1}{k!} d^k f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)^k\,,$$

con la convenzione che $d^0 f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)^0$, il primo addendo della somma, sia $f(\boldsymbol{x}_0)$. Si ha quindi

$$\begin{aligned} p_n(\boldsymbol{x}) &= \sum_{k=0}^{n} \frac{1}{k!}[(x_1-x_1^0)D_{x_1}+(x_2-x_2^0)D_{x_2}]^k f(\boldsymbol{x}_0) \\ &= \sum_{k=0}^{n} \frac{1}{k!}\left(\sum_{j=0}^{k}\binom{k}{j}\frac{\partial^k f}{\partial^{k-j}x_1\partial^j x_2}(\boldsymbol{x}_0)\;(x_1-x_1^0)^{k-j}(x_2-x_2^0)^j\right). \end{aligned}$$

Può essere utile la seguente espressione per il polinomio di secondo grado:

$$p_2(\boldsymbol{x}) = f(\boldsymbol{x}_0)+\nabla f(\boldsymbol{x}_0)\cdot(\boldsymbol{x}-\boldsymbol{x}_0)+\tfrac{1}{2}\Big(Hf(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)\Big)\cdot(\boldsymbol{x}-\boldsymbol{x}_0)\,.$$

Il teorema sopra dimostrato resta valido per qualsiasi dimensione N, pur di interpretare correttamente le notazioni: ad esempio, per un vettore $\boldsymbol{h}=(h_1,h_2,\dots,h_N)$, si dovrà leggere

$$d^k f(\boldsymbol{x}_0)\boldsymbol{h}^k = [h_1D_{x_1}+h_2D_{x_2}+\ldots+h_ND_{x_N}]^k f(\boldsymbol{x}_0)\,.$$

In questo caso, volendo esplicitare il polinomio di Taylor, sarà utile utilizzare la formula di Leibniz

$$(a_1+a_2+\ldots+a_N)^k = \sum_{m_1+m_2+\ldots+m_N=k}\frac{k!}{m_1!\,m_2!\cdots m_N!}\,a_1^{m_1}a_2^{m_2}\cdots a_N^{m_N}\,.$$

6 La ricerca di massimi e minimi

Come sopra, consideriamo un insieme aperto $E\subseteq\mathbb{R}^N$ e una funzione $f:E\to\mathbb{R}$. Diremo che $\boldsymbol{x}_0\in E$ è un "punto di massimo locale" per la funzione f se esiste un intorno U di x_0 contenuto in E per cui x_0 è punto di massimo della restrizione di f a U. Equivalentemente, se

$$\exists\delta>0:\ \forall\boldsymbol{x}\in E\quad d(\boldsymbol{x},\boldsymbol{x}_0)<\delta\ \Rightarrow\ f(\boldsymbol{x})\le f(\boldsymbol{x}_0)\,.$$

Analogamente per "punto di minimo locale".

Teorema (di Fermat). *Se $\boldsymbol{x}_0$ è un punto di massimo o di minimo locale e f è differenziabile in $\boldsymbol{x}_0$, allora $\nabla f(\boldsymbol{x}_0)=0$.*

Dimostrazione. Se $\boldsymbol{x}_0$ è punto di massimo locale, per ogni direzione $\boldsymbol{v}\in\mathbb{R}^N$ avremo che

$$\frac{f(\boldsymbol{x}_0+t\boldsymbol{v})-f(\boldsymbol{x}_0)}{t}\quad\begin{cases}\ge 0 & \text{se } t<0\,,\\ \le 0 & \text{se } t>0\,.\end{cases}$$

Siccome f è differenziabile in $\boldsymbol{x}_0$, ne deduciamo che

$$\frac{\partial f}{\partial\boldsymbol{v}}(\boldsymbol{x}_0)=\lim_{t\to0}\frac{f(\boldsymbol{x}_0+t\boldsymbol{v})-f(\boldsymbol{x}_0)}{t}=0\,.$$

In particolare, sono nulle tutte le derivate parziali, per cui $\nabla f(\boldsymbol{x}_0)=0$. Nel caso in cui $\boldsymbol{x}_0$ sia un punto di minimo locale, si procede in modo analogo. ■

Un punto il cui il gradiente si annulli è detto "punto stazionario". Naturalmente un tale punto potrebbe non essere nè di massimo nè di minimo.

Mostreremo ora come la formula di Taylor possa essere usata per stabilire un criterio affinchè un punto stazionario sia di massimo, o di minimo. Iniziamo con una definizione. Diremo che una matrice $\mathbb{A}$ simmetrica $N \times N$ è *definita positiva* se

$$[\mathbb{A}\boldsymbol{h}] \cdot \boldsymbol{h} > 0 \,, \qquad \text{per ogni } \boldsymbol{h} \in \mathbb{R}^N \setminus \{\boldsymbol{0}\} \,.$$

Diremo che $\mathbb{A}$ è *definita negativa* se vale la disuguaglianza opposta, ossia se $-\mathbb{A}$ è definita positiva.

Teorema. *Se $\boldsymbol{x}_0$ è un punto stazionario e f è di classe C^2, con matrice hessiana $Hf(\boldsymbol{x}_0)$ definita positiva, allora $\boldsymbol{x}_0$ è un punto di minimo locale. Se invece $Hf(\boldsymbol{x}_0)$ è definita negativa, allora $\boldsymbol{x}_0$ è un punto di massimo locale.*

Dimostrazione. Per la formula di Taylor, per $\boldsymbol{x} \neq \boldsymbol{x}_0$ in un intorno di $\boldsymbol{x}_0$ esiste un $\boldsymbol{\xi} \in]\boldsymbol{x}_0, \boldsymbol{x}[$ per cui

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0) \cdot (\boldsymbol{x} - \boldsymbol{x}_0) + \tfrac{1}{2}\Big(Hf(\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{x}_0)\Big) \cdot (\boldsymbol{x} - \boldsymbol{x}_0) \,.$$

Se $\mathbb{A} = Hf(\boldsymbol{x}_0)$ è definita positiva, esiste un $c > 0$ tale che, per ogni $\boldsymbol{v} \in \mathbb{R}^N$ con $\|\boldsymbol{v}\| = 1$,

$$[\mathbb{A}\boldsymbol{v}] \cdot \boldsymbol{v} \geq c \,.$$

(Abbiamo qui usato il teorema di Weierstrass, e il fatto che la sfera $\{v \in \mathbb{R}^N : \|\boldsymbol{v}\| = 1\}$ è un insieme compatto.) Quindi

$$\Big(Hf(\boldsymbol{x}_0)\frac{\boldsymbol{x} - \boldsymbol{x}_0}{\|\boldsymbol{x} - \boldsymbol{x}_0\|}\Big) \cdot \frac{\boldsymbol{x} - \boldsymbol{x}_0}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} \geq c \,.$$

Per la continuità delle derivate seconde, se $\boldsymbol{x}$ è sufficientemente vicino a $\boldsymbol{x}_0$,

$$\Big(Hf(\boldsymbol{\xi})\frac{\boldsymbol{x} - \boldsymbol{x}_0}{\|\boldsymbol{x} - \boldsymbol{x}_0\|}\Big) \cdot \frac{\boldsymbol{x} - \boldsymbol{x}_0}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} \geq \tfrac{1}{2}c > 0 \,.$$

(Lo si vede per assurdo, usando di nuovo laa compattezza della sfera.) Essendo $\nabla f(\boldsymbol{x}_0) = \boldsymbol{0}$, per tali $\boldsymbol{x}$ abbiamo che

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \tfrac{1}{2}\Big(Hf(\boldsymbol{\xi})(\boldsymbol{x} - \boldsymbol{x}_0)\Big) \cdot (\boldsymbol{x} - \boldsymbol{x}_0) \\ &\geq f(\boldsymbol{x}_0) + \tfrac{1}{2}c\|\boldsymbol{x} - \boldsymbol{x}_0\|^2 > f(\boldsymbol{x}_0) \,, \end{aligned}$$

per cui $\boldsymbol{x}_0$ è un punto di minimo locale.

La dimostrazione della seconda affermazione è analoga. ■

Enunciamo ora (senza dimostrazione) due criteri utili a stabilire quando una matrice $\mathbb{A}$ simmetrica $N \times N$ è definita positiva o negativa. Ricordiamo che gli autovalori di una matrice simmetrica sono tutti reali.

Primo criterio. *La martrice $\mathbb{A}$ è definita positiva se tutti i suoi autovalori sono positivi. Essa è definita negativa se tutti i suoi autovalori sono negativi.*

Secondo criterio. *La martrice $\mathbb{A} = (a_{ij})_{ij}$ è definita positiva se*

$$
\begin{aligned}
& a_{11} > 0\,, \\
& \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0\,, \\
& \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} > 0\,,\ \ldots \\
& \det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix} > 0\,.
\end{aligned}
$$

Essa è definita negativa se i determinanti scritti sopra hanno segno alternato: quelli delle sottomatrici con un numero dispari di righe e di colonne sono negativi, mentre quelli delle sottomatrici con un numero pari di righe e di colonne sono positivi.