Riassunto schematico sul differenziale

Luciano Battaia(*)

18 ottobre 2010

In questo articolo sono proposti, in forma sintetica, i risultati connessi col concetto di differenziale per le funzioni di $\mathbb{R}^n$ in $\mathbb{R}^m$, anche con l'intento di mostrare come si possa trattare in maniera sostanzialmente identica il caso di funzioni di una o più variabili(1).

1 Notazioni

Gli spazi $\mathbb{R}^n$, $n \geq 1$, saranno pensati della struttura vettoriale solita, del prodotto scalare "canonico" e della norma dedotta da questo prodotto scalare. Gli elementi (punti) di $\mathbb{R}^n$ saranno denotati con $\vec{x}$: $\vec{x} = (x_1, x_2, \dots, x_n)^T$; di norma, inoltre, scriveremo semplicemente $(x_1, x_2, \dots, x_n)$ invece di $(x_1, x_2, \dots, x_n)^T$. Nel caso $n = 1$, il "vettore" $\vec{x}$ si riduce ad avere una sola componente, $\vec{x} = (x_1)$, che si può chiamare semplicemente x, omettendo anche la coppia di parentesi tonde: $\vec{x} = x$. In due o tre variabili spesso le componenti si chiamano (x, y) o (x, y, z), anziché (x_1, x_2) o (x_1, x_2, x_3). In questo articolo non useremo queste "semplificazioni", allo scopo di favorire una trattazione unitaria dei vari casi.

Ricordiamo anche che il prodotto scalare canonico e la norma da esso dedotta sono, rispettivamente,

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \,, \qquad \|\vec{x}\| = \sqrt{\vec{x} \cdot \vec{x}} \,. \tag{1}$$

La base canonica di $\mathbb{R}^n$ sarà indicata con $(\vec{e}_1,\, \vec{e}_2,\, \dots,\, \vec{e_n})$. Nei casi $n = 1,\, 2,\, 3$ si usano anche, secondo tradizione, i simboli $(\vec{i})$, $(\vec{i},\, \vec{j})$, $(\vec{i},\, \vec{j},\, \vec{k})$ rispettivamente.

2 Funzioni $\mathbb{R}^n \to \mathbb{R}$

Sia ora

$$f \colon A \subseteq \mathbb{R}^n \to \mathbb{R} \tag{2}$$

una funzione definita in un aperto (cosa che riterremo vera in tutto il seguito) A di $\mathbb{R}^n$. Fissato un punto $\vec{x}_0$ di $\mathbb{R}^n$, sia $f(\vec{x}_0)$ il corrispondente valore della funzione (che potremo denotare con $\vec{y}_0$ all'occorrenza (anche se la freccia su y sarebbe inutile trattandosi, in questo caso, di funzioni a valori in $\mathbb{R}$). Se consideriamo un generico punto $\vec{x}$ di A, porremo

$$\vec{h} = \vec{x} - \vec{x}_0 \,, \tag{3}$$

*http://www.batmath.it

[1] Una trattazione più dettagliata del concetto di differenziale si può trovare in *Osservazioni sul differenziale*, reperibile al link http://www.batmath.it/matematica/0-appunti_uni/differenziale.pdf

e diremo che $\vec{h}$ è l'*incremento* della variabile indipendente al passaggio da $\vec{x}_0$ a $\vec{x}$.

Porremo poi

$$\Delta f_{\vec{x}_0}(\vec{h}) = f(\vec{x}_0 + \vec{h}) - f(\vec{x}_0) \tag{4}$$

e chiameremo $\Delta f_{\vec{x}_0}(\vec{h})$ *incremento* della funzione relativo all'incremento $\vec{h}$ della variabile e al punto $\vec{x}_0$. Osserviamo che $\Delta f_{\vec{x}_0}$ è una funzione dell'incremento $\vec{h}$, una volta fissato $\vec{x}_0$ e che si ha, ovviamente, $\Delta f_{\vec{x}_0}(\vec{0}) = f(\vec{x}_0 + \vec{0}) - f(\vec{x}_0) = f(\vec{x}_0) - f(\vec{x}_0) = 0$. Segnaliamo esplicitamente che si ha, invece,

$$\lim_{\vec{h}\to\vec{0}} \Delta f_{\vec{x}_0}(\vec{h}) = 0 \tag{5}$$

solo se f è *continua*.

Consideriamo ora una arbitraria funzione *lineare* $L_{\vec{x}_0} \colon \mathbb{R}^n \to \mathbb{R}$ (*forma* lineare). Il pedice x_0 che compare nella scrittura di questa funzione è del tutto pleonastico e lo usiamo solo perché saremo interessati a valutare la funzione su un incremento $\vec{h} = \vec{x} - \vec{x}_0$. Come è noto la funzione lineare $L_{\vec{x}_0}$ è sempre[2] del tipo

$$L_{\vec{x}_0}(\vec{h}) = L_{\vec{x}_0}(h_1, h_2, \dots, h_n) = a_1h_1 + a_2h_2 + \cdots + a_nh_n = \vec{a} \cdot \vec{h}. \tag{6}$$

Nel caso $n = 1$ la formula (6) diventa semplicemente $L_{\vec{x}_0}(\vec{h}) = L_{x_0}(h) = ah$, e la funzione L_{x_0} ha come grafico, nel piano O_{hy}, una retta per l'origine, di cui a è il coefficiente angolare.

Nel caso $n = 2$ la formula (6) diventa semplicemente $L_{\vec{x}_0}(\vec{h}) = a_1h_1 + a_2h_2$, e la funzione $L_{\vec{x}_0}$ ha come grafico, nello spazio $O_{h_1h_2y}$, un piano per l'origine. Nei casi $n \geq 3$, come è ben noto, non si può dare una rappresentazione cartesiana della funzione $L_{\vec{x}_0}(\vec{h})$: si usa parlare, estendendo la terminologia precedente, di *iperpiano* per l'origine.

Sia l'incremento $\Delta f_{\vec{x}_0}$ che la funzione $L_{\vec{x}_0}$ si annullano per $\vec{h} = \vec{0}$ ("passano" per l'origine), ovvero se poniamo

$$\omega(\vec{h}) = \Delta f_{\vec{x}_0}(\vec{h}) - L_{\vec{x}_0}(\vec{h}) \tag{7}$$

avremo che $\omega(\vec{0}) = 0$ e anzi, se la funzione f è continua in $\vec{x}_0$, avremo anche che

$$\lim_{\vec{h}\to\vec{0}} \omega(\vec{h}) = 0\,. \tag{8}$$

La funzione ω misura, al variare di $\vec{h}$, l'*errore* che si commette se si sostituisce $\Delta f_{\vec{x}_0}(\vec{h})$ con $L_{\vec{x}_0}(\vec{h})$. La formula (8) dice che, nell'ipotesi di continuità per f, tale errore è *infinitesimo* con $\vec{h}$[3]

La domanda che dà origine all'introduzione del concetto di differenziale è la seguente: è possibile scegliere la funzione $L_{\vec{x}_0}$ in modo da "ottimizzare" (in un senso che andrà opportunamente precisato) l'errore ω?

Poiché la (8) implica, come già osservato, che, almeno per funzioni continue in $\vec{x}_0$, ω è infinitesima con $\vec{h}$, una prima ottimizzazione dell'errore si ottiene richiedendo che ω sia, addirittura, infinitesima di ordine superiore rispetto a $\|\vec{h}\|$, cioè che si abbia

$$\lim_{\vec{h}\to\vec{0}} \frac{\omega(\vec{h})}{\|\vec{h}\|} = 0\,. \tag{9}$$

[2] Teorema di Riesz

[3] Si presti bene attenzione al fatto che *infinitesimo* non è sinonimo di "piccolo" (aggettivo che del resto non ha alcun significato assoluto): dire che ω è infinitesima con $\vec{h}$, significa solo dire che vale la (8).

Si noti che la frazione che compare nella (9) ha senso solo se il denominatore è un numero reale: è per questo che vi compare la norma di $\vec{h}$.

Proprio da qui nasce il concetto di differenziale: una funzione $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}$ si dice *differenziabile* in un punto $\vec{x}_0 \in A$ se esiste una forma lineare $L_{\vec{x}_0}\colon \mathbb{R}^n \to \mathbb{R}$ tale che

$$\lim_{\vec{h}\to\vec{0}} \frac{f(\vec{x}_0+\vec{h})-f(\vec{x}_0)-L_{\vec{x}_0}(\vec{h})}{\|\vec{h}\|}=0\,. \tag{10}$$

Se una tal forma lineare esiste essa si chiama *il differenziale di f nel punto x_0* e si indica con il simbolo

$$\mathrm{d}f_{\vec{x}_0}\,. \tag{11}$$

Anche se è ovviamente contenuto nella definizione che abbiamo appena dato, è opportuno evidenziare il fatto che il differenziale di una funzione $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}$ in un punto $\vec{x}_0$ è una nuova funzione (lineare però!), definita in tutto[4] $\mathbb{R}^n$ e a valori in $\mathbb{R}$. Tenendo conto della (6), potremo dire che si avrà

$$\mathrm{d}f_{\vec{x}_0}(\vec{h})=a_1h_1+a_2h_2+\cdots+a_nh_n=\vec{a}\cdot\vec{h}\,, \tag{12}$$

con opportuni valori di $(a_1,\,a_2,\,\ldots,\,a_n)$.

Il risultato fondamentale relativo al differenziale è contenuto nel seguente teorema: se una funzione f è differenziabile in un punto $\vec{x}_0$, allora è continua, ha tutte le derivate parziali in $\vec{x}_0$ e per i coefficienti a_i della (12) si ha

$$a_i=\frac{\partial f}{\partial x_i}(\vec{x}_0) \tag{13}$$

In sostanza si può affermare che l'esistenza di tutte le derivate parziali in un punto è *condizione necessaria* per la differenziabilità di una funzione nel punto.

Per questo motivo, data un funzione $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}$ e un punto x_0 in cui essa ha tutte le derivate parziali, si definisce il seguente vettore di $\mathbb{R}^n$, detto *gradiente* di f in $\vec{x}_0$,

$$\nabla f(\vec{x}_0)=\left(\frac{\partial f}{\partial x_1}(\vec{x}_0)\,,\ \frac{\partial f}{\partial x_2}(\vec{x}_0)\,,\ \ldots\,,\ \frac{\partial f}{\partial x_n}(\vec{x}_0)\right)=\frac{\partial f}{\partial x_1}(\vec{x}_0)\,\vec{e}_1+\cdots+\frac{\partial f}{\partial x_n}(\vec{x}_0)\,\vec{e}_n\,, \tag{14}$$

definizione che permette di scrivere la (12) nel seguente modo:

$$\mathrm{d}f_{\vec{x}_0}(\vec{h})=\nabla f(\vec{x}_0)\cdot\vec{h}\,, \tag{15}$$

e di concludere che, se una funzione f è differenziabile in un punto $\vec{x}_0$ si ha

$$\Delta f_{\vec{x}_0}(\vec{h})=\mathrm{d}f_{\vec{x}_0}(\vec{h})+\omega(\vec{h})=\nabla f(\vec{x}_0)\cdot\vec{h}+\omega(\vec{h})\,, \tag{16}$$

con ω dotata della proprietà (9) (cioè di essere infinitesimo di ordine superiore rispetto a $\|\vec{h}\|$).

Sostituendo $\vec{x}-\vec{x}_0$ al posto di $\vec{h}$ e tenendo conto della definizione di $\Delta f_{\vec{x}_0}$, la (16) si riscrive nel seguente modo:

$$f(\vec{x})=f(\vec{x}_0)+\mathrm{d}f(\vec{x}_0)(\vec{x}-\vec{x}_0)+\omega(\vec{x}-\vec{x}_0)=f(\vec{x}_0)+\nabla f(\vec{x}_0)\cdot(\vec{x}-\vec{x}_0)+\omega(\vec{x}-\vec{x}_0)\,. \tag{17}$$

Nel caso $n=1$ per il gradiente di f in x_0 si ha semplicemente $\nabla f(x_0)=f'(x_0)\vec{e}_1$ e la (17) si scrive

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\omega(x-x_0) \tag{18}$$

[4] Le funzioni lineari hanno sempre come dominio naturale tutto $\mathbb{R}^n$.

Nel caso $n = 2$ per il gradiente di f in x_0 si ha $\nabla f(x_0) = f'_{x_1}(x_0)\vec{e}_1 + f'_{x_2}(x_0)\vec{e}_2$ e la (17) si scrive

$$f(x_1, x_2) = f(x_{1,0}, x_{2,0}) + \frac{\partial f}{\partial x_1}(x_{1,0}, x_{2,0})(x_1 - x_{1,0}) + \frac{\partial f}{\partial x_2}(x_{1,0}, x_{2,0})(x_2 - x_{2,0}) + \omega(\vec{x} - \vec{x}_0) \tag{19}$$

Il luogo di equazione

$$\vec{y} = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) \tag{20}$$

si chiama (proprio in virtù del fatto che ω è infinitesimo di ordine superiore rispetto a $\|\vec{h}\|$), *iperpiano tangente* al grafico di f in $(\vec{x}_0, f(\vec{x}_0))$. Nel caso $n = 1$ si tratta della retta tangente alla curva-grafico della funzione, nel caso $n = 2$ del piano tangente alla superficie-grafico. Si vedano le due figure seguenti, relative a questi casi.

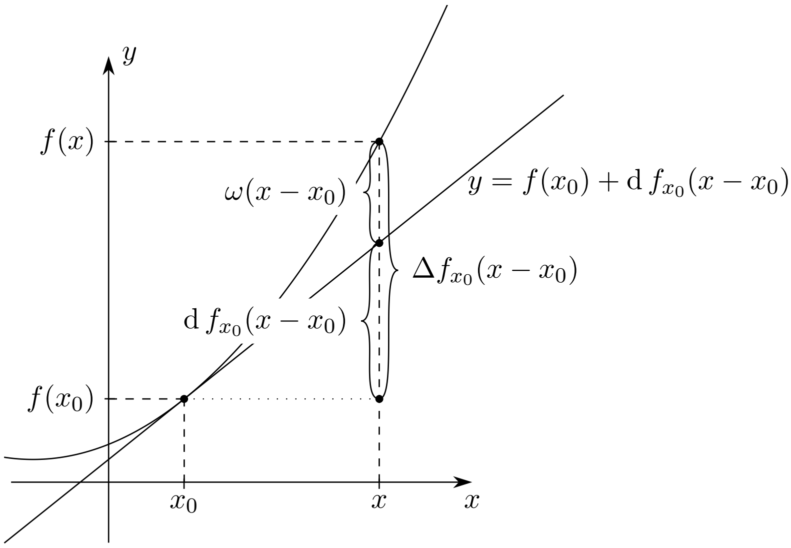

Figura 1 *Differenziale in una variabile e significato geometrico*

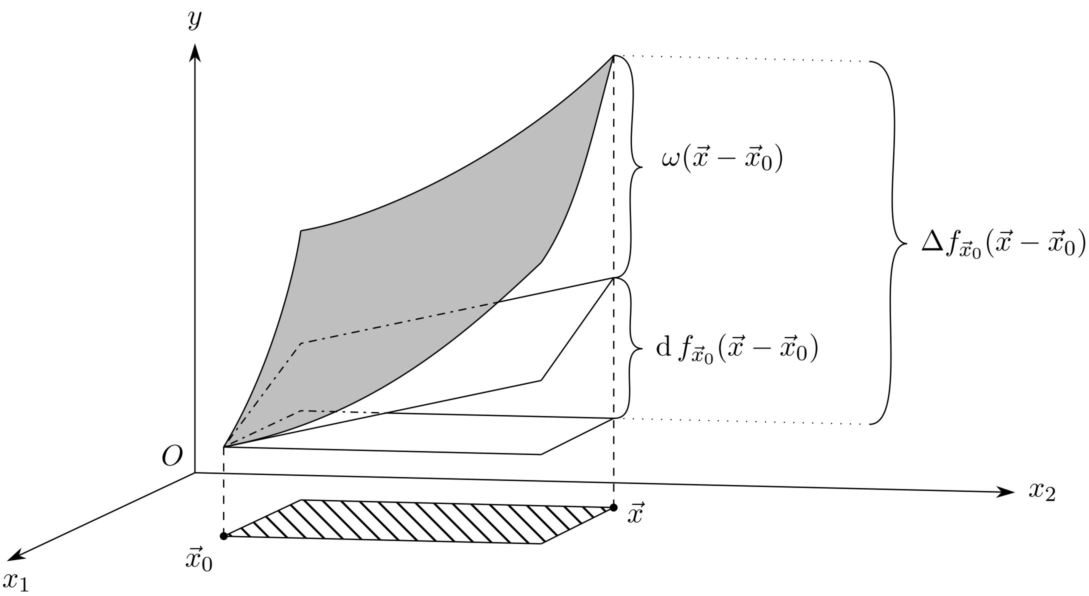

Figura 2 *Differenziale in due variabili e significato geometrico*

Come si può ben vedere dalla presentazione fatta, i casi $n = 1$ e $n > 1$ possono essere trattati in maniera assolutamente identica. Tuttavia esiste una fondamentale differenza tra i due casi, relativa alla possibilità di invertire il teorema fondamentale sopra citato. Precisamente si ha quanto segue.

Per $n = 1$ la derivabilità (rispetto all'unica variabile) è anche *condizione sufficiente* per la differenziabilità di una funzione in un punto.	Per $n \geq 2$ l'esistenza di tutte le derivate parziali *non è affatto condizione sufficiente* per la differenziabilità di una funzione in un punto (anzi, peggio ancora, non è nemmeno condizione sufficiente per la continuità).

Si dimostra poi che se una funzione è differenziabile in un punto allora è parzialmente derivabile lungo ogni direzione per quel punto[5] e si ha

$$\frac{\partial f}{\partial \vec{u}}(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \vec{u}\,, \quad \forall \vec{u},\ \text{versore di } \mathbb{R}^n. \tag{21}$$

Si tenga comunque presente che nemmeno questo risultato può essere invertito: una funzione può benissimo essere derivabile lungo ogni direzione in un punto di $\mathbb{R}^n$ ($n \geq 2$), senza essere differenziabile nel punto (anzi, ancora una volta, senza essere nemmeno continua nel punto).

3 Funzioni $\mathbb{R}^n \to \mathbb{R}^m$

Il passaggio da campi scalari a campi vettoriali (cioè a funzioni con valori in $\mathbb{R}^m$ anziché in $\mathbb{R}$) è più questione di forma che di sostanza. Basterà tenere presente quanto segue.

— Una funzione $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ può sempre essere pensata come una m-upla di funzioni $f_i\colon A \subseteq \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \dots, m$, dette *funzioni componenti*.

— Una funzione lineare $L\colon \mathbb{R}^n \to \mathbb{R}^m$ è individuata da una matrice A, con m righe e n colonne: infatti L sarà una m-upla di funzioni lineari di $\mathbb{R}^n$ in $\mathbb{R}^m$, e ciascuna di esse sarà individuata (vedi la formula (6) nella pagina 2) da una n-upla di numeri che costituiranno una delle righe della matrice A.

— Se una funzione $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ si potranno scrivere le derivate parziali di tutte le funzioni componenti e disporle in una matrice $m \times n$, detta *matrice jacobiana*, come segue

$$J_f(\vec{x}_0) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1}(\vec{x}_0) & \dots & \dfrac{\partial f_1}{\partial x_n}(\vec{x}_0) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\vec{x}_0) & \dots & \dfrac{\partial f_m}{\partial x_n}(\vec{x}_0) \end{pmatrix}, \tag{22}$$

cioè in sostanza scrivendo sulle righe i gradienti (ovviamente calcolati in $\vec{x}_0$) delle funzioni componenti di f.

Ripetendo quasi parola per parola quanto detto per campi scalari si giungerà alla conclusione che ora il differenziale è individuato dalla matrice jacobina anziché dal gradiente e si otterrà l'analogo della formula (17),

$$f(\vec{x}) = f(\vec{x}_0) + \mathrm{d}f(\vec{x}_0)(\vec{x} - \vec{x}_0) + \omega(\vec{x} - \vec{x}_0) = f(\vec{x}_0) + J_f(\vec{x}_0)(\vec{x} - \vec{x}_0)^T + \omega(\vec{x} - \vec{x}_0)\,, \tag{23}$$

dove abbiamo voluto esplicitamente evidenziare il fatto che il vettore $\vec{x} - \vec{x}_0$ è un vettore-colonna (si deve fare il prodotto righe per colonne della matrice J_f per il vettore $\vec{x} - \vec{x}_0$).

[5] Ovviamente la cosa ha interesse solo nel caso $n \geq 2$.

Il caso $n = 1$, cioè funzioni di $\mathbb{R}$ in $\mathbb{R}^m$ è particolarmente importante (si tratta delle *curve* di $\mathbb{R}^m$). In questo caso ciascuna delle funzioni f_i è funzione di una sola variabile e la matrice jacobiana sarà costituita da una sola colonna. In questo caso varrà naturalmente quanto già detto per le funzioni di $\mathbb{R}$ in $\mathbb{R}$: la derivabilità di ciascuna delle f_i in un punto sarà anche condizione *sufficiente* per la differenziabilità della funzione f.

4 Funzioni composte e differenziabilità

È estremamente importante la seguente proprietà delle funzioni differenziabili, che costituisce l'analogo della derivabilità delle funzioni composte per il caso di funzioni reali di una variabile reale:

Siano $f\colon A \subset \mathbb{R}^n \to \mathbb{R}^p$ una funzione differenziabile in $\vec{x}_0 \in A$ e $g\colon B \subset \mathbb{R}^p \to \mathbb{R}^m$, una funzione differenziabile in $\vec{y}_0 = f(\vec{x}_0) \in B$; sia inoltre $h = g \circ f\colon A \subset \mathbb{R}^n \to \mathbb{R}^m$ la funzione composta di f e g. Allora h è differenziabile in $\vec{x}_0$ e si ha

$$\mathrm{d}h_{\vec{x}_0}(\vec{x} - \vec{x}_0) = \mathrm{d}g_{\vec{y}_0}(\vec{y} - \vec{y}_0) \circ \mathrm{d}f_{\vec{x}_0}(\vec{x} - \vec{x}_0). \tag{24}$$

In termini di matrici jacobiane la formula (24) si scrive

$$J_h(\vec{x}_0) = J_g(\vec{y}_0) J_f(\vec{x}_0)\,, \tag{25}$$

con ovvio significato dei simboli.

Si tenga ben presente che, invece, la composta di due funzioni derivabili, magari lungo ogni direzione, può non essere affatto una funzione derivabile lungo qualche direzione.

5 Il differenziale secondo per campi scalari

Sia ora $f\colon A \subseteq \mathbb{R}^n \to \mathbb{R}$ un campo scalare definito in un aperto A di $\mathbb{R}?n$ e consideriamo l'insieme A' dei punti x in cui f è differenziabile. Supporremo che A' sia ancora un aperto. Risulta allora definita una funzione $g\colon A' \subseteq \mathbb{R}^n \to \mathbb{R}^n$ nel seguente modo:

$$g(\vec{x}) = \nabla f(\vec{x})\,,\ \text{ovvero}\ g_i(\vec{x}) = \frac{\partial f}{\partial x_i}(\vec{x})\,, \quad i = 1, 2, \dots, n. \tag{26}$$

Se g (attenzione: g è un campo vettoriale!) risulta differenziabile in $\vec{x}_0$, allora si dice che f è *due volte differenziabile* in $\vec{x}_0$ e la matrice jacobiana di g si chiama *matrice hessiana* di f. Essa è ovviamente costituita dalle derivate seconde di f[6]:

$$H_f(\vec{x}_0) = J_g(\vec{x}_0) = \begin{pmatrix} \dfrac{\partial g_1}{\partial x_1} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial g_n}{\partial x_1} & \cdots & \dfrac{\partial g_n}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{pmatrix}. \tag{27}$$

La differenziabilità di f implica che la matrice Hessiana è simmetrica, come si potrebbe dimostrare[7].

Se una funzione f è due volte differnziabile in un punto, costruita la matrice Hessiana, si può costruire una funzione[8] di $\mathbb{R}^n$ in $\mathbb{R}$ nel seguente modo:

$$\forall \vec{u} \in \mathbb{R}^n, \quad \mathrm{d}^2 f_{\vec{x}_0}(\vec{u}) = H_f(\vec{x}_0)\vec{u} \cdot \vec{u}. \tag{28}$$

[6] Tralasciamo l'indicazione esplicita che le derivate sono calcolate in $\vec{x}_0$.

[7] Si tratta di una proprietà un po' più generale di quanto previsto dal Teorema di Schwartz, dove si richiede che le derivate seconde miste siano continue per poter garantire che siano uguali.

[8] Si tenga conto che $H_f(\vec{x}_0)\vec{u}$ è un vettore di $\mathbb{R}^n$.

Ricordiamo ora la formula di Taylor, arrestata al secondo ordine, per funzioni di $\mathbb{R}$ in $\mathbb{R}$:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \omega(x - x_0)\,, \tag{29}$$

dove ω è infinitesima, per $x \to x_0$, di ordine superiore al secondo rispetto a $x - x_0$. Riscriviamo la formula in una forma adatta ad essere estesa a campi scalari definiti in $\mathbb{R}^n$ (cosa che abbiamo già fatto per il primo ordine dello sviluppo). In questo caso osserviamo preliminarmente che la matrice Hessiana di f è una matrice a una sola riga e una sola colonna: $H_f(x_0) = \big(f''(x_0)\big)$.

$$f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + \frac{1}{2} H_f(x_0)(x - x_0) \cdot (x - x_0) + \omega(x - x_0)\,. \tag{30}$$

Per passare al caso di campi scalari definiti in $\mathbb{R}^n$ basterà “mettere le freccette” sopra alle x:

$$f(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + \frac{1}{2} H_f(\vec{x}_0)(\vec{x} - \vec{x}_0) \cdot (\vec{x} - \vec{x}_0) + \omega(\vec{x} - \vec{x}_0)\,. \tag{31}$$

Come si vede, anche in questo caso un opportuno uso delle notazioni può facilitare di molto la scrittura delle formule (e soprattutto la comprensione del loro significato!).