# CAPITOLO 7
# Calcolo differenziale 2. Funzioni di più variabili

In questo capitolo estendiamo il calcolo differenziale a funzioni reali di $n$ variabili reali (Sez. 1) e a funzioni vettoriali di $n$ variabili reali (Sez. 2).

Sia $\mathbf{f} : \mathbb{R}^n \supseteq A \to \mathbb{R}^m$; l'obiettivo è quello di studiare le variazioni di $\mathbf{f}$ relativamente a variazioni del suo argomento nell'intorno di un punto interno ad $A$. Per questa ragione, $A$ sarà, salvo avviso contrario, quasi sempre un aperto di $\mathbb{R}^n$.

Nel Paragrafo 2.4 è presentato l'importante teorema di inversione locale, generalizzazione al caso di trasformazioni non lineari del teorema di Cramer.

La Sezione 3 è dedicata alle funzioni implicite. Il principale risultato (Teorema 3.8) è l'analogo non lineare del teorema di Rouché-Capelli.

Il teorema di inversione locale e il Teorema 3.8 giocano un ruolo essenziale in geometria differenziale e nella teoria delle equazioni differenziali, argomenti che saranno trattati nel secondo volume.

## 1. FUNZIONI DA $\mathbb{R}^n$ IN $\mathbb{R}$

### 1.1 Derivate direzionali e derivate parziali

Siano $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, $A$ aperto di $\mathbb{R}^n$, $\mathbf{x} \in A$.

Se $n = 1$, la derivata prima di $f$ assegna il tasso di incremento di $f$ relativamente a $x$. Essendoci una sola dimensione, è unica la direzione lungo la quale $x$ viene incrementato. In dimensione $n > 1$ il modo in cui $f$ varia dipende dalla direzione nella quale avviene la variazione di $\mathbf{x}$. Ciò conduce in modo naturale al concetto di derivata in una data direzione o derivata direzionale, che ora descriviamo.

Introduciamo un versore $\mathbf{v} \in \mathbb{R}^n$ ($|\mathbf{v}| = 1$) che assegna una direzione e un verso in $\mathbb{R}^n$ e, per ogni $t \in \mathbb{R}$ tale che $\mathbf{x} + t\mathbf{v} \in A$, consideriamo il rapporto

$$\frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}, \tag{1.1}$$

detto *rapporto incrementale di $f$ nella direzione* $\mathbf{v}$, che rappresenta il tasso di variazione medio di $f$ in quella direzione e verso.

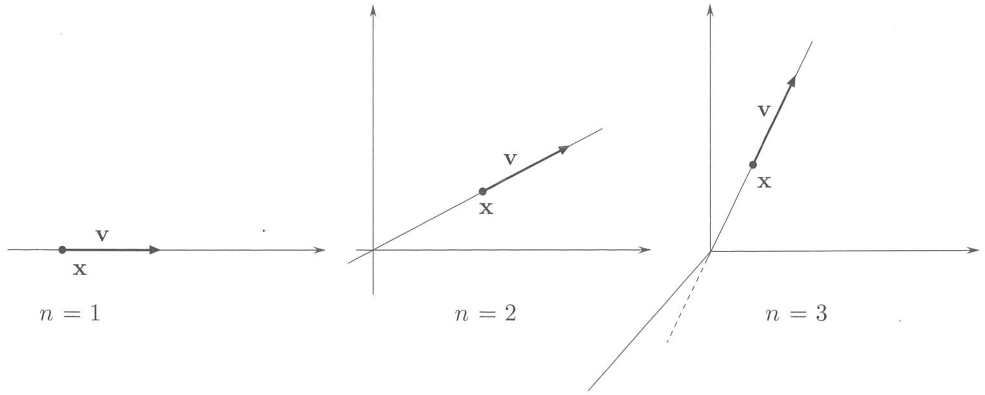


**Figura 7.1.** Incremento di $\mathbf{x}$ nella direzione $\mathbf{v}$.

**Definizione 1.1** *Quando esiste finito, il*

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

*si chiama* ***derivata nella direzione*** $\mathbf{v}$ *di $f$ nel punto $x$ e si indica con $D_{\mathbf{v}}f(\mathbf{x})$.*
*In questo caso $f$ si dice* derivabile nella direzione $\mathbf{v}$ *in* $\mathbf{x}$.

La Definizione 1.1 è essenzialmente uni-dimensionale; ciò si può mettere in luce osservando che, essendo $\mathbf{x}$ e $\mathbf{v}$ fissati, posto $\varphi(t) = f(\mathbf{x} + t\mathbf{v})$, $\varphi$ è una funzione della variabile reale $t$, definita in un intorno di $t = 0$, e si ha:

$$D_{\mathbf{v}}f(\mathbf{x}) = \varphi'(0).$$

**Esempio 1.1** Si voglia calcolare la derivata lungo una direzione $\mathbf{v} \in \mathbb{R}^3$ della funzione $f : \mathbf{x} \mapsto |\mathbf{x}|^2 : \mathbb{R}^3 \to \mathbb{R}$, nel punto generico $\mathbf{x}$.
Posto $\varphi(t) = |\mathbf{x} + t\mathbf{v}|^2$ si ha:

$$\varphi'(t) = \frac{d}{dt} \sum_{j=1}^{3} (x_j + tv_j)^2 = 2 \sum_{j=1}^{3} (x_j + tv_j) v_j = 2\langle \mathbf{x} + t\mathbf{v}, \mathbf{v} \rangle.$$

Dunque $D_{\mathbf{v}}f(\mathbf{x}) = \varphi'(0) = 2\langle \mathbf{x}, \mathbf{v} \rangle$ e $f$ è derivabile lungo qualunque direzione in qualunque punto.

Nella definizione di $D_{\mathbf{v}}f(\mathbf{x})$ intervengono solo i valori di $f$ lungo la retta $\mathbf{x}+t\mathbf{v}$ (ovvero la restrizione di $f$ lungo tale retta), cosicché l'esistenza della derivata in una direzione non dà informazioni circa l'esistenza della derivata in un'altra direzione.
Per esempio la funzione

$$f(x_1, x_2) = \frac{x_1^2}{x_1^2 + x_2^2} \quad \text{se} \quad (x_1, x_2) \neq (0, 0)$$

e $f(0,0) = 0$, ha $D_{\mathbf{e}^2} f(0,0) = 0$ mentre $D_{\mathbf{e}^1} f(0,0)$ non esiste, come si può facilmente verificare.



Particolare importanza rivestono le derivate lungo le direzioni degli assi coordinati, individuate dai versori $\mathbf{e}^1, \mathbf{e}^2, \ldots, \mathbf{e}^n$.
In questo caso $D_{\mathbf{e}^j} f(\mathbf{x})$ prende il nome di **derivata parziale rispetto a** $x_j$, di $f$ nel punto $\mathbf{x}$.
Per le derivate parziali si usano le seguenti notazioni:

$$\frac{\partial f}{\partial x_j}, \quad D_j f, \quad D_{x_j} f, \quad \partial_{x_j} f, \quad f_{x_j}.$$

In una derivata parziale, quella rispetto a $x_j$, per esempio, viene incrementata la sola variabile $x_j$; ne segue che per il calcolo di $f_{x_j}$ si può pensare alle altre variabili come costanti e utilizzare le regole di derivazione note per le funzioni di una variabile.
Per esempio, se

$$f(x_1, x_2, x_3) = x_1 \sqrt{x_2^2 + x_3^2}$$

si ha:

$$f_{x_1}(x_1, x_2, x_3) = \sqrt{x_2^2 + x_3^2}$$
$$f_{x_2}(x_1, x_2, x_3) = \frac{x_1 x_2}{\sqrt{x_2^2 + x_3^2}}$$
$$f_{x_3}(x_1, x_2, x_3) = \frac{x_1 x_3}{\sqrt{x_2^2 + x_3^2}}.$$

*Se una funzione $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$ ammette $n$ derivate parziali in un punto $\mathbf{x} \in A$, è definito un vettore che si chiama* ***gradiente*** *di $f$ in $\mathbf{x}$, le cui componenti sono le $n$ derivate parziali di $f$ in $\mathbf{x}$ e che si indica con i simboli $\nabla f(\mathbf{x})$ o* grad $f(\mathbf{x})$*:*

$$\nabla f(\mathbf{x}) := (f_{x_1}(\mathbf{x}), f_{x_2}(\mathbf{x}), \ldots, f_{x_n}(\mathbf{x})).$$

Il significato geometrico di derivata direzionale nel caso $n = 2$ è illustrato in Figura 7.2 dove il piano $\pi$ è perpendicolare al piano $x_1$, $x_2$ e lo interseca nella retta $\mathbf{x}+t\mathbf{v}$. Si ha che $D_{\mathbf{v}}f(\mathbf{x}) = \operatorname{tg}\theta =$ pendenza della sezione del piano $\pi$ col grafico di $f$, nel punto di coordinate $(\mathbf{x}, f(\mathbf{x}))$.
Si possono definire derivate destre o sinistre lungo una data direzione individuata da un versore $\mathbf{v}$ in modo ovvio:

$$D_{\mathbf{v}}^{\pm} f(\mathbf{x}) := \lim_{t \to 0_{\pm}} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}.$$

quando i limiti esistano finiti.
In molte questioni di ottimizzazione si ha a che fare con funzioni $f : \overline{A} \to \mathbb{R}$, dove $\overline{A}$ è la chiusura di un aperto $A \subseteq \mathbb{R}^n$, e occorre calcolare derivate direzionali in punti $\mathbf{x} \in \partial A$.
Queste derivate sono necessariamente solo destre o sinistre, inoltre è evidente che, diversamente dal caso dei punti interni, avremo solo determinate direzioni (o versori) ammissibili, dipendenti dal punto $\mathbf{x}$.

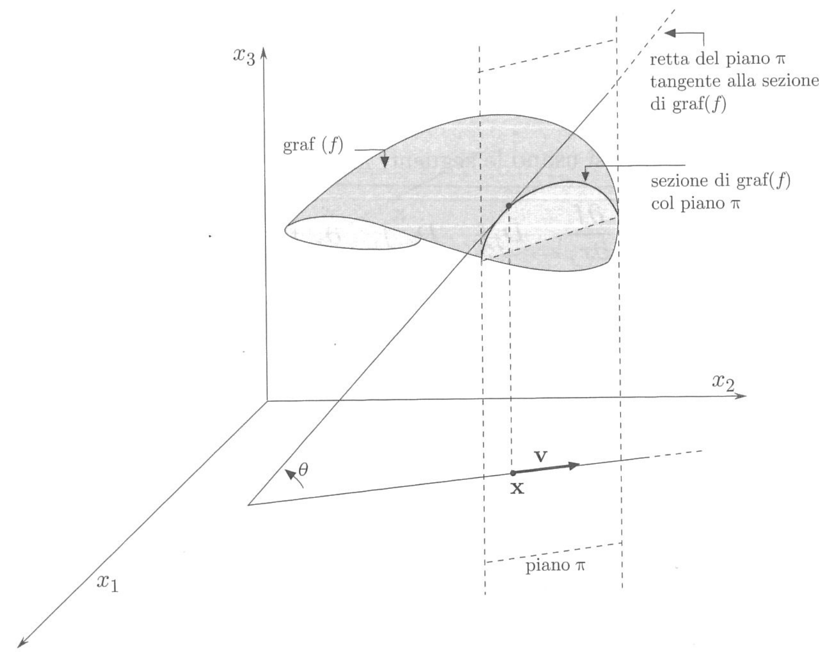


**Figura 7.2.** Significato geometrico di derivata direzionale: $D_{\mathbf{v}} f(\mathbf{x}) = \operatorname{tg}\theta$.

Infatti per calcolare $D^+_{\mathbf{v}} f(\mathbf{x})$ (risp. $D^-_{\mathbf{v}} f(\mathbf{x})$) per $\mathbf{x} \in \partial A$, occorre che, per $t$ in un intorno destro (risp. sinistro) di zero, si abbia $\mathbf{x} + t\mathbf{v} \in A$. Nei casi più comuni i versori $\mathbf{v} \in \mathbb{R}^n$ ammissibili costituiscono nel loro insieme un *cono* con vertice in $\mathbf{x}$.

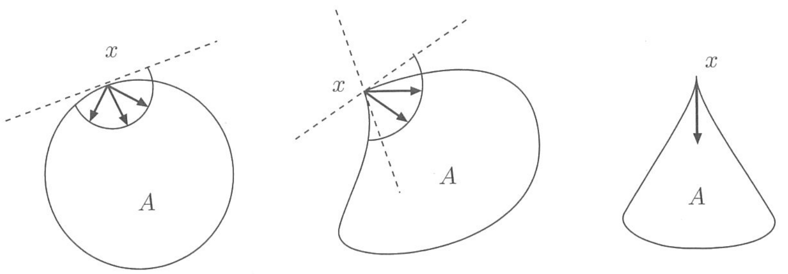


**Figura 7.3.** Esempi di direzioni ammissibili in vari casi.

Concludiamo il paragrafo illustrando la relazione tra derivabilità e continuità.

Evidentemente se $D_{\mathbf{v}} f(\mathbf{x})$ esiste, allora $f$ è continua lungo la direzione $\mathbf{v}$ in $\mathbf{x}$.

Il seguente esempio mostra che l'esistenza di *tutte* le derivate direzionali in un punto *non implica* la continuità in quel punto, in dimensione maggiore di 1.

**Esempio 1.2** Sia

$$f(x_1, x_2) = \begin{cases} x_1 e^{x_1/x_2} & \text{se } x_2 \neq 0 \\ 0 & \text{se } x_2 = 0. \end{cases}$$



Poiché siamo in $\mathbb{R}^2$, una qualunque direzione è individuata dal versore $\mathbf{v}_\theta$ di componenti $(\cos\theta, \sin\theta)$ con $\theta \in [0, 2\pi)$.

Proviamo che $f$ è derivabile lungo qualunque direzione in $(x_1, x_2) = (0, 0)$. Si ha

$$D_{\mathbf{v}_\theta} f(0,0) = \lim_{t\to 0} \frac{f(t\cos\theta, t\sin\theta)}{t} = \lim_{t\to 0} \frac{t\cos\theta e^{\operatorname{cotg}\theta}}{t} = \cos\theta e^{\operatorname{cotg}\theta} \quad (\theta \neq 0)$$
$$D_{\mathbf{v}_0} f(0,0) = f_{x_1}(0,0) = 0.$$

D'altra parte $f$ *non* è continua in $(0,0)$ poiché il limite per $(x_1, x_2) \to (0,0)$ per esempio lungo la cubica di equazione $x_2 = x_1^3$ è $\infty$ (infatti $\lim\limits_{x_1 \to 0^\pm} x_1 e^{1/x_1^2} = \pm\infty$).

L'Esempio 1.2 indica che, se $n > 1$, per studiare proprietà relative a un intorno $n$-dimensionale di un punto occorre un concetto più potente di quello di derivabilità.

Tale concetto è quello di differenziabilità (oggetto del prossimo paragrafo) che, dunque, in dimensione maggiore di 1, non risulta più equivalente a quello di derivabilità.

## 1.2 Differenziale

L'idea di differenziabilità è essenzialmente quella di poter approssimare l'incremento $\Delta f = f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})$ con una funzione (reale) *lineare* in $\mathbf{h}$ a meno di infinitesimi di ordine superiore a $|\mathbf{h}|$.

Ricordiamo che ogni funzione lineare $l : \mathbb{R}^n \to \mathbb{R}$ è "rappresentata" da un unico vettore $\mathbf{a} \in \mathbb{R}^n$, nel senso che[1]

$$l(\mathbf{h}) = \langle \mathbf{a}, \mathbf{h} \rangle \qquad \forall\, \mathbf{h} \in \mathbb{R}^n.$$

Siamo così condotti alla seguente definizione.

**Definizione 1.2** *Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, $A$ aperto; $f$ si dice differenziabile in $\mathbf{x} \in A$ se esiste un vettore $\mathbf{a} \in \mathbb{R}^n$ tale che*

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{h} \rangle + o(|\mathbf{h}|) \quad \textit{per } |\mathbf{h}| \to 0 \tag{1.2}$$

*per ogni $\mathbf{h} \in \mathbb{R}^n$ con $\mathbf{x} + \mathbf{h} \in A$.*

*L'applicazione lineare da $\mathbb{R}^n$ in $\mathbb{R}$ data da*

$$\mathbf{h} \mapsto \langle \mathbf{a}, \mathbf{h} \rangle$$

*prende il nome di **differenziale** di $f$ in $\mathbf{x}$ e viene indicata col simbolo $df(\mathbf{x})$.*

Se $f$ è differenziabile in ogni punto di $A$, diremo semplicemente $f$ differenziabile in $A$.

Nel caso $n = 1$ il differenziale si riduce alla funzione $h \mapsto ah$, con $a$, $h \in \mathbb{R}$ e in questo caso abbiamo visto che $a = f'(x)$. Possiamo "identificare" il vettore $\mathbf{a}$ che compare nella (1.2) anche nel caso pluridimensionale?

La risposta è contenuta nel prossimo teorema insieme ad altre importanti conseguenze della differenziabilità.

[1] $\mathbb{R}^n$ si intende riferito alla base canonica.

■ **TEOREMA 1.1** *Sia* $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, *$A$ aperto; se $f$ è differenziabile in* $\mathbf{x} \in A$ *allora*

i) *$f$ è continua in* $\mathbf{x}$;

ii) *$f$ è derivabile in $\mathbf{x}$ lungo ogni direzione; in particolare esistono tutte le derivate parziali di $f$ in $\mathbf{x}$ e, se $\mathbf{a}$ è il vettore in* (1.2), *si ha* $\mathbf{a} = \nabla f(\mathbf{x})$. *Inoltre vale la formula*

$$D_{\mathbf{v}} f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{v} \rangle. \tag{1.3}$$

*Dimostrazione.* i) Passando al limite per $\mathbf{h} \to \mathbf{0}$ nella (1.2) si ottiene $f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) \to \mathbf{0}$ ovvero $f(\mathbf{x}+\mathbf{h}) \to f(\mathbf{x})$.
ii) Identifichiamo prima il vettore $\mathbf{a} = (a_1, a_2, \ldots, a_n)$; scegliendo nella (1.2) $\mathbf{h} = t\mathbf{e}^j$ si ha:

$$f(\mathbf{x} + t\mathbf{e}^j) - f(\mathbf{x}) = \langle \mathbf{a}, t\mathbf{e}^j \rangle + o(|t\mathbf{e}^j|) = t\langle \mathbf{a}, \mathbf{e}^j \rangle + o(t). \tag{1.4}$$

Dividendo entrambi i membri della (1.4) per $t$ e passando al limite per $t \to 0$ si deduce che

$$f_{x_j}(\mathbf{x}) = a_j.$$

Dunque esistono tutte le derivate parziali in $\mathbf{x}$ e $\mathbf{a} = \nabla f(\mathbf{x})$.
Scegliendo ora nella (1.2) $\mathbf{h} = t\mathbf{v}$ dove $\mathbf{v}$ è un generico versore in $\mathbb{R}^n$ si ha:

$$f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x}) = \langle \mathbf{a}, t\mathbf{v} \rangle + o(|t\mathbf{v}|) = t\langle \mathbf{a}, \mathbf{v} \rangle + o(t). \tag{1.5}$$

Dividendo entrambi i membri di (1.5) per $t$ e passando al limite per $t \to 0$ si deduce che

$$D_{\mathbf{v}} f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{v} \rangle = \langle \nabla f(\mathbf{x}), \mathbf{v} \rangle. \qquad \square$$

Possiamo dunque scrivere

$$df(\mathbf{x})(\mathbf{h}) = \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle \quad \text{per ogni } \mathbf{h} \in \mathbb{R}^n \tag{1.6}$$

Osserviamo poi che, se $g(\mathbf{x}) = x_j$, abbiamo $dg(\mathbf{x})(\mathbf{h}) = \langle \mathbf{e}_j, \mathbf{h} \rangle$; o, in altri termini, $dx_j : \mathbf{h} \to h_j$. Dalla (1.6) deduciamo l'uguaglianza (tra funzioni lineari)

$$df(\mathbf{x}) = \sum_{j=1}^{n} f_{x_j}(\mathbf{x}) dx_j.$$

**OSSERVAZIONE 1.1** Spesso nelle applicazioni del calcolo differenziale, il differenziale $dx_j$ è interpretato come incremento "infinitesimo" della variabile $x_j$. In tal caso, si dovrebbe scrivere

$$df(\mathbf{x})(d\mathbf{x}) = (\nabla f(\mathbf{x}), d\mathbf{x}),$$

ma si continua a scrivere $df(\mathbf{x}) = \langle \nabla f(\mathbf{x}), d\mathbf{x} \rangle$, con leggero abuso di notazione, che anche noi adotteremo più avanti.

La formula (1.3) permette di individuare le *direzioni di massima e minima crescita* di una funzione differenziabile. Infatti si può scrivere

$$D_{\mathbf{v}} f(\mathbf{x}) = |\nabla f(\mathbf{x})| \cos \beta \tag{1.7}$$

dove $\beta$ è l'angolo formato dai vettori $\mathbf{v}$ e $\nabla f(\mathbf{x})$.



La (1.7) indica che $D_{\mathbf{v}} f(\mathbf{x})$ è massima quando $\beta = 0$ e quindi $\mathbf{v}_{\max} = \frac{\nabla f(\mathbf{x})}{|\nabla f(\mathbf{x})|}$ ed è minima quando $\beta = \pi$ e quindi $\mathbf{v}_{\min} = -\frac{\nabla f(\mathbf{x})}{|\nabla f(\mathbf{x})|}$.
In conclusione

$$\max_{|\mathbf{v}|=1} D_{\mathbf{v}} f(\mathbf{x}) = |\nabla f(\mathbf{x})| \quad \text{e} \quad \min_{|\mathbf{v}|=1} D_{\mathbf{v}} f(\mathbf{x}) = -|\nabla f(\mathbf{x})|.$$

L'aspetto geometrico della differenziabilità è legato all'esistenza del piano (iperpiano se $n > 2$) tangente.
Sia $f$ differenziabile in un punto $\mathbf{x}^0$; ponendo $\mathbf{h} = \mathbf{x} - \mathbf{x}^0$ scriviamo la (1.2) nella forma seguente:

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \langle \nabla f(\mathbf{x}^0), \mathbf{x} - \mathbf{x}^0 \rangle + o(|\mathbf{x} - \mathbf{x}^0|). \tag{1.8}$$

La funzione $z = f(\mathbf{x}^0) + \langle \nabla f(\mathbf{x}^0), \mathbf{x} - \mathbf{x}^0 \rangle$, ha come grafico un iperpiano e la (1.8) equivale ad affermare che essa è la funzione lineare (o meglio, affine) che meglio approssima $f$ in un intorno di $\mathbf{x}^0$. Tale piano si chiama *piano tangente*.
In dimensione 2, se $\mathbf{x}^0 = (x_0, y_0)$, $\mathbf{x} = (x, y)$, $z_0 = f(x_0, y_0)$, la sua equazione si scrive esplicitamente così:

$$z - z_0 - f_x(x_0, y_0)(x - x_0) - f_y(x_0, y_0)(y - y_0) = 0. \tag{1.9}$$

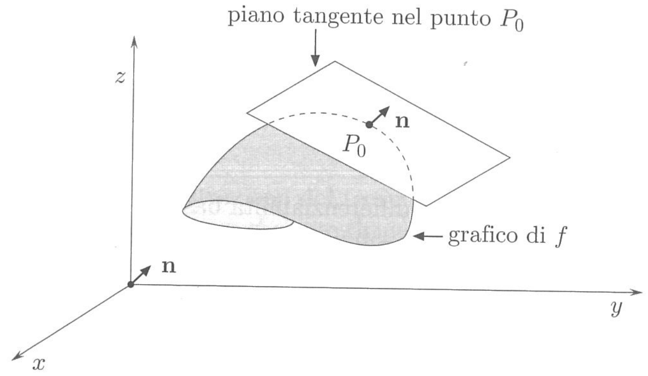


**Figura 7.4.** Piano tangente a una superficie.

La (1.9) indica che il vettore $\mathbf{n} = (-f_x(x_0, y_0), -f_y(x_0, y_0), 1) \in \mathbb{R}^3$ è un vettore normale al piano tangente nel punto $P_0$ di coordinate $(x_0, y_0, z_0)$, dunque, per definizione, normale al grafico di $f$ nello stesso punto.

**Esempio 1.3** Sia

$$f(x, y) = \begin{cases} |y|^\alpha e^{-x^2/y^2} & y \neq 0 \\ 0 & y = 0. \end{cases}$$

Esaminiamo le proprietà di $f$ in $(0, 0)$ al variare di $\alpha$.
Se $\alpha \leq 0$, $f(x, y) \not\to 0$ lungo la direzione $y = x$; dunque $f$ non è continua in $(0, 0)$.

Sia $\alpha > 0$. In questo caso, essendo $0 < e^{-x^2/y^2} \leq 1$, abbiamo

$$0 \leq f(x,y) \leq |y|^\alpha$$

e quindi $f(x,y) \to 0$ per $(x,y) \to (0,0)$; cioè $f$ è continua in $(0,0)$.

Sia $\mathbf{v} \in \mathbb{R}^2$, $|\mathbf{v}| = 1$; possiamo scrivere $\mathbf{v}_\theta = (\cos\theta, \sin\theta)$, $\theta \in [0, 2\pi)$.
Calcoliamo le derivate direzionali in $(0,0)$:

$$\frac{f(t\cos\theta, t\sin\theta)}{t} = \frac{|t\sin\theta|^\alpha e^{-(\operatorname{cotg}\theta)^2}}{t}, \quad \text{per } \theta \neq 0, \pi,$$

$$\frac{f(t,0)}{t} = 0, \quad \text{per } \theta = 0, \pi.$$

Dunque per $\theta = 0$, $\pi$ otteniamo $D_{\mathbf{v}_\theta} f(0,0) = \pm f_x(0,0) = 0$.
Per $\theta \neq 0$, $\pi$, abbiamo:

- se $0 < \alpha \leq 1$, $D_{\mathbf{v}_\theta} f(0,0)$ non esiste;
- se $\alpha > 1$, $D_{\mathbf{v}_\theta} f(0,0) = 0$.

Sia ora $\alpha > 1$; cerchiamo i valori di $\alpha$ per i quali $f$ è differenziabile in $(0,0)$.
Poiché $\nabla f(0,0) = \mathbf{0}$, si ha

$$f(x,y) - f(0,0) - f_x(0,0)x - f_y(0,0)y = |y|^\alpha e^{-x^2/y^2}.$$

Poiché

$$0 < \frac{|y|^\alpha e^{-x^2/y^2}}{\sqrt{x^2+y^2}} = \frac{|y|}{\sqrt{x^2+y^2}} |y|^{\alpha-1} e^{-x^2/y^2} \leq |y|^{\alpha-1},$$

essendo $\alpha > 1$, si ha che $\lim_{(x,y)\to(0,0)} |y|^\alpha e^{-x^2/y^2} = 0$, dunque $f$ è differenziabile in $(0,0)$ e $df(0,0) = 0$.[2]

Una condizione sufficiente per la differenziabilità basata sulla conoscenza delle sole derivate parziali è contenuta nel prossimo teorema.

■ **TEOREMA 1.2** *Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, $A$ aperto; se in un intorno di $\mathbf{x} \in A$ esistono tutte le derivate parziali di $f$ e sono continue in $\mathbf{x}$ allora $f$ è differenziabile in $\mathbf{x}$.*

*Dimostrazione.* Consideriamo il caso $n = 2$; senza difficoltà si può estendere la dimostrazione al caso generale.
Siano $\mathbf{x} = (x_1, x_2)$ e $\mathbf{h} = (h_1, h_2)$; scriviamo:

$$\begin{aligned} f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) &= f(x_1+h_1, x_2+h_2) - f(x_1,x_2) = \\ &= f(x_1+h_1, x_2+h_2) - f(x_1+h_1, x_2) + f(x_1+h_1, x_2) - f(x_1,x_2). \end{aligned} \tag{1.10}$$

Applichiamo il teorema del valor medio ai primi due termini dell'ultima somma, osservando che la sola variabile incrementata è $x_2$, si trova:

$$f(x_1+h_1, x_2+h_2) - f(x_1+h_1, x_2) = f_{x_2}(x_1+h_1, x_2+\theta_2 h_2)h_2 \tag{1.11}$$

dove $\theta_2 \in (0,1)$ dipende da $x_1$, $x_2$ e $h_1, h_2$.

[2] Osserviamo che $df(0,0) = 0$ è equivalente a dire che $df(0,0)$ è l'applicazione lineare nulla da $\mathbb{R}^n$ in $\mathbb{R}$ ovvero che $\nabla f(0,0) = \mathbf{0}$.



Se usiamo ora la continuità di $f_{x_2}$ in $x_1$, $x_2$ abbiamo

$$f_{x_2}(x_1+h_1, x_2+\theta_2 h_2) \to f_{x_2}(x_1,x_2) \quad \text{se} \quad (h_1,h_2) \to (0,0);$$

ovvero

$$f_{x_2}(x_1+h_1, x_2+\theta_2 h_2) = f_{x_2}(x_1,x_2) + \varepsilon(h_1,h_2),$$

dove $\varepsilon(h_1,h_2) \to 0$ se $(h_1,h_2) \to (0,0)$.
Si può dunque scrivere la (1.11) nella forma

$$f(x_1+h_1, x_2+h_2) - f(x_1+h_1, x_2) = f_{x_2}(x_1,x_2)h_2 + h_2\varepsilon(h_1,h_2). \tag{1.12}$$

Per gli ultimi due termini della (1.10), usando la definizione di $f_{x_1}(x_1,x_2)$, possiamo scrivere

$$f(x_1+h_1, x_2) - f(x_1,x_2) = f_{x_1}(x_1,x_2)h_1 + h_1\eta(h_1) \tag{1.13}$$

dove $\eta(h_1) \to 0$ se $h_1 \to 0$.
Sostituiamo ora (1.12) e (1.13) nella (1.10); si ottiene:

$$f(x_1+h_1, x_2+h_2) - f(x_1,x_2) = f_{x_1}(x_1,x_2)h_1 + f_{x_2}(x_1,x_2)h_2 + h_1\eta(h_1) + h_2\varepsilon(h_1,h_2).$$

Per concludere che $f$ è differenziabile in $(x_1,x_2)$ occorre mostrare che

$$h_1\eta(h_1) + h_2\varepsilon(h_1,h_2) = o\left(\sqrt{h_1^2+h_2^2}\right), \quad \text{per} \quad (h_1,h_2) \to (0,0).$$

Si ha:

$$\begin{aligned} 0 \leq \frac{|h_1\eta(h_1) + h_2\varepsilon(h_1,h_2)|}{\sqrt{h_1^2+h_2^2}} &\leq \frac{|h_1|}{\sqrt{h_1^2+h_2^2}}|\eta(h_1)| + \frac{|h_2|}{\sqrt{h_1^2+h_2^2}}|\varepsilon(h_1,h_2)| \leq \\ &\leq |\eta(h_1)| + |\varepsilon(h_1,h_2)|. \end{aligned}$$

Poiché $\eta(h_1) \to 0$ se $h_1 \to 0$ ed $\varepsilon(h_1,h_2) \to 0$ se $(h_1,h_2) \to (0,0)$ la dimostrazione è conclusa. □

**OSSERVAZIONE 1.2** Nella dimostrazione del teorema si è usata la continuità della sola derivata $f_{x_2}$. Nel caso generale è sufficiente richiedere la continuità di tutte le derivate parziali di $f$ tranne una.

**OSSERVAZIONE 1.3** La condizione espressa nel Teorema 1.2 è solo sufficiente per la differenziabilità. Già nel caso $n = 1$ si vede che esistono funzioni differenziabili con derivata *non* continua. Un esempio è il seguente:

$$f(x) = x^2 \sin\frac{1}{x} \quad \text{se} \quad x \neq 0, \quad f(0) = 0.$$

Dunque la classe delle funzioni differenziabili in un aperto $A$ contiene strettamente quella delle funzioni con derivate parziali continue in $A$ (o, come si usa dire, delle funzioni *differenziabili con continuità*). Quest'ultima classe di funzioni si indica con il simbolo $C^1(A)$. Evidentemente $C^1(A)$ è strutturato come spazio vettoriale su $\mathbb{R}$, rispetto alla somma usuale di funzioni e al prodotto di una funzione per un numero reale.
Il Teorema 1.2 può essere enunciato nella forma seguente:

*se $f \in C^1(A)$, allora $f$ è differenziabile in $A$.*

L'operazione di differenziazione si comporta come nel caso unidimensionale rispetto alle operazioni aritmetiche; valgono infatti le seguenti semplici proprietà:

- siano $f, g : \mathbb{R}^n \supseteq A \to \mathbb{R}$, differenziabili in $\mathbf{x} \in A$; allora: $f+g, fg, f/g$ (se $g(\mathbf{x}) \neq 0$) sono differenziabili e valgono le formule

$$\begin{aligned} d(f+g)(\mathbf{x}) &= df(\mathbf{x}) + dg(\mathbf{x}) \\ d(fg)(\mathbf{x}) &= df(\mathbf{x})g(\mathbf{x}) + f(\mathbf{x})dg(\mathbf{x}) \\ d(f/g)(\mathbf{x}) &= \frac{g(\mathbf{x})df(\mathbf{x}) - f(\mathbf{x})dg(\mathbf{x})}{(g(\mathbf{x}))^2}. \end{aligned}$$

Segnaliamo anche l'importanza pratica del differenziale come conveniente approssimazione per l'incremento di una funzione in genere non lineare. Per esempio, nel calcolo degli errori, se una quantità $q$ è funzione di altre quantità $m_1, m_2, \ldots, m_k$ la cui misura è affetta da un errore $dm_1, dm_2, \ldots, dm_k$, il valore di $q$ risulterà affetto da un errore con buona approssimazione dato da

$$dq = \frac{\partial q}{\partial m_1}dm_1 + \frac{\partial q}{\partial m_2}dm_2 + \ldots + \frac{\partial q}{\partial m_k}dm_k.$$

Anche rispetto alla composizione di funzioni il differenziale si comporta come nel caso unidimensionale; nel caso generale però occorre far intervenire funzioni a valori vettoriali. Per questa ragione tratteremo la differenziazione delle funzioni composte nella Sezione 2.

### 1.3 Derivate e differenziali di ordine superiore

Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, $A$ aperto. Supponiamo che, fissato il versore $\mathbf{v} \in \mathbb{R}^n$, $D_{\mathbf{v}}f$ esista in un intorno di un punto $\mathbf{x} \in A$, che denotiamo con $U(\mathbf{x})$.

Allora è definita la funzione

$$D_{\mathbf{v}}f : U(\mathbf{x}) \to \mathbb{R}.$$

Se ora $\mathbf{w} \in \mathbb{R}^n$ è un altro versore, ci si può chiedere se esiste $D_{\mathbf{w}}D_{\mathbf{v}}f(\mathbf{x})$, che prende il nome di *derivata seconda* di $f$ nelle direzioni $\mathbf{v}$ e $\mathbf{w}$ (nell'ordine) e si indica con $D^2_{\mathbf{w}\mathbf{v}}f(\mathbf{x})$.

Nel caso particolare in cui $\mathbf{v} = \mathbf{e}^j$ e $\mathbf{w} = \mathbf{e}^k$, $D^2_{\mathbf{e}^k\mathbf{e}^j}f(\mathbf{x})$ si chiama *derivata parziale seconda rispetto* a $x_j$ e $x_k$ e si indica con uno dei simboli seguenti:

$$\frac{\partial^2 f}{\partial x_k \partial x_j}(\mathbf{x}), \quad f_{x_k x_j}(\mathbf{x}), \quad D_{x_k x_j}f(\mathbf{x}), \quad D^2_{kj}f(\mathbf{x}), \quad \partial_{x_k x_j}f(\mathbf{x}).$$

Se $k \neq j$, le derivate $\frac{\partial^2 f}{\partial x_k \partial x_j}$ si chiamano *miste*; se $k = j$ si chiamano *pure* e il primo simbolo si semplifica in $\frac{\partial^2 f}{\partial x_j^2}(\mathbf{x})$.

**Esempio 1.4** Sia $f : \mathbf{x} \mapsto \frac{1}{|\mathbf{x}|} : \mathbb{R}^n \backslash \{\mathbf{0}\} \to \mathbb{R}$. Calcoliamo $f_{x_k x_j}(\mathbf{x})$, $\mathbf{x} \neq \mathbf{0}$. Si possono usare le ordinarie regole di derivazione, come abbiamo già osservato. Si ha:

$$f_{x_j}(\mathbf{x}) = -\frac{x_j}{|\mathbf{x}|^3}, \quad f_{x_k x_j}(\mathbf{x}) = -\frac{\delta_{kj}}{|\mathbf{x}|^3} + 3\frac{x_k x_j}{|\mathbf{x}|^5}$$

dove $\delta_{jk}$ è il simbolo di Kronecker.



Notiamo che nell'Esempio 1.4, $f_{x_k x_j} = f_{x_j x_k}$ per ogni $j$ e $k$.

Sarà sempre vero che $D^2_{\mathbf{w}\mathbf{v}}f = D^2_{\mathbf{v}\mathbf{w}}f$, per ogni funzione due volte derivabile lungo le direzioni $\mathbf{v}$ e $\mathbf{w}$?

Il prossimo esempio fornisce una risposta negativa.

**Esempio 1.5** Sia

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

Calcoliamo $f_{xy}(0,0)$ e $f_{yx}(0,0)$. Occorre naturalmente calcolare prima $f_x(0,y)$ e $f_y(x,0)$. Si ha:

$$\begin{aligned} f_x(0,y) &= \lim_{x\to 0}\frac{f(x,y)-f(0,y)}{x} = \lim_{x\to 0} y\frac{x^2-y^2}{x^2+y^2} = -y; \\ f_y(x,0) &= \lim_{y\to 0}\frac{f(x,y)-f(x,0)}{y} = \lim_{y\to 0} x\frac{x^2-y^2}{x^2+y^2} = x. \end{aligned}$$

Si conclude che

$$f_{yx}(0,0) = -1 \quad \text{e} \quad f_{xy}(0,0) = 1,$$

e le derivate miste in $(0,0)$ sono diverse.

Dunque, in generale

$$f_{x_k x_j}(\mathbf{x}) \neq f_{x_j x_k}(\mathbf{x}).$$

Il prossimo teorema indica una condizione sufficiente per l'uguaglianza delle derivate seconde miste.

■ **Teorema 1.3** (di Schwarz) *Se $f_{x_k x_j}$ e $f_{x_j x_k}$ esistono in un intorno di $\mathbf{x}$ e sono continue in $\mathbf{x}$ allora*

$$f_{x_k x_j}(\mathbf{x}) = f_{x_j x_k}(\mathbf{x}).$$

*Dimostrazione.* Indichiamo con $U$ l'intorno di $\mathbf{x}$ in cui esistono $f_{x_k x_j}$ e $f_{x_j x_k}$; consideriamo l'espressione

$$\Delta = f(\mathbf{x} + t\mathbf{e}^k + t\mathbf{e}^j) - f(\mathbf{x} + t\mathbf{e}^k) - f(\mathbf{x} + t\mathbf{e}^j) + f(\mathbf{x})$$

con $|t|$ abbastanza piccolo in modo che i punti $\mathbf{x} + t\mathbf{e}^k$, $\mathbf{x} + t\mathbf{e}^j$ non escano da $U$.

Consideriamo la funzione

$$g(\tau) = f(\mathbf{x} + t\mathbf{e}^k + \tau\mathbf{e}^j) - f(\mathbf{x} + \tau\mathbf{e}^j)$$

dove si pensa $t$ fisso e $\tau \in [0,t]$. Allora

$$\Delta = g(t) - g(0),$$

come facilmente si verifica. D'altra parte, variando $\tau$, varia soltanto l'argomento di $f$ nella direzione $\mathbf{e}^j$, ovvero varia soltanto $x_j$; di conseguenza

$$g'(\tau) = f_{x_j}(\mathbf{x} + t\mathbf{e}^k + \tau\mathbf{e}^j) - f_{x_j}(\mathbf{x} + \tau\mathbf{e}^j).$$

Applicando il teorema di Lagrange a $g$, otteniamo:

$$\Delta = g(t) - g(0) = \{f_{x_j}(\mathbf{x} + t\mathbf{e}^k + \theta t\mathbf{e}^j) - f_{x_j}(\mathbf{x} + \theta t\mathbf{e}^j)\}t \tag{1.14}$$

dove $\theta$ è un opportuno numero tra 0 e 1 (dipendente da $\mathbf{x}$, $t$).

Introduciamo ora la funzione

$$\varphi(\tau) = f_{x_j}(\mathbf{x} + \tau t\mathbf{e}^k + \theta t\mathbf{e}^j),$$

dove ancora $t$ è fisso e $\tau \in [0, t]$. Si può scrivere, dalla (1.14):

$$\Delta = \{\varphi(t) - \varphi(0)\}t.$$

Questa volta, variando $\tau$ varia solo $x_k$ e, per ipotesi, $\varphi(\tau)$ è derivabile (poiché esiste $f_{x_k x_j}$ in $U$). Applicando ancora il teorema di Lagrange otteniamo:

$$\Delta = \varphi'(\eta t)t^2 = f_{x_k x_j}(\mathbf{x} + \eta t\mathbf{e}^k + \theta t\mathbf{e}^j)t^2 \tag{1.15}$$

dove $\eta \in (0, 1)$ è un numero opportuno, dipendente da $\mathbf{x}$ e $t$.

Osserviamo ora che l'espressione di $\Delta$ è simmetrica rispetto a $x_j$ e $x_k$; ripetendo lo stesso procedimento con i ruoli di $x_j$ e $x_k$ scambiati, si ottiene

$$\Delta = f_{x_j x_k}(\mathbf{x} + \eta' t\mathbf{e}^k + \theta' t\mathbf{e}^j)t^2. \tag{1.16}$$

dove $\eta'$ e $\theta'$ hanno le stesse proprietà di $\eta$ e $\theta$.

Dalla (1.15) e (1.16) otteniamo, dopo aver diviso per $t^2$:

$$f_{x_k x_j}(\mathbf{x} + \eta t\mathbf{e}^k + \theta t\mathbf{e}^j) = f_{x_j x_k}(\mathbf{x} + \eta' t\mathbf{e}^k + \theta' t\mathbf{e}^j). \tag{1.17}$$

Essendo $f_{x_k x_j}$ e $f_{x_j x_k}$ continue in $\mathbf{x}$, passando al limite per $t \to 0$ nella (1.17) si deduce $f_{x_k x_j}(\mathbf{x}) = f_{x_j x_k}(\mathbf{x})$. □

**OSSERVAZIONE 1.4** Il teorema vale per derivate direzionali seconde qualunque, non solo per le derivate seconde miste.

Inoltre si potrebbe dimostrare che, se $f_{x_k}$, $f_{x_j}$, $f_{x_k x_j}$ esistono in un intorno di un punto $\mathbf{x}$ e $f_{x_k x_j}$ è continua in $\mathbf{x}$, allora esiste anche $f_{x_j x_k}(\mathbf{x})$ ed è uguale a $f_{x_k x_j}(\mathbf{x})$.

Oltre alle derivate seconde potremo considerare in maniera ovvia derivate di ordine superiore. Per esempio le notazioni

$$\frac{\partial^3}{\partial x_k \partial x_j \partial x_i} f(\mathbf{x}), \quad f_{x_k x_j x_i}(\mathbf{x}), \quad D^3_{kji} f(\mathbf{x}), \quad \partial_{x_k x_j x_i} f(\mathbf{x})$$

indicano la derivata parziale terza di $f$ rispetto a $x_i$, $x_j$, $x_k$ nell'ordine. Nulla di particolare v'è da aggiungere riguardo al loro calcolo.

Passiamo ora alla definizione di *differenziale secondo*.

Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$ differenziabile in $A$. Allora per ogni $\mathbf{x} \in A$ esistono le derivate parziali $f_{x_j}(\mathbf{x})$, $j = 1, \ldots, n$.

Se queste derivate sono a loro volta differenziabili in $\mathbf{x}$ diremo che $f$ è *due volte differenziabile* in $\mathbf{x}$ e si chiama *differenziale secondo* di $f$ in $\mathbf{x}$ la forma quadratica nell'incremento $\mathbf{h} = (h_1, \cdots, h_n)$ data da $d^2 f(\mathbf{x}) : \mathbf{h} \mapsto \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}) h_i h_j$. In altri termini:

$$d^2 f(\mathbf{x}) := \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}) dx_i dx_j. \tag{1.18}$$



Il differenziale secondo giocherà un ruolo fondamentale nella teoria dell'ottimizzazione per funzioni di $n$ variabili.

La matrice quadrata di ordine $n$ i cui elementi sono $f_{x_i x_j}(\mathbf{x})$ si chiama *matrice hessiana*[3] di $f$ nel punto $\mathbf{x}$ e viene indicata col simbolo $\mathbf{H}_f(\mathbf{x})$, cioè:

$$\mathbf{H}_f(\mathbf{x}) := \begin{pmatrix} f_{x_1x_1}(\mathbf{x}) & f_{x_1x_2}(\mathbf{x}) & \ldots & f_{x_1x_n}(\mathbf{x}) \\ f_{x_2x_1}(\mathbf{x}) & f_{x_2x_2}(\mathbf{x}) & \ldots & f_{x_2x_n}(\mathbf{x}) \\ \vdots & \vdots & \ldots & \vdots \\ f_{x_nx_1}(\mathbf{x}) & f_{x_nx_2}(\mathbf{x}) & \ldots & f_{x_nx_n}(\mathbf{x}) \end{pmatrix}.$$

Potremo allora scrivere

$$d^2 f(\mathbf{x}) = \langle \mathbf{H}_f(\mathbf{x}) d\mathbf{x}, d\mathbf{x} \rangle$$

dove $\mathbf{H}_f(\mathbf{x}) d\mathbf{x}$ indica il prodotto righe per colonne della matrice $\mathbf{H}_f(\mathbf{x})$ per il vettore $d\mathbf{x}$.

Se $f$ è due volte differenziabile in $\mathbf{x}$, come immediata conseguenza abbiamo che esistono le derivate $D^2_{\mathbf{vw}} f(\mathbf{x})$ per *ogni coppia* di versori $\mathbf{v}$ e $\mathbf{w} \in \mathbb{R}^n$. Inoltre vale la formula

$$D^2_{\mathbf{vw}} f(\mathbf{x}) = \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}) v_i w_j = \langle \mathbf{H}_f(\mathbf{x})\mathbf{w}, \mathbf{v} \rangle. \tag{1.19}$$

Si noti la somiglianza tra le formule (1.19) e (1.18).

La (1.19) si dimostra osservando che, essendo $f$ differenziabile in $\mathbf{x}$, esiste $D_{\mathbf{w}} f(\mathbf{x})$ e si ha per ogni $\mathbf{x} \in A$:

$$D_{\mathbf{w}} f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{w} \rangle = \sum_{j=1}^{n} f_{x_j}(\mathbf{x}) w_j$$

dove $w_j (j = 1, \ldots, n)$ sono le componenti di $\mathbf{w}$.

Essendo le $f_{x_j}$ differenziabili in $\mathbf{x}$ si avrà

$$D_{\mathbf{v}} f_{x_j}(\mathbf{x}) = \langle \nabla f_{x_j}(\mathbf{x}), \mathbf{v} \rangle = \sum_{i=1}^{n} f_{x_i x_j}(\mathbf{x}) v_i$$

e quindi

$$D^2_{\mathbf{vw}} f(\mathbf{x}) = \sum_{j=1}^{n} D_{\mathbf{v}} f_{x_j}(\mathbf{x}) w_j = \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}) v_i w_j.$$

**Esempio 1.6** Siano $\mathbf{v} = (\cos\theta, \sin\theta)$, $\mathbf{w} = (\cos\varphi, \sin\varphi)$ e $f(x, y) = xe^{2y}$. Allora $f_{xx} = 0$, $f_{yy} = 4xe^{2y}$, $f_{yx} = 2e^{2y}$ e

$$D^2_{\mathbf{vw}} f(0, 0) = 2\cos\theta \sin\varphi + 2\sin\theta \cos\varphi.$$

[3] Otto Hesse (1811-1874).

Se $f$ è una funzione di 2 variabili differenziabile due volte in un punto $(x, y) \in A \subseteq \mathbb{R}^n$ il suo differenziale secondo in $(x, y)$ si scrive

$$d^2f(x,y) = f_{xx}(x,y)dx^2 + f_{xy}(x,y)dxdy + f_{yx}(x,y)dxdy + f_{yy}(x,y)dy^2.$$

Il fatto che $f$ sia due volte differenziabile non implica la continuità di $f_{xy}$ e $f_{yx}$: non si può dunque applicare il Teorema 1.3 per concludere che $f_{xy}(x,y) = f_{yx}(x,y)$. Che ciò sia vero segue però dal seguente teorema:

■ **TEOREMA 1.4** *Se $f$ è due volte differenziabile in $\mathbf{x}$ l'ordine di derivazione nelle derivate miste è invertibile.*

*Dimostrazione.* Facciamo vedere che $f_{x_kx_j}(\mathbf{x}) = f_{x_jx_k}(\mathbf{x})$. Il ragionamento è simile a quello della dimostrazione del Teorema 1.3. Sia dunque $\Delta$ come in quella dimostrazione; usando il teorema di Lagrange possiamo ancora scrivere, poiché le derivate prime esistono in un intorno di $\mathbf{x}$:

$$\Delta = \{f_{x_j}(\mathbf{x} + t\mathbf{e}^k + \theta t\mathbf{e}^j) - f_{x_j}(\mathbf{x} + \theta t\mathbf{e}^j)\}t$$

per $\theta \in (0,1)$ opportuno.

Non possiamo ora utilizzare il teorema di Lagrange poiché sappiamo che le derivate seconde esistono in $\mathbf{x}$ soltanto.

Sappiamo però che $f_{x_j}$ è differenziabile in $\mathbf{x}$, essendo $f$ due volte differenziabile. Possiamo allora scrivere:

$$f_{x_j}(\mathbf{x} + t\mathbf{e}^k + \theta t\mathbf{e}^j) = f_{x_j}(\mathbf{x}) + f_{x_kx_j}(\mathbf{x})t + f_{x_jx_j}(\mathbf{x})\theta t + \eta(\mathbf{x}, t)$$
$$f_{x_j}(\mathbf{x} + \theta t\mathbf{e}^j) = f_{x_j}(\mathbf{x}) + f_{x_jx_j}(\mathbf{x})\theta t + \eta_1(\mathbf{x}, t)$$

dove $\eta$ ed $\eta_1$ sono infinitesimi di ordine superiore a $t$ per $t \to 0$.

Allora:

$$\Delta = \{f_{x_kx_j}(\mathbf{x})t + \eta(\mathbf{x}, t) - \eta_1(\mathbf{x}, t)\}t. \tag{1.20}$$

Analogamente, sfruttando la simmetria di $\Delta$ rispetto a $x_k$ e $x_j$, si ha:

$$\Delta = \{f_{x_jx_k}(\mathbf{x})t + \overline{\eta}(\mathbf{x}, t) - \overline{\eta}_1(\mathbf{x}, t)\}t \tag{1.21}$$

dove ancora $\overline{\eta}$ e $\overline{\eta}_1$ sono infinitesimi di ordine superiore a $t$ per $t \to 0$.

Dalla (1.20) si deduce

$$\lim_{t\to 0} \frac{\Delta}{t^2} = f_{x_kx_j}(\mathbf{x}),$$

mentre dalla (1.21) si ha:

$$\lim_{t\to 0} \frac{\Delta}{t^2} = f_{x_jx_k}(\mathbf{x}).$$

Di conseguenza

$$f_{x_jx_k}(\mathbf{x}) = f_{x_kx_j}(\mathbf{x}).$$ □

In base al Teorema 1.4 si può scrivere, per una funzione due volte differenziabile di due variabili

$$d^2f(x,y) = f_{xx}(x,y)dx^2 + 2f_{xy}(x,y)dxdy + f_{yy}(x,y)dy^2. \tag{1.22}$$

È evidente l'analogia della (1.22) con la formula per il quadrato di un binomio. Questa analogia può essere spinta più in profondità e, come vedremo, generalizzata.



A tale scopo, data la solita funzione $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, riguardiamo le derivate $\frac{\partial f}{\partial x_j}$ come effetto dell'azione dell'*operatore* $\frac{\partial}{\partial x_j}$ sulla funzione $f$:

$$f \longrightarrow \boxed{\frac{\partial}{\partial x_j}} \longrightarrow \frac{\partial f}{\partial x_j}.$$

Incidentalmente osserviamo che l'azione dell'operatore $\frac{\partial}{\partial x_j}$ può essere automatizzata per mezzo dei cosiddetti "linguaggi formali" che l'informatica ci mette oggi a disposizione.

Così come si può definire la composizione di funzioni, si può definire la composizione di operatori come $\frac{\partial}{\partial x_j}$, che vengono chiamati *operatori differenziali.*

In questo modo $\frac{\partial^2}{\partial x_k \partial x_j}$ è interpretabile come composizione di $\frac{\partial}{\partial x_j}$ e $\frac{\partial}{\partial x_k}$, nell'ordine, ovvero come "prodotto di $\frac{\partial}{\partial x_j}$ e $\frac{\partial}{\partial x_k}$":

$$f \longrightarrow \boxed{\frac{\partial}{\partial x_j}} \longrightarrow \frac{\partial f}{\partial x_j} \longrightarrow \boxed{\frac{\partial}{\partial x_k}} \longrightarrow \frac{\partial^2 f}{\partial x_k \partial x_j};$$

$\frac{\partial^2}{\partial x_k^2}$ è interpretabile come composizione di $\frac{\partial}{\partial x_k}$ con sé stesso, ovvero come il "quadrato" di $\frac{\partial}{\partial x_k}$ che possiamo indicare con $\left(\frac{\partial}{\partial x_k}\right)^2$.

Con questa convenzione di scrittura la (1.22) diventa

$$d^2 f = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^2 f \qquad \text{(si legge: quadrato formale di...)}$$

dove il quadrato agisce nel modo usuale sui differenziali $dx$ e $dy$.

In generale, per una funzione 2 volte differenziabile in $\mathbf{x}$, si avrà:

$$d^2 f = \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + \ldots + \frac{\partial}{\partial x_n}dx_n\right)^2 f. \tag{1.23}$$

I differenziali di ordine superiore al secondo si possono definire facilmente. Sia $k > 2$ e sia $f$ una funzione dotata di *tutte le derivate parziali di ordine* $k-1$ in un intorno di $\mathbf{x}$.

Se ognuna di queste derivate è differenziabile in $\mathbf{x}$, allora si dice che $f$ è *$k$ volte differenziabile* in $\mathbf{x}$. Si noti che, in base ai Teoremi 1.3 e 1.4, se $f$ è $k$ volte differenziabile in $\mathbf{x}$, esistono in un intorno di $\mathbf{x}$ tutte le derivate parziali dall'ordine 1 all'ordine $k-1$ e sono ivi differenziabili; inoltre per ognuna di esse vale il teorema relativo all'inversione dell'ordine di derivazione.

Il differenziale di ordine $k$ di $f$ in $\mathbf{x}$ è assegnato dalla formula seguente:

$$d^k f(\mathbf{x}) := \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + \ldots + \frac{\partial}{\partial x_n}dx_n\right)^k f =$$
$$= \sum_{i_1, i_2, \ldots, i_k = 1}^{n} \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_k}}(\mathbf{x}) dx_{i_1} dx_{i_2} \ldots dx_{i_k}.$$

Si vede che $d^k f(\mathbf{x})$ è un polinomio omogeneo di grado $k$ nelle componenti di $d\mathbf{x}$.

Importanti classi di funzioni sono $C^k(A)$ e $C^\infty(A)$; $C^k(A)$ indica la classe delle funzioni $k$ volte differenziabili con continuità, ovvero dotate di derivate parziali fino all'ordine $k$ incluso, continue in $A$; esse sono dunque $k$ volte differenziabili in base al Teorema 1.3 (applicato alle derivate di ordine $k-1$); $C^\infty(A)$ indica l'insieme delle funzioni infinite volte differenziabili con continuità.

Concludiamo il paragrafo con una serie di osservazioni riguardanti gli operatori differenziali.

Consideriamo $\frac{\partial}{\partial x_j}$; le proprietà dell'operazione di derivata implicano immediatamente che, per ogni $\alpha, \beta \in \mathbb{R}$ e per ogni $f, g$ derivabili:

$$\frac{\partial}{\partial x_j}(\alpha f + \beta g) = \alpha \frac{\partial f}{\partial x_j} + \beta \frac{\partial g}{\partial x_j}$$

ovvero che l'*operatore* $\frac{\partial}{\partial x_j}$ *è lineare*.

Se inoltre consideriamo $\frac{\partial}{\partial x_k}$, in base al teorema di Schwarz, per funzioni di classe $C^2(A)$ abbiamo $\frac{\partial^2 f}{\partial x_k \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_k}$, formula che possiamo interpretare come *commutatività* degli operatori $\frac{\partial}{\partial x_i}$ e $\frac{\partial}{\partial x_k}$.

Anche le derivate direzionali possono essere interpretate come operatori differenziali con le stesse proprietà di linearità. Fissato $\mathbf{v} \in \mathbb{R}^n$, $|\mathbf{v}| = 1$, l'operatore $D_\mathbf{v}$ associa a $f$ la derivata $D_\mathbf{v} f$.

Un altro importante operatore differenziale è il seguente:

$$\Delta : f \mapsto \Delta f := \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \ldots + \frac{\partial^2 f}{\partial x_n^2} \quad (\textit{operatore di Laplace}).$$

Le funzioni $f \in C^2(A)$ tali che $\Delta f = 0$ in $A$ si dicono *armoniche*. Per esempio $f(x, y) = x^2 - y^2$ è armonica in $\mathbb{R}^2$. Queste funzioni hanno un ruolo particolarmente importante in fisica-matematica.

## 1.4 Formula di Taylor

Per una funzione differenziabile in un punto $\mathbf{x}$, $df(\mathbf{x})$ costituisce la migliore approssimazione lineare in un intorno di $\mathbf{x}$; approssimazioni locali più accurate si ottengono facendo intervenire i differenziali di ordine superiore, come nel caso unidimensionale. Ciò che si ottiene è l'estensione della formula di Taylor. Il prossimo teorema estende tale formula con il resto nella forma di Lagrange. Interpretiamo $d\mathbf{x}$ come un piccolo incremento della variabile $\mathbf{x}$ (cfr. Oss. 1.1)

■ **TEOREMA 1.5** *Sia* $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$ *e supponiamo che il segmento chiuso* $[\mathbf{x}, \mathbf{x} + d\mathbf{x}]$ *sia contenuto in* $A$.

*Se* $f$ *è differenziabile con continuità* $k-1$ *volte nel segmento chiuso e* $k$ *volte differenziabile nel segmento aperto, allora esiste* $\theta \in (0,1)$ *tale che*

$$(1.24)\quad f(\mathbf{x}+d\mathbf{x}) - f(\mathbf{x}) = df(\mathbf{x}) + \frac{1}{2} d^2 f(\mathbf{x}) + \ldots + \frac{1}{(k-1)!} d^{k-1} f(\mathbf{x}) + \frac{1}{k!} d^k f(\mathbf{x} + \theta d\mathbf{x}).$$

La (1.24) si chiama Formula di Taylor di ordine $k$ con centro nel punto $\mathbf{x}$. Se $\mathbf{x} = \mathbf{0}$ si chiama ancora formula di Mac Laurin.



Per la dimostrazione facciamo uso della seguente proposizione, che risulta essere un caso particolare del Teorema 2.1, che verrà presentato nella Sezione 2.

**PROPOSIZIONE 1.6** *Sia* $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, *differenziabile in* $\mathbf{x} \in A$; *siano inoltre* $x_1 = x_1(t)$, $x_2 = x_2(t), \ldots, x_n = x_n(t)$ *funzioni da* $(a,b) \to A$, *differenziabili in* $(a,b)$ *e tali che* $\mathbf{x} = (x_1(t_0), x_2(t_0), \ldots x_n(t_0))$. *Allora la funzione composta*

$$F(t) = f(x_1(t), x_2(t), \ldots, x_n(t))$$

*è differenziabile in* $t_0$ *e vale la formula*

$$(1.25)\quad F'(t_0) = \frac{\partial f}{\partial x_1}(\mathbf{x}) \cdot x_1'(t_0) + \frac{\partial f}{\partial x_2}(\mathbf{x}) \cdot x_2'(t_0) + \ldots + \frac{\partial f}{\partial x_n}(\mathbf{x}) \cdot x_n'(t_0).$$

*Dimostrazione del Teorema* 1.5. L'idea è di ricondursi al caso unidimensionale introducendo la funzione $g(t) = f(\mathbf{x} + t d\mathbf{x})$ e osservando che:

i) $f(\mathbf{x} + d\mathbf{x}) - f(\mathbf{x}) = g(1) - g(0)$;

ii) in base alla Proposizione 1.6, con $x_j(t) = x_j + t dx_j$ $(j = 1, \ldots, n)$, $g$ risulta $k-1$ volte differenziabile in $[0,1]$ e $k$ volte differenziabile in $(0,1)$;

iii) inoltre, poiché $x_j'(t) = dx_j$ per ogni $j = 1, \ldots, n$ si ha, in base alla formula (1.25)

$$g'(t) = \frac{\partial f}{\partial x_1}(\mathbf{x} + t d\mathbf{x}) dx_1 + \frac{\partial f}{\partial x_2}(\mathbf{x} + t d\mathbf{x}) dx_2 + \ldots + \frac{\partial f}{\partial x_n}(\mathbf{x} + t d\mathbf{x}) dx_n$$

cosicché $g'(0) = df(\mathbf{x})$.

Analogamente si ricava che

$$(1.26)\quad g''(0) = d^2 f(\mathbf{x}), \quad \ldots, \quad g^{(k-1)}(0) = d^{k-1} f(\mathbf{x})$$

e infine

$$(1.27)\quad g^{(k)}(t) = d^k f(\mathbf{x} + t d\mathbf{x}).$$

Applicando la formula di Taylor a $g$ per l'intervallo $[0,1]$ possiamo scrivere:

$$(1.28)\quad g(1) - g(0) = g'(0) + \frac{1}{2} g''(0) + \ldots + \frac{1}{(k-1)!} g^{(k-1)}(0) + \frac{1}{k!} g^{(k)}(\theta)$$

dove $\theta \in (0,1)$, opportuno.

Tenendo conto di i), ii) e iii) e di (1.26), (1.27), la (1.28) coincide esattamente con la (1.24). □

Il caso particolare $k = 1$ nella (1.24) è l'estensione al caso di funzioni reali di $n$ variabili del teorema del valor medio di Lagrange:

$$(1.29)\quad f(\mathbf{x} + d\mathbf{x}) - f(\mathbf{x}) = df(\mathbf{x} + \theta d\mathbf{x}) \quad (\theta \in (0,1) \text{ opportuno}).$$

Come già nel caso unidimensionale, tramite la (1.29) possiamo caratterizzare le funzioni costanti in un *aperto connesso*.

**PROPOSIZIONE 1.7** *Sia* $A$ *aperto connesso e* $df(\mathbf{x}) = 0$ $\forall \mathbf{x} \in A$. *Allora* $f$ *è costante.*

La dimostrazione della Proposizione 1.7 si ricava facilmente ricordando che $A$ aperto connesso è connesso per segmenti. Lasciamo i dettagli al lettore.

Come vedremo nella Sezione 2, la formula di Taylor con il resto di Lagrange *non* è estendibile al caso di funzioni vettoriali e quindi neppure il teorema del valor medio.

Si potrà invece estendere anche a queste funzioni la formula di Taylor con il resto di Peano, oggetto del prossimo teorema.

■ **Teorema 1.8** *Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, $k$ volte differenziabile in $\mathbf{x}$. Allora*

$$f(\mathbf{x}+d\mathbf{x}) - f(\mathbf{x}) = df(\mathbf{x}) + \frac{1}{2}d^2f(\mathbf{x}) + \ldots + \frac{1}{k!}d^k(\mathbf{x}) + o(|d\mathbf{x}|^k) \tag{1.30}$$

*per* $|d\mathbf{x}| \to 0$.

*Dimostrazione.* Ne diamo solo l'idea nel caso $n = 2$, $k = 2$. Poniamo $d\mathbf{x} = (h, k)$, $\mathbf{x} = (x, y)$ e

$$\begin{aligned} g(h,k) = {} & f(x+h, y+k) - f(x,y) - f_x(x,y)h - f_y(x,y)k + \\ & - \frac{1}{2}\{f_{xx}(x,y)h^2 + 2f_{xy}(x,y)hk + f_{yy}(x,y)k^2\}. \end{aligned}$$

Occorre dimostrare che $g(h,k)/(h^2+k^2) \to 0$ quando $h^2 + k^2 \to 0$.
Usiamo il teorema del valor medio, osservando che $g(0,0) = 0$:

$$g(h,k) = g(h,k) - g(0,0) = g_h(\theta h, \theta k)h + g_k(\theta h, \theta k)k \tag{1.31}$$

con $\theta$ opportuno tra 0 e 1, $\theta = \theta(h,k)$. Ora abbiamo:

$$g_h(h,k) = f_x(x+h, y+k) - f_x(x,y) - f_{xx}(x,y)h - f_{xy}(x,y)k$$

e quindi, essendo $f_x$ differenziabile in $(x,y)$, si ha $g_h(h,k) = o(\sqrt{h^2+k^2})$.

Analogamente,

$$g_k(h,k) = f_y(x+h, y+k) - f_y(x,y) - f_{yy}(x,y)k - f_{xy}(x,y)h = o(\sqrt{h^2+k^2})$$

in virtù della differeziabilità di $f_y$.

Sostituendo nella (1.31) otteniamo, essendo $0 < \theta = \theta(h,k) < 1$:

$$|g(h,k)| \le (|h| + |k|) \cdot |o(\sqrt{h^2+k^2})|$$

da cui $|g(h,k)|/(h^2+k^2) \le |o(\sqrt{h^2+k^2})|/\sqrt{h^2+k^2}$ e quindi la tesi. □

## 1.5 Funzioni omogenee; funzioni convesse e concave

Presentiamo in questo paragrafo due particolari classi di funzioni spesso impiegate nelle applicazioni.

La prima è quella delle funzioni *positivamente omogenee di grado $\mu$*, $\mu \in \mathbb{R}$. Queste funzioni sono definite in *coni* con vertice in $\mathbf{0}$, vale a dire che, se sono definite in un punto $\mathbf{x}$, allora sono definite su tutta la semiretta $\rho\mathbf{x}$ per ogni $\rho > 0$. Il cono non è necessariamente convesso e, naturalmente, può coincidere con tutto $\mathbb{R}^n$.

**Definizione 1.3** *Sia $\mathcal{C} \subseteq \mathbb{R}^n$ un cono e $f : \mathcal{C} \to \mathbb{R}$; $f$ si dice positivamente omogenea di grado $\mu$, $\mu \in \mathbb{R}$, se $\forall \mathbf{x} \in \mathcal{C}$ e $\forall \rho > 0$ risulta*

$$f(\rho\mathbf{x}) = \rho^\mu f(\mathbf{x}). \tag{1.32}$$



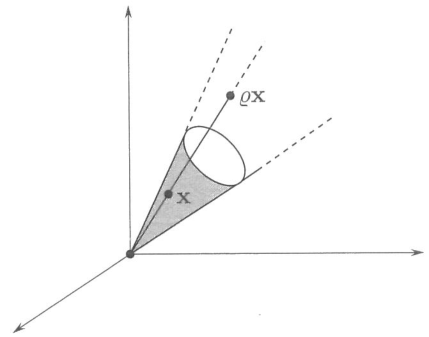


**Figura 7.5.** Un cono convesso in $\mathbb{R}^3$.

### Esempi

**1.7** Qualunque *polinomio omogeneo* di grado $\mu$, $\mu \in \mathbb{N}$, in $k$ variabili (ovvero ogni termine ha grado $\mu$) è omogeneo dello stesso grado nel senso della Definizione 1.3.

Per esempio,

$$f(\mathbf{x}) = a_1x_1 + a_2x_2 + \ldots + a_kx_k \qquad (\mu = 1)$$

$$f(\mathbf{x}) = \sum_{i,j=1}^{k} a_{ij}x_ix_j. \qquad (\mu = 2)$$

La prima funzione si dice anche linearmente omogenea; la seconda si chiama *forma quadratica* e si può scrivere nel modo seguente:

$$\sum_{i,j=1}^{k} a_{ij}x_ix_j = \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle$$

dove $\mathbf{A}$ è la matrice quadrata di ordine $k$ i cui elementi sono $a_{ij}$ e $\mathbf{x} = (x_1, x_2, \ldots, x_k)$.

In particolare il differenziale secondo è una forma quadratica (nelle componenti di $d\mathbf{x}$). Per questa ragione le forme quadratiche avranno un ruolo speciale nella teoria dell'ottimizzazione per funzioni di più variabili.

**1.8** La norma di un vettore $\mathbf{x} \in \mathbb{R}^n$ è positivamente omogenea di grado 1. Infatti, se $\rho > 0$

$$|\rho\mathbf{x}| = \rho|\mathbf{x}|.$$

**1.9** Siano $\mathbf{x}$ e $\mathbf{y}$ vettori non nulli di $\mathbb{R}^3$ e sia $f$ la funzione che associa a ogni coppia $(\mathbf{x}, \mathbf{y})$ il coseno dell'angolo tra $\mathbf{x}$ e $\mathbf{y}$, in formule:

$$f(\mathbf{x}, \mathbf{y}) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{|\mathbf{x}| \cdot |\mathbf{y}|}.$$

Allora $f$ è positivamente omogenea di *grado zero* ($\mu = 0$).

Il seguente teorema caratterizza le funzioni positivamente omogenee differenziabili in coni aperti.

■ **Teorema 1.9** (di Eulero) *Sia $\mathcal{C} \subseteq \mathbb{R}^n$ un cono aperto e $f : \mathcal{C} \to \mathbb{R}$ differenziabile in $\mathcal{C}$. Allora $f$ è positivamente omogenea di grado $\mu$ se e solo se, per ogni $\mathbf{x} \in \mathcal{C}$, risulta:*

$$\langle \mathbf{x}, \nabla f(\mathbf{x}) \rangle = \mu f(\mathbf{x}). \tag{1.33}$$

*Dimostrazione.* Se $f$ è positivamente omogenea di grado $\mu$, allora, derivando l'identità (1.32) rispetto a $\rho$ e usando la Proposizione 1.6, si ottiene

$$\frac{\partial f}{\partial x_1}(\rho\mathbf{x})x_1 + \frac{\partial f}{\partial x_2}(\rho\mathbf{x})x_2 + \ldots + \frac{\partial f}{\partial x_n}(\rho\mathbf{x})x_n = \mu\rho^{\mu-1} f(\mathbf{x}).$$

Ponendo $\rho = 1$ si deduce la (1.33).

Viceversa, sia valida la (1.33); poniamo

$$\varphi(\rho) = f(\rho\mathbf{x})/\rho^\mu$$

e facciamo vedere che $\varphi'(\rho) = 0$ in $\mathbb{R}_+$. Si ha:

$$\varphi'(\rho) = \frac{\langle \mathbf{x}, \nabla f(\rho\mathbf{x})\rangle \rho^\mu - \mu\rho^{\mu-1} f(\rho\mathbf{x})}{\rho^{2\mu}} = \frac{1}{\rho^{\mu+1}}\{\langle \rho\mathbf{x}, \nabla f(\rho\mathbf{x})\rangle - \mu f(\rho\mathbf{x})\} = 0$$

per la (1.33), con $\rho\mathbf{x}$ al posto di $\mathbf{x}$. Dunque $\varphi(\rho)$ è costante in $\mathbb{R}_+$ e quindi $\varphi(\rho) = \varphi(1)$ ovvero $f(\rho\mathbf{x}) = \rho^\mu f(\mathbf{x})$. □

Estendiamo ora al caso di funzioni reali di più variabili la nozione di funzione *convessa* e *concava*, che, del resto, è formalmente identica a quella del caso unidimensionale.

Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$. Indichiamo con $\mathrm{Epi}(f)$ l'*epigrafico* di $f$, cioè

$$\mathrm{Epi}(f) := \{(\mathbf{x}, z) \in \mathbb{R}^{n+1} : z \geq f(\mathbf{x}),\ \mathbf{x} \in A\}.$$

**Definizione 1.4** *Sia $f$ definita su un insieme convesso $A \subseteq \mathbb{R}^n$; $f$ si dice convessa in $A$ se $\mathrm{Epi}(f)$ è convesso in $\mathbb{R}^{n+1}$; $f$ si dice concava se $-f$ è convessa.*

Alternativamente: $f$ è convessa in $A$ se per ogni coppia di punti $\mathbf{x}$, $\mathbf{y} \in A$ e per ogni $t \in (0, 1)$ risulta

$$f(t\mathbf{y} + (1-t)\mathbf{x}) \leq tf(\mathbf{y}) + (1-t)f(\mathbf{x}). \tag{1.34}$$

La (1.34) significa che i punti del segmento di estremi $(\mathbf{x}, f(\mathbf{x}))$ e $(\mathbf{y}, f(\mathbf{y}))$ stanno al di sopra del grafico di $f$ o, al più, sul grafico di $f$ (vedi Fig. 7.6).

Se la disuguaglianza (1.33) vale in senso stretto, $f$ si dice *strettamente convessa*.

Per le funzioni concave (strettamente concave) la (1.34) vale con $\geq (>)$ anziché $\leq (<)$.

Ci riferiremo d'ora in poi alle funzioni convesse.



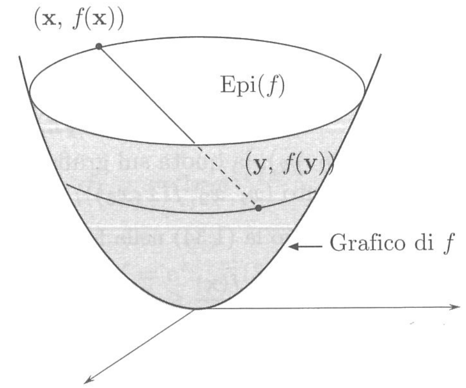


**Figura 7.6.** Funzione strettamente convessa.

## Esempi

**1.10** $f(x, y) = x^2 + y^2$ è strettamente convessa in $\mathbb{R}^2$; $f(\mathbf{x}) = \sum_{j=1}^{n} x_j^2 = |\mathbf{x}|^2$ è strettamente convessa in $\mathbb{R}^n$.

**1.11** $f(\mathbf{x}) = \sqrt{x^2 + y^2}$ è convessa in $\mathbb{R}^2$. $f(\mathbf{x}) = |\mathbf{x}|$ è convessa in $\mathbb{R}^n$.

**1.12** Se $f$ è convessa in $A$, allora anche $f^\alpha$ ($\alpha > 1$, $f > 0$) ed $e^f$ sono convesse. Si veda in riferimento a questo esempio l'Esercizio 23 del presente paragrafo.

Come per le funzioni convesse di una variabile, quelle di più variabili hanno, per il solo fatto di essere convesse, notevoli proprietà di regolarità, come risulta dal seguente teorema che ci limitiamo a enunciare.

■ **Teorema 1.10** *Sia $f$ convessa in $A$, aperto convesso in $\mathbb{R}^n$. Allora:*

i) *$f$ è continua in $A$;*

ii) *in ogni punto di $A$, $f$ ammette derivate parziali destre e sinistre;*

iii) *nei punti in cui esistono tutte le derivate parziali, $f$ è differenziabile.*

Se una funzione convessa di una variabile è derivabile in un punto $x_0$, la tangente al grafico nel punto di ascissa $x_0$ sta sotto (o almeno non sopra) il grafico stesso. L'equivalente di questa affermazione per funzioni convesse di due variabili sarebbe: se $f = f(x, y)$ è *differenziabile* in $(x_0, y_0)$ e convessa, allora il piano tangente sta sotto (o almeno non sopra) il grafico di $f$.

L'affermazione è vera, come mostra il seguente teorema.

■ **Teorema 1.11** *Sia $f$ differenziabile in $A$ aperto convesso in $\mathbb{R}^n$; $f$ è convessa (strettamente convessa) in $A$ se e solo se per ogni $\mathbf{x}$, $\mathbf{y} \in A$:*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + df(\mathbf{x}) \quad (f(\mathbf{y}) > f(\mathbf{x}) + df(\mathbf{x}),\ \mathbf{x} \neq \mathbf{y}). \tag{1.35}$$