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.

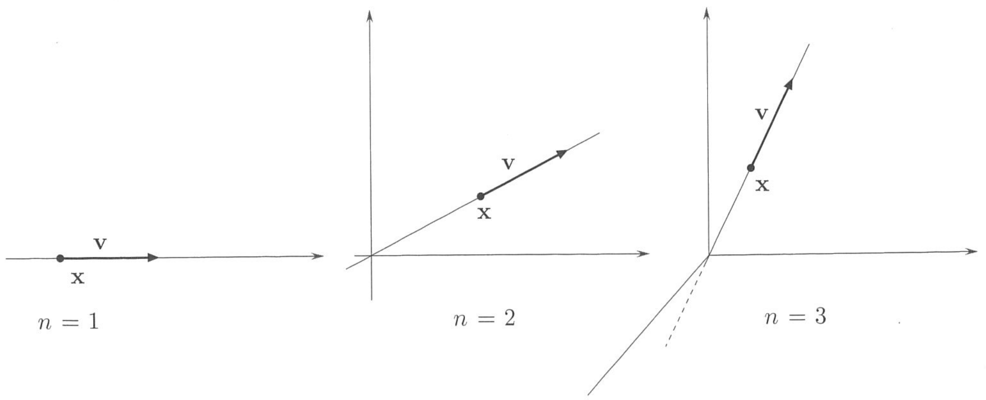

Figura 7.1. Incremento di $\mathbf{x}$ nella direzione $\mathbf{v}$.

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 $\varphi(t) = f(\mathbf{x} + t\mathbf{v})$, φ 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* $\operatorname{grad} f(\mathbf{x})$*:*

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

Il significato geometrico di derivata direzionale nel caso $n = 2$ è illustrato in Figura 7.2 dove il piano π 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 π 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 $\mathbf{x}$.

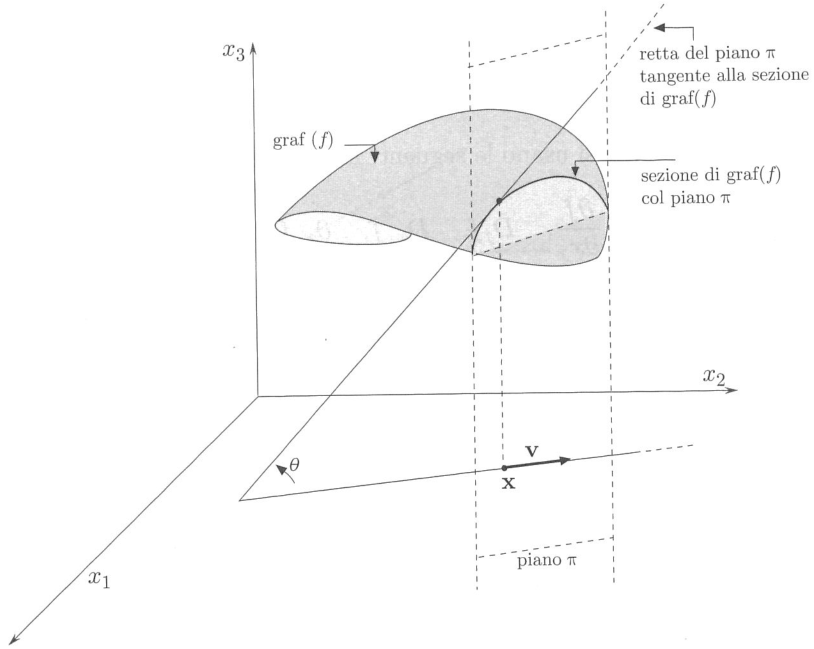

Figura 7.2. Significato geometrico di derivata direzionale: $D_{\mathbf{v}} f(\mathbf{x}) = \operatorname{tg}\theta$.

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 $\mathbf{x}$.

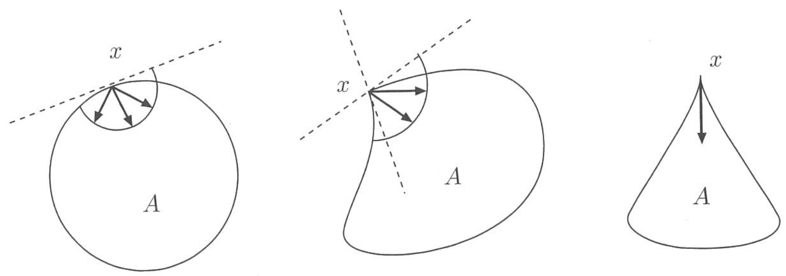

Figura 7.3. Esempi di direzioni ammissibili in vari casi.

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 ∞ (infatti $\lim_{x_1 \to 0^\pm} x_1 e^{1/x_1^2} = \pm\infty$).

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*

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

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

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

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 β 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:

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

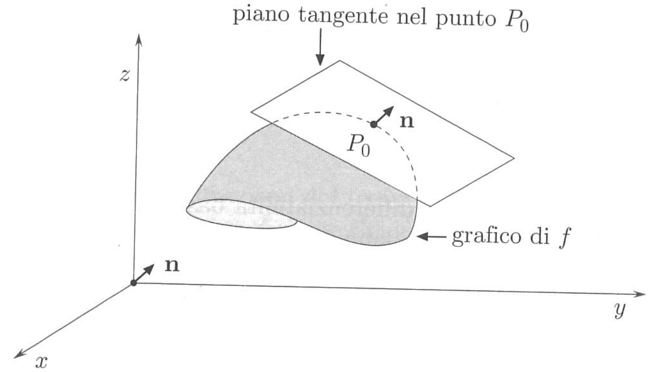

Figura 7.4. Piano tangente a una superficie.

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 α.

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 $\theta = 0$, π otteniamo $D_{\mathbf{v}_\theta} f(0,0) = \pm f_x(0,0) = 0$.

Per $\theta \neq 0$, π, 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 α 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 δ_{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 τ, 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 θ è un opportuno numero tra 0 e 1 (dipendente da $\mathbf{x}$, t).

Introduciamo ora la funzione

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

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 τ 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 Δ 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 η' e θ' hanno le stesse proprietà di η e θ.

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:

$$\mathbf{H}_f(\mathbf{x}) := \begin{pmatrix} f_{x_1 x_1}(\mathbf{x}) & f_{x_1 x_2}(\mathbf{x}) & \ldots & f_{x_1 x_n}(\mathbf{x}) \\ f_{x_2 x_1}(\mathbf{x}) & f_{x_2 x_2}(\mathbf{x}) & \ldots & f_{x_2 x_n}(\mathbf{x}) \\ \vdots & \vdots & \ldots & \vdots \\ f_{x_n x_1}(\mathbf{x}) & f_{x_n x_2}(\mathbf{x}) & \ldots & f_{x_n x_n}(\mathbf{x}) \end{pmatrix}.$$

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

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

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 $f_{x_k x_j}(\mathbf{x}) = f_{x_j x_k}(\mathbf{x})$. Il ragionamento è simile a quello della dimostrazione del Teorema 1.3. Sia dunque Δ 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:

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

dove η ed η_1 sono infinitesimi di ordine superiore a t per $t \to 0$.

Allora:

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

Analogamente, sfruttando la simmetria di Δ rispetto a x_k e x_j, si ha:

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

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

Dalla (1.20) si deduce

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

mentre dalla (1.21) si ha:

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

Di conseguenza

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

□

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

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

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 $\frac{\partial}{\partial x_k}$":

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

$\frac{\partial^2}{\partial x_k^2}$ 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 $\mathbf{x} = 0$ 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*

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

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

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

e infine

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

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

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

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:

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

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

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

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 θ 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* μ, $\mu \in \mathbb{R}$. 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* μ, $\mu \in \mathbb{R}$, *se* $\forall \mathbf{x} \in \mathcal{C}$ *e* $\forall \rho > 0$ *risulta*

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

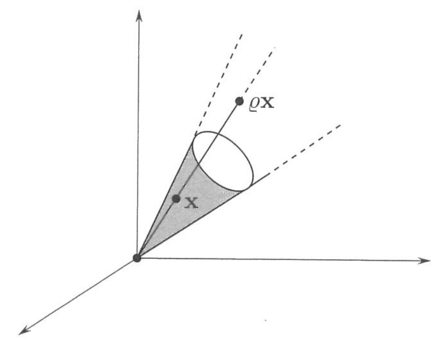

Figura 7.5. Un cono convesso in $\mathbb{R}^3$.

Esempi

1.7 Qualunque *polinomio omogeneo* di grado μ, $\mu \in \mathbb{N}$, in k variabili (ovvero ogni termine ha grado μ) 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.

23. Sia f convessa in A, $F : \mathbb{R} \to \mathbb{R}$ convessa e *crescente*; dimostrare che $F \circ f$ è convessa in A.

24.* Diciamo che una forma quadratica $q(\mathbf{y})$ è semidefinita positiva se $q(\mathbf{y}) \geq 0$ per ogni $\mathbf{y} \in \mathbb{R}^n$.

Dimostrare che, se f è due volte differenziabile in A, aperto convesso in $\mathbb{R}^n$, allora f è convessa in A se e solo se $d^2 f(\mathbf{x})$ è semidefinita positiva $\forall \mathbf{x} \in A$.

[Usare la formula di Taylor con il resto di Peano per il "solo se".]

2. FUNZIONI A VALORI VETTORIALI

2.1 Derivate e differenziali

Le nozioni di derivata e differenziale si estendono in modo naturale a funzioni con valori in $\mathbb{R}^m$, $m > 1$.

In generale penseremo i vettori scritti come vettori riga. D'altra parte se $\mathbf{x}$ è un vettore in $\mathbb{R}^n$ ed $\mathbf{M}$ una matrice $m \times n$, nel prodotto (righe per colonne) $\mathbf{Mx}$, $\mathbf{x}$ è pensato come vettore colonna. Anche se dovrebbe risultare chiaro dal contesto, preciseremo quando in una formula i vettori devono essere pensati come colonne.

Sia $\mathbf{f} : \mathbb{R}^n \supseteq A \to \mathbb{R}^m$, A aperto. Per ogni $\mathbf{x} \in A$, $\mathbf{f}(\mathbf{x})$, è un vettore $(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ le cui componenti f_j, $j = 1, \ldots, m$ sono funzioni da A in $\mathbb{R}$.

Fissato un versore $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{f}$ è derivabile lungo la direzione $\mathbf{v}$ nel punto $\mathbf{x}$ se e solo se esistono $D_{\mathbf{v}} f_j(\mathbf{x})$ per ogni $j = 1, \ldots, m$ e

$$D_{\mathbf{v}}\mathbf{f} := (D_{\mathbf{v}} f_1, D_{\mathbf{v}} f_2, \ldots, D_{\mathbf{v}} f_m).$$

Per quanto riguarda la differenziabilità, l'idea è sempre la stessa e cioè quella di approssimare l'incremento $\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x})$ (vettore di $\mathbb{R}^m$) con una trasformazione lineare in $\mathbf{h}$, da $\mathbb{R}^n$ in $\mathbb{R}^m$. Ora, ogni trasformazione di questo tipo è della forma $\mathbf{Mh}$ dove $\mathbf{M}$ è una matrice $m \times n$ e il prodotto è quello "righe per colonne"[4]. Siamo così condotti alla seguente definizione:

Definizione 2.1 $\mathbf{f}$ *si dice differenziabile in* $\mathbf{x}$ *se esiste una matrice* M *di ordine* $m \times n$ *tale che (vettori colonna):*

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = M\mathbf{h} + o(|\mathbf{h}|) \quad \textit{per } |\mathbf{h}| \to 0 \tag{2.1}$$

per ogni $\mathbf{h} \in \mathbb{R}^n$ *con* $\mathbf{x} + \mathbf{h} \in A$. *L'applicazione lineare da* $\mathbb{R}^n$ *in* $\mathbb{R}^m$ *data da:*

$$\mathbf{h} \mapsto M\mathbf{h}$$

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

Il Teorema 1.1 continua a valere; per quanto riguarda "l'identificazione" della matrice $\mathbf{M}$ osserviamo che scrivendo la (2.1) per la componente f_k di $\mathbf{f}$ si ottiene

$$f_k(\mathbf{x} + \mathbf{h}) - f_k(\mathbf{x}) = m_{k1}h_1 + m_{k2}h_2 + \ldots + m_{kn}h_n + o(|\mathbf{h}|). \tag{2.2}$$

[4] $\mathbb{R}^n$ ed $\mathbb{R}^m$ si intendono riferiti alla base canonica.

La (2.2) indica che f_k è differenziabile e che la k-esima riga di $\mathbf{M}$ è $\nabla f_k(\mathbf{x})$, in base alla (1.3). Dunque si ha:

$$\mathbf{M} = \begin{pmatrix} \nabla f_1(\mathbf{x}) \\ \nabla f_2(\mathbf{x}) \\ \vdots \\ \nabla f_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} D_{x_1} f_1(\mathbf{x}) & D_{x_2} f_1(\mathbf{x}) & \ldots & D_{x_n} f_1(\mathbf{x}) \\ D_{x_1} f_2(\mathbf{x}) & D_{x_2} f_2(\mathbf{x}) & \ldots & D_{x_n} f_2(\mathbf{x}) \\ \vdots & & & \\ D_{x_1} f_m(\mathbf{x}) & D_{x_2} f_m(\mathbf{x}) & \ldots & D_{x_n} f_m(\mathbf{x}) \end{pmatrix}. \tag{2.3}$$

La matrice (2.3) prende il nome di *matrice jacobiana*[5] di $\mathbf{f}$ in $\mathbf{x}$ e si indica con i simboli

$$\mathbf{J}_f(\mathbf{x}), \quad \mathbf{Df}(\mathbf{x}), \quad \mathbf{f}'(\mathbf{x}).$$

Scrivendo la (2.1) per f_k, $k = 1, \ldots, m$, si vede che $\mathbf{f}$ *è differenziabile in* $\mathbf{x}$ *se e solo se lo è ognuna delle sue componenti.* Continua dunque a valere il Teorema 1.2.

Ragionando sulle componenti è poi possibile definire derivate e differenziali di ordine superiore e verificare la validità dei Teoremi 1.3 e 1.4. Per quanto riguarda la formula di Taylor, per $\mathbf{f}$ vettore continua a valere il Teorema 1.8; non può invece essere valido il Teorema 1.5 (e di conseguenza anche il teorema del valor medio) come mostra il seguente semplice esempio.

Esempio 2.1 Sia $\mathbf{f}(x) = (\cos x, \sin x)$, $x \in [0, 2\pi]$. Allora $\mathbf{f}(2\pi) - \mathbf{f}(0) = \mathbf{0}$ ma non esiste alcun $\overline{x} \in [0, 2\pi]$ per cui $\mathbf{f}'(\overline{x}) = (-\sin \overline{x}, \cos \overline{x})$ sia nullo.

Concludiamo il paragrafo con una serie di esempi di funzioni a valori vettoriali evidenziandone il significato geometrico e fisico.

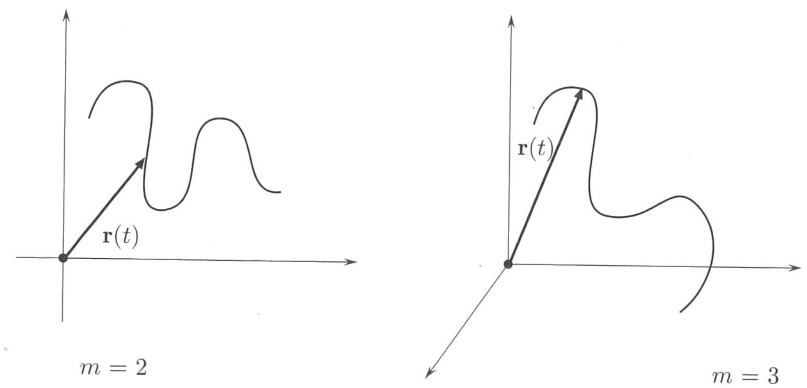

Figura 7.8. Curve nel piano e nello spazio.

Esempio 2.2 *Funzioni vettoriali di variabile reale.* Sia $I \subseteq \mathbb{R}$ un intervallo. Per funzioni da I in $\mathbb{R}^m$ conviene usare t come variabile indipendente e $\mathbf{r}(t) = (x_1(t), x_2(t), \ldots, x_m(t))$ come notazione per la funzione e le sue componenti. Nei casi $n = 2$

[5] Dal nome del matematico Carl Gustave Jacobi.

e $n = 3$ useremo la notazione più comune $\mathbf{r}(t) = (x(t),\, y(t))$ e $\mathbf{r}(t) = (x(t),\, y(t),\, z(t))$, rispettivamente.

In questi casi $\mathbf{r}(t)$ è interpretabile come posizione di un punto al tempo t nel piano o nello spazio rispettivamente; al variare di t in I, $\mathbf{r}(t)$ descrive una "linea" (Fig. 7.8).

Per esempio:

$$\mathbf{r}(t) = (x_0 + \alpha t, y_0 + \beta t, z_0 + \gamma t), \quad t \in \mathbb{R}, \quad \alpha^2 + \beta^2 + \gamma^2 = 1,$$

al variare di t in $\mathbb{R}$ descrive una retta passante per (x_0, y_0, z_0), con coseni direttori α, β, γ.

$\mathbf{r}(t) = (x_0 + R\cos t,\ y_0 + R\sin t)$, $t \in [0, 2\pi]$, $R > 0$, descrive una circonferenza con centro (x_0, y_0) e raggio R.

$\mathbf{r}(t) = (x_0 + R\cos t,\ y_0 + R\sin t,\ kt)$, $t \in \mathbb{R}$, $R > 0$, $k \in \mathbb{R}$, descrive un'elica cilindrica (Fig. 7.9).

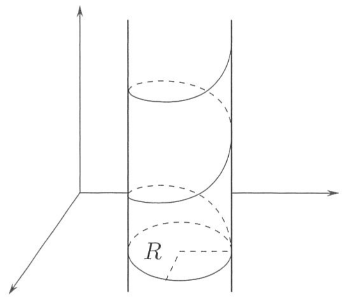

Figura 7.9. Elica cilindrica.

$\mathbf{r}(t)$ è differenziabile se tutte le sue componenti $x_1, \ldots, x_n$ sono derivabili. La matrice jacobiana di $\mathbf{r}$ è il vettore

$$\mathbf{r}'(t) = (x_1'(t), x_2'(t), \ldots, x_n'(t)).$$

Cinematicamente $\mathbf{r}'(t)$ rappresenta la velocità del punto mentre $|\mathbf{r}'(t)|$ è la velocità scalare. Così $\mathbf{r}''(t)$ sarà la sua accelerazione.

Geometricamente $\mathbf{r}'(t)$ è un vettore *tangente* alla curva nel punto $\mathbf{r}(t)$ come è illustrato in Figura 7.10 nel caso $m = 2$.

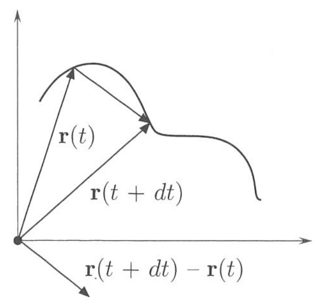

Figura 7.10. Per $dt \to 0$ il vettore $\frac{\mathbf{r}(t+dt)-\mathbf{r}(t)}{dt}$ tende a una posizione limite, tangente alla linea in $\mathbf{r}(t)$.

Se $\mathbf{r}$, $\mathbf{u}$: $I \to \mathbb{R}^m$ sono differenziabili in I, risulta differenziabile anche la funzione $f(t) = \langle \mathbf{r}(t), \mathbf{u}(t) \rangle$ da I in $\mathbb{R}$, prodotto scalare di $\mathbf{r}$ e $\mathbf{u}$; infatti

$$f(t) = \sum_{j=1}^{n} x_j(t) u_j(t) \quad (\mathbf{u}(t) = (u_1(t) \ldots u_m(t)).$$

Vale la formula, come facilmente si verifica

$$f'(t) = \langle \mathbf{r}'(t), \mathbf{u}(t) \rangle + \langle \mathbf{r}(t), \mathbf{u}'(t) \rangle$$

che generalizza la formula per la derivata del prodotto di due funzioni reali.

Torneremo estesamente sulle funzioni da I in $\mathbb{R}^3$ nel secondo volume.

Esempio 2.3 *Funzioni da $\mathbb{R}^2$ in $\mathbb{R}^3$.* Sia $A \subseteq \mathbb{R}^2$, A aperto. Per le funzioni da A in $\mathbb{R}^3$ usiamo la notazione $\mathbf{r}(u, v) = (x(u, v), y(u, v), z(u, v))$.

Al variare di (u, v) in A, $\mathbf{r}(u, v)$ descrive una "superficie"[6] nello spazio (Fig. 7.11).

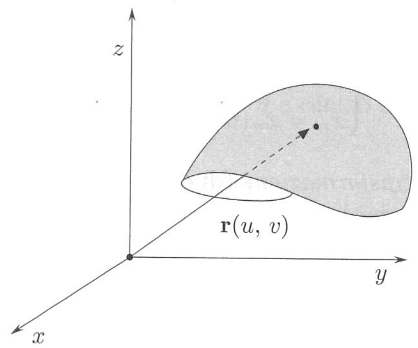

Figura 7.11. Il vettore $\mathbf{r} = \mathbf{r}(u, v)$ descrive una superficie.

Se $\mathbf{r}$ è differenziabile in A, la sua matrice jacobiana è:

$$\mathbf{r}' = \begin{pmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{pmatrix}.$$

I vettori $\mathbf{r}_u = (x_u, y_u, z_u)$ ed $\mathbf{r}_v = (x_v, y_v, z_v)$ sono vettori tangenti alle linee coordinate $\mathbf{r}(u, \overline{v})$ e $\mathbf{r}(\overline{u}, v)$ sulla superficie, corrispondenti a $\overline{v}$ = costante e $\overline{u}$ = costante nel piano (u, v) (Fig. 7.12).

Sulla teoria delle superfici ritorneremo ampiamente in seguito.

[6] I concetti di linea e di superficie verranno trattati rigorosamente in seguito.

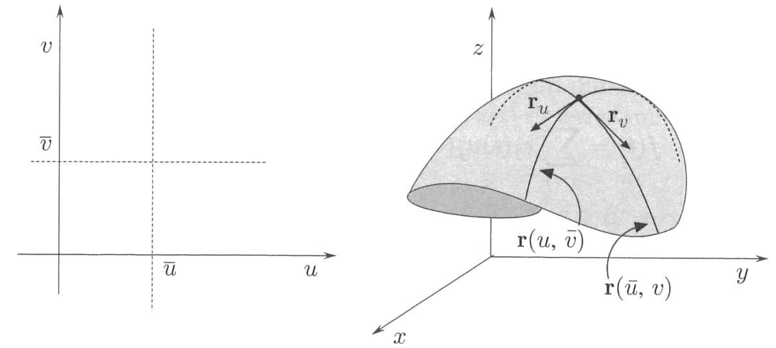

Figura 7.12. Linee coordinate su una superficie.

Esempio 2.4 *Funzioni da* $\mathbb{R}^n$ *in* $\mathbb{R}^n$. Se si scrive $\mathbf{f} = (x_1(u_1, \ldots, u_n), x_2(u_1, \ldots, u_n), \ldots, x_n(u_1, \ldots, u_n))$, $\mathbf{f}$ è interpretabile come un cambiamento di variabili in $\mathbb{R}^n$; in questo caso si preferisce scrivere

$$\begin{cases} x_1 = x_1(u_1, u_2, \ldots, u_n) \\ x_2 = x_2(u_1, u_2, \ldots, u_n) \\ \vdots \\ x_n = x_n(u_1, u_2, \ldots, u_n) \end{cases} \tag{2.4}$$

e si usa per $\mathbf{f}$ il termine *trasformazione* (di coordinate) in $\mathbb{R}^n$.

Un esempio è costituito dai cambiamenti di variabili lineari, esprimibili nella forma

$$\mathbf{x} = \mathbf{M}\mathbf{u} \text{ (vettori colonna) oppure } \mathbf{x} = \mathbf{u}\mathbf{M}^\top \text{ (vettori riga)}, \tag{2.5}$$

dove $\mathbf{M}$ è una matrice quadrata di ordine n; $\mathbf{M}^\top$ indica la matrice trasposta di $\mathbf{M}$, cioè la matrice che si ottiene da $\mathbf{M}$ scambiando le righe con le colonne.

In questo caso la matrice jacobiana della trasformazione coincide con $\mathbf{M}$, infatti se m_{jk} sono gli elementi di $\mathbf{M}$, la (2.5) implica che

$$x_j = \sum_{k=1}^{n} m_{jk} u_k$$

e quindi $\partial x_j / \partial u_k = m_{jk}$.

Si noti che la (2.5) è $1 \leftrightarrow 1$ da $\mathbb{R}^n$ in $\mathbb{R}^n$ se e solo se $\mathbf{M}$ è non singolare (cioè $\det \mathbf{M} \neq 0$) ovvero se e solo se la matrice jacobiana della trasformazione è non singolare.

Anche per una trasformazione generale come la (2.4) è importante, come vedremo, il determinante della matrice jacobiana (o *determinante jacobiano*); per esso è preferibile usare la notazione

$$\frac{\partial(x_1, x_2, \ldots, x_n)}{\partial(u_1, u_2, \ldots, u_n)}.$$

Esempi notevoli di trasformazioni sono i seguenti.

Esempio 2.5 Trasformazioni lineari particolarmente importanti sono quelle rappresentate da una matrice $\mathbf{M}$ *ortogonale*, cioè tale che $\mathbf{M}^{-1} = \mathbf{M}^\top$.

In altri termini $\mathbf{M} \cdot \mathbf{M}^\top = \mathbf{M}^\top \cdot \mathbf{M} = \mathbf{I}_n$ dove $\mathbf{I}_n$ è la matrice identità di ordine n:

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Esempi di matrici ortogonali nel piano sono

$$\mathbf{M}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{e} \quad \mathbf{M}_2 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

La prima rappresenta una riflessione rispetto alla bisettrice $x_1 = x_2$; infatti

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix};$$

la seconda una rotazione antioraria di angolo θ e centro nell'origine (Fig. 7.13).

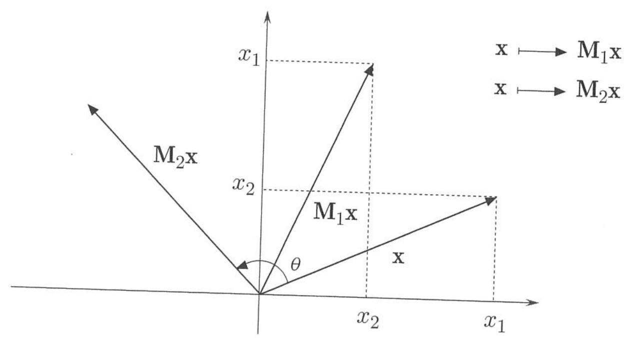

Figura 7.13. Riflessione rispetto alla bisettrice e notazione antioraria di un angolo θ.

Notevoli proprietà delle matrici ortogonali sono le seguenti (lasciamo la verifica al lettore):

i) $\det \mathbf{M} = \pm 1$;

ii) l'insieme dei vettori riga (colonna) costituisce una base ortogonale per $\mathbb{R}^n$;

iii) $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\langle \mathbf{M}\mathbf{x}, \mathbf{M}\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$.

e in particolare $|\mathbf{M}\mathbf{x}| = |\mathbf{x}|$ (la lunghezza di un vettore non varia per trasformazioni ortogonali).

Esempi

2.6 *Coordinate polari nel piano rispetto al polo* (x_0, y_0).

$$\begin{cases} x = x_0 + \rho\cos\theta \\ y = y_0 + \rho\sin\theta \end{cases} \qquad (\rho,\theta) \in (0,+\infty) \times [0, 2\pi).$$

Si ha:

$$\frac{\partial(x,y)}{\partial(\rho,\theta)} = \begin{vmatrix} x_\rho & x_\theta \\ y_\rho & y_\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix} = \rho.$$

2.7 *Coordinate cilindriche nello spazio rispetto all'asse* $x = x_0$, $y = y_0$.

$$\begin{cases} x = x_0 + \rho\cos\theta \\ y = y_0 + \rho\sin\theta \\ z = t \end{cases} \qquad (\rho,\theta,t) \in (0,+\infty) \times [0, 2\pi) \times \mathbb{R}.$$

Si ha:

$$\frac{\partial(x,y,z)}{\partial(\rho,\theta,t)} = \begin{vmatrix} x_\rho & x_\theta & x_t \\ y_\rho & y_\theta & y_t \\ z_\rho & z_\theta & z_t \end{vmatrix} = \begin{vmatrix} \cos\theta & -\rho\sin\theta & 0 \\ \sin\theta & \rho\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = \rho.$$

2.8 *Coordinate polari o sferiche nello spazio rispetto al polo* $(0,0,0)$.

$$(\rho,\psi,\theta) \in (0,+\infty) \times [0,\pi] \times [0,2\pi).$$

$$\begin{cases} x = \rho\sin\psi\cos\theta \\ y = \rho\sin\psi\sin\theta \\ z = \rho\cos\psi \end{cases}.$$

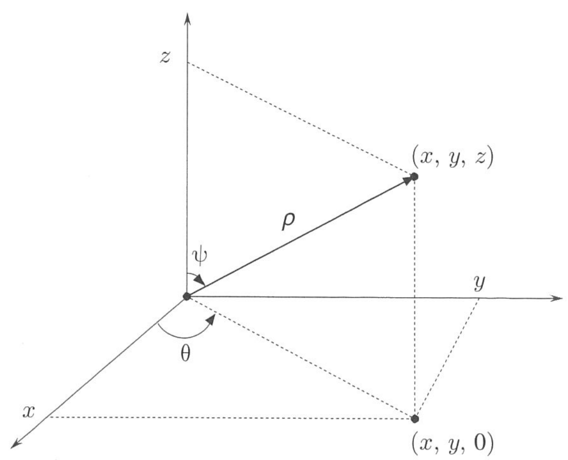

Figura 7.14. Coordinate sferiche.

Si ha:

$$\frac{\partial(x,y,z)}{\partial(\rho,\psi,\theta)} = \begin{vmatrix} \sin\psi\cos\theta & \rho\cos\psi\cos\theta & -\rho\sin\psi\sin\theta \\ \sin\psi\sin\theta & \rho\cos\psi\sin\theta & \rho\sin\psi\cos\theta \\ \cos\psi & -\rho\sin\psi & 0 \end{vmatrix} = \rho^2\sin\psi.$$

2.9 *Ancora funzioni da* $\mathbb{R}^n$ *a* $\mathbb{R}^n$. Oltre che come trasformazioni di variabili, le funzioni $\mathbf{f} : \mathbb{R}^n \supseteq A \to \mathbb{R}^n$ si possono interpretare come *campi di vettori*.

Un esempio per $n = 2$ è il campo di velocità per il moto piano stazionario di un fluido:

$$\mathbf{v}(x,y) = (v_1(x,y), v_2(x,y)),$$

che assegna la velocità della particella di fluido che si trova nel punto di coordinate (x,y).

Un altro esempio è il campo elettrostatico

$$\mathbf{E}(x,y,z) = (E_1(x,y,z), E_2(x,y,z), E_3(x,y,z)).$$

Per i campi vettoriali derivabili è definito un operatore differenziale che si chiama *divergenza* e si indica con "div" e che in coordinate cartesiane è assegnato dalla formula

$$\operatorname{div}\mathbf{f} := \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \ldots + \frac{\partial f_n}{\partial x_n}.$$

Un'altra notazione usata per la divergenza di $\mathbf{f}$ è $\nabla \cdot \mathbf{f}$.

Ai campi vettoriali $\mathbf{f}$ derivabili in $\mathbb{R}^3$ si può associare un altro vettore, il *rotore* di $\mathbf{f}$, denotato con $\operatorname{rot}\mathbf{f}$, $\operatorname{curl}\mathbf{f}$ oppure $\nabla \times \mathbf{f}$, le cui componenti in coordinate cartesiane sono le seguenti:

$$\operatorname{rot}\mathbf{f} := \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}, \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}, \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right).$$

I campi vettoriali a divergenza nulla sono detti *solenoidali*; quelli a rotore nullo, *irrotazionali*.

L'importanza di divergenza e rotore sarà più chiara in seguito.

2.2 Differenziale delle funzioni composte

Vediamo ora come si comporta il differenziale rispetto all'operazione di composizione.

Vale il seguente importante risultato.

■ **TEOREMA 2.1** *Siano* $A \subseteq \mathbb{R}^n$, $B \subseteq \mathbb{R}^m$ *aperti e*

$$\mathbf{f} : A \to B, \quad \mathbf{g} : B \to \mathbb{R}^p.$$

Se $\mathbf{f}$ *è differenziabile in* $\mathbf{x} \in A$ *e* $\mathbf{g}$ *è differenziabile in* $\mathbf{f}(\mathbf{x})$, *allora la funzione composta* $\mathbf{g} \circ \mathbf{f} : A \to \mathbb{R}^p$ *è differenziabile in* $\mathbf{x}$ *e vale la formula:*

$$\mathbf{J}_{\mathbf{g}\circ\mathbf{f}}(\mathbf{x}) = \mathbf{J}_{\mathbf{g}}(\mathbf{f}(\mathbf{x})) \cdot \mathbf{J}_{\mathbf{f}}(\mathbf{x}). \tag{2.6}$$

Notiamo che $\mathbf{J}_{\mathbf{g}\circ\mathbf{f}}$ è di ordine $p \times n$, $\mathbf{J}_{\mathbf{g}}$ di ordine $p \times m$ e $\mathbf{J}_{\mathbf{f}}$ di ordine $m \times n$.

La dimostrazione del teorema ricalca quella del caso $n = m = p = 1$, per cui la omettiamo. Ricaviamo invece dalla (2.6) una formula esplicita per le derivate di ogni componente di $\mathbf{g} \circ \mathbf{f}$.

A tale scopo poniamo

$$\mathbf{y} = \mathbf{f}(\mathbf{x}), \quad \varphi(\mathbf{x}) = (\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = \mathbf{g}(\mathbf{y}).$$

Dalla (2.6) si ricava:

$$\frac{\partial \varphi_k}{\partial x_j} = \sum_{i=1}^{m} \frac{\partial g_k}{\partial y_i} \cdot \frac{\partial y_i}{\partial x_j} \qquad \begin{array}{l} k = 1, \ldots, p \\ j = 1, \ldots, n \end{array}. \tag{2.7}$$

Usando le altre notazioni per la matrice jacobiana si può scrivere la (2.6) in una delle forme seguenti

$$\mathbf{D}\varphi(\mathbf{x}) = \mathbf{D}\mathbf{g}(\mathbf{y}) \cdot \mathbf{D}\mathbf{f}(\mathbf{x})$$

oppure

$$\varphi'(\mathbf{x}) = \mathbf{g}'(\mathbf{y}) \cdot \mathbf{f}'(\mathbf{x}).$$

Esempio 2.10 Sia $f : \mathbb{R}^2 \backslash \{0, 0\} \to \mathbb{R}$, differenziabile; se $f = f(x, y)$ poniamo $\tilde{f}(\rho, \theta) = f(\rho \cos\theta, \rho \sin\theta)$, effettuando così un passaggio a coordinate polari.

Vogliamo esprimere le derivate di $\tilde{f}$ rispetto a ρ e θ mediante quelle di f rispetto a x e y.

Dalla (2.7) abbiamo:

$$\frac{\partial \tilde{f}}{\partial \rho} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial \rho} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial \rho} = \frac{\partial f}{\partial x} \cos\theta + \frac{\partial f}{\partial y} \sin\theta$$

$$\frac{\partial \tilde{f}}{\partial \theta} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial \theta} = \frac{\partial f}{\partial x}(-\rho \sin\theta) + \frac{\partial f}{\partial y}(\rho \cos\theta),$$

dove le derivate parziali di f si intendono calcolate in $x = \rho \cos\theta$ e $y = \rho \sin\theta$.

Le stesse formule si trovano applicando direttamente la (2.6), tenendo presente che

$$\begin{pmatrix} \cos\theta & -\rho \sin\theta \\ \sin\theta & \rho \cos\theta \end{pmatrix}$$

è la matrice jacobiana della trasformazione in polari:

$$\left(\frac{\partial \tilde{f}}{\partial \rho}, \frac{\partial \tilde{f}}{\partial \theta} \right) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \cdot \begin{pmatrix} \cos\theta & -\rho \sin\theta \\ \sin\theta & \rho \cos\theta \end{pmatrix}.$$

Esempio 2.11 Sia $f : \mathbb{R}^n \to \mathbb{R}$ due volte differenziabile. Operiamo il cambio di variabile $\mathbf{x} = \mathbf{M}\mathbf{y}$ (vettori colonna) dove $\mathbf{M}$ è una matrice quadrata di ordine n.

Poniamo

$$\tilde{f}(\mathbf{y}) = f(\mathbf{M}\mathbf{y}).$$

Vogliamo stabilire quale relazione sussiste tra le matrici hessiane di f ed $\tilde{f}$.

Usando la formula (2.7) abbiamo (ricordando l'Esempio 2.4):

$$\frac{\partial \tilde{f}}{\partial y_j}(\mathbf{y}) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{x}) \cdot \frac{\partial x_k}{\partial y_j} = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{x}) m_{kj}.$$

Riapplicando la stessa formula per $\frac{\partial \tilde{f}}{\partial y_i}$:

$$\frac{\partial^2 \tilde{f}}{\partial y_i \partial y_j}(\mathbf{y}) = \sum_{s=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_s \partial x_k}(\mathbf{x}) m_{si} m_{kj}. \tag{2.8}$$

Se $\mathbf{M}^\top$ indica la trasposta di $\mathbf{M}$, la (2.8) si riscrive nella forma seguente:

$$\mathbf{H}_{\tilde{f}}(\mathbf{y}) = \mathbf{M}^\top \cdot \mathbf{H}_f(\mathbf{x}) \cdot \mathbf{M}, \tag{2.9}$$

che è la relazione cercata.

2.3 Funzioni da $\mathbb{C}$ in $\mathbb{C}$

In questo paragrafo ci occuperemo di funzioni complesse di variabile complessa. Sia dunque $w = f(z)$ una funzione definita in un aperto $A \subseteq \mathbb{C}$ a valori in $\mathbb{C}$.

Introducendo parte reale e parte immaginaria di z ed f possiamo identificare $\mathbb{C}$ con $\mathbb{R}^2$ e considerare f come una funzione da $\mathbb{R}^2$ in $\mathbb{R}^2$. In altri termini, posto $z = x + iy$, $u(x, y) = \operatorname{Re} f(z)$, $v(x, y) = \operatorname{Im} f(z)$ avremo $f(z) = \mathbf{F}(x, y) = (u(x, y), v(x, y))$. Per esempio se $f(z) = z^2$, $u(x, y) = x^2 - y^2$ e $v(x, y) = 2xy$.

Per $\mathbf{F}$ sono dunque definite tutte le nozioni introdotte nei paragrafi precedenti e in particolare la differenziabilità.

D'altra parte l'identificazione tra $\mathbb{C}$ ed $\mathbb{R}^2$ è possibile solo a livello di spazio vettoriale metrico (su $\mathbb{R}$) mentre sappiamo che $\mathbb{C}$ è dotato altresì della struttura di campo, che permette moltiplicazioni e divisioni tra numeri complessi.

È quindi ragionevole pensare di estendere la nozione di derivata, così come è stata introdotta per le funzioni reali di *una* variabile reale, alle funzioni complesse di *una* variabile complessa.

Definizione 2.2 *Sia $f : \mathbb{C} \supseteq A \to \mathbb{C}$, A aperto e $z_0 \in A$. Se esiste finito il limite:*

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} \qquad (z \in A)$$

questo viene detto derivata complessa di f in z_0 e indicato con uno dei simboli $f'(z_0)$ o $Df(z_0)$. In tal caso la funzione f si dice derivabile in senso complesso in z_0.

Se f ha derivata complessa in ogni punto di A, f si dice **olomorfa** in A.

Esempio 2.12 Sia $f(z) = z^n$. Allora

$$\frac{f(z) - f(z_0)}{z - z_0} = z^{n-1} + z^{n-2} z_0 + \cdots + z z_0^{n-2} + z_0^{n-1} \longrightarrow n z_0^{n-1}$$

per $z \to z_0$ e quindi $f'(z_0) = n z_0^{n-1}$, esattamente come nel caso reale.

La funzione $f(z) = z^n$ è dunque olomorfa in $\mathbb{C}$.