APOGEO education

Matematica

per le scienze economiche e sociali

Algebra lineare, funzioni di più variabili e ottimizzazione statica

Claudio Mattalia
Fabio Privileggi

MAGGIOLI EDITORE

PARTE II

CALCOLO DIFFERENZIALE E OTTIMIZZAZIONE STATICA

Capitolo 4

Funzioni in più variabili e calcolo differenziale

Questo capitolo esordisce con alcuni richiami alle proprietà *topologiche* degli spazi con dimensione maggiore di 1. Ci occuperemo innanzitutto di "punti delicati" (di accumulazione, interni ecc.) e di alcuni insiemi particolari (intorni, insiemi compatti ecc.) estendendo concetti già visti nel primo volume per le funzioni di una variabile. Introdurremo quindi formalmente la nozione di *funzione di più variabili* e vedremo che se $n = 2$, cioè se ci sono solo due variabili indipendenti, è ancora possibile una rappresentazione grafica della funzione. Per questo motivo nel seguito del testo soffermeremo la nostra attenzione soprattutto sui fenomeni descrivibili mediante *funzioni di due variabili*.

Successivamente ci dedicheremo alla generalizzazione dello strumento *derivata*, già ampiamente studiato nel primo volume per quanto riguarda le funzioni di una sola variabile, al mondo delle *funzioni reali di n variabili*. Ci occuperemo inoltre di alcuni accessori specifici del caso in più variabili, strumenti che non assumono significato nel caso di funzioni di una variabile. Rassicuriamo il lettore: le novità rispetto al primo volume riguardano esclusivamente aspetti teorici e interpretativi, l'utilizzo pratico dei nuovi strumenti richiederà semplicemente il calcolo di derivate in una sola variabile, regole che abbiamo già digerito; non sarà dunque necessario imparare nuove formule complicate.

La speciale attenzione dedicata alle funzioni di due variabili non è solo giustificata dall'avere a disposizione una rappresentazione grafica esplicativa, ma anche dal fatto che la generalizzazione alle funzioni di $n > 2$ variabili è puramente formale e non aggiunge nulla di concettualmente importante. Il salto di qualità vero e proprio è passare dalle funzioni di una sola variabile alle funzioni di due variabili: le seconde introducono delle novità sostanziali rispetto alle prime, laddove aumentare il numero di variabili (ovvero la dimensione dello spazio euclideo che contiene il loro dominio) a più di due richiede semplicemente l'applicazione degli stessi criteri un numero maggiore di volte, con in più lo svantaggio di non poter 'visualizzare' nulla.

4.1 Sottoinsiemi particolari di $\mathbb{R}^n$

Studiamo ora alcuni importanti sottoinsiemi di $\mathbb{R}^n$ le cui proprietà torneranno utili nel seguito. Vedremo prima la struttura topologica di $\mathbb{R}^n$, poi introdurremo il concetto di insieme convesso.

4.1.1 Elementi di topologia

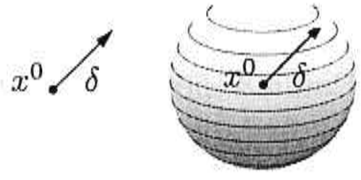
Utilizzeremo il concetto di *intorno* per costruire la *topologia* standard di $\mathbb{R}^n$. Tutte le definizioni di seguito generalizzano quelle già viste per $\mathbb{R}$ nel primo volume.

Notazione 4.1 Per coerenza con la Notazione 3.1 usata in precedenza, indicheremo con $x^0 \in \mathbb{R}^n$, con lo zero ad apice anziché a pedice, un *punto di riferimento* dello spazio $\mathbb{R}^n$ con coordinate $x_1^0, x_2^0, \ldots, x_n^0$. La scrittura '$x_0$' potrebbe creare confusione fra il punto stesso (elemento di $\mathbb{R}^n$) e le sue coordinate (elementi di $\mathbb{R}$).

Definizione 4.1 (Intorno circolare) Si dice **intorno (circolare)** di centro $x^0 \in \mathbb{R}^n$ e raggio $\delta > 0$ ($\delta \in \mathbb{R}$) l'insieme:

$$I_{x^0}^{\delta} = \{x \in \mathbb{R}^n : \|x - x^0\| < \delta\} \tag{4.1}$$

La definizione è analoga a quella degli intorni in $\mathbb{R}$ – in cui la norma euclidea prende il posto del valore assoluto – e anche l'interpretazione è la stessa: $I_{x^0}^{\delta}$ è l'*insieme dei punti di* $\mathbb{R}^n$ *che distano da* x^0 *meno di* δ. La Figura 4.1 rappresenta graficamente un intorno di un punto x^0 rispettivamente (a) in $\mathbb{R}^2$ e (b) in $\mathbb{R}^3$; nel primo caso è un *cerchio* con centro $x^0 \in \mathbb{R}^2$ e raggio δ che non contiene la circonferenza $\{x \in \mathbb{R}^2 : \|x - x^0\| = \delta\}$, mentre nel secondo caso è una *palla* (piena) che non contiene la superficie sferica $\{x \in \mathbb{R}^3 : \|x - x^0\| = \delta\}$ (il suo involucro, la 'buccia'...). Come gli intorni di $\mathbb{R}$, anche in $\mathbb{R}^n$ l'intorno è un *insieme aperto*: *non contiene la frontiera*.



(a) (b)

Figura 4.1: intorni, (a) in $\mathbb{R}^2$ e (b) in $\mathbb{R}^3$.

Definizione 4.2

1. $x^0 \in \mathbb{R}^n$ si dice **punto di accumulazione** per l'insieme $X \subseteq \mathbb{R}^n$ se ogni suo intorno $I^\delta_{x^0}$ contiene punti di X diversi da x^0, cioè se ogni $I^\delta_{x^0}$ è tale che $(I^\delta_{x^0} \setminus \{x^0\}) \cap X \neq \varnothing$.
2. $x^0 \in \mathbb{R}^n$ si dice **punto interno** all'insieme $X \subseteq \mathbb{R}^n$ se esiste un suo intorno $I^\delta_{x^0}$ tutto contenuto in X [indicheremo con $\mathrm{Int}\,(X)$ l'**insieme dei punti interni** di X].
3. $x^0 \in \mathbb{R}^n$ si dice **punto esterno** all'insieme $X \subseteq \mathbb{R}^n$ se è interno al complementare X^c (cioè se $x^0 \in \mathrm{Int}\,(X^c)$).
4. $x^0 \in \mathbb{R}^n$ si dice **punto di frontiera** per l'insieme $X \subseteq \mathbb{R}^n$ se ogni suo intorno $I^\delta_{x^0}$ contiene contemporaneamente punti di X e punti del suo complemento X^c, cioè se ogni $I^\delta_{x^0}$ è tale che $(I^\delta_{x^0} \cap X \neq \varnothing) \wedge (I^\delta_{x^0} \cap X^c \neq \varnothing)$.
5. $x^0 \in \mathbb{R}^n$ si dice **punto isolato** dell'insieme $X \subseteq \mathbb{R}^n$ se esiste $I^\delta_{x^0}$ tale che $I^\delta_{x^0} \cap X = \{x^0\}$.

La Figura 4.2 illustra tre casi in $\mathbb{R}^2$, evidenziando il ruolo dell'intorno nella definizione di ciascun punto notevole.

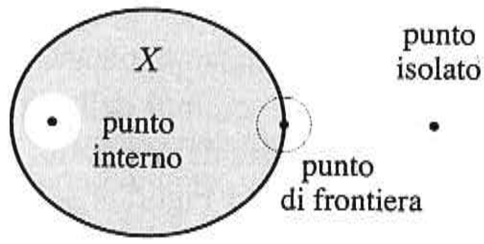


FIGURA 4.2: punti interni, di frontiera e isolati.

Definizione 4.3

1. $X \subseteq \mathbb{R}^n$ si dice **insieme aperto** se contiene solo punti interni (non contiene punti di frontiera), cioè se $X = \mathrm{Int}\,(X)$.
2. $X \subseteq \mathbb{R}^n$ si dice **insieme chiuso** se contiene *tutti* i suoi punti di frontiera.

Si vede immediatamente che

$$X \text{ aperto} \iff X^c \text{ chiuso}$$

La Figura 4.3 mostra qualche esempio in $\mathbb{R}^2$: in (a) è raffigurata un'ellisse con incluso il bordo, si tratta quindi di un insieme chiuso; in (b) il rettangolo $X = \{x = (x_1, x_2) \in \mathbb{R}^2 : (a \leq x_1 < b) \wedge (c \leq x_2 < d)\}$, che contiene solo i due lati di sud-ovest, e non è né aperto né chiuso; in (c) è riportato l'intorno di un punto x^0, che è aperto per definizione.

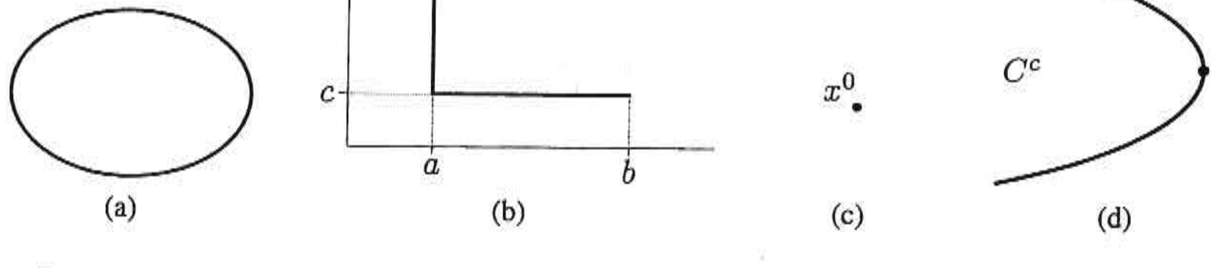


FIGURA 4.3: (a) insieme chiuso; (b) insieme né aperto né chiuso; (c) insieme aperto; (d) una curva (l'insieme C) è un insieme chiuso perché contiene solo punti di frontiera.

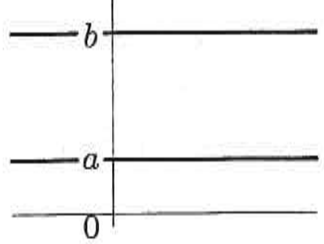


FIGURA 4.4: insieme chiuso ma non limitato (non compatto).

Le *curve nel piano* $\mathbb{R}^2$ sono esempi particolari (e importanti) di insiemi chiusi: esse sono costituite *esclusivamente da punti di frontiera*, non possiedono punti interni. In altre parole, una curva $C \subset \mathbb{R}^2$ contiene tutti i suoi punti di frontiera (è un insieme chiuso) ed è tale che $\mathrm{Int}\,(C) = \varnothing$. La Figura 4.3(d) mostra che qualsiasi intorno di un punto di una curva necessariamente contiene contemporaneamente punti della curva stessa e punti che non stanno sulla curva, ovvero punti del suo complemento C^c.

DEFINIZIONE 4.4 (INSIEME LIMITATO) $X \subset \mathbb{R}^n$ si dice **insieme limitato** se esiste un intorno che lo contiene (cioè non "scappa" verso infinito in alcuna direzione).

DEFINIZIONE 4.5 (INSIEME COMPATTO) $X \subset \mathbb{R}^n$ si dice **insieme compatto** se è chiuso e limitato.

I primi tre esempi della Figura 4.3 sono insiemi limitati, ma solamente il primo, in (a), è un insieme compatto; l'ultimo caso, in Figura 4.3(d), è un insieme chiuso ma non limitato. La Figura 4.4 mostra un altro esempio di insieme di $\mathbb{R}^2$ chiuso ma non compatto: l'insieme $X = \{x = (x_1, x_2) \in \mathbb{R}^2 : a \leq x_2 \leq b\}$, rappresentato dall'area grigia, è chiuso ma non è limitato in quanto contiene punti di $\mathbb{R}^2$ la cui prima coordinata, x_1, tende a $\pm\infty$. Come per l'ottimizzazione in una variabile, la nozione di compattezza giocherà un ruolo fondamentale per l'esistenza del massimo e del minimo assoluto anche per le funzioni di n variabili, come vedremo nei prossimi capitoli.

4.1.2 Insiemi convessi

Introduciamo una tipologia di insiemi che avevamo già incontrato, senza chiamarli specificamente in questo modo, studiando la retta reale $\mathbb{R}$: gli insiemi convessi. Essi assumono pieno significato solamente in $\mathbb{R}^n$, con $n \geq 2$, e saranno un ingrediente necessario per definire funzioni di più variabili concave e convesse.

DEFINIZIONE 4.6 (INSIEME CONVESSO) $X \subseteq \mathbb{R}^n$ si dice **insieme convesso** se per ogni coppia di punti $x, y \in X$ il segmento congiungente x a y è tutto contenuto in X.

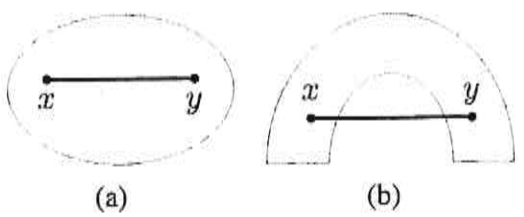


FIGURA 4.5: (a) insieme convesso; (b) insieme non convesso.

La Figura 4.5(a) illustra un esempio di insieme convesso in $\mathbb{R}^2$, mentre la Figura 4.5(b) mostra un esempio di insieme non convesso in cui la Definizione 4.6 è violata perché un tratto del segmento congiungente i punti x e y "esce" dall'insieme stesso.

Applicando la Definizione 4.6 agli insiemi della retta reale $\mathbb{R}$, ci rendiamo conto che *gli insiemi* *convessi di $\mathbb{R}$ sono tutti e soli gli intervalli* (cioè tutti e soli gli insiemi di $\mathbb{R}$ che non hanno "buchi"). In $\mathbb{R}$, dunque, non è necessaria la definizione di convessità: la nozione di intervallo è più che sufficiente. Diversamente, in $\mathbb{R}^n$, per $n \geq 2$, la convessità assume una connotazione ben precisa che consentirà di definire le funzioni concave (convesse).

OSSERVAZIONE 4.1

1. La convessità è del tutto indipendente dalla limitatezza e dalla compattezza: insiemi convessi possono essere chiusi o aperti, limitati o illimitati, mentre insiemi compatti possono essere convessi o non convessi.
2. Non confondere *insiemi convessi* con *funzioni convesse*: sono oggetti ben diversi!

ESEMPIO 4.1 Illustriamo graficamente le proprietà di chiusura, limitatezza e convessità dei seguenti insiemi dello spazio bidimensionale $\mathbb{R}^2$.

1. Si verifica facilmente che **l'ortante positivo** di $\mathbb{R}^2$, cioè l'insieme $\mathbb{R}^2_+ = \{x = (x_1, x_2) \in \mathbb{R}^2 : (x_1 \geq 0) \wedge (x_2 \geq 0)\}$ è *chiuso, illimitato* e *convesso.*
2. Disegnando l'insieme $X = \mathbb{R}^2_+ \cup \{(x_1, x_2) \in \mathbb{R}^2 : (x_1 \leq 0) \wedge (x_2 \leq 0)\}$, si vede agevolmente che si tratta di un insieme *chiuso* e *illimitato* che non è convesso.
3. A partire dal disegno dell'insieme X dell'esempio precedente, verifichiamo che l'insieme $Y = X \cap \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \leq 1 - x_1\} \cap \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \geq -1 - x_1\}$ è *compatto* ma non è *convesso* (suggerimento: le disuguaglianze $x_2 \leq 1 - x_1$ e $x_2 \geq -1 - x_1$ definiscono gli insiemi delimitati dalle funzioni di una variabile $x_2 = f(x_1) = 1 - x_1$ e $x_2 = g(x_1) = -1 - x_1$, i cui grafici sono delle rette).
4. L'insieme $B = \mathbb{R}^2_+ \cap \{(x_1, x_2) \in \mathbb{R}^2 : p_1 x_1 + p_2 x_2 \leq w;\ p_1, p_2, w > 0\}$, dove i parametri p_1, p_2 e w vengono interpretati come il prezzo del bene 1, il prezzo del bene 2 (di cui x_1 e x_2 rappresentano le quantità) e la ricchezza disponibile[1] rispettivamente, è un esempio di **vincolo di bilancio**: l'insieme delle coppie di quantità (x_1, x_2) dei due beni che è possibile acquistare con una disponibilità monetaria pari a w e dati i prezzi dei due beni p_1 e p_2. Si disegni per esercizio il grafico di tale insieme e si verifichi che è *compatto* e *convesso.*
5. L'insieme $\widehat{B} = \mathbb{R}^2_+ \cap \{(x_1, x_2) \in \mathbb{R}^2 : p_1 x_1 + p_2 x_2 = w;\ p_1, p_2, w > 0\}$ è un segmento (con gli estremi inclusi!) in $\mathbb{R}^2$ ed è *compatto* e *convesso.* Lo chiameremo **retta di bilancio**. Dati i prezzi p_1 e p_2 dei due beni, $\widehat{B}$ rappresenta l'insieme delle quantità (x_1, x_2) di beni che è possibile acquistare in modo da spendere l'intera disponibilità monetaria pari a w.
6. Consideriamo l'insieme $X = \{(x_1, x_2) \in \mathbb{R}^2 :\ x_2 > |x_1 - 2| - 3\}$. Disegniamo il grafico della funzione di una variabile $x_2 = f(x_1) = |x_1 - 2| - 3$, corrispondente al bordo dell'area grigia in Figura 4.6(a); la stessa area grigia (escluso il bordo) è l'insieme X, ovvero l'insieme dei punti (x_1, x_2) tali che $x_2 > |x_1 - 2| - 3$. Tale insieme è *convesso, illimitato* (verso l'alto) e *aperto* in quanto nessun punto della frontiera, descritta dal grafico di $f(x_1)$, è contenuto in X.
7. L'insieme $X = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \geq |e^{2x_1} - 2| - 1\}$ è formato da tutti i punti del piano che stanno sopra il grafico della funzione $x_2 = f(x_1) = |e^{2x_1} - 2| - 1$. Disegnando il grafico di tale funzione [la curva nera in Figura 4.6(b)] si vede che

[1] La lettera w abbrevia la parola inglese *wealth*, che significa, appunto, *ricchezza*.

X non è convesso (il segmento AB è esterno a X) e che è *illimitato* verso l'alto. Poiché la disuguaglianza $x_2 \geq \left|e^{2x_1} - 2\right| - 1$ è debole, la frontiera [il grafico di $f(x_1)$] è contenuta in X, che dunque è un insieme *chiuso*.

4.2 Funzioni reali di n variabili

È finalmente giunto il momento di sfruttare la notazione vettoriale per definire le funzioni di n variabili.

DEFINIZIONE 4.7 (FUNZIONE REALE DI n VARIABILI) Si dice che $f : X \to \mathbb{R}$, dove $X \subseteq \mathbb{R}^n$, è una **funzione reale di n variabili** $x_1, x_2, \ldots, x_n$, e si indica con $y = f(x_1, x_2, \ldots, x_n)$, oppure, in forma sintetica, con $y = f(x)$ dove resta inteso che $x \in X \subseteq \mathbb{R}^n$ è un generico *vettore* $x = (x_1, x_2, \ldots, x_n)$, se f rappresenta una legge che associa a ogni punto (vettore) $x \in X$ **uno e un solo** numero reale y.

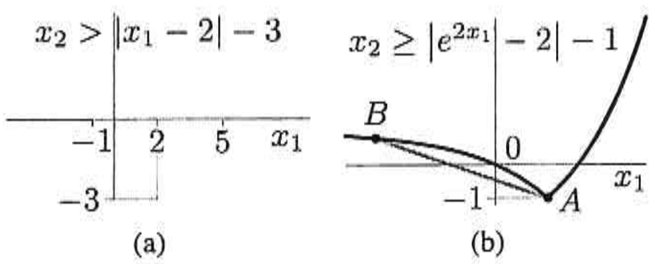


FIGURA 4.6: casi 6 e 7 dell'Esempio 4.1.

In analogia alle funzioni di una variabile, $x_1, x_2, \ldots, x_n$ si dicono *variabili indipendenti*, y si dice *variabile dipendente*, X è il *dominio* (o campo di esistenza) di f, che si indica con $\mathrm{Dom}(f)$ (oppure con CE), e l'insieme dei valori che la variabile dipendente y può assumere si chiama *codominio* o *immagine* e si indica con $\mathrm{Im}(f)$.

La Definizione 4.7 è simile a quella di funzione di una variabile: l'unica differenza è che il *dominio* di f, nel caso di n variabili, è un sottoinsieme dello spazio euclideo $\mathbb{R}^n$ (o al più tutto $\mathbb{R}^n$) anziché essere un sottoinsieme della retta reale $\mathbb{R}$. Si tratta di un meccanismo che trasforma n-uple di numeri (vettori), anziché singoli numeri, in un numero. La Figura 4.7 illustra schematicamente la trasformazione di terne di numeri in numeri di una funzione di tre variabili, $f : \mathbb{R}^3 \to \mathbb{R}$.

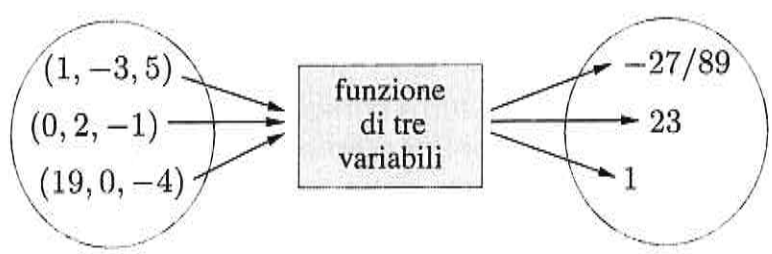


FIGURA 4.7: una funzione di tre variabili trasforma terne di numeri in numeri.

NOTAZIONE 4.2 Per quanto riguarda le **funzioni di due** sole **variabili** spesso si usa la notazione:

$$z = f(x, y)$$

dove x e y sono le variabili indipendenti e z è la variabile dipendente; naturalmente, x e y vanno sempre e comunque interpretate come *la prima e la seconda coordinata di un* *vettore di* $\mathbb{R}^2$. Questa notazione è comunemente adottata nei *testi di economia* e noi ne faremo abbondantemente uso.

Analogamente, alle volte indicheremo con $f(x, y, z)$ una **funzione di tre variabili**, intendendo che x, y e z sono tutte variabili indipendenti, da interpretare, rispettivamente, come *la prima, la seconda e la terza coordinata di un vettore di* $\mathbb{R}^3$.

Ad esempio, $f(x_1, x_2) = \sqrt{x_1} + e^{3x_2}$ è funzione delle due variabili x_1 e x_2, scriveremo quindi $f : \mathbb{R}^2 \to \mathbb{R}$. Più precisamente, scriveremo $f : X \to \mathbb{R}$, dove $X = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0\}$, il dominio di f, è l'unione del primo e quarto ortante di $\mathbb{R}^2$ incluso l'asse delle ordinate. Naturalmente, è più comodo scrivere $f(x, y) = \sqrt{x} + e^{3y}$ e $X = \{(x, y) \in \mathbb{R}^2 : x \geq 0\}$. Analogamente, $f(x_1, x_2, x_3, x_4) = x_1 - \ln(x_2 x_3) + 1/x_4$ è una funzione di quattro variabili, x_1, x_2, x_3 e x_4; scriveremo dunque $f : \mathbb{R}^4 \to \mathbb{R}$, definita su $X = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : (x_2 x_3 > 0) \wedge (x_4 \neq 0)\}$.

4.2.1 Esempi economici

Se x e y rappresentano le quantità di due fattori produttivi (ad esempio, capitale e lavoro) impiegati in un processo di produzione, possiamo indicare con $c(x, y)$ la funzione che associa a ciascuna coppia di fattori (x, y) il loro **costo complessivo**; ad esempio, se $c(x, y) = x^2 + y^2$ e vengono impiegate 2 unità del primo fattore e 3 unità del secondo, il costo complessivo è $c(2, 3) = 2^2 + 3^2 = 13$. Se invece x, y e z rappresentassero le quantità consumate di tre beni (ad esempio, pizza, birra e tiramisù), potremmo indicare con $u(x, y, z)$ il livello di **utilità totale** generato dal consumo complessivo dei tre beni; se, ad esempio, $u(x, y, z) = \sqrt{x + y + z}$ e vengono consumate 5 unità del primo bene, 3 del secondo e 1 del terzo, l'utilità totale è $u(5, 3, 1) = \sqrt{5 + 3 + 1} = 3$.

In generale, la **funzione di domanda** di un certo bene non dipende solo dal prezzo p del bene in questione, bensì anche dal prezzo di altri n beni sostituti,[2] $p_1, p_2, \ldots, p_n$ e dalla ricchezza disponibile w: $q = f(p, p_1, p_2, \ldots, p_n, w)$, ovvero $f : \mathbb{R}^{n+2} \to \mathbb{R}$. Ad esempio, se consideriamo solamente i prezzi di altri due beni sostituti, l'espressione (lineare) $f(p, p_1, p_2, w) = aw - bp + cp_1 + dp_2$, dove $a, b, c, d > 0$ sono parametri, definisce una funzione di domanda plausibile. Una forma funzionale di questo tipo assume significato economico perché è crescente nella ricchezza e nei prezzi dei beni sostituti, mentre è decrescente nel prezzo del bene che si vuole acquistare.

Volendo essere pignoli, anche la funzione (12.4) del Paragrafo 12.3.3 del primo volume, che esprime il **montante al tempo t** di un capitale C investito al tasso annuo nominale convertibile k volte all'anno, j_k, è *funzione delle quattro variabili*, t, C, k e j_k; scriveremo quindi $M(t, C, k, j_k) = C(1 + j_k/k)^{kt}$.

Terminiamo con un esempio la cui scrittura sfrutta la *notazione del prodotto scalare*. Se indichiamo con $y = f(x_1, x_2, \ldots, x_n)$ la **quantità prodotta** di un bene finale in funzione delle quantità impiegate degli n fattori produttivi e beni intermedi, con p

[2] I beni sostituti presentano caratteristiche simili a quelle del bene considerato, come, ad esempio, il tè per il caffè.

il prezzo del bene prodotto (e venduto), e con $p_1, p_2, \ldots, p_n$ i prezzi di ciascun fattore e bene intermedio impiegato, la **funzione del profitto** π nel caso di mercati perfettamente concorrenziali sia per il bene finale sia per i fattori e beni intermedi assume la forma tipica della 'differenza tra ricavo e costi' ampiamente utilizzata per funzioni di una sola variabile: $\pi(x_1, x_2, \ldots, x_n) = py - \sum_{i=1}^{n} p_i x_i$, dove il prodotto 'prezzo p × quantità venduta y' è il ricavo, mentre la **struttura dei costi** è espressa come somma di prodotti 'prezzo p_i × quantità di fattore x_i', per $i = 1, \ldots, n$. L'espressione dei costi non è altro che la Definizione 1.4 di prodotto scalare; pertanto, se indichiamo con $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ il *vettore della combinazione di fattori e beni intermedi impiegati* e con $\widehat{p} = (p_1, p_2, \ldots, p_n) \in \mathbb{R}^n$ il *vettore dei prezzi dei fattori e beni intermedi* (il simbolo $\widehat{p}$ serve per distinguere il vettore dei prezzi dei fattori dal numero p che indica il prezzo del bene finale), l'espressione del profitto diventa:

$$\pi(x) = pf(x) - \widehat{p} \cdot x$$

in cui il primo prodotto (il ricavo) è un *prodotto fra numeri*, p e $y = f(x)$, mentre il secondo prodotto è a tutti gli effetti un *prodotto scalare fra i vettori* $\widehat{p}$ e x, entrambi elementi di $\mathbb{R}^n$. Si noti che, a differenza dell'espressione del profitto in una sola variabile studiata nel primo volume, in cui la variabile indipendente era la quantità prodotta q, in questo caso la funzione del profitto è pensata come funzione degli n fattori produttivi e beni intermedi $x_1, x_2, \ldots, x_n$. L'obiettivo dei prossimi capitoli sarà trovare il (o i) vettore(i) $x^* \in \mathbb{R}^n$ che definisce la combinazione (ottima) di fattori produttivi e beni intermedi che massimizza il profitto π.

4.2.2 Campi di esistenza

La determinazione dei *campi di esistenza* di funzioni di n variabili segue le stesse regole viste per le funzioni di una variabile, si tratta di studiare le limitazioni imposte dalle funzioni elementari.

ESEMPIO 4.2

1. Nella funzione $f(x, y) = e^x + xy - \ln y$ c'è solamente una limitazione dovuta all'ultimo addendo, il logaritmo, che è definito per $y > 0$, mentre tutti gli altri addendi sono definiti per ogni coppia di numeri $(x, y) \in \mathbb{R}^2$. Il campo di esistenza di f è *l'insieme più ampio possibile entro il quale tutti i "pezzi" di f assumono significato*; pertanto è l'insieme $CE = \{(x, y) \in \mathbb{R}^2 : y > 0\}$, è cioè il semipiano che si trova al di sopra dell'asse delle ascisse, escluso l'asse stesso, corrispondente all'area grigia in Figura 4.8(a). Si noti che l'asse x [cioè l'insieme $\{(x, y) \in \mathbb{R}^2 : y = 0\}$] non è contenuto nel dominio di f.
2. La funzione $f(x, y) = 1/x - 1/y$ ha per dominio l'insieme $CE = \{(x, y) \in \mathbb{R}^2 : x, y \neq 0\}$ rappresentato dall'area grigia in Figura 4.8(b): tutto $\mathbb{R}^2$ esclusi gli assi cartesiani.

3. La funzione $f(x, y) = \ln(xy)$, che può essere pensata come la funzione composta di $h(t) = \ln t$ e $t = g(x, y) = xy$, ha per dominio l'insieme $CE = \{(x, y) \in \mathbb{R}^2 : xy > 0\} = \{(x, y) \in \mathbb{R}^2 : [(x > 0) \wedge (y > 0)] \cup [(x < 0) \wedge (y < 0)]\}$, cioè l'unione del primo ortante con il terzo ortante, esclusi gli assi cartesiani, corrispondente all'area grigia in Figura 4.8(c).
4. La funzione $f(x, y) = \ln(y - x)$, che può essere pensata come la funzione composta di $h(t) = \ln t$ e $t = g(x, y) = y - x$, ha per dominio l'insieme $CE = \{(x, y) \in \mathbb{R}^2 : y - x > 0\} = \{(x, y) \in \mathbb{R}^2 : y > x\}$ rappresentato dall'area grigia disegnata in Figura 4.8(d); è l'insieme dei punti di $\mathbb{R}^2$ che stanno sopra la bisettrice $y = x$, esclusa la bisettrice stessa.
5. La funzione $f(x, y) = \ln(x - e^y)$, che può essere pensata come la funzione composta di $h(t) = \ln t$ e $t = g(x, y) = x - e^y$, ha per dominio l'insieme $CE = \{(x, y) \in \mathbb{R}^2 : x - e^y > 0\} = \{(x, y) \in \mathbb{R}^2 : y < \ln x\}$, cioè l'insieme dei punti di $\mathbb{R}^2$ che si trovano al di sotto del grafico della funzione logaritmo naturale, escluso il grafico stesso, come mostra la Figura 4.8(e).
6. La funzione di tre variabili $f : \mathbb{R}^3 \to \mathbb{R}$ definita da $f(x, y, z) = \sqrt{1 - x^2 - y^2 - z^2}$, interpretabile come la funzione composta di $h(t) = \sqrt{t}$ e $t = g(x, y, z) = 1 - x^2 - y^2 - z^2$, ha per dominio l'insieme $CE = \{(x, y, z) \in \mathbb{R}^3 : 1 - x^2 - y^2 - z^2 \geq 0\} = \{(x, y, z) \in \mathbb{R}^2 : x^2 + y^2 + z^2 \leq 1\}$, cioè l'insieme dei punti di $\mathbb{R}^3$ contenuti nella palla di raggio 1, inclusa la superficie sferica $\{(x, y, z) \in \mathbb{R}^2 : x^2 + y^2 + z^2 = 1\}$, come mostra la Figura 4.8(f).

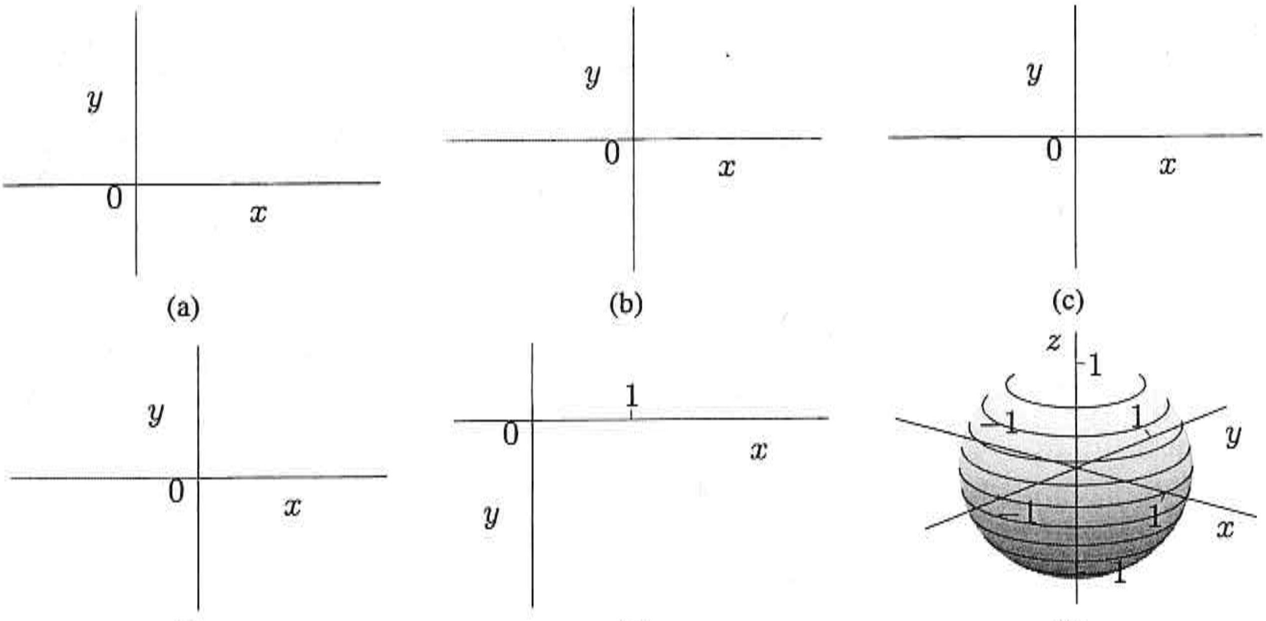


FIGURA 4.8: rappresentazione grafica del domino di alcune funzioni di due variabili: (a) $f(x, y) = e^x + xy - \ln y$; (b) $f(x, y) = 1/x - 1/y$; (c) $f(x, y) = \ln(xy)$; (d) $f(x, y) = \ln(y - x)$; (e) $f(x, y) = \ln(x - e^y)$; (f) $f(x, y, z) = \sqrt{1 - x^2 - y^2 - z^2}$.

Lo scopo dell'esempio precedente è mostrare che, per stabilire il campo di esistenza, si tratta semplicemente di *scomporre* la funzione data in funzioni più semplici a cui applicare i principi già noti per le funzioni di una variabile. La rappresentazione grafica di tali insiemi risulta complicata per le funzioni di tre variabili ed è impossibile per funzioni di più di tre variabili.

4.2.3 Funzioni di due variabili e loro grafico

Le funzioni di due variabili meritano una trattazione a parte perché è possibile disegnare il loro grafico. Sfruttando questa proprietà saremo in grado di giustificare i principali risultati del calcolo differenziale e dell'ottimizzazione in modo geometrico (e quindi intuitivo), basando le nostre argomentazioni su una falsariga molto simile a quella seguita nel primo volume.

La prossima definizione generalizza quella data nel primo volume per funzioni di una variabile alle funzioni di n variabili.

DEFINIZIONE 4.8 (GRAFICO DI UNA FUNZIONE DI n VARIABILI) Data $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, si dice **grafico** di f l'insieme:

$$G_f = \{(x_1, x_2, \ldots, x_n, y) \in \mathbb{R}^{n+1} : y = f(x_1, x_2, \ldots, x_n)\}$$

Anche se questa definizione vale in astratto per funzioni di n variabili, siamo particolarmente interessati al caso in cui $n = 2$, cioè al **grafico di una funzione di due variabili**, $f(x, y)$, che vale la pena specificare con la notazione semplificata (Notazione 4.2):

$$G_f = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\} \tag{4.2}$$

Si tratta di un oggetto tridimensionale perché tiene conto di tre variabili: le due indipendenti, x e y, e la dipendente, z. Questo consente di disegnare il grafico (4.2), anche se ciò non è facile, dal momento che tracciare un oggetto immerso nello spazio 'tridimensionale' richiede un po' di "senso della prospettiva". Lo scopo di questi disegni è soprattutto quello di fornire una rappresentazione intuitiva; quando avremo bisogno di un maggior dettaglio, trasformeremo "pezzi" di grafici a tre dimensioni in grafici a due dimensioni, a noi più familiari.

Se le variabili indipendenti sono più di due il grafico è un insieme con più di tre dimensioni, oggetto impossibile da disegnare.

Il grafico delle funzioni di una variabile $f : \mathbb{R} \to \mathbb{R}$ è una *curva nel piano* $\mathbb{R}^2$ (un sottoinsieme del piano cartesiano). Analogamente, per le funzioni di due variabili, esso è una *superficie nello spazio* $\mathbb{R}^3$ (un sottoinsieme dello spazio cartesiano). Si tratta di una specie di "lenzuolo", come illustra la Figura 4.9. Nella stessa figura viene evidenziato il dominio di f: il rettangolo grigio sotto la superficie "ondulata". La superficie che rappresenta il grafico di $f(x, y)$ in $\mathbb{R}^3$ non può essere verticale in alcun punto, né "piegarsi" in modo da sovrapporsi, perché altrimenti esisterebbero due numeri distinti z_1 e z_2 associati entrambi alla stessa coppia (x, y), contraddicendo la Definizione 4.7 di funzione. In altre parole si tratta di un lenzuolo che non può essere "piegato e messo in un cassetto"!

ESEMPIO 4.3 La funzione:

$$f(x, y) = x^2 + y^2 \qquad (4.3)$$

è la generalizzazione a due variabili della parabola $f(x) = x^2$ in una variabile; è definita su tutto $\mathbb{R}^2$. La Figura 4.10(a) mostra il grafico in tre dimensioni [l'insieme (4.2)] di questa funzione. Le Figure 4.10(b) e 4.10(c) mostrano le sezioni verticali del grafico di f in (4.3) lungo l'asse x e y rispettivamente [le curve tracciate in grassetto nella Figura 4.10(a)]: sono i grafici in due dimensioni delle funzioni di una variabile ottenute ponendo $y = 0$ e $x = 0$ in (4.3) rispettivamente, si tratta della stessa parabola. Possiamo pensare di ottenere il grafico tridimensionale di f in (4.3) partendo da uno dei due grafici bidimensionali nelle Figure 4.10(b) o 4.10(c) (o entrambi) e facendolo(li) ruotare orizzontalmente intorno all'asse verticale z fino a riempire tutta la superficie in grigio sfumato in Figura 4.10(a).

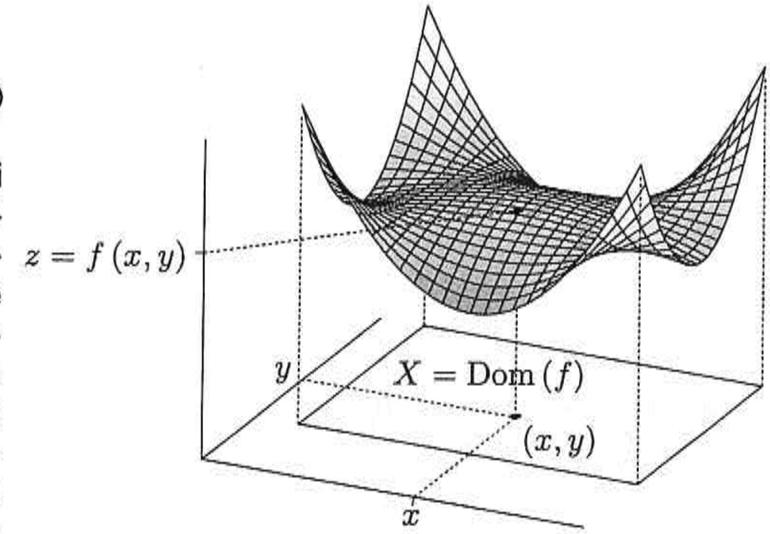


FIGURA 4.9: il grafico di $f(x, y)$ è una superficie (una specie di "lenzuolo") in $\mathbb{R}^3$.

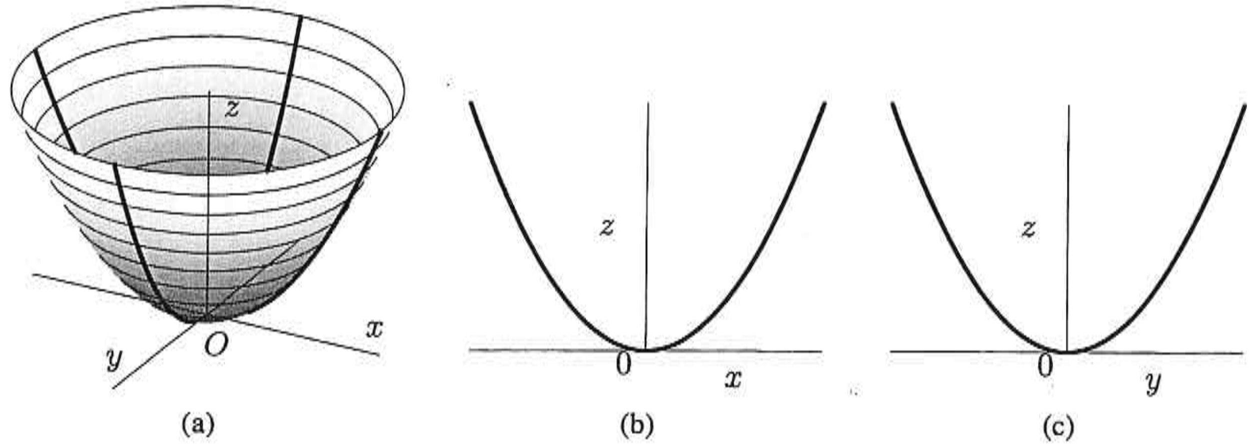


FIGURA 4.10: (a) grafico in tre dimensioni di $f(x, y) = x^2 + y^2$; (b) grafico in due dimensioni di $z = x^2$; (c) grafico in due dimensioni di $z = y^2$.

ESEMPIO 4.4 La funzione:

$$f(x, y) = -x^2 - y^2$$

è la generalizzazione a due variabili della parabola rovesciata $f(x) = -x^2$ in una variabile, come mostra la Figura 4.11. Anch'essa è definita su tutto $\mathbb{R}^2$.

ESEMPIO 4.5 La funzione:

$$f(x, y) = ax + by + c \qquad (4.4)$$

dove a, b e c sono parametri, è la **funzione affine** in due variabili; è cioè la generalizzazione della *retta* $f(x) = mx + q$ in una variabile ed è rappresentata graficamente da un *piano* in $\mathbb{R}^3$. È definita su tutto $\mathbb{R}^2$. La Figura 4.12(c) mostra un esempio di tale piano ottenuto mediante una costruzione simile a quella sviluppata nell'Esempio 2.1: le Figure 4.12(a) e 4.12(b) mostrano il grafico bidimensionale delle funzioni di una sola variabile ottenute ponendo $y = 0$ e $x = 0$ rispettivamente in (4.4), ossia $z = f(x, 0) = ax + c$ nel piano (x, z) e $z = f(0, y) = by + c$ nel piano (y, z). La Figura 4.12(c) riporta i grafici bidimensionali appena disegnati sopra gli assi x e y (le due rette in grassetto) nello spazio tridimensionale $\mathbb{R}^3$; trattandosi di due rette oblique fra loro che si intersecano nel punto $(0, 0, c)$, ossia in prossimità dell'intercetta c con l'asse verticale z, esse generano il piano in $\mathbb{R}^3$ indicato con la superficie in grigio sfumato. Dalla Figura 4.12(a) osserviamo che $a < 0$ (la pendenza della prima retta è negativa), mentre dalla Figura 4.12(b) si vede che $b > 0$ (la pendenza della seconda retta è positiva), mentre è evidente che l'intercetta con l'asse z è positiva, $c > 0$. Le intercette con gli altri due assi cartesiani, x e y, sono i numeri $-c/a$ e $-c/b$ rispettivamente. Chiaramente, se $c = 0$, f in (4.4) diventa a tutti gli effetti una *funzione lineare* in due variabili (Definizione 2.2) che assume la stessa forma dell'espressione (2.8) con due soli coefficienti, $a_1 = a$ e $a_2 = b$: $f(x, y) = ax + by$. Dal Paragrafo 2.2.2 sappiamo che il suo grafico è un piano passante per l'origine, come illustrato nella Figura 2.1 dell'Esempio 2.1.

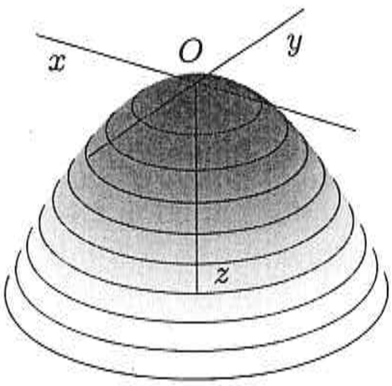


FIGURA 4.11: grafico di $f(x, y) = -x^2 - y^2$.

I grafici delle funzioni negli Esempi 4.3 e 4.4 si chiamano, non senza una certa suggestione, *paraboloidi*. L'espressione analitica generale di un paraboloide in due variabili è:

$$Q(x, y) = ax^2 + by^2 + 2cxy \tag{4.5}$$

con a, b e c parametri qualsiasi, e la scelta di porre '$2c$' al posto della più semplice opzione 'c' come coefficiente dell'ultimo addendo diventerà chiara successivamente. Queste funzioni sono molto importanti per lo studio di massimi e minimi relativi interni, tanto da meritarsi un nome proprio: **forme quadratiche**,[3] termine che ricorda la proprietà che *la somma degli esponenti delle variabili che compaiono in ciascun addendo è sempre* 2. Spesso le forme quadratiche hanno grafici simili ai paraboloidi delle Figure 4.10(a) e 4.11, che risulteranno più o meno deformati a seconda dei valori assunti da a, b e c: il vertice è sempre l'origine [se $x = y = 0$ allora $Q(x, y) = 0$] ma al posto dei cerchi evidenziati nelle suddette figure, avranno ellissi come sezioni orizzontali.[4]

[3]Indicare questa famiglia di funzioni con la lettera maiuscola Q, come in (4.5), è una pratica comunemente adottata nella maggior parte dei libri di testo perché ricorda appunto il nome 'forma *quadratica*'.

[4]Una *sezione orizzontale* del grafico di una funzione di due variabili è l'intersezione tra il grafico e un piano orizzontale.

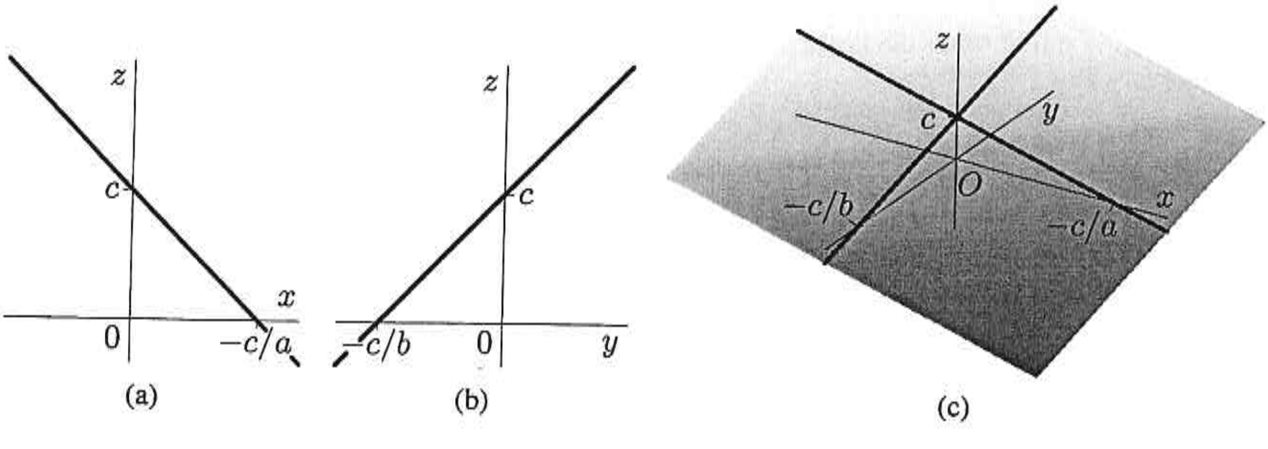


FIGURA 4.12: (a) grafico bidimensionale di $z = f(x, 0) = ax + c$; (b) grafico bidimensionale di $z = f(0, y) = by + c$; (c) grafico tridimensionale di $z = f(x, y) = ax + by + c$.

Più avanti vedremo le espressioni delle funzioni affini e delle forme quadratiche in più di due variabili. Ricordando le approssimazioni di Taylor di primo grado e secondo grado per funzioni di una variabile utilizzate nel primo volume per illustrare i punti di ottimo relativo interno, è facile immaginare il ruolo che i piani (possibilmente orizzontali...) definiti da (4.4) e i paraboloidi definiti da (4.5) avranno nel caratterizzare i punti di ottimo relativo interno per le funzioni di due variabili mediante i polinomi di Taylor di primo e secondo grado (opportunamente ridefiniti per le funzioni di due variabili).

4.3 Derivata e differenziale

Non approfondiamo le definizioni rigorose di *limite* e *continuità* per funzioni di n variabili. Ci limitiamo a osservare che un punto (vettore) $x \in \mathbb{R}^n$ *tende* a un altro punto $x^0 \in \mathbb{R}^n$ (si ricordi la Notazione 4.1), che dovrà essere di accumulazione (punto 1 della Definizione 4.2), e si scrive $x \to x^0$, se ciascuna coordinata di x tende singolarmente a ciascuna coordinata di x^0, cioè se $x_1 \to x_1^0$, $x_2 \to x_2^0, \ldots, x_n \to x_n^0$ simultaneamente. Pertanto, la definizione di limite per funzioni di più variabili, che si scrive sempre $\lim_{x \to x^0} f(x)$, è equivalente alla definizione data nel primo volume, l'unica differenza è che $x \to x^0$ presuppone lo studio del comportamento dell'immagine $f(x)$ al tendere di ciascuna coordinata x_i a x_i^0 separatamente, $x_i \to x_i^0$, per $1 = 1, \ldots, n$. In ogni caso, non ci occuperemo del calcolo esplicito di limiti per funzioni di tante variabili.

Con questa premessa la definizione di *funzione continua* diventa un'ovvia estensione delle definizioni date nel primo volume a più variabili. Per noi è sufficiente pensare ad una funzione di due variabili il cui grafico non si interrompe mai sul suo dominio (la superficie che individua il grafico della funzione – il "lenzuolo" descritto nel Paragrafo 4.2.3 – non ha "strappi"!) come idea di funzione continua.

Per quanto riguarda la *derivata*, invece, il discorso è più delicato. L'equivalente della derivata per funzioni di n variabili, $f : \mathbb{R}^n \to \mathbb{R}$, si chiama *differenziale* e viene costruito a partire da n particolari derivate di una sola variabile, dette *derivate parziali*.

4.3.1 Derivate parziali

Data una funzione $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, fissiamo un *punto* (vettore) *interno* $x^0 \in \operatorname{Int}(X)$ e definiamo *n funzioni di una variabile*, $f_1 : \mathbb{R} \to \mathbb{R}$, $f_2 : \mathbb{R} \to \mathbb{R}$,..., $f_n : \mathbb{R} \to \mathbb{R}$, ciascuna dipendente dalla funzione di partenza f e dal vettore di riferimento x^0, nel modo seguente:

$$\begin{aligned} f_1(x_1) &= f\left(x_1, x_2^0, \ldots, x_n^0\right) \\ f_2(x_2) &= f\left(x_1^0, x_2, x_3^0 \ldots, x_n^0\right) \\ &\vdots \\ f_i(x_i) &= f\left(x_1^0, x_2^0, \ldots, x_{i-1}^0, x_i, x_{i+1}^0, \ldots, x_n^0\right) \\ &\vdots \\ f_n(x_n) &= f\left(x_1^0, \ldots, x_{n-1}^0, x_n\right) \end{aligned} \tag{4.6}$$

Per comprendere il significato delle (4.6) distinguiamo le variabili (le coordinate) sulle quali compare lo zero ad apice, x_i^0, da quelle in cui non compare, x_i: in ogni termine a destra delle uguaglianze compare solamente una variabile senza l'apice, la stessa che compare come unico argomento del rispettivo termine a sinistra. Questo significa che ciascuna funzione f_i, per $i = 1, \ldots, n$, *dipende dalla sola variabile* x_i mentre le altre $n-1$ variabili rimangono *costanti*, pari ai valori $x_1^0, \ldots, x_{i-1}^0, x_{i+1}^0, \ldots, x_n^0$; queste ultime diventano quindi dei numeri (dei parametri), ovvero non sono più 'variabili'. Poiché rimane un'unica 'variabile' degna di questo nome, la x_i (quella senza lo zero ad apice), *ciascuna* f_i in (4.6) *è a tutti gli effetti una funzione della sola variabile* x_i.

OSSERVAZIONE 4.2 Ciascuna f_i in (4.6) dipende dal punto prescelto $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$: se cambiamo punto anche le f_i risulteranno diverse. Inoltre, dovrebbe essere evidente la differenza fra la funzione originale $f : X \to \mathbb{R}$ e le f_i, basti pensare che sono definite su spazi totalmente diversi, la prima su $X \subseteq \mathbb{R}^n$ e le seconde su $\mathbb{R}$, anche se le f_i sono costruite a partire dalla f.

ESEMPIO 4.6 Così come faremo negli esempi successivi, **concentriamo la nostra attenzione sulle funzioni di due e di tre variabili**, $f : \mathbb{R}^2 \to \mathbb{R}$ e $f : \mathbb{R}^3 \to \mathbb{R}$, che converrà indicare con $f(x, y)$ e $f(x, y, z)$ rispettivamente; in questi casi scriveremo i *punti interni di riferimento* (x_0, y_0) e (x_0, y_0, z_0), senza dimenticare che si tratta di vettori di $\mathbb{R}^2$ e di $\mathbb{R}^3$ rispettivamente.

1. $f(x, y) = x^2 + y$ è definita su tutto $\mathbb{R}^2$. Fissiamo $(x_0, y_0) = (2, 1) \in \mathbb{R}^2$ e definiamo $f_1(x) = x^2 + y_0 = x^2 + 1$ e $f_2(y) = x_0^2 + y = 2^2 + y = 4 + y$. La prima funzione è una banale parabola verso l'alto mentre la seconda è una retta. Se, diversamente, fissiamo il punto di riferimento $(x_0, y_0) = (1, 2) \in \mathbb{R}^2$, diverso dal precedente,

otteniamo altre due funzioni diverse dalle precedenti: $f_1(x) = x^2 + y_0 = x^2 + 2$ e $f_2(y) = x_0^2 + y = 1^2 + y = 1 + y$. Ancora una volta si tratta di una parabola e di una retta, ma sono diverse dalle precedenti.

2. $f(x,y) = e^x + xy - \ln y$ è definita sull'insieme $CE = \{(x,y) \in \mathbb{R}^2 : y > 0\}$, come abbiamo visto nell'Esempio 4.2 al punto 1. Se $(x_0, y_0) = (1,1) \in CE$, allora $f_1(x) = e^x + xy_0 - \ln y_0 = e^x + x - 0 = e^x + x$ e $f_2(y) = e^{x_0} + x_0 y - \ln y = e + y - \ln y$.
3. $f(x,y,z) = \ln(xyz)$ è definita su $CE = \{(x,y,z) \in \mathbb{R}^3 : xyz > 0\}$. Ponendo $(x_0, y_0, z_0) = (1, 2, 1/2) \in CE$, riesce $f_1(x) = \ln x$, $f_2(y) = \ln(y/2)$ e $f_3(z) = \ln(2z)$.

Le funzioni di una variabile f_i introdotte in (4.6) ci permettono di definire lo strumento più importante del calcolo differenziale in più variabili: la *derivata parziale*.

DEFINIZIONE 4.9 (DERIVATE PARZIALI) Data $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, si dice che f è **derivabile parzialmente rispetto a** x_i nel punto $x^0 = (x_1^0, x_2^0, \ldots, x_n^0) \in \operatorname{Int}(X)$ se f_i [definita in (4.6)] è derivabile in x_i^0 e si pone:

$$\frac{\partial}{\partial x_i} f(x^0) = f_i'(x_i^0)$$

Le funzioni di una variabile $(\partial/\partial x_i) f(x^0)$, per $i = 1, \ldots, n$, si chiamano **derivate parziali** di f in x^0.

NOTAZIONE 4.3 La notazione più comune per le derivate parziali è $f_{x_i}(x^0)$, che nel caso di funzioni di due variabili si esplicita con $f_x(x_0, y_0)$, $f_y(x_0, y_0)$, e, nel caso di funzioni di tre variabili, con $f_x(x_0, y_0, z_0)$, $f_y(x_0, y_0, z_0)$ e $f_z(x_0, y_0, z_0)$. Scritture alternative sono $D_{x_i} f(x^0)$, oppure $D_i f(x^0)$. Purtroppo molti testi usano anche la notazione $f_i(x^0)$, che è una scritture infelice perché si confonde con le funzioni (che non sono le derivate!) f_i definite in (4.6). Noi useremo soprattutto la notazione f_{x_i}, che reputiamo inequivocabile, e, sporadicamente, la notazione $\partial f / \partial x_i$.

Le derivate parziali sono dunque comunissime derivate di funzioni di una variabile.

Per le *funzioni di due variabili*, $z = f(x,y)$, è possibile un'illuminante *interpretazione geometrica* delle derivate parziali. Nella Figura 4.13(a) le 'funzioni parziali' f_1 e f_2 definite in (4.6) sono la restrizione della f alle rette $x = x_0$ e $y = y_0$ [le due rette parallele agli assi x e y e passanti per il punto (x_0, y_0), disegnate sul piano orizzontale]; pertanto i loro grafici sono le curve nere che si trovano sul grafico di f (la superficie grigia, secondo la descrizione del Paragrafo 4.2.3). In altre parole, i grafici delle funzioni parziali f_1 e f_2 si ottengono sezionando il grafico di $f(x,y)$ con due piani verticali paralleli agli assi x e y e passanti per il punto (x_0, y_0). L'interpretazione geometrica delle derivate in una sola variabile ci dice che le derivate parziali f_x e f_y nel punto (x_0, y_0) corrispondono alle pendenze delle rette tangenti ai grafici di f_1 e di f_2 nei punti x_0 e y_0 rispettivamente [le rette tangenti alla superficie grigia nella Figura 4.13(a)]. Le Figure 4.13(b) e 4.13(c) mostrano le proiezioni delle funzioni parziali (di una variabile) f_1 e f_2 sui piani di coordinate x, z e y, z rispettivamente, dove sono riportate anche le rette tangenti al loro grafico nei punti x_0 e y_0.

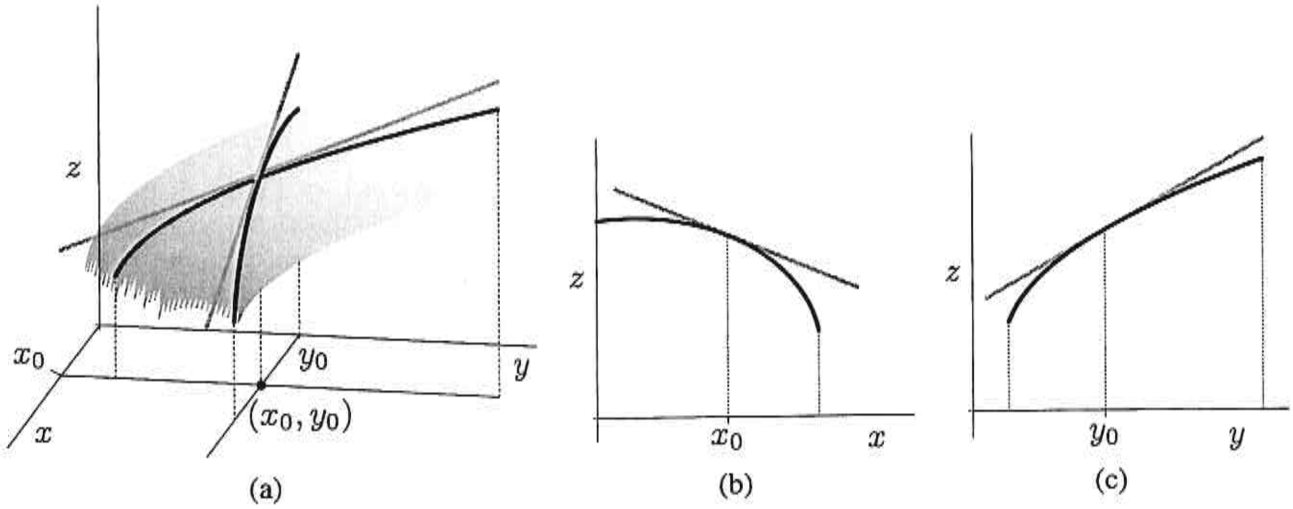


FIGURA 4.13: visione geometrica delle derivate parziali; (a) situazione in tre dimensioni; (b) grafico in due dimensioni della funzione parziale $f_1(x)$ e interpretazione della sua derivata in x_0, $f_x(x_0, y_0) = f_1'(x_0)$, come pendenza della retta tangente al suo grafico in x_0; (c) grafico in due dimensioni della funzione parziale $f_2(y)$ e interpretazione della sua derivata in y_0, $f_y(x_0, y_0) = f_2'(y_0)$, come pendenza della retta tangente al suo grafico in y_0.

Le derivate parziali misurano quanto ciascuna variabile contribuisce separatamente alla variazione infinitesima del valore della funzione f. La derivata parziale f_{x_i} nel punto x^0 di una funzione di n variabili misura la variazione infinitesima di f prodotta da una variazione infinitesima della variabile x_i a partire dal punto x^0 mantenendo le altre $n-1$ variabili, $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$, fisse ai valori $x_1^0, \ldots, x_{i-1}^0, x_{i+1}^0, \ldots, x_n^0$. Dovrebbe essere chiara a questo punto l'interpretazione delle derivate parziali come *grandezze marginali*, sulla falsariga di quanto detto nel primo volume. Riprendendo gli esempi del Paragrafo 4.2.1, nel caso di una funzione di costo $c(x, y)$ che dipende da due fattori, c_x indica il costo marginale del primo fattore produttivo per una quantità fissata del secondo fattore, mentre c_y indica il costo marginale del secondo fattore tenendo fisso il primo. Analogamente, f_{x_i} è il prodotto marginale dell'i-esimo fattore produttivo per una funzione di produzione $f(x_1, x_2, \ldots, x_n)$, calcolato mantenendo fissa la quantità degli altri $n-1$ fattori, mentre u_{x_i} rappresenta l'utilità marginale dell'i-esimo bene di consumo per una funzione di utilità $u(x_1, x_2, \ldots, x_n)$, calcolata per una quantità fissata dei rimanenti $n-1$ beni.

Come facevamo per le derivate in $\mathbb{R}$, penseremo le derivate parziali come *funzioni derivate*: al variare di x^0 ci sono n derivate, ciascuna funzione di n variabili, $f_{x_i} : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, n$.

OSSERVAZIONE 4.3 Attenzione: per il calcolo di ciascuna f_{x_i}, per $i = 1, \ldots, n$, bisogna considerare le altre variabili, $x_1^0, \ldots, x_{i-1}^0, x_{i+1}^0, \ldots, x_n^0$, come parametri (costanti!) e *derivare solamente rispetto a x_i* (l'unica variabile che rimane tale). Così facendo, otteniamo una funzione di n variabili, $f_{x_i}(x_1, x_2, \ldots, x_n)$, che è la nostra funzione derivata parziale.

ESEMPIO 4.7

1. Da $f(x,y) = x^2 + y$, ponendo y = costante, ricaviamo $f_x(x,y) = (\partial/\partial x)(x^2+y) = 2x$, mentre ponendo x = costante ricaviamo $f_y(x,y) = (\partial/\partial y)(x^2+y) = 1$. Per una coincidenza, la prima espressione è funzione della sola variabile x, mentre la seconda è addirittura costante (non dipende né da x né da y); in generale, come vedremo nei prossimi esempi, le espressioni delle derivate parziali sono a tutti gli effetti funzioni di tutte le variabili in gioco. Volendo calcolare il valore delle derivate parziali nel punto $(x_0, y_0) = (2,1)$, sostituiamo gli argomenti ottenendo $f_x(2,1) = 2 \cdot 2 = 4$ e $f_y(2,1) = 1$.
2. Le derivate parziali di $f(x,y) = e^x + xy - \ln y$ avranno senso nei punti interni del dominio di f. Ponendo y = costante, ricaviamo $f_x(x,y) = e^x + y$, mentre ponendo x = costante ricaviamo $f_y(x,y) = x - 1/y$. In questo caso sia f_x sia f_y sono genuinamente funzioni di entrambe le variabili x e y.
3. Le derivate parziali di $f(x,y) = 1/x - 1/y$ esistono su tutto il dominio di f, essendo un insieme aperto (Esempio 4.2 al punto 2). Ponendo y = costante ricaviamo $f_x = -1/x^2$ [poiché si sottintende che le *derivate parziali sono funzioni*, gli argomenti (x,y) vengono spesso omessi per comodità], mentre ponendo x = costante ricaviamo $f_y = 1/y^2$.
4. $f(x,y) = (1+x)^\alpha e^y$, dove $\alpha \in \mathbb{R}$ è un parametro, ha dominio $\mathbb{R}^2$. Ponendo y = costante, ricaviamo $f_x = \alpha(1+x)^{\alpha-1} e^y$, e ponendo x = costante ricaviamo $f_y = (1+x)^\alpha e^y$.
5. $f(x,y,z) = \ln(xyz)$ ha derivate parziali su tutto il dominio che è un insieme aperto. Ponendo y, z = costante, ricaviamo $f_x = yz/(xyz) = 1/x$, ponendo x, z = costante ricaviamo $f_y = xz/(xyz) = 1/y$ e ponendo x, y = costante ricaviamo $f_z = xy/(xyz) = 1/z$.
6. Le derivate parziali di $f(x,y,z) = x^2y^2z$ esistono su tutto $\mathbb{R}^3$. Ponendo y, z = costante, $f_x = 2xy^2z$, ponendo x, z = costante, $f_y = 2x^2yz$, e ponendo x, y = costante, $f_z = x^2y^2$.

4.3.2 Il differenziale

Purtroppo le derivate parziali non sono l'equivalente della derivata delle funzioni di una variabile. Nella Figura 4.13 abbiamo rappresentato geometricamente le derivate parziali di una funzione di due variabili $f(x,y)$ come rette tangenti ai grafici delle due 'funzioni parziali' di una variabile, f_1 e f_2 definite in (4.6), nei punti x_0 e y_0 rispettivamente. Da ciò deduciamo che le derivate parziali forniscono informazioni estremamente limitate (da cui, appunto, l'appellativo 'parziali') su cosa succede *in tutto un intorno di* (x_0, y_0); esse ci raccontano una storia che vale solamente lungo le due *direzioni canoniche*, quelle parallele agli assi x e y, e nulla dicono su cosa accade su punti altrettanto vicini a (x_0, y_0) ma che non giacciono sulle rette parallele agli assi passanti per (x_0, y_0).

L'obiettivo finale è costruire un oggetto che abbia per le funzioni di due variabili lo stesso ruolo della retta tangente al grafico di f per funzioni di una variabile. Poiché i grafici delle funzioni di due variabili sono superfici nello spazio tridimensionale, è naturale aspettarsi che tale ruolo ora venga assunto da un *piano* che dovrà essere *tangente* al grafico di f nel punto (x_0, y_0); il concetto di tangenza ora però coinvolge *tutto il grafico* di f in un *intorno completo* del punto (x_0, y_0).

Un argomento convincente per illustrare l'insufficienza delle derivate parziali ai nostri scopi è costituito dal fatto che, diversamente da quanto avevamo osservato nel Paragrafo 11.3.1 del primo volume per le funzioni di una variabile, l'esistenza delle derivate parziali non garantisce affatto la continuità di f.

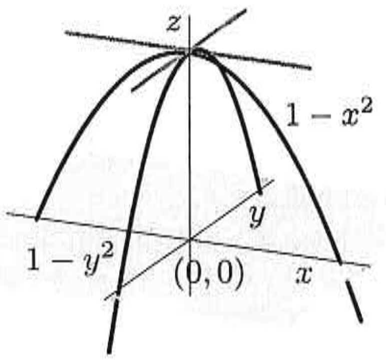


FIGURA 4.14: grafico della $f(x, y)$ discussa nell'esempio 4.8.

ESEMPIO 4.8 Consideriamo la funzione $f : \mathbb{R}^2 \to \mathbb{R}$ definita per parti da:

$$f(x, y) = \begin{cases} 0 & \text{se } xy \neq 0 \\ 1 - x^2 & \text{se } y = 0 \\ 1 - y^2 & \text{se } x = 0. \end{cases}$$

Si tratta della funzione costante 0 su tutto $\mathbb{R}^2$ tranne che sugli assi cartesiani, su ciascuno dei quali ha come grafico una parabola orientata verso il basso con vertice nel punto $(0, 0, 1) \in \mathbb{R}^3$, ossia assume il valore (massimo) 1 nell'origine. In altre parole, il grafico di f, rappresentato in Figura 4.14, ricorda il telaio di un ombrello da cui è stato tolto il tessuto impermeabile e sono rimaste le quattro stecche metalliche (le due curve in nero); per essere più precisi, fuori dagli assi cartesiani il tessuto è stato sostituito con il piano orizzontale a livello 0 indicato in grigio chiaro in figura. È evidente che f è *discontinua* nel punto $(x_0, y_0) = (0, 0)$, dal momento che $f(0, 0) = 1$ ma *ogni intorno di* $(x_0, y_0) = (0, 0)$ contiene punti tali che $x \neq 0$ e $y \neq 0$ contemporaneamente e sui quali vale $f(x, y) = 0$. Il grafico di f ha dunque un "salto" dal valore 0 al valore 1 nel punto $(x_0, y_0) = (0, 0)$. Per verifica, si consideri, ad esempio, la direzione della bisettrice del piano $\mathbb{R}^2$, cioè si ponga $y = x$ e si calcoli $\lim_{x \to 0} f(x, x)$: chiaramente vale[5] $\lim_{x \to 0} f(x, x) = 0$, mentre $f(0, 0) = 1$, e quindi f risulta discontinua[6] in $(x_0, y_0) = (0, 0)$.

Cionondimeno, nel punto $(x_0, y_0) = (0, 0)$ *le derivate parziali esistono* e sono entrambe nulle; esse sono rappresentate dai due segmenti orizzontali in grigio scuro in Figura 4.14. Infatti, poiché sugli assi cartesiani i prodotti $xy = x0$ e $xy = 0y$ sono sempre nulli, le funzioni parziali definite sugli assi sono rispettivamente le parabole $f_1(x) = 1 - x^2$ e $f_2(y) = 1 - y^2$, con derivate parziali date da $f_x(x, 0) = f_1'(x) = -2x$ e $f_y(0, y) = f_2'(y) = -2y$, entrambe nulle nel punto $(x_0, y_0) = (0, 0)$: $f_x(0, 0) = f_y(0, 0) = 0$.

[5]Applicando le tecniche sviluppate nei Capitoli 8 e 9 del primo volume, poiché per costruzione $f(x, x) \equiv 0$ per ogni $x \neq 0$, vale necessariamente $\lim_{x \to 0} f(x, x) = 0$ lungo qualsiasi direzione che non sia parallela agli assi cartesiani.

[6]Non avendo formulato esplicitamente la definizione di funzione continua in n variabili, ispirandoci alla Definizione 9.1 nel Capitolo 9 del primo volume, accontentiamoci di immaginare il "salto" del grafico di f come "prova tangibile" di discontinuità. Si osservi che non si tratta di una discontinuità eliminabile (cosiddetta 'di terza specie', secondo la terminologia introdotta nel Paragrafo 9.1.1 del primo volume) dal momento che nell'origine si intersecano entrambi i vertici delle due parabole (le "stecche" dell'ombrello) definite sugli assi cartesiani: non possiamo pensare di eliminare l'intero telaio dell'ombrello...

Volendo generalizzare uno strumento a spazi più ampi di quelli in cui è stato definito la prima volta, è buona norma mantenere le proprietà originali dello strumento stesso. Serve dunque un concetto di 'derivabilità' per funzioni di due variabili che garantisca senz'altro (almeno) la continuità di f.

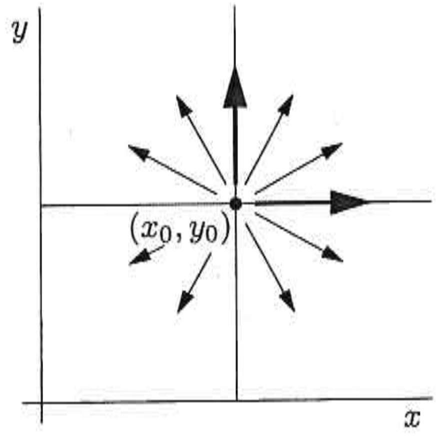


FIGURA 4.15: direzioni di uscita da (x_0, y_0) in un suo intorno.

Osservando la Figura 4.13(a) "dall'alto", cioè dalla direzione parallela all'asse z, possiamo immaginare un ipotetico intorno del punto (x_0, y_0), come quello rappresentato in Figura 4.15, in cui sono evidenziate in grassetto le due direzioni canoniche rispetto alle quali vengono calcolate le derivate parziali. Ci rendiamo facilmente conto che ciò di cui abbiamo bisogno è qualcosa che rappresenti le *derivate rispetto a tutte le direzioni di uscita da* (x_0, y_0) (come, ad esempio, le direzioni individuate da alcune frecce oblique in Figura 4.15), non solo quelle parallele agli assi cartesiani.

In realtà esistono infinite direzioni di uscita e pertanto questa strada non sembra percorribile.[7] Fortunatamente è stato dimostrato il prossimo Teorema 4.1 che garantisce l'esistenza del piano tangente al grafico di f in (x_0, y_0), ovvero la 'derivabilità completa' di f in tutto un intorno di (x_0, y_0) (rispetto a tutte le direzioni), a partire da una proprietà che riguarda solamente le derivate parziali (rispetto alle direzioni canoniche), senza bisogno di scomodare derivate rispetto ad altre direzioni.

TEOREMA 4.1 (DEL DIFFERENZIALE TOTALE) Sia $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^2$, dotata di derivate parziali in un intorno del punto $(x_0, y_0) \in \operatorname{Int}(X)$. Se *entrambe le derivate parziali sono continue* in (x_0, y_0), allora:

1. f è **differenziabile in** (x_0, y_0), ovvero esiste il *piano tangente* al grafico di f nel punto (x_0, y_0), piano che chiameremo **differenziale di f in** (x_0, y_0);
2. f è *continua* in (x_0, y_0).

La Figura 4.16 mostra il differenziale come piano tangente al grafico di f nel punto (x_0, y_0). Tale piano è individuato dalle due rette ortogonali[8] tangenti al grafico di f evidenziate in grassetto, una con pendenza $f_x(x_0, y_0)$ e l'altra con pendenza $f_y(x_0, y_0)$. Se il differenziale esiste, esso è dunque definito proprio dalle due derivate parziali.

Diremo che f *è differenziabile* se ammette piano tangente al suo grafico su tutti i punti interni del suo dominio (nell'Esempio 4.8 questo scenario si verifica se il tessuto

[7] A voler essere precisi fino in fondo, neanche la derivabilità rispetto a tutte le direzioni di uscita *rettilinee* sarebbe sufficiente; per assicurare un comportamento coerente della funzione in un intorno del punto (x_0, y_0) dovremmo verificare che la pendenza del grafico di f sia la stessa rispetto a *tutte le traiettorie possibili* che intersecano il punto (x_0, y_0), anche quelle rappresentate da *curve*. Appare evidente che questa via è del tutto improponibile.

[8] Si ricordi che due rette (non parallele) che si intersecano nello spazio individuano un unico piano.

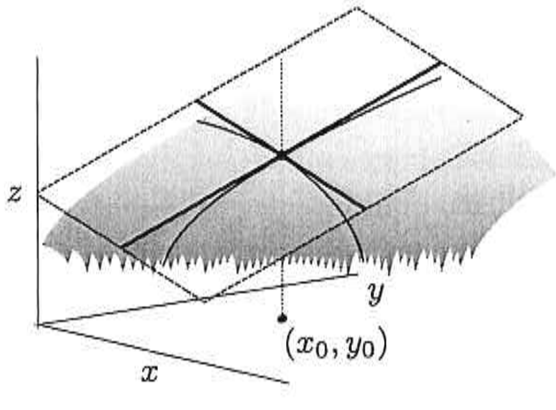


FIGURA 4.16: il differenziale è rappresentato dal piano tangente al grafico di f nel punto (x_0, y_0).

impermeabile rimane al suo posto...). Il Teorema 4.1 stabilisce che ciò avviene quando le derivate parziali di f sono *ovunque* continue.

L'estensione del Teorema 4.1 alle funzioni di $n > 2$ variabili è ovvia: se una funzione $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, ha tutte le derivate parziali, f_{x_i}, continue in $x^0 \in \text{Int}(X)$, per $i = 1, \dots, n$, allora essa è differenziabile in x^0, ovvero ammette 'iperpiano' tangente al suo grafico in x^0; in altre parole, essa è a tutti gli effetti 'derivabile' in x^0.

4.3.3 Approssimazione lineare e differenziale

Un modo intuitivo di pensare al differenziale (piano tangente) è quello di intenderlo come approssimazione dei valori assunti dalla funzione $f(x, y)$ in un intorno del punto (x_0, y_0) di $\mathbb{R}^2$, esattamente come avevamo fatto per approssimare i valori di una funzione di una variabile $f(x)$ mediante la retta tangente nelle vicinanze di un punto x_0 sulla retta $\mathbb{R}$ nel Paragrafo 11.1.1 del primo volume.

La nostra percezione della superficie terrestre indicherebbe che essa è piana; in verità sappiamo che la terra è rotonda. Questa percezione è un'approssimazione dovuta al fatto che la nostra vista termina su un orizzonte limitato che stabilisce il raggio di un intorno molto piccolo rispetto alla superficie totale del pianeta, sul quale è naturale 'approssimare' una superficie rotonda con un piano.

Consideriamo una funzione $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^2$, differenziabile in (x_0, y_0), supponiamo cioè che esista il piano tangente al suo grafico nel punto (x_0, y_0), e vediamo in dettaglio come si costruisce tale piano. Nell'Esempio 4.5 avevamo visto che un piano nello spazio $\mathbb{R}^3$ è rappresentato dall'espressione analitica della funzione affine (che impropriamente chiameremo 'lineare'):

$$\ell(x, y) = ax + by + c \tag{4.7}$$

che è la generalizzazione della retta in quanto presuppone l'esistenza di *due pendenze*, i parametri a e b riferiti alle variabili x e y rispettivamente, e il parametro c che individua l'intercetta con l'asse z. Trattandosi di un piano, dovrà avere un'espressione analitica del tipo in (4.7); il nostro obiettivo è quindi calcolare i valori dei parametri a, b e c per tale piano.

Se manteniamo fissa la seconda variabile al valore $y = y_0$ possiamo studiare la variazione di f dovuta al solo variare dell'altra variabile, la x, intorno a x_0. Poiché in questo caso si tratta di una funzione di una sola variabile, $f(x, y_0)$, la (11.3) nel Paragrafo 11.1.1 del primo volume si può riscrivere come:

$$f(x, y_0) = f(x_0, y_0) + f_x(x_0, y_0)\,\mathrm{dx} + o(\mathrm{dx}) \qquad \text{per dx} \to 0 \tag{4.8}$$

dove $\mathrm{dx} = x - x_0$ indica la variazione finita della (sola) variabile indipendente x e $o(\mathrm{dx})$ indica l'errore commesso, che possiamo considerare *trascurabile* se dx è sufficientemente piccolo, cioè se x è 'sufficientemente vicino' a x_0.[9] La (4.8) ci racconta che la variazione della variabile dipendente $\Delta f = f(x, y_0) - f(x_0, y_0)$ lungo l'asse x è approssimata dal prodotto $f_x(x_0, y_0)\,\mathrm{dx}$ dove $f_x(x_0, y_0)$ è la derivata parziale rispetto a x di f in (x_0, y_0). Analogamente, se manteniamo fissa la prima variabile, $x = x_0$, la variazione di f dovuta al solo variare della variabile y intorno a y_0 è esprimibile come

$$f(x_0, y) = f(x_0, y_0) + f_y(x_0, y_0)\,\mathrm{dy} + o(\mathrm{dy}) \qquad \text{per } \mathrm{dy} \to 0$$

dove $\mathrm{dy} = y - y_0$ indica la variazione finita della (sola) variabile indipendente y e, di nuovo, $o(\mathrm{dy})$ indica l'errore commesso, che possiamo considerare *trascurabile* se y è 'sufficientemente vicino' a y_0.

Ma cosa succede se facciamo variare *contemporaneamente* entrambe le variabili x e y? Come nel caso di una variabile, l'idea sottostante a questa costruzione è quella di osservare gli effetti dovuti a variazioni delle variabili indipendenti supponendo di "muoversi" sul piano tangente anziché sul grafico della funzione f. Il piano tangente ha un'espressione analitica come in (4.7), in cui gli spostamenti delle variabili x e y agiscono in modo additivo; infatti $\ell(x, y)$ è il risultato della somma dei termini ax e by più una costante c. In altre parole, le variazioni di x e y intervengono separatamente e il loro effetto globale è pari alla *somma dei due effetti*; possiamo pertanto scrivere:

$$\begin{aligned} f(x, y) &= f(x_0 + \mathrm{dx}, y_0 + \mathrm{dy}) \\ &= f(x_0, y_0) + f_x(x_0, y_0)\,\mathrm{dx} + f_y(x_0, y_0)\,\mathrm{dy} + o[\|(\mathrm{dx}, \mathrm{dy})\|] \\ &\quad \text{per } (\mathrm{dx}, \mathrm{dy}) \to 0 \end{aligned} \tag{4.9}$$

dove l'argomento dell'errore ora è la *distanza euclidea fra i punti* (x, y) *e* (x_0, y_0) di $\mathbb{R}^2$, ossia, utilizzando la Formula (1.3), $\|(\mathrm{dx}, \mathrm{dy})\| = \sqrt{(x - x_0)^2 + (y - y_0)^2}$. La scrittura $o[\|(\mathrm{dx}, \mathrm{dy})\|]$ significa che ora l'errore tende a zero più velocemente della distanza fra (x, y) e (x_0, y_0) *rispetto a qualsiasi direzione* di avvicinamento a (x_0, y_0).

La (4.9) indica che la variazione della variabile dipendente, $\Delta f = f(x, y) - f(x_0, y_0)$, rispetto a qualsiasi direzione di uscita da (x_0, y_0) è approssimata dalla somma dei prodotti $f_x(x_0, y_0)\,\mathrm{dx}$ e $f_y(x_0, y_0)\,\mathrm{dy}$, cioè dalla *somma dei differenziali in una sola variabile* lungo le direzioni canoniche. La Figura 4.17 illustra la scomposizione lungo le direzioni parallele agli assi cartesiani dello spostamento dal punto $f(x_0, y_0)$ al punto $\ell(x, y) = \ell(x_0 + \mathrm{dx}, y_0 + \mathrm{dy})$ ottenuto muovendosi sul piano tangente al grafico di f in (x_0, y_0): poiché ci stiamo muovendo su di un piano, possiamo indicare lo spostamento complessivo con il *vettore* congiungente $f(x_0, y_0)$ con $\ell(x, y)$, che è pari alla *somma dei*

[9] Il simbolo $o(\mathrm{dx})$ è definito nel Paragrafo 9.4.2 del primo volume e indica una funzione che è un infinitesimo di ordine superiore al suo argomento, specificamente, $\lim_{\mathrm{dx} \to 0} [o(\mathrm{dx}) / \mathrm{dx}] = 0$ (si vedano anche gli approfondimenti del Paragrafo 11.1.3). In questo caso questa proprietà si legge "$o(\mathrm{dx})$ è un errore che tende a zero più velocemente della distanza dx"; in questo senso parliamo di un errore *trascurabile*.

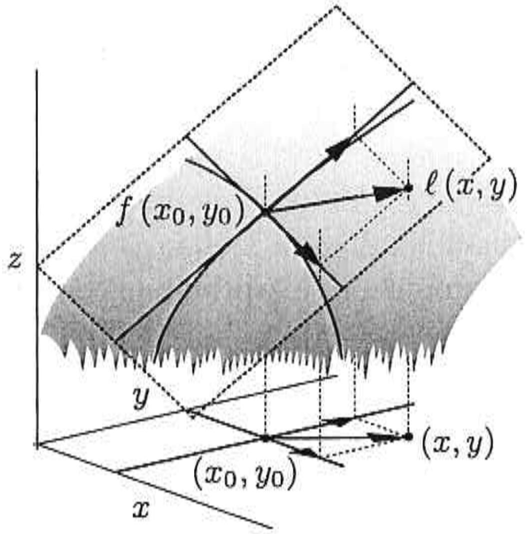


FIGURA 4.17: scomposizione in somma di direzioni canoniche dello spostamento da $f(x_0, y_0)$ a $\ell(x, y) = \ell(x_0 + \mathrm{dx}, y_0 + \mathrm{dy})$ sul differenziale.

due vettori che esprimono ciascun spostamento lungo le direzioni canoniche (si veda la Figura 1.4).[10]

Dalla (4.9) otteniamo immediatamente l'espressione analitica del piano tangente come funzione di x e y:

$$\ell(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \qquad (4.10)$$

da cui ricaviamo i valori dei parametri della (4.7) cercati: $a = f_x(x_0, y_0)$, $b = f_y(x_0, y_0)$ e $c = f(x_0, y_0) - [f_x(x_0, y_0)]\, x_0 - [f_y(x_0, y_0)]\, y_0$. Come nel caso di una sola variabile, il differenziale viene formalmente definito come funzione delle *variazioni* $\mathrm{dx} = x - x_0$ e $\mathrm{dy} = y - y_0$, anziché direttamente in funzione delle variabili indipendenti x e y.

DEFINIZIONE 4.10 (DIFFERENZIALE TOTALE) La variazione lungo il piano tangente al grafico di f nel punto (x_0, y_0):

$$\mathrm{df} = f_x(x_0, y_0)\,\mathrm{dx} + f_y(x_0, y_0)\,\mathrm{dy} \qquad (4.11)$$

si chiama **differenziale totale**, nome utilizzato anche per le funzioni di $n > 2$ variabili, $f : \mathbb{R}^n \to \mathbb{R}$, per le quali scriveremo:

$$\mathrm{df} = \sum_{i=1}^{n} f_{x_i}(x^0)\,\mathrm{dx}_i \qquad (4.12)$$

per indicare la variazione lungo l'iperpiano tangente.

L'espressione (4.11) definisce il differenziale totale in senso proprio; in realtà, con abuso di notazione, parleremo di differenziale tutte le volte che avremo in mente il piano tangente, quindi, impropriamente, chiameremo differenziale anche la funzione $\ell(x, y)$, definita dall'espressione (4.10) direttamente come funzione di x e y anziché delle variazioni dx e dy. Per essere precisi il differenziale è una funzione di quattro variabili: $x_0, y_0, \mathrm{dx} = x - x_0$ e $\mathrm{dy} = y - y_0$; esso è cioè funzione delle due coordinate del punto di riferimento (x_0, y_0) e delle due variazioni dx e dy.

[10] Possiamo anche pensare al piano tangente come un *sottospazio vettoriale di dimensione* 2 (nel senso discusso nel Capitolo 3) traslato rispetto all'origine, per il quale vale la struttura algebrica di $\mathbb{R}^2$ e quindi ha perfettamente senso considerare in esso la somma di vettori.

NOTAZIONE 4.4 Il *vettore*[11] *delle derivate parziali* si chiama **gradiente** e si indica con[12] ∇f:

$$\nabla f = (f_{x_1}, f_{x_2}, \ldots, f_{x_n}) \tag{4.13}$$

Utilizzando la (4.13) possiamo riscrivere la (4.12) sotto forma di *prodotto scalare*:

$$\mathrm{df} = \nabla f\left(x^0\right) \cdot \mathrm{dx} \tag{4.14}$$

dove $(\mathrm{dx}) = (\mathrm{dx}_1, \mathrm{dx}_2, \ldots, \mathrm{dx}_n) \in \mathbb{R}^n$ è il vettore delle variazioni $\mathrm{dx}_i = x_i - x_i^0$. La (4.14) assomiglia alla Notazione 11.1 usata nel Paragrafo 11.1.1 del primo volume: ancora una volta il prodotto scalare consente di scrivere una proprietà che riguarda le funzioni di n variabili in modo analogo alla proprietà equivalente riguardante le funzioni di una sola variabile. La (4.14) ci autorizza inoltre ad *interpretare l'intero vettore delle derivate parziali, il gradiente ∇f, come l'omologo in n variabili della derivata prima, f', in una variabile* (non dimentichiamo che ciò è lecito solo se tutte le f_{x_i} sono continue!).

Vale la pena scrivere la (4.9) nel caso di n variabili sfruttando la notazione vettoriale (4.14):

$$f(x) = f\left(x^0 + \mathrm{dx}\right) = f\left(x^0\right) + \nabla f\left(x^0\right) \cdot \mathrm{dx} + o\left(\|\mathrm{dx}\|\right) \qquad \text{per } \mathrm{dx} \to 0 \tag{4.15}$$

dove il termine $f\left(x^0\right) + \nabla f\left(x^0\right) \cdot \mathrm{dx}$ indica l'*approssimazione lineare di f* in un intorno del punto[13] $x^0 = \left(x_1^0, \ldots, x_n^0\right)$ e $\|\mathrm{dx}\| = \left\|x - x^0\right\|$ è la distanza euclidea fra il punto di riferimento x^0 e il punto $x = x^0 + \mathrm{dx}$.

ESEMPIO 4.9 Si consideri la funzione $f(x,y) = \sqrt{x+y}$ definita su $CE = \{(x,y) \in \mathbb{R}^2 : y \geq -x\}$. Prendiamo il punto $(x_0, y_0) = (6,3)$, sul quale conosciamo il valore assunto da f, $f(6,3) = \sqrt{9} = 3$. Approssimiamo il valore $f(11/2, 5/2) = \sqrt{8}$ usando la (4.15). Vale $\mathrm{dx} = 11/2 - 6 = -1/2$ e $\mathrm{dy} = 5/2 - 3 = -1/2$; le derivate parziali sono uguali, $f_x = f_y = 1/\left(2\sqrt{x+y}\right)$, e valgono entrambe $1/6$ in $(x_0, y_0) = (6,3)$, pertanto il gradiente in $(x_0, y_0) = (6,3)$ è $\nabla f(6,3) = (1/6, 1/6)$. Quindi:

$$\begin{aligned}\sqrt{8} &\approx f(6,3) + f_x(6,3)\,\mathrm{dx} + f_y(6,3)\,\mathrm{dy} \\ &= \sqrt{9} + (1/6)(-1/2) + (1/6)(-1/2) = 3 - 1/12 - 1/12 \simeq 2.8333\end{aligned}$$

[11] Poiché il gradiente si trova coinvolto quasi esclusivamente in operazioni di prodotto scalare con altri vettori e non viene mai considerato come matrice di dimensione $1 \times n$ (o $n \times 1$), per semplicità esso viene definito nella (4.13) direttamente come '*vettore riga*' anziché come vettore colonna, in apparente contraddizione con la convenzione di considerare tutti i vettori come vettori colonna adottata nella prima parte di questo volume. Tale scelta è coerente con la generalizzazione del gradiente alla matrice Jacobiana che introdurremo nel Paragrafo 6.3.2 (Definizione 6.5).

[12] Per curiosità, il simbolo '∇', che si scrive come una delta maiuscola rovesciata, si chiama *Nabla*.

[13] Si noti la differenza con la notazione usata in (4.9): ora abbiamo utilizzato la notazione x^0, con lo zero ad apice, per indicare un vettore di $\mathbb{R}^n$, mentre nella (4.9) i simboli x_0 e y_0 indicano numeri.

4.4 Derivata seconda e matrice Hessiana

Come si può generalizzare il concetto di derivata seconda al caso di funzioni di n variabili? Poiché la derivata seconda in una variabile è la derivata della derivata della funzione, un punto di partenza promettente è chiedersi che cosa succede se deriviamo (parzialmente) a loro volta le derivate parziali. Ricordando che le derivate parziali $f_{x_i}(x_1, x_2, \ldots, x_n)$ sono funzioni di n variabili, ha senso calcolare le n derivate parziali di ciascuna derivata parziale f_{x_i}, per un totale di $n \times n = n^2$ derivate parziali seconde. Notiamo subito un aspetto inquietante: l'"esplosione" del numero di derivate parziali! Corre il rischio che la situazione ci sfugga di mano... Continuando di questo passo (derivando parzialmente ciascuna derivata parziale seconda) otterremmo n^3 derivate parziali terze, n^4 derivate parziali quarte ecc., un gran numero di oggetti che diventa difficile da gestire e soprattutto difficile da interpretare. Fortunatamente il nostro obiettivo è stabilire condizioni del secondo ordine, che coinvolgono solamente le derivate seconde, pertanto ci fermeremo alle $n \times n$ derivate parziali seconde, senza correre il rischio di "intasare" la nostra mente con una folla di derivate.

Superato il primo shock, ci accorgiamo che, dopotutto, $n \times n$ derivate parziali seconde non sono poi così tante; ad esempio, per le funzioni di due e tre variabili, sono 4 e 9 rispettivamente, quantità tutto sommato controllabili. Inoltre, guarda caso, $n \times n$ è il numero dei coefficienti di una *matrice quadrata* A $(n \times n)$: una coincidenza? Non solo, il prossimo Teorema 4.2 attenuerà ulteriormente il nostro disagio stabilendo che quasi sempre molte di queste derivate sono dei "doppioni"; ovvero, di fatto, le derivate parziali seconde da calcolare sono meno di $n \times n$.

DEFINIZIONE 4.11 (DERIVATE PARZIALI SECONDE) Data una funzione $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, derivabile parzialmente su $\operatorname{Int}(X)$, se le funzioni derivate parziali $f_{x_i} : \operatorname{Int}(X) \to \mathbb{R}$, per $i = 1, 2, \ldots, n$, sono tutte a loro volta parzialmente derivabili in un punto $x^0 \in \operatorname{Int}(X)$, cioè se esistono tutti i numeri:

$$\frac{\partial}{\partial x_j}\left[\frac{\partial}{\partial x_i} f\left(x^0\right)\right] \quad \text{per ogni coppia di indici } i = 1, 2, \ldots, n \text{ e } j = 1, 2, \ldots, n \qquad (4.16)$$

allora diremo che f **è derivabile parzialmente due volte** nel punto x^0.

NOTAZIONE 4.5 Le $n \times n$ derivate parziali definite in (4.16) si scrivono anche:

$$\frac{\partial^2}{\partial x_i^2} f\left(x^0\right) \quad \text{se } i = j \qquad \text{e} \qquad \frac{\partial^2}{\partial x_j \partial x_i} f\left(x^0\right) \quad \text{se } i \neq j$$

dove la prima notazione indica che deriviamo due volte rispetto alla stessa variabile x_i, mentre la seconda che deriviamo *prima rispetto a x_i e poi rispetto a x_j*. Scritture alternative (che useremo più frequentemente) sono $f_{x_i x_i}\left(x^0\right)$ e $f_{x_j x_i}\left(x^0\right)$ rispettivamente. Il secondo tipo di derivate parziali seconde (quelle che presuppongono di derivare prima rispetto alla variabile x_i e poi rispetto alla variabile x_j) hanno il suggestivo appellativo di

derivate parziali seconde miste e rimangono distinte dal primo tipo, le quali, in contrapposizione, vengono battezzate **derivate parziali seconde pure.** Si osservi che in tutte le notazioni viste il riferimento alla variabile di destra corrisponde alla prima derivazione mentre il riferimento alla variabile di sinistra corrisponde alla seconda derivazione (che può essere pensata come la derivazione "più esterna"), in altre parole, $\partial^2 f / (\partial x_j \partial x_i)$ e $f_{x_j x_i}$ indicano entrambe che si deriva prima rispetto a x_i e poi rispetto a x_j.

Se f è derivabile parzialmente due volte su tutto $\text{Int}(X)$ abbiamo $n \times n$ *funzioni derivate parziali seconde*, $f_{x_j x_i}(x_1, x_2, \ldots, x_n)$, per $i, j = 1, 2, \ldots, n$, ciascuna funzione di n variabili.

OSSERVAZIONE 4.4 La Definizione 4.11 non presuppone che f sia differenziabile, è sufficiente che esistano le derivate parziali e che siano a loro volta derivabili parzialmente. Alla stregua delle derivate parziali prime, la derivabilità parziale due volte è del tutto indipendente dal fatto che la funzione f sia *regolare* – ovvero che abbia un grafico liscio, senza spigoli, strappi o salti, ovvero che ammetta piano tangente – e pertanto fornisce uno strumento estremamente debole (ancora una volta, molto 'parziale') per lo studio della f stessa. Per un proficuo utilizzo delle derivate parziali seconde avremo bisogno di condizioni aggiuntive, come richiederà il Teorema 4.2.

Per quanto riguarda l'aspetto puramente tecnico del calcolo delle derivate parziali seconde non c'è nulla di nuovo da imparare. Bisogna soltanto derivare, derivare, derivare, ... e ancora derivare!

ESEMPIO 4.10 La funzione $f(x, y) = \ln(x - 2y)$ definita su $CE = \{(x, y) \in \mathbb{R}^2 : y < (1/2)x\}$ ha derivate parziali prime $f_x = (x - 2y)^{-1}$ e $f_y = -2/(x - 2y)$. Le 4 derivate parziali seconde sono

$$\begin{array}{ll} f_{xx} = \dfrac{\partial}{\partial x}\left(\dfrac{1}{x-2y}\right) = -\dfrac{1}{(x-2y)^2} & f_{xy} = \dfrac{\partial}{\partial x}\left(-\dfrac{2}{x-2y}\right) = \dfrac{2}{(x-2y)^2} \\ f_{yx} = \dfrac{\partial}{\partial y}\left(\dfrac{1}{x-2y}\right) = \dfrac{2}{(x-2y)^2} & f_{yy} = \dfrac{\partial}{\partial y}\left(-\dfrac{2}{x-2y}\right) = -\dfrac{4}{(x-2y)^2} \end{array} \tag{4.17}$$

Osserviamo un dettaglio interessante: nell'esempio le derivate parziali seconde miste sono uguali. È possibile una simile coincidenza? Per scrupolo facciamo un altro tentativo.

ESEMPIO 4.11 La funzione $f(x, y) = e^x + xy - 3\ln y$ ha $CE = \{(x, y) \in \mathbb{R}^2 : y > 0\}$ e vale $f_x = e^x + y$ e $f_y = x - 3/y$. Le derivate parziali seconde sono $f_{xx} = e^x$, $f_{yx} = 1$, $f_{xy} = 1$ e $f_{yy} = 3/y^2$.

Ancora una volta $f_{yx} = f_{xy} = 1$. Potremmo continuare a lungo, la verità è che le derivate miste sono (quasi) sempre uguali, come stabilisce il prossimo importante teorema che non dimostriamo.

TEOREMA 4.2 (SCHWARZ) Sia $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, dotata di derivate parziali prime e seconde in un intorno del punto $x^0 \in \operatorname{Int}(X)$. Se *tutte le derivate parziali prime e seconde*, f_{x_i} e $f_{x_j x_i}$, per $i, j = 1, 2, \ldots, n$, sono *continue* in x^0, allora:

$$f_{x_j x_i}\left(x^0\right) = f_{x_i x_j}\left(x^0\right) \qquad \text{per ogni } i, j = 1, 2, \ldots, n$$

Deduciamo che se f ammette derivate parziali prime e seconde continue su tutti i punti di $\operatorname{Int}(X)$, allora le *funzioni derivate parziali seconde miste sono uguali* su tutti i punti interni. Poiché tutti gli esempi ed esercizi che affronteremo soddisferanno sempre le ipotesi del Teorema 4.2, le derivate parziali seconde da calcolare saranno solamente n (le pure) $+ n(n-1)/2$ (la metà delle miste), per un totale di $n(n+1)/2$ ($< n^2$), come si verifica facilmente, a parziale conforto della nostra avversione al calcolo di derivate! Si noti infine che il teorema, richiedendo la continuità delle derivate parziali prime, implica automaticamente che f sia differenziabile.

NOTAZIONE 4.6 La proprietà che tutte le f_{x_i}, per $i = 1, \ldots, n$, esistono e sono continue si dice '*f è derivabile con continuità*' (è differenziabile) ed equivale a dire succintamente che **f è di classe C^1**. Analogamente, se tutte le derivate parziali seconde $f_{x_i x_j}$, per $i, j = 1, \ldots, n$, esistono e sono continue si dice '*f è due volte derivabile con continuità*' ed equivale a dire che **f è di classe C^2**.

Le (4.17), scritte in quell'ordine, suggeriscono il modo con cui rappresentare l'insieme delle derivate parziali seconde: in una matrice quadrata $n \times n$ sulla cui diagonale disporremo le derivate seconde pure, $f_{x_i x_i}$, e fuori dalla diagonale le derivate seconde miste, $f_{x_i x_j}$, rispecchiando l'ordine (la posizione individuata degli indici i, j) standard dei coefficienti a_{ij} di qualsiasi matrice.

DEFINIZIONE 4.12 (MATRICE HESSIANA) La matrice quadrata $n \times n$ i cui coefficienti a_{ij} sono le derivate parziali seconde $f_{x_i x_j}$ si chiama **matrice Hessiana** e si indica con $D^2 f$, oppure $\nabla^2 f$ (poiché è calcolata a partire dal gradiente di f, ∇f, ovvero il vettore delle derivate parziali prime), o più semplicemente con H quando non ci sono dubbi su quale sia la funzione f a cui si riferisce:

$$H = \begin{bmatrix} f_{x_1 x_1} & f_{x_1 x_2} & \cdots & f_{x_1 x_n} \\ f_{x_2 x_1} & f_{x_2 x_2} & & f_{x_2 x_n} \\ \vdots & & \ddots & \vdots \\ f_{x_n x_1} & f_{x_n x_1} & \cdots & f_{x_n x_n} \end{bmatrix}.$$

Quando dovremo indicare i valori specifici assunti dalla matrice H su un preciso punto $x^0 \in \operatorname{Int}(X)$, $X \subseteq \mathbb{R}^n$, scriveremo $H_f\left(x^0\right)$, oppure $D^2 f\left(x^0\right)$ o $\nabla^2 f\left(x^0\right)$.

Il Teorema 4.2 aggiunge una proprietà importante alla matrice Hessiana H: non solo è una matrice quadrata, ma, se f è di classe C^2, è anche una **matrice simmetrica**.

Abbiamo quindi trovato un'importante applicazione dei vettori e delle matrici, argomenti discussi nei capitoli precedenti: le derivate (parziali) prime di una funzione

$f : \mathbb{R}^n \to \mathbb{R}$ si rappresentano con un *vettore* di $\mathbb{R}^n$, il gradiente ∇f, e le derivate (parziali) seconde si rappresentano con una *matrice simmetrica* $n \times n$, la matrice Hessiana H. Con riferimento all'Esempio 4.10, scriveremo:

$$\nabla f = \left(\frac{1}{x-2y}, -\frac{2}{x-2y} \right) \qquad \text{e} \qquad H = \begin{bmatrix} -\dfrac{1}{(x-2y)^2} & \dfrac{2}{(x-2y)^2} \\ \dfrac{2}{(x-2y)^2} & -\dfrac{4}{(x-2y)^2} \end{bmatrix}$$

4.5 Approssimazione non lineare

Dal primo volume sappiamo che il segno della derivata seconda individua il *tipo di curvatura* di una funzione (verso l'alto o verso il basso) mentre il suo valore assoluto determina il *grado di curvatura* di una funzione (curvatura più o meno accentuata); questi aspetti hanno rilevanza locale nel caso dell'approssimazione non lineare e nella ricerca di punti estremi, mentre hanno rilevanza globale nel caso della concavità/convessità. Vediamo ora un'interpretazione analoga delle derivate parziali seconde (la matrice Hessiana) incominciando dal loro ruolo nell'approssimazione non lineare.

L'approssimazione non lineare (di secondo grado) di una *funzione di due variabili*, $f(x, y)$, *di classe* C^2 in un intorno di un punto interno (x_0, y_0) è il punto di partenza ideale per studiare il tipo di curvatura della f data *nelle vicinanze*[14] *di* (x_0, y_0), poiché in questo caso si può fare affidamento su di una rappresentazione grafica. Se per le funzioni di una variabile eravamo alla ricerca di un polinomio di secondo grado come approssimante, ora cerchiamo l'analogo in due variabili, ovvero una *forma quadratica* del tipo in (4.5) che qui riportiamo:

$$Q(x, y) = ax^2 + by^2 + 2cxy \tag{4.18}$$

con a, b e c parametri. Q ha spesso come grafico un oggetto simile (più o meno deformato) a quelli riportati nelle Figure 4.10(a) e 4.11, che chiamiamo 'paraboloidi' poiché costituiscono la generalizzazione delle parabole in due variabili; essi sono ciò che di meglio abbiamo a disposizione per farci un'idea geometrica (approssimata) del tipo di curvatura del grafico di una funzione di due variabili $f(x, y)$ in un intorno di (x_0, y_0). A differenza del mondo in una variabile, però, già la determinazione del tipo di curvatura delle stesse forme quadratiche non è affatto scontata: ponendo, ad esempio, $c = 0$ e imponendo segni discordi ai parametri a e b in (4.18) $Q(x, y)$ risulta concava rispetto a una variabile e convessa rispetto all'altra, generando una certa confusione.[15] Rinviamo la discussione specifica sullo studio della curvatura delle forme quadratiche e, più in generale,

[14] Si ricordi che l'analisi locale è significativa soltanto se non ci si allontana troppo dal punto di riferimento (x_0, y_0).

[15] Questo è il tipico caso dei *punti di sella*, anch'essi caratterizzabili mediante forme quadratiche, che approfondiremo nel Capitolo 5; non ci soffermiamo ora su questi casi particolari.

delle funzioni di n variabili, nei capitoli successivi; qui ci limitiamo a prendere un po' di confidenza con le forme quadratiche in sé e con il loro utilizzo nell'approssimazione non lineare.

4.5.1 Forme quadratiche

L'estensione della (4.18) a tre variabili si fa in modo naturale; conviene in questo caso, però, utilizzare la notazione x_1, x_2 e x_3, anziché x, y e z, per indicare le variabili indipendenti:

$$Q(x_1, x_2, x_3) = a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 \quad (4.19)$$

La notazione x_1, x_2 e x_3 per le variabili indipendenti aiuta a identificare i parametri (coefficienti) a_{ij} con gli indici delle variabili moltiplicate in ciascun addendo; il coefficiente a_{ii} viene associato alla variabile 'x_i^2', ovvero nell'addendo in cui x_i compare '2 volte' (x_i che moltiplica se stessa), mentre il coefficiente a_{ij} viene associato al prodotto tra le variabili 'x_i' e 'x_j'. A voler essere pignoli, per completare il quadro dovremmo includere da qualche parte, oltre ai coefficienti a_{12}, a_{13} e a_{23}, anche i coefficienti a_{21}, a_{31} e a_{32}. Ma, se vogliamo associare gli indici i, j alle coppie di variabili x_i, x_j che compongono ciascun addendo, ci rendiamo conto che i coefficienti a_{21}, a_{31} e a_{32} ci sono già: poiché ad essi corrispondono i prodotti x_2x_1, x_3x_1 e x_3x_2, che, per la proprietà commutativa del prodotto (fra numeri!) coincidono rispettivamente con x_1x_2, x_1x_3 e x_2x_3, è il caso di porre $a_{21} = a_{12}$, $a_{31} = a_{13}$ e $a_{32} = a_{23}$. Questo, per inciso, spiega la presenza del fattore '2' che moltiplica ciascun addendo: tale fattore indica semplicemente che in esso si tiene conto di entrambi i prodotti, x_ix_j e x_jx_i, ovvero $2a_{ij}x_ix_j$ comprende sia $a_{ij}x_ix_j$ sia $a_{ji}x_jx_i$ sotto l'ipotesi che $a_{ij} = a_{ji}$.

Questa discussione un po' pedante sui coefficienti a_{ij} lascia presagire che la definizione di forma quadratica per n variabili coinvolgerà una 'matrice'; addirittura, poiché $a_{ij} = a_{ji}$, siamo in grado di prevedere che si tratterà di una *matrice simmetrica* (e quindi necessariamente quadrata). Procediamo con ordine: cominciamo con la definizione generale di forma quadratica per n variabili.

DEFINIZIONE 4.13 (FORMA QUADRATICA) La **forma quadratica** in n variabili, $Q : \mathbb{R}^n \to \mathbb{R}$, è definita da:

$$Q(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_ix_j \quad (4.20)$$

Cerchiamo di controllare lo spavento: la 'doppia' sommatoria in (4.20) non significa altro che, partendo da $i = 1$, sommiamo prima tutti gli addendi $a_{1j}x_1x_j$ rispetto a j, poi ripetiamo l'operazione per $i = 2$, poi per $i = 3$, e così via fino a $i = n$, e infine sommiamo il tutto rispetto[16] a i. La (4.20) è la forma più generale possibile: tiene conto dei termini al quadrato quando $i = j$ e considera distinti gli addendi $a_{ij}x_ix_j$ e $a_{ji}x_jx_i$, cosa che funziona anche, ma non solo, quando $a_{ij} = a_{ji}$.

Il lettore attento, osservando la (4.20), dovrebbe già sentire "odore" di prodotti tra matrici, odore che si avverte ogni volta che ci troviamo di fronte a somme di prodotti fra numeri. Consideriamo il caso più semplice di due sole variabili e riscriviamo la (4.18) nella forma generale (4.20):

$$Q(x_1, x_2) = a_{11}x_1^2 + a_{22}x_2^2 + a_{12}x_1x_2 + a_{21}x_2x_1 \tag{4.21}$$

Tenendo conto della *distinzione tra vettore colonna e vettore riga*, possiamo scrivere l'argomento di Q in forma vettoriale ponendo $x^T = (x_1, x_2) \in \mathbb{R}^2$; a questo punto è facile verificare (si invita il lettore a farlo) che la (4.21) è equivalente ai seguenti prodotti:

$$Q(x) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}x_ix_j = x \cdot Ax = x^T Ax \tag{4.22}$$

dove la matrice quadrata $A\,(2 \times 2)$ è:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Il primo prodotto in (4.22) (dopo la seconda uguaglianza) è il *prodotto scalare* fra il vettore x e il vettore Ax, entrambi vettori di $\mathbb{R}^2$, dove Ax è il prodotto matriciale fra la matrice $A\,(2 \times 2)$ e la matrice $x\,(2 \times 1)$ (cioè, nuovamente il vettore x), che dà come risultato un vettore colonna di $\mathbb{R}^2$ (matrice 2×1), mentre il secondo prodotto (dopo la terza uguaglianza) è la stessa cosa però scritta in termini di prodotto fra la matrice $x^T\,(1 \times 2)$ (il vettore x riga) e la matrice (o vettore) $Ax\,(2 \times 1)$.

La (4.22) fornisce dunque una versione matriciale della (4.21) per esprimere una forma quadratica di due variabili. Prima di enunciare il caso generale di n variabili, osserviamo che fin qui non abbiamo affatto ipotizzato che $a_{12} = a_{21}$. In effetti, esistono infinite coppie di coefficienti a_{12} e a_{21} diverse tra loro che hanno somma, $a_{12} + a_{21}$, costante; questo lascia aperto un problema perché significa che potenzialmente esistono infinite matrici diverse per rappresentare la stessa forma quadratica. Ad esempio, le tre matrici:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 4 \\ 0 & -1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix} \tag{4.23}$$

sono diverse una dall'altra ma rappresentano la stessa forma quadratica $Q(x, y) = x^2 - y^2 + 4xy$ (lo si verifichi). Ma noi siamo consapevoli che esiste una

[16] Se il lettore continua ad avvertire un seppur minimo disagio è invitato a scrivere questa procedura per esteso.

via d'uscita a questo impasse: ponendo $a_{12} = a_{21}$ la rappresentazione matriciale diventa unica; così, oltre all'unicità della rappresentazione matriciale, imponiamo anche la simmetria della matrice rappresentante, proprietà che tornerà utile in seguito. Stabiliamo dunque per convenzione che l'unica vera rappresentante di $Q(x, y) = x^2 - y^2 + 4xy$ è la matrice A in (4.23), l'unica tale che $a_{12} = a_{21} = 2$.

TEOREMA 4.3 La forma quadratica $Q : \mathbb{R}^n \to \mathbb{R}$ ammette la rappresentazione matriciale:

$$Q(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x^T A x \quad (4.24)$$

dove $A\,(n \times n)$ è l'*unica matrice simmetrica* che soddisfa le uguaglianze. Viceversa, ogni matrice simmetrica $A\,(n \times n)$ individua una e una sola forma quadratica $Q : \mathbb{R}^n \to \mathbb{R}$ definita dalla (4.24).

I prodotti nei due termini a destra della (4.24) assumono significato perché x è un vettore colonna di $\mathbb{R}^n$ (ovvero è una matrice $n \times 1$, così come x^T è una matrice $1 \times n$) e A è $n \times n$. Si consiglia il lettore di verificare la (4.24) per qualche forma quadratica con $n > 2$ variabili.

Il Teorema 4.3 assomiglia al Teorema 2.1: sia le funzioni lineari sia le forme quadratiche ammettono una rappresentazione matriciale; per le prime è sempre unica mentre per le seconde l'unicità è garantita solo se la matrice rappresentante è simmetrica. Poiché, in generale, qualsiasi matrice quadrata rappresenta una forma quadratica, il seguente è un criterio per individuare l'*unica matrice simmetrica* che rappresenta formalmente la forma quadratica in questione.

REGOLA 4.1 Data una qualsiasi matrice quadrata $B\,(n \times n)$, la matrice

$$A = \frac{1}{2}\left(B + B^T\right) \quad (4.25)$$

1. rappresenta la stessa forma quadratica di B, $Q(x) = x^T A x = x^T B x$,
2. è simmetrica.

Le matrici ammissibili per definire forme quadratiche sono tutte e sole le matrici simmetriche.

ESEMPIO 4.12 La forma quadratica $Q : \mathbb{R}^2 \to \mathbb{R}$ definita da $Q(x, y) = x^2 - xy + 5y^2$ ammette come unica rappresentazione matriciale quella definita dalla matrice simmetrica:

$$A = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 5 \end{bmatrix}$$

ottenuta ponendo sulla diagonale principale i coefficienti dei termini x^2 e y^2 così come stanno e ponendo la metà del coefficiente del termine xy fuori dalla diagonale.

La tecnica descritta dalla Regola 4.1 e adottata nell'Esempio 4.12 vale per qualsiasi numero di variabili.

D'altro canto, per qualsiasi matrice simmetrica è possibile ricostruire la formula esplicita della forma quadratica da essa definita. Ad esempio la forma quadratica $Q : \mathbb{R}^3 \to \mathbb{R}$ definita da:

$$A = \begin{bmatrix} -1 & 2 & 1 \\ 2 & -3 & -5/2 \\ 1 & -5/2 & 0 \end{bmatrix}$$

è $Q(x_1, x_2, x_3) = -x_1^2 - 3x_2^2 + 4x_1x_2 + 2x_1x_3 - 5x_2x_3$.

Se la matrice non è simmetrica, si deve prima utilizzare la Formula (4.25).

4.5.2 Forme quadratiche definite da matrici Hessiane

La disquisizione sul significato di 'approssimazione di secondo grado' su cui ci siamo dilungati nei Paragrafi 11.1.4 e 11.1.5 del primo volume mantiene inalterata la sua validità nel caso di funzioni di n variabili, $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$: fissato un punto di riferimento $x^0 \in \text{Int}(X)$ basta sostituire una forma quadratica $Q : \mathbb{R}^n \to \mathbb{R}$ al posto del termine di secondo grado del polinomio definito nella (11.14) del primo volume per introdurre una *componente di curvatura che si somma all'approssimazione lineare* $f(x^0) + \nabla f(x^0) \cdot \mathrm{dx}$ definita in (4.15).

Riguardando l'espressione del polinomio di Taylor di secondo grado per funzioni di una variabile nella (11.14) del primo volume, e sapendo che le derivate parziali seconde miste di una *funzione* $f : \mathbb{R}^n \to \mathbb{R}$ *di classe* C^2 sono uguali, il colpo d'occhio offerto dalla (4.20) [o, se si preferisce, dalla (4.19), che è scritta per esteso] è stimolante: la tentazione naturale è quella di porre $a_{ii} = (1/2) f_{x_i x_i}$ e $a_{ij} = a_{ji} = (1/2) f_{x_i x_j} = (1/2) f_{x_j x_i}$, per ogni $i, j = 1, \ldots, n$. Arrangiando tutti questi coefficienti in una matrice, il Teorema 4.3 ci dice che abbiamo costruito la *forma quadratica* $Q : \mathbb{R}^n \to \mathbb{R}$ *definita dalla matrice Hessiana* H (simmetrica) della funzione f moltiplicata per lo scalare $1/2$:

$$Q(x) = \frac{1}{2} x^T H x \tag{4.26}$$

Il nostro entusiasmo viene ancora una volta premiato dal prossimo risultato, il Teorema 4.4.

4.5.3 Il polinomio di Taylor

Il Teorema 11.2 del primo volume aveva definito il polinomio di Taylor come *funzione della variazione*, $\mathrm{dx} = x - x_0 \in \mathbb{R}$; ora dobbiamo fare altrettanto in n variabili: definiamo il *vettore variazione* $(\mathrm{dx}) = (\mathrm{dx}_1, \mathrm{dx}_2, \ldots, \mathrm{dx}_n) \in \mathbb{R}^n$, dove $\mathrm{dx}_i = x_i - x_i^0$, per $i = 1, 2, \ldots, n$. Ricordiamo inoltre che $\|\mathrm{dx}\| = \|x - x^0\|$ indica la *distanza euclidea fra il punto di riferimento* x^0 *e il punto* $x = x^0 + \mathrm{dx}$.

TEOREMA 4.4 (APPROSSIMAZIONE DI TAYLOR DI SECONDO GRADO) Se la funzione $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, è di classe C^2 nel punto $x^0 \in \operatorname{Int}(X)$, allora:

$$f(x) = f\left(x^0 + \mathrm{dx}\right) = f\left(x^0\right) + \nabla f\left(x^0\right) \cdot \mathrm{dx} + \frac{1}{2}(\mathrm{dx})^T H\left(x^0\right)(\mathrm{dx}) + o\left(\|\mathrm{dx}\|^2\right) \quad (4.27)$$

dove il termine $o\left(\|\mathrm{dx}\|^2\right)$ indica un errore che tende a zero più rapidamente del *quadrato della distanza (variazione)* $\|\mathrm{dx}\| = \left\|x - x^0\right\|$.

Si noti come la notazione matriciale/vettoriale determini la somiglianza della (4.27) con la (11.14) del primo volume. Possiamo quindi migliorare l'approssimazione lineare (4.15) scrivendo:

$$f(x) = f\left(x^0 + \mathrm{dx}\right) \approx f\left(x^0\right) + \nabla f\left(x^0\right) \cdot \mathrm{dx} + \frac{1}{2}(\mathrm{dx})^T H\left(x^0\right)(\mathrm{dx}) \quad (4.28)$$

dove il termine a destra è il **polinomio di Taylor di secondo grado** definito come funzione del vettore variazione dx, che indicheremo con $T_2(\mathrm{dx})$. Analogamente, indicheremo con $T_1(\mathrm{dx}) = f\left(x^0\right) + \nabla f\left(x^0\right) \cdot \mathrm{dx}$ il **polinomio di Taylor di primo grado** [la (4.15)], anch'esso definito come funzione della variazione dx. Tutte le considerazioni sulla riduzione dell'errore commesso considerando T_2 al posto di T_1 rimangono valide anche nel caso di n variabili; in particolare, $o\left(\|\mathrm{dx}\|^2\right)$ è decisamente inferiore all'errore $o(\|\mathrm{dx}\|)$ introdotto con l'approssimazione lineare perché dipende dal quadrato della distanza (variazione), $\|\mathrm{dx}\|^2$, anziché dalla distanza semplice $\|\mathrm{dx}\|$.

ESEMPIO 4.13 Studiamo la funzione di tipo Cobb-Douglas $f(x,y) = x^{1/5}y^{4/5}$ in un intorno del punto $(x_0, y_0) = (1,1)$. Vale:

$$\nabla f = \left((1/5)\, x^{-4/5}y^{4/5}, (4/5)\, x^{1/5}y^{-1/5}\right) \quad \text{e}$$
$$H = \begin{bmatrix} -(4/25)\, x^{-9/5}y^{4/5} & (4/25)\, x^{-4/5}y^{-1/5} \\ (4/25)\, x^{-4/5}y^{-1/5} & -(4/25)\, x^{1/5}y^{-6/5} \end{bmatrix}$$

e quindi:

$$\nabla f(1,1) = (1/5, 4/5) \quad \text{e} \quad H(1,1) = \begin{bmatrix} -4/25 & 4/25 \\ 4/25 & -4/25 \end{bmatrix}$$

Poiché $f(1,1) = 1$, i polinomi di Taylor di primo e secondo grado sono $T_1(\mathrm{dx}, \mathrm{dy}) = 1 + (1/5)\,\mathrm{dx} + (4/5)\,\mathrm{dy}$ e $T_2(\mathrm{dx}, \mathrm{dy}) = 1 + (1/5)\,\mathrm{dx} + (4/5)\,\mathrm{dy} - (2/25)(\mathrm{dx})^2 - (2/25)(\mathrm{dy})^2 + (4/25)(\mathrm{dx})(\mathrm{dy})$.

Ponendo $\mathrm{dx} = -0.1$ e $\mathrm{dy} = 0.1$ possiamo approssimare il numero $(0.9)^{1/5}(1.1)^{4/5}$ mediante T_1 e T_2: si vede facilmente che vale $T_1(0.9, 1.1) = 1.06$ e $T_2(0.9, 1.1) = 1.0568$, entrambi da confrontare con il valore vero (ottenuto dalla calcolatrice) $(0.9)^{1/5}(1.1)^{4/5} \simeq 1.05673$, per cui riesce $o(\|(\mathrm{dx}, \mathrm{dy})\|) \simeq |1.06 - 1.05673| = 0.00327$ mentre $o\left(\|(\mathrm{dx}, \mathrm{dy})\|^2\right) \simeq |1.0568 - 1.05673| = 0.00007$.

4.6 Aspetti peculiari delle funzioni differenziabili

Concentriamo ora la nostra attenzione su proprietà peculiari delle funzioni di n variabili che non assumono significato per le funzioni di una variabile e che si basano sul concetto di 'direzione', che ha senso solo in spazi di dimensione $n \geq 2$. Iniziamo studiando la *derivata delle funzioni composte* di più variabili, strumento che ci servirà poi per analizzare gli *"aspetti direzionali" della derivata.*

4.6.1 Funzione composta e regola della catena

In questo paragrafo studiamo la *funzione composta* nei casi in cui la funzione esterna dipende da n variabili mentre le funzioni interne dipendono da una sola variabile,[17] con l'obiettivo di definire la regola di derivazione di tale funzione sulla falsariga di quanto visto nel Paragrafo 10.6.6 del primo volume.

Indichiamo con $x_i(t)$ n funzioni[18] della stessa variabile indipendente t, ovvero $x_i : \mathbb{R} \to \mathbb{R}$, per $i = 1, 2, \ldots, n$, e consideriamo il vettore[19] $x(t) = (x_1(t), x_2(t), \ldots, x_n(t)) \in \mathbb{R}^n$. Poi consideriamo la solita funzione di n variabili $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$. Possiamo dunque definire la funzione composta di una sola variabile $h : \mathbb{R} \to \mathbb{R}$ come:

$$h(t) = f[x(t)] = f[x_1(t), x_2(t), \ldots, x_n(t)]$$

Il valore immagine $y = h(t)$ dipende dalla sola variabile t attraverso una complicata alchimia che presuppone che i valori delle n variabili indipendenti, x_i, della funzione esterna f dipendano tutti dalla stessa variabile, t, e a loro volta determinano $y = h(t)$ "attraverso" f. In altre parole, fissato t, rimangono determinati n numeri $x_1(t), x_2(t), \ldots, x_n(t)$ che, se $(x_1(t), x_2(t), \ldots, x_n(t)) \in X$, a loro volta determinano il valore immagine $y = h(t) = f[x_1(t), x_2(t), \ldots, x_n(t)]$. La Figura 4.18 illustra la situazione mediante i diagrammi di Venn; si noti che il vettore $(x_1(t), x_2(t), \ldots, x_n(t))$ deve stare nel dominio X di f.

La costruzione appena vista è importante perché consente di trasformare una funzione di n variabili, la f, in una funzione di una variabile sola, la h, semplicemente introducendo un legame tra le n variabili x_i [legame determinato dalle funzioni $x_i(t)$] che rimangono così tutte "controllate" da un'unica variabile t. Per essere precisi, il vettore $x(t)$ definisce una curva nello spazio $\mathbb{R}^n$.

---

[17] Si tratta di un caso molto particolare, dal momento che, in generale, le funzioni interne potrebbero dipendere anch'esse da più variabili. Per i nostri scopi è sufficiente saper maneggiare questa situazione specifica.

[18] Usare i simboli x_i per indicare funzioni di una variabile è chiaramente un'anomalia rispetto alla simbologia usata finora; il lettore attento però si sarà sicuramente già fatto un'idea su dove vogliamo andare a parare...

[19] Possiamo pensare al vettore $x(t)$ come ad una *funzione vettoriale* $x : \mathbb{R} \to \mathbb{R}^n$, secondo i dettami del Paragrafo 2.2.1.

DEFINIZIONE 4.14 (CURVA NELLO SPAZIO) Se le funzioni $x_i(t)$, per $i = 1, 2, \ldots, n$, *sono tutte continue*, al variare di $t \in \mathbb{R}$ il vettore $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))$ definisce una **curva nello spazio $\mathbb{R}^n$**.

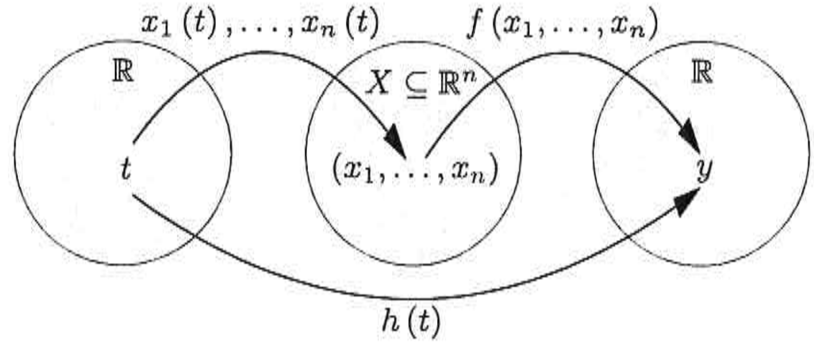


FIGURA 4.18: la funzione composta $h(t) = f[x_1(t), x_2(t), \ldots, x_n(t)]$ con i diagrammi di Venn.

Se le $x_i(t)$ sono continue, valori di t prossimi a t_0 individuano punti $x(t)$ di $\mathbb{R}^n$ prossimi a $x(t_0)$, ovvero ciò che si ottiene al variare di t è una curva continua, senza interruzioni o strappi.

Quindi $h(t)$ determina una *restrizione* della f di partenza alla curva $x(t)$; il grafico di $h(t)$ è dunque il grafico di una funzione di una sola variabile "immerso" nello spazio $n+1$-dimensionale $\mathbb{R}^{n+1}$; si tratta di una curva in $\mathbb{R}^{n+1}$ che "si sviluppa in $n+1$ dimensioni". Naturalmente, se $n > 2$, si tratta di una (iper)curva in senso lato (astratto), poiché non siamo in grado di disegnarla.

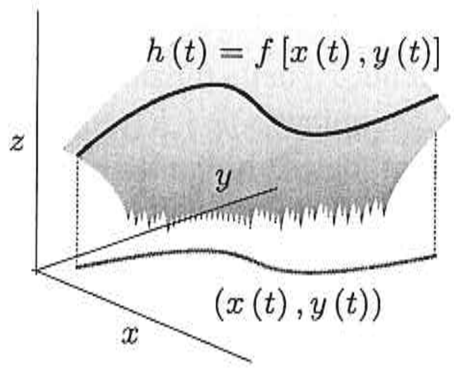


FIGURA 4.19: il grafico di $h(t) = f[x(t), y(t)]$ è la restrizione del grafico di f alla curva $(x(t), y(t))$.

Il caso $n = 2$ chiarisce le idee su questo tipo di costruzione poiché consente una rappresentazione grafica. Se $n = 2$, $(x(t), y(t))$ definisce una *curva nel piano*[20] $\mathbb{R}^2$. La Figura 4.19 mostra un esempio: il grafico di $h(t)$ è la curva nera che giace sul grafico di $f(x, y)$ (sulla superficie grigia), tale curva è ottenuta come proiezione verticale della curva $(x(t), y(t))$ disegnata sotto, sul piano orizzontale. La cosa importante è che $h(t)$ rimane a tutti gli effetti una funzione della sola variabile t, variabile che governa integralmente il suo andamento, anche se il suo grafico assomiglia più a un serpente che si muove nello spazio tridimensionale (una curva in $\mathbb{R}^3$) piuttosto che alla sua ombra proiettata sul pavimento (una curva in $\mathbb{R}^2$).[21]

[20] Avevamo già parlato di curve nel piano studiando le *funzioni implicite* nel Paragrafo 3.5 del primo volume. In realtà la nozione di curva fornita dalla Definizione 4.14 per $n = 2$ non è altro che una generalizzazione del concetto di curva definita implicitamente da una funzione: se poniamo $x(t) = t$, cioè assumiamo che la prima funzione sia la funzione identica, la curva che otteniamo è $(x, y(x))$, cioè la curva definita implicitamente dalla funzione $y(x)$. In effetti le curve che costruiremo nel seguito si baseranno proprio sulla funzione implicita, come approfondiremo nel Paragrafo 6.1.2.

[21] Una suggestione truculenta è immaginare la curva in $\mathbb{R}^3$ come un serpente "vivo" e in movimento a un certo istante, e la sottostante curva in $\mathbb{R}^2$ come lo stesso serpente però "morto", schiacciato da un rullo compressore un istante dopo.

La costruzione di $h(t)$ costituirà il passaggio chiave nello studio di problemi di ottimo vincolato in n variabili, i quali verranno in questo modo trasformati in problemi in una sola variabile e verranno risolti con i metodi già visti nel Capitolo 13 del primo volume.

La regola per derivare funzioni composte di una sola variabile prescrive il calcolo del prodotto tra la derivata della funzione esterna e la derivata della funzione interna. Sfruttando l'idea che la derivata della funzione esterna è rappresentata dal differenziale, il quale sintetizza in modo additivo la variazione lungo le direzioni canoniche, la prossima regola estende il risultato alla funzione composta $h(t) = f[x(t)]$, dove $x : \mathbb{R} \to \mathbb{R}^n$ e $f : \mathbb{R}^n \to \mathbb{R}$.

REGOLA 4.2 (DERIVATA DELLA FUNZIONE COMPOSTA) Siano $x_i : \mathbb{R} \to \mathbb{R}$, per $i = 1, 2, \ldots, n$, funzioni derivabili e $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^2$, una funzione differenziabile su $\operatorname{Int}(X)$. Definiamo la funzione composta $h : \mathbb{R} \to \mathbb{R}$ come $h(t) = f[x(t)]$ e fissiamo $t_0 \in \mathbb{R}$ tale che $x(t_0) \in \operatorname{Int}(X)$. Allora:

$$\begin{aligned} h'(t_0) &= f_{x_1}[x(t_0)]\, x_1'(t_0) + f_{x_2}[x(t_0)]\, x_2'(t_0) + \cdots + f_{x_n}[x(t_0)]\, x_n'(t_0) \\ &= \nabla f[x(t_0)] \cdot x'(t_0) \end{aligned} \tag{4.29}$$

dove $\nabla f[x(t_0)]$ è il gradiente di f in $x(t_0)$ e $x'(t_0)$ denota il *vettore delle n derivate di una sola variabile,* $(x_1'(t_0), x_2'(t_0), \ldots, x_n'(t_0))$, in cui ciascuna derivata è calcolata in t_0.

La derivata di $h(t) = f[x(t)]$ è dunque la somma dei prodotti di ciascuna derivata parziale della funzione esterna f con la derivata della funzione interna corrispondente a ciascuna variabile x_i.

Naturalmente, la regola appena enunciata fornisce la *funzione derivata* $h'(t)$ per tutti i valori $t \in \mathbb{R}$ tali che $x(t) \in \operatorname{Int}(X)$. Richiamiamo la Notazione 10.2 introdotta nel primo volume che aiuta a memorizzare la Regola 4.2 ed è molto popolare nella letteratura economica.

NOTAZIONE 4.7 (REGOLA DELLA CATENA) La derivata della funzione composta si scrive talvolta nella seguente forma succinta e facile da memorizzare:

$$\frac{\mathrm{df}}{\mathrm{dt}} = \frac{\partial f}{\partial x_1}\frac{\mathrm{dx}_1}{\mathrm{dt}} + \frac{\partial f}{\partial x_2}\frac{\mathrm{dx}_2}{\mathrm{dt}} + \cdots + \frac{\partial f}{\partial x_n}\frac{\mathrm{dx}_n}{\mathrm{dt}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\mathrm{dx}_i}{\mathrm{dt}} \tag{4.30}$$

dove abbiamo mescolato le notazioni infinitesimali per le derivate parziali, $\partial f / \partial x_i$, e le notazioni differenziali df, dx_i e dt. Il motivo di questo ibrido è dettato da ragioni di coerenza con quanto visto nei paragrafi precedenti; infatti, se moltiplichiamo tutti i termini della (4.30) per dt (ricordiamo che, trattandosi di notazione differenziale, $\mathrm{dt} \neq 0$) otteniamo proprio il differenziale totale (4.12).

ESEMPIO 4.14 Siano $f(x, y) = x \ln y$, $x(t) = -2t$ e $y(t) = 3t^2$. Calcoliamo la derivata della funzione composta $h(t) = f[x(t), y(t)]$ mediante la Regola 4.2: vale

$\nabla f = (\ln y, x/y)$, mentre $x'(t) = -2$ e $y'(t) = 6t$, pertanto:

$$h'(t) = \ln[y(t)](-2) + \frac{x(t)}{y(t)}6t = -2\ln(3t^2) + -\frac{2t}{3t^2}6t = -2\ln(3t^2) - 4$$

Verifichiamo il risultato calcolando direttamente la derivata di $h(t)$: sostituendo $x(t) = -2t$ e $y(t) = 3t^2$ nell'espressione di $f(x,y)$, otteniamo $h(t) = -2t\ln(3t^2)$, che, derivando rispetto a t, fornisce esattamente:

$$h'(t) = -2\ln(3t^2) - \frac{2t}{3t^2}6t = -2\ln(3t^2) - 4$$

L'esempio precedente suggerisce che per calcolare la derivata di una funzione composta $h(t) = f[x(t)]$ è più facile (e veloce) sostituire le espressioni delle $x_i(t)$ nella forma funzionale di f, ottenendo così un'espressione in una sola variabile che si può derivare direttamente. In effetti, la Regola 4.2 non viene utilizzata nel calcolo diretto; essa, come vedremo, assume un'importanza più teorica che pratica. Il prossimo esempio tratta una situazione un po' più complicata in cui il calcolo diretto della derivata non è possibile perché non viene esplicitata una delle funzioni componenti e bisogna quindi per forza ricorrere alla Regola 4.2.

ESEMPIO 4.15 Data una funzione di due variabili $f(x,y)$ differenziabile, consideriamo la composizione $g(t) = \ln\left[f\left(\sqrt{t}, t^3\right)\right]$ e calcoliamo $g'(t)$. Poniamo $x(t) = \sqrt{t}$, $y(t) = t^3$, $h(t) = f[x(t), y(t)]$ e $\phi(z) = \ln z$. Allora possiamo scrivere $g(t) = \phi[h(t)]$, la cui derivata è immediata da calcolare con la regola della catena in una variabile:

$$g'(t) = \phi'[h(t)]\,h'(t) = \frac{h'(t)}{h(t)} = \frac{h'(t)}{f[x(t), y(t)]} = \frac{h'(t)}{f\left(\sqrt{t}, t^3\right)}$$

Rimane da ricavare $h'(t)$, per la quale applichiamo la (4.30):

$$h'(t) = \frac{\partial f}{\partial x}\frac{\mathrm{dx}}{\mathrm{dt}} + \frac{\partial f}{\partial y}\frac{\mathrm{dy}}{\mathrm{dt}} = \frac{\partial f}{\partial x}\frac{1}{2\sqrt{t}} + \frac{\partial f}{\partial y}3t^2$$

pertanto possiamo scrivere:

$$g'(t) = \frac{f_x\left(\sqrt{t}, t^3\right)/2\sqrt{t} + f_y\left(\sqrt{t}, t^3\right)\left(3t^2\right)}{f\left(\sqrt{t}, t^3\right)}$$

4.6.2 Derivata direzionale

Una prima applicazione della regola della catena ci permette di soddisfare una curiosità che era rimasta in sospeso dal Paragrafo 4.3.2, nel quale avevamo accennato al fatto che, almeno in linea di principio, una definizione soddisfacente di derivabilità per funzioni di due variabili deve tenere conto delle derivate rispetto a *tutte le direzioni uscenti dal punto* (x_0, y_0), come schematizzato dalla Figura 4.15. Poiché tali direzioni sono infinite, è irrimediabilmente vano ogni tentativo di calcolarle tutte; mediante la regola della catena però ora possiamo toglierci lo sfizio di calcolare quante e quali derivate vogliamo, rispetto a qualsiasi direzione uscente da (x_0, y_0) a nostra scelta. Ricordiamo che l'esistenza di tali *derivate*, dette *direzionali*, è assicurata dal Teorema 4.1.

Vediamo il caso di una funzione f di n variabili. Sfruttando la Definizione 4.14, si tratta di trovare n funzioni di una variabile, $x_1(t), x_2(t), \ldots, x_n(t)$, che generino una curva particolare: la retta nello spazio $\mathbb{R}^n$ passante per il punto di riferimento x^0 che abbia la direzione desiderata. Poi basta calcolare la derivata di f ristretta a tale retta, cioè la derivata di $h(t) = f[x(t)]$.

Una retta è individuata in modo univoco da un segmento; ricordiamo che la *notazione vettoriale di un punto* $x \in \mathbb{R}^n$ diverso dall'origine è proprio il segmento orientato che unisce l'origine con il punto x (Paragrafo 1.1). Una scelta opportuna di coordinate $x_1, x_2, \ldots, x_n$ per il vettore x, dunque, definisce una e una sola retta. Traslando l'origine del nostro spazio in modo da farla coincidere con il punto x^0, il gioco è fatto: abbiamo stabilito una direzione di uscita dal punto x^0 stesso.

La 'lunghezza' del vettore x, la norma $\|x\|$, non conta: due vettori proporzionali (sulla stessa retta) con norma diversa individuano la stessa retta. Pertanto, da un lato per semplificarci la vita e dall'altro per mantenere l'interpretazione grafica della derivata direzionale come pendenza della retta tangente al grafico di f lungo la direzione generata dal vettore[22] x (come avevamo già fatto per le derivate parziali nella Figura 4.13), stabiliamo fin d'ora la convenzione di usare sempre e solo *vettori di norma unitaria* allo scopo di stabilire direzioni d'uscita; in altre parole, il nostro vettore sarà sempre un **versore** (Definizione 1.6) che indicheremo con v, a ricordare che $\|v\| = 1$.

Una volta scelto il versore v che individuata la retta (la direzione) che ci interessa, dobbiamo decidere come "ci si muove" lungo questa direzione (ricordiamoci che dobbiamo calcolare una derivata...). Poiché vogliamo governare questo "movimento" mediante la variabile t, dobbiamo individuare n funzioni della sola variabile t, $x_1(t), x_2(t), \ldots, x_n(t)$, che determinano la nostra retta.

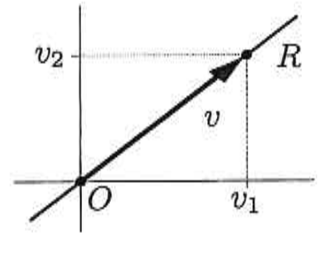


FIGURA 4.20: il versore v e la retta R.

La Figura 4.20 mostra un versore $v = (v_1, v_2)$ che individua una specifica retta, che indichiamo con R, *nel caso di due variabili*. Nel Paragrafo 1.2.2 abbiamo appreso che è possibile rappresentare tutti i punti della retta individuata da un vettore come *moltiplicazione scalare* fra il vettore stesso e tutti i possibili valori (reali) di uno scalare. Pertanto, se indichiamo con (la variabile) t lo scalare in questione, possiamo definire la retta R mediante la moltiplicazione scalare $tv = (tv_1, tv_2)$. In altre parole, la curva $(x(t), y(t))$ generata dalle *funzioni lineari*[23] di una variabile $x(t) = v_1 t$ e $y(t) = v_2 t$ è proprio la

[22]Vedremo il perché nella prossima Osservazione 4.5.

[23]Lineari nel senso che, come si verifica immediatamente, soddisfano entrambe la Definizione 2.2.

retta R cercata; essa passa per i punti $O = (0, 0)$ (quando $t = 0$) e $v = (v_1, v_2)$ (quando $t = 1$), e ha *pendenza* pari al rapporto fra le coordinate di v, v_2/v_1. La retta R definita da $(v_1 t, v_2 t)$ determina dunque *la direzione che ci interessa.*

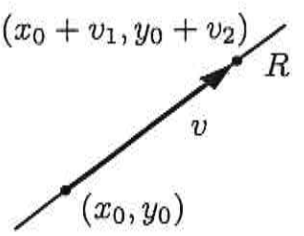


FIGURA 4.21: v e la direzione di uscita da (x_0, y_0).

Dobbiamo ora 'traslare' la retta appena individuata in modo che l'origine O coincida con il punto (x_0, y_0) sul quale vogliamo calcolare la derivata di $f(x, y)$ rispetto alla direzione definita da R. A tal fine è sufficiente porre:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} x_0 + v_1 t \\ y_0 + v_2 t \end{bmatrix} \tag{4.31}$$

La retta definita in (4.31) dalle funzioni di una variabile $x(t) = x_0 + v_1 t$ e $y(t) = y_0 + v_2 t$ è parallela alla retta R – mantiene, cioè, la direzione stabilita dal versore v – e passa per il punto (x_0, y_0). Si verifichi per esercizio che la pendenza di questa nuova retta è ancora pari a v_2/v_1, e che essa passa per il punto (x_0, y_0) (quando $t = 0$) e per il punto $(x_0 + v_1, y_0 + v_2)$ (quando $t = 1$). La Figura 4.21 mostra che questa costruzione equivale ad "applicare" il versore v al punto (x_0, y_0) anziché all'origine.

Infine, per poter calcolare la derivata di f lungo la direzione individuata dalla retta (4.31), introduciamo la funzione composta $h(t) = f[x(t), y(t)] = f(x_0 + v_1 t, y_0 + v_2 t)$ e osserviamo che $h(0) = f(x_0, y_0)$, ovvero per $t = 0$ la funzione composta h ha lo stesso valore assunto da f in (x_0, y_0), punto in cui vogliamo calcolare la derivata di f rispetto alla direzione scelta. Quindi la derivata direzionale di f cercata corrisponde alla derivata di h nel punto $t = 0$, $h'(0)$. Applicando la Regola 4.2 otteniamo:

$$h'(t) = f_x(x_0 + v_1 t, y_0 + v_2 t)\, v_1 + f_y(x_0 + v_1 t, y_0 + v_2 t)\, v_2$$

che in $t = 0$ vale:

$$h'(0) = f_x(x_0, y_0)\, v_1 + f_y(x_0, y_0)\, v_2$$

A parole, abbiamo stabilito che la derivata di una funzione di due variabili f in un punto (x_0, y_0) rispetto alla direzione definita da un versore v è data dalla somma dei prodotti tra le derivate parziali di f in (x_0, y_0) e le coordinate di v. Riassumiamo il risultato nella seguente regola generale.

REGOLA 4.3 (DERIVATA DIREZIONALE) Data $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, differenziabile su $\text{Int}(X)$, e un versore $v \in \mathbb{R}^n$, cioè un vettore tale che $\|v\| = 1$, la *derivata* di f in un punto $x^0 \in \text{Int}(X)$ *lungo la direzione determinata da* v, che si indica con $(\partial/\partial v) f(x^0)$ [oppure $f_v(x^0)$ o $D_v f(x^0)$], è data da:

$$\frac{\partial}{\partial v} f(x^0) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(x^0)\, v_i = \nabla f(x^0) \cdot v \tag{4.32}$$

è pari, cioè, al prodotto scalare fra il gradiente $\nabla f(x^0)$ e il versore v che determina la direzione.

Il principio di fondo rimane sempre lo stesso: ogni variazione della funzione f viene scomposta nella somma delle variazioni rispetto alle n direzioni canoniche (quelle parallele a ciascun asse cartesiano); nel caso della derivata direzionale tali variazioni vengono "pesate" dai valori assunti dalle coordinate del versore v, che determina la direzione rispetto a cui deriviamo. È immediato verificare che applicando la Regola 4.3 sulle direzioni canoniche, ovvero utilizzando un versore del tipo $v = (0, \ldots, 0, 1, 0, \ldots, 0)$, che ha tutte le coordinate nulle tranne la i-esima, che vale $v_i = 1$ (si noti che $\|v\| = 1$), otteniamo proprio la derivata parziale $f_{x_i}(x^0)$, per qualsiasi $i = 1, \ldots, n$.

La Figura 4.22 descrive l'usuale interpretazione della derivata direzionale come pendenza della retta tangente al grafico di una *funzione di due variabili* $f(x, y)$ ristretta alla retta $(x_0 + v_1 t, y_0 + v_2 t)$ generata dal versore v e passante per il punto (x_0, y_0); essa coincide con la retta tangente al grafico (in $\mathbb{R}^3$) della funzione di una sola variabile $h(t)$ nel punto $t = 0$. In questo caso specifico il grafico di $h(t)$ coincide con l'intersezione tra il grafico di f e il piano verticale passante per la retta generata da v.

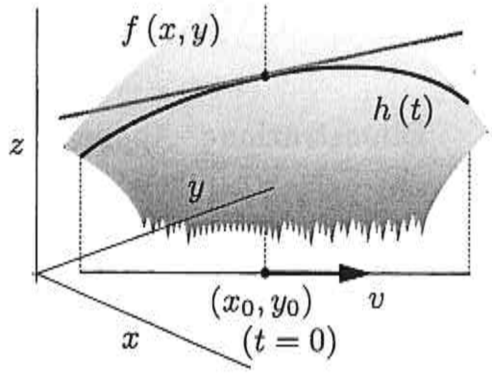


FIGURA 4.22: derivata direzionale come pendenza della retta tangente al grafico di $h(t)$ in $t = 0$.

OSSERVAZIONE 4.5 È evidente che il valore di $(\partial/\partial v) f(x^0)$ dato dalla Formula (4.32) dipende dalla norma del vettore v che stabilisce la direzione, ossia dalle sue coordinate $(v_1, \ldots, v_n)$: poiché il gradiente $\nabla f(x^0)$ nel punto x^0 è costante, valori per $v_1, \ldots, v_n$ diversi da quelli che rendono v un versore, ossia quando $\|v\| \neq 1$, fornirebbero un valore per la derivata direzionale in (4.32) diverso da quello pari alla pendenza dalla retta tangente al grafico di f nel punto x^0 lungo la direzione generata da v.

ESEMPIO 4.16 Studiamo la derivata di $f(x, y) = 5x^{4/5}y^{1/5}$ rispetto alla direzione $v = (2/\sqrt{5}, 1/\sqrt{5})$ nel punto $(x_0, y_0) = (1, 1)$. Osserviamo che $\|v\| = \sqrt{(2/\sqrt{5})^2 + (1/\sqrt{5})^2} = \sqrt{4/5 + 1/5} = 1$, ovvero v è un versore. Quindi applichiamo la Regola 4.3 per ottenere:

$$\begin{aligned} f_v(x_0, y_0) &= f_x(x_0, y_0)\frac{2}{\sqrt{5}} + f_y(x_0, y_0)\frac{1}{\sqrt{5}} \\ &= 4(y_0/x_0)^{1/5}\frac{2}{\sqrt{5}} + (x_0/y_0)^{4/5}\frac{1}{\sqrt{5}} \end{aligned}$$

che valutata in $(x_0, y_0) = (1, 1)$ vale $f_v(1, 1) = 9/\sqrt{5}$.

4.6.3 Gradiente e direzione di massima pendenza

Concludiamo questo capitolo introducendo un'importante proprietà del gradiente che ci permetterà di giustificare alcune tecniche di risoluzione di problemi di ottimo vincolato nel seguito. Fissato un punto interno x^0, il gradiente $\nabla f\left(x^0\right)$ è a tutti gli effetti un vettore costituito da n coordinate, le derivate parziali $f_{x_i}\left(x^0\right)$, come mostra la Figura 4.23 nel caso di due variabili. Dunque, lo stesso *gradiente* individua a pieno titolo una *direzione in uscita* dal punto x^0. Si dimostra facilmente che tale direzione ha una peculiarità importante, come riporta il seguente teorema.

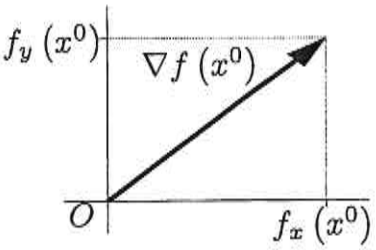


FIGURA 4.23: il vettore $\nabla f\left(x^0\right)$.

TEOREMA 4.5 Se il gradiente, ∇f, di una funzione $f : X \to \mathbb{R}$, $X \subseteq \mathbb{R}^n$, differenziabile, calcolato in un punto $x^0 \in \operatorname{Int}(X)$ è tale che $\nabla f\left(x^0\right) \neq (0,0,\ldots,0)$, allora esso indica la direzione di uscita da x^0 verso cui f aumenta più rapidamente.

Dimostrazione. Applicando una proprietà del valore assoluto, la disuguaglianza di Cauchy-Schwarz (punto 3 della Proprietà 1.4) e il fatto che $\|v\| = 1$ alla Formula (4.32) otteniamo il seguente valore maggiorante per $(\partial/\partial v)\, f\left(x^0\right)$:

$$\frac{\partial}{\partial v} f\left(x^0\right) = \nabla f\left(x^0\right) \cdot v \leq \left|\nabla f\left(x^0\right) \cdot v\right| \leq \left\|\nabla f\left(x^0\right)\right\| \|v\| = \left\|\nabla f\left(x^0\right)\right\| \tag{4.33}$$

A parole, la derivata direzionale di f nel punto x^0 rispetto a qualsiasi direzione (qualsiasi versore v) non può superare il valore della norma del gradiente nello stesso punto, $\left\|\nabla f\left(x^0\right)\right\|$; pertanto, anche il suo valore massimo sul cerchio unitario (si veda la Figura 1.8) non può superarlo: $\max\left\{(\partial/\partial v)\, f\left(x^0\right) : \|v\| = 1\right\} \leq \left\|\nabla f\left(x^0\right)\right\|$. Se consideriamo la direzione definita proprio dal vettore gradiente, ossia, omettendo l'argomento x^0 per semplicità, consideriamo il versore $v = (1/\|\nabla f\|)\, \nabla f$, che ha la stessa direzione di ∇f ma norma 1 [si riveda la Formula (1.2)], possiamo scrivere:

$$\frac{\partial}{\partial v} f = \nabla f \cdot v = \nabla f \cdot \left[\left(\frac{1}{\|\nabla f\|}\right) \nabla f\right] = \frac{\nabla f \cdot \nabla f}{\|\nabla f\|} = \frac{\|\nabla f\|^2}{\|\nabla f\|} = \|\nabla f\|$$

da cui impariamo che, recuperando l'argomento x^0, quando la direzione è definita dal gradiente $\nabla f\left(x^0\right)$, la disuguaglianza (4.33) vale con uguaglianza e pertanto il valore massimo per la derivata direzionale è raggiunto, $\max\left\{(\partial/\partial v)\, f\left(x^0\right) : \|v\| = 1\right\} = \left\|\nabla f\left(x^0\right)\right\|$, e il teorema è dimostrato. ■

Vedremo più avanti, nell'Esercizio 6.6, una dimostrazione alternativa di questo teorema nel caso di due variabili.

La Figura 4.24 illustra il teorema per una funzione di due variabili $f(x, y)$: il vettore gradiente ∇f è la proiezione sul piano $\mathbb{R}^2$ della freccia riportata sul grafico di f che indica la direzione di massima pendenza del grafico stesso a partire dal punto (x_0, y_0).

ESEMPIO 4.17 Consideriamo di nuovo la funzione $f(x,y) = 5x^{4/5}y^{1/5}$ che interpretiamo come una funzione di produzione che dipende da due fattori produttivi. Supponiamo che la produzione attuale impieghi le quantità $(x_0, y_0) = (1,1)$ di fattori. Per stabilire con quale proporzione dobbiamo aumentare entrambi i fattori per aumentare più rapidamente il prodotto, calcoliamo il gradiente:

$$\nabla f(x_0, y_0) = \left(4\left(\frac{y_0}{x_0}\right)^{1/5}, \left(\frac{x_0}{y_0}\right)^{4/5}\right)$$

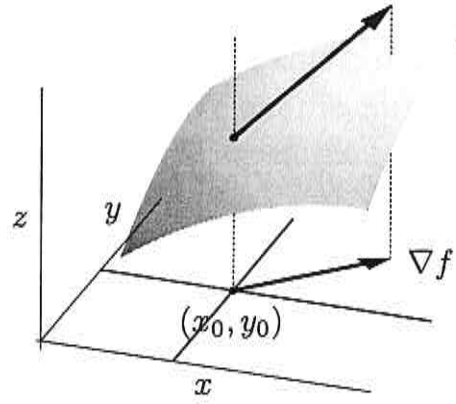


FIGURA 4.24: ∇f e direzione di massima pendenza da (x_0, y_0).

e quindi $\nabla f(1,1) = (4,1)$. Per ottenere il massimo aumento di prodotto dobbiamo perciò aumentare i fattori in modo che a ogni unità del secondo fattore corrispondano quattro unità del primo.

Esercizi

ESERCIZIO 4.1 Individuare il dominio naturale (campo di esistenza) delle seguenti funzioni e calcolarne le derivate parziali:
1. $f(x,y) = x^2y + 1$ **2.** $f(x,y) = x^{4/5}y^{1/5}$ **3.** $f(x,y) = 2x^3y + 5xy^2 - 3$
4. $f(x,y) = x + 2ey - e^x - e^{2y}$ **5.** $f(x,y) = x^\alpha y^\beta$, $\alpha, \beta \in \mathbb{R}$
6. $f(x,y) = \ln(3x^2 + 4y)$ **7.** $f(x,y) = \ln(x + \ln y)$ **8.** $f(x,y) = \ln(1 + x/y)$
9. $f(x,y) = xy\ln(x+y)$ **10.** $f(x,y) = \ln[(x+2y)/(x-3y)]$
11. $f(x,y) = \sqrt{xy}$ **12.** $f(x,y) = \sqrt{x^2y + 2x}$ **13.** $f(x,y) = \sqrt{x + \sqrt{y}}$
14. $f(x,y,z) = (x+y+z)^{-2}$ **15.** $f(x,y,z) = (x+z)/(x-y)$
16. $f(x,y,z) = z/\sqrt{x^2+y^2+z^2}$ **17.** $f(x,y,z) = (x^3+y^3)/(x^2+z^2)$
18. $f(x,y,z) = x/y + y/z + y/x$ **19.** $f(x,y,z) = ze^{-x/y}$
20. $f(x,y,z) = e^{xyz(x^2+y^2)}$ **21.** $f(x,y,z) = e^{(x^3+y^2+z)^2}$
22. $f(x,y,z) = e^{3x^2+2y^2-xyz}$ **23.** $f(x,y,z) = \ln(x+10y) - ye^z$

ESERCIZIO 4.2 Usare il differenziale per approssimare le seguenti funzioni nel punto indicato:
1. $f(x,y) = x^3 + x(y^2)$ su $(x,y) = (9,11)$
2. $f(x,y) = \sqrt{x^{2/3} + y^{1/3}}$ su $(x,y) = (6/5, 26)$

ESERCIZIO 4.3 Un'impresa produce con l'ausilio di una tecnologia di tipo Cobb-Douglas, $f(x,y) = 10x^{1/2}y^{1/3}$, e attualmente sta utilizzando la combinazione di fattori $(x_0, y_0) = (16, 27)$.

1. Quanto sta producendo?
2. Approssimare la quantità prodotta se $x = 16.1$ e $y = 26.7$.
3. Ripetere il punto precedente nel caso in cui $\mathrm{dx} = \mathrm{dy} = 0.2$.

ESERCIZIO 4.4 Calcolare le matrici Hessiane delle seguenti funzioni verificando che vale il Teorema di Schwarz, cioè che ciascuna matrice è simmetrica:
1. $f(x,y) = xy^2 + 1$ **2.** $f(x,y) = x^2y - 3y$ **3.** $f(x,y) = 2x^3y + 5xy^2 - 3$
4. $f(x,y) = x + 2ey - e^x - e^{2y}$ **5.** $f(x,y) = (x - y^2)\, e^{-x}$ **6.** $f(x,y) = x \ln y - y\sqrt{x}$
7. $f(x,y) = 2\ln(y/\sqrt{x})$ **8.** $f(x,y,z) = xyz$ **9.** $f(x,y,z) = x/y + y/z + y/x$
10. $f(x,y,z) = e^{xyz}$ **11.** $f(x,y,z) = \ln x + \ln y - e^{-yz}$
12. $f(x,y,z) = \ln(x + 10y) - ye^z$

ESERCIZIO 4.5 Scrivere le matrici che definiscono le seguenti forme quadratiche:
1. $Q(x_1,x_2) = x_1^2 + 2x_1x_2 + x_2^2$ **2.** $Q(x,y) = -3x^2 + 6xy - 2y^2$
3. $Q(x_1,x_2,x_3) = 3x_1^2 - 2x_2^2 + x_3^2 - 4x_1x_3 + 8x_2x_3$
4. $Q(x,y,z) = 3x^2 + xy + y^2 - xz - yz + 5z^2$

ESERCIZIO 4.6 Scrivere in forma esplicita le forme quadratiche definite dalle seguenti matrici:
1. $\begin{bmatrix} 0 & 1/4 \\ 1/4 & -1 \end{bmatrix}$ **2.** $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ **3.** $\begin{bmatrix} -1 & 0 & 1/3 \\ 0 & 0 & -1/2 \\ 1/3 & -1/2 & 3 \end{bmatrix}$ **4.** $\begin{bmatrix} -1 & 0 & 3 \\ -2 & 2 & 1 \\ 0 & 1 & -3 \end{bmatrix}$

ESERCIZIO 4.7 Scrivere i polinomi di Taylor del primo e secondo grado, $T_1(\mathrm{dx},\mathrm{dy})$ e $T_2(\mathrm{dx},\mathrm{dy})$, delle seguenti funzioni nei punti di riferimento (x_0,y_0) indicati, calcolare i valori di tali approssimazioni per $(\mathrm{dx},\mathrm{dy}) = (0.1,-0.1)$ e $(\mathrm{dx},\mathrm{dy}) = (0.01, 0.01)$ e determinare gli errori commessi da T_1 e T_2 rispetto ai valori veri di f calcolati con una calcolatrice:
1. $f(x,y) = e^x\sqrt{1+y}$ in $(x_0,y_0) = (0,0)$
2. $f(x,y) = (\ln y)\,/\,(1+x)$ in $(x_0,y_0) = (0,1)$

ESERCIZIO 4.8 Calcolare la derivata delle seguenti funzioni composte:
1. $h(t) = f[x(t), y(t)]$ con $f(x,y) = e^{x+2y}$, $x(t) = t^2 - 1$ e $y(t) = \ln(t+1)$
2. $g(t) = \sqrt{f(\ln t, e^t)}$ con $f(x,y)$ differenziabile

ESERCIZIO 4.9 Calcolare le derivate direzionali delle seguenti funzioni:
1. $f(x,y) = \ln(1 + x^2 + y^2)$ lungo la direzione $v = (1/\sqrt{2}, 1/\sqrt{2})$ in $(0,0)$ e in $(1,1)$
2. $f(x,y) = x^2y + xy^2$ lungo la direzione $v = (1/\sqrt{10}, 3/\sqrt{10})$ in $(4,-2)$

ESERCIZIO 4.10 Determinare le direzioni di massima pendenza per le seguenti funzioni nei punti indicati, esprimendo i risultati mediante un versore (vettore di lunghezza unitaria):
1. $f(x,y) = 3y^3 \ln x$ in $(x_0,y_0) = (1,2)$
2. $f(x,y) = \sqrt{2x - e^x + y^{2/3}}$ su $(x_0,y_0) = (0,8)$

Cosa bisogna ricordare di questo capitolo

1. È necessario acquisire familiarità con le principali nozioni di topologia relative a $\mathbb{R}^n$.

2. Occorre comprendere il concetto di funzione di n variabili e di grafico di tale funzione, con particolare riferimento al caso di $n = 2$.
3. È importante familiarizzare con la nozione di derivate parziali e di differenziale per funzioni di più variabili, e capire l'idea di approssimazione lineare attraverso il differenziale.
4. Si deve ricordare la definizione di gradiente, derivate seconde e matrice Hessiana.
5. Occorre comprendere il concetto di forma quadratica e la sua rappresentazione in forma matriciale.
6. Si deve capire l'idea di approssimazione non lineare attraverso il polinomio di Taylor arrestato al secondo ordine.
7. Occorre imparare la regola della catena per la derivazione di funzioni composte e la nozione di derivata direzionale.