

Capitolo 16

Programmazione non lineare

Il termine ‘programmazione non lineare’ farebbe pensare che rientri in questo ambito tutto ciò che non è lineare. In realtà il termine va inteso in un senso più ristretto, quando cioè la funzione obiettivo e le funzioni che definiscono i vincoli sono sufficientemente lisce da permettere l’uso di tecniche differenziali ed i metodi di risoluzione fanno esclusivamente uso di queste tecniche. Quindi nella programmazione non lineare si cercano solo minimi locali.

Si usa distinguere la programmazione non lineare a seconda se siano presenti o meno i vincoli, che a loro volta possono essere di eguaglianza o di disequaglianza. Anche se è normale che nei casi reali siano presenti dei vincoli, tuttavia il caso non vincolato riveste una speciale importanza, perché i metodi che trattano i casi vincolati vengono spesso riformulati come problemi svincolati.

La programmazione non lineare è una disciplina molto complessa in cui non sempre i principi e i metodi generali si traducono immediatamente in un algoritmo efficace a fronte di una determinata istanza reale. È normalmente richiesto un gran lavoro di sintonia fine dei metodi per adattarli con successo alle singole istanze. Siccome inoltre i metodi risolutivi impiegati sono di natura diversa da quelli per i problemi lineari o discreti, la programmazione non lineare costituisce normalmente una branca separata dell’ottimizzazione.

Tuttavia, in questi ultimi anni, si è assistito ad un sorprendente riavvicinamento fra l’ottimizzazione lineare e quella non lineare dovuto allo sviluppo degli algoritmi ai punti interni, che verranno esposti nel prossimo capitolo. Si è visto infatti come la programmazione lineare possa anche essere efficacemente formulata e risolta come un problema di programmazione non lineare. Non solo, si è anche notato come delle tecniche di non linearizzazione di un problema combinatorio (vedi sezione 13.5) possano produrre dei significativi vantaggi computazionali. A questo riguardo, il recente impetuoso sviluppo della disciplina nota come ‘programmazione semidefinita’, tipicamente ‘non lineare’ nei metodi di risoluzione, ma ‘lineare’ nella formulazione, ha fornito ulteriori strumenti alla risoluzione di difficili problemi combinatori.

La programmazione non lineare ‘classica’ richiederebbe, per essere assimilata in modo da essere utilizzabile con efficacia, un approfondimento tale da non poter trovare spazio in questo testo. Ci limitiamo pertanto a fornire il quadro generale della disciplina con una maggiore attenzione al caso svincolato rimandando il lettore alla letteratura specializzata per ulteriori approfondimenti. Citiamo ad esempio Ortega e Rheinboldt [1970], Avriel [1976], Gill et al. [1981], Dennis e Schnabel [1983], McCormick [1983], Luenberger [1984], Fletcher [1987], Lucidi [1994].

16.1. Introduzione

Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ almeno due volte differenziabile con gradiente $Df(x)$ ed Hessiano $H(x)$. I minimi locali di una funzione $f(x)$ differenziabile sono caratterizzati dalle seguenti note condizioni:

16.1 TEOREMA. *Se x è minimo locale allora $Df(x) = 0$ e $H(x)$ è positivo semidefinito.* ■

16.2 TEOREMA. *Se $Df(x) = 0$ e $H(x)$ è positivo definito allora x è minimo locale.* ■

La minimizzazione di una funzione viene effettuata nella maggioranza dei casi tramite un metodo iterativo, cioè viene generata una successione x^0, x^1, \dots che converge all'ottimo (locale) x^* . La successione viene troncata quando si è raggiunta una soluzione che dista dall'ottimo meno di un errore prefissato ε .

L'efficienza di un metodo iterativo viene valutata in base alle seguenti definizioni:

16.3 DEFINIZIONE. *Sia $x^k \rightarrow x^*$. Se per una certa norma $\|\cdot\|$ esistono K e $\alpha \in [0, 1)$ tali che*

$$\|x^{k+1} - x^*\| \leq \alpha \|x^k - x^*\| \quad \forall k \geq K$$

allora si dice che la successione converge linearmente nella norma $\|\cdot\|$ con tasso di convergenza α . Se esiste una successione $\alpha_k \rightarrow 0$ tale che

$$\|x^{k+1} - x^*\| \leq \alpha_k \|x^k - x^*\| \quad \forall k$$

allora si dice che la successione converge superlinearmente. Se esistono K , $\alpha > 0$ e $p > 0$ tali che

$$\|x^{k+1} - x^*\| \leq \alpha \|x^k - x^*\|^p \quad \forall k \geq K$$

allora si dice che la successione converge con ordine almeno p . In particolare se $p = 2$ si dice che la successione converge in modo quadratico. ■

I casi di maggior interesse pratico sono quelli in cui la convergenza è lineare, superlineare, oppure quadratica. Le convergenze superlineare e quadratica possono essere considerate 'veloci'. Si noti ad esempio che con la convergenza quadratica il numero di cifre significative di x^k uguali a quelle di x^* raddoppia ad ogni iterazione. Invece la convergenza lineare può non essere soddisfacente se il tasso di convergenza è prossimo ad uno. Si noti ancora che la convergenza lineare dipende dal tipo di norma. Il numero di iterazioni richiesto nel caso di convergenza lineare è $O(\log(M/\varepsilon)/\log \alpha^{-1})$ dove M è una limitazione superiore di $\|x^0 - x^*\|$ e $\varepsilon > 0$ è l'errore di approssimazione prefissato (mentre il numero di iterazioni è approssimativamente $O(\log \log \varepsilon^{-1})$ per la convergenza quadratica). Si noti che, non essendo noto x^* , non ci si può basare su $\|x^k - x^*\| < \varepsilon$ per la terminazione del metodo. Se la convergenza è lineare allora è facile vedere che $\|x^{k+1} - x^k\| \leq (1-\alpha)\varepsilon$ implica $\|x^k - x^*\| < \varepsilon$.

Tipicamente i metodi iterativi si basano sul seguente schema ricorsivo:

$$x' := x + \alpha h \tag{16.1}$$

dove h è un'opportuna direzione e α è la lunghezza del passo lungo tale direzione, per cui ogni iterazione consiste di due fasi, il calcolo della direzione e successivamente il calcolo della lunghezza del passo.

Nel calcolo di h e di α si può far uso di un'approssimazione lineare di f o di un'approssimazione quadratica. Nel primo caso si usa il gradiente $Df(x)$ mentre nel secondo si usa anche l'Hessiano $H(x)$. In generale si può dire che l'uso dell'Hessiano migliora la velocità di convergenza, ma rende anche computazionalmente più oneroso il calcolo di h e α . Essenziale è in ogni caso l'informazione fornita dal gradiente.

16.4 DEFINIZIONE. Una direzione h si dice di discesa in x se $Df(x)h < 0$. ■

Nella maggior parte dei metodi si fa uso di direzioni di discesa. Ovviamente l'opposto del gradiente $-Df(x)$ costituisce una naturale scelta come direzione di discesa, ma non sempre, come vedremo, si tratta di una direzione di discesa vantaggiosa.

Un aspetto cruciale di un metodo iterativo è costituito dalla garanzia che ci sia effettivamente convergenza verso il minimo. Diremo che in questo caso il metodo possiede *convergenza globale*. Il semplice fatto che una direzione sia di discesa non è sufficiente. Bisogna anche scegliere il passo d'iterazione in modo opportuno.

16.2. Scelta del passo d'iterazione

In questa sezione si mostra come calcolare il passo d'iterazione una volta fissata la direzione h , che supponiamo di discesa. Valutiamo dapprima sotto quali condizioni la scelta del passo garantisce la convergenza globale. Consideriamo le seguenti condizioni:

$$f(x) - f(x + \alpha h) > -\beta Df(x)(\alpha h) \quad (16.2)$$

$$-\gamma Df(x)(\alpha h) \geq -Df(x + \alpha h)(\alpha h) \quad (16.3)$$

$$-Df(x)h > \delta \|Df(x)\| \|h\| \quad (16.4)$$

dove $0 < \beta < \gamma < 1$ e $\delta > 0$. Si noti che tutti i membri delle disequazioni sono positivi (h è di discesa). La condizione (16.2) impone che il decremento della funzione sia almeno una frazione fissata del decremento della derivata direzionale. Si noti che (16.2) è sempre soddisfatta per valori piccoli di α . La condizione (16.3) invece vincola il passo ad essere abbastanza lungo da percepire un apprezzabile diminuzione nella derivata direzionale, fatto questo che indica la maggior vicinanza al minimo. Si noti che per valori piccoli di α (16.3) non può essere soddisfatta. In particolare non può essere soddisfatta se $Df(x) = Df(x + \alpha h)$. Infine la condizione (16.4) impone alla direzione di discesa qualcosa di più: si vuole che l'antigradiente e la direzione di discesa non siano 'quasi' ortogonali. Valgono allora i seguenti teoremi che riportiamo senza dimostrazione (per una dimostrazione si veda ad esempio Dennis e Schnabel [1983]):

16.5 TEOREMA. Sia f limitata verso il basso. Sia h una direzione di discesa in x . Allora se $0 < \beta < \gamma < 1$, esistono $0 < \alpha_1 < \alpha_2$ tali che per ogni $\alpha \in [\alpha_1, \alpha_2]$ le condizioni (16.2) e (16.3) sono soddisfatte. ■

16.6 TEOREMA. Sia f limitata verso il basso e sia Df continuo secondo Lipschitz. Sia $x^{k+1} := x^k + \alpha_k h^k$ con x^0 arbitrario, e h^k, α_k soddisfino (16.2), (16.3) e (16.4) per ogni k . Allora o $Df(x^k) = 0$ per qualche k oppure $Df(x^k) \rightarrow 0$. ■

È importante notare che nel precedente teorema non si garantisce la convergenza verso il minimo ma solo verso punti che soddisfano la condizione necessaria di primo ordine e non è difficile costruire esempi in cui l'iterazione converge verso punti non di minimo ma con gradiente nullo (ad esempio sia $f(x) = x_1^2 - x_2^2$, $x^0 = (1, 0)$, $h^k = (-1, 0)$, $\forall k$; si trova $x^k \rightarrow (0, 0)$ che non è minimo). Tuttavia è lecito attendersi che il 'rumore' causato dall'errore di arrotondamento faccia allontanare l'iterazione da punti stazionari (cioè con $Df(x) = 0$) ma non di minimo. Le costanti β e γ possono essere fissate arbitrariamente (purché ovviamente $0 < \beta < \gamma < 1$).

Si tratta in ogni caso di considerare la funzione unidimensionale

$$\varphi(\alpha) : \alpha \mapsto f(x + \alpha h)$$

e di valutarla in modi opportuni per calcolare il passo. Si noti innanzitutto che il valore α^* per il quale $\varphi(\alpha)$ è minima soddisfa le condizioni (16.2) e (16.3). Quindi un primo approccio al calcolo del passo consiste nel calcolare il minimo di $\varphi(\alpha)$. Questo approccio può essere raccomandato in metodi a convergenza lineare con tasso di convergenza prossimo a uno. In metodi a convergenza superlineare o quadratica il vantaggio di avere il minimo di φ non compensa l'onere dell'elevato numero di valutazioni di $f(x)$ che viene richiesto per il calcolo del minimo. In questi casi è sufficiente solo una conoscenza approssimata del minimo.

Presentiamo dapprima un metodo di minimizzazione di φ . Se h è una direzione di discesa e il problema non è illimitato, allora φ è una funzione unimodale su un opportuno intervallo $I = [0, \eta]$. La minimizzazione di una funzione unimodale $\varphi : I \rightarrow \mathbb{R}$ è un problema abbastanza semplice. Sia $[\alpha_1, \alpha_2]$ un intervallo contenente il minimo α^* e siano β_1 e β_2 due punti tali che $\alpha_1 < \beta_1 < \beta_2 < \alpha_2$. L'unimodalità di φ implica che:

$$\begin{aligned} \varphi(\beta_1) \leq \varphi(\beta_2) &\implies \alpha^* \in [\alpha_1, \beta_2] \\ \varphi(\beta_1) \geq \varphi(\beta_2) &\implies \alpha^* \in [\beta_1, \alpha_2] \end{aligned}$$

Quindi si è ridotto l'intervallo in cui si sa essere contenuto il minimo. Procedendo in modo iterativo si continua finché l'intervallo d'incertezza è minore di una tolleranza prefissata.

La riduzione dell'intervallo è tanto più grande quanto più β_1 e β_2 si avvicinano al punto medio dell'intervallo. Si noti che, se φ è differenziabile, questo corrisponde a valutare la derivata nel punto intermedio e le iterazioni sono equivalenti al metodo di bisezione applicato alla derivata.

Tuttavia una migliore efficienza del metodo si può raggiungere cercando piuttosto di ridurre il numero di valutazioni di φ . A questo fine si noti che, una volta determinato il sottointervallo (ad esempio sia $[\alpha_1, \beta_2]$) al suo interno c'è un punto (β_1 nel caso specifico, si veda anche la figura 16.1) in cui φ è già stata calcolata, per cui risulta conveniente alla successiva iterazione aggiornare i valori nel seguente modo:

$$\alpha'_1 := \alpha_1 \quad \alpha'_2 := \beta_2 \quad \beta'_2 := \beta_1 \quad \text{se } \varphi(\beta_1) \leq \varphi(\beta_2)$$

oppure

$$\alpha'_1 := \beta_1 \quad \alpha'_2 := \alpha_2 \quad \beta'_1 := \beta_2 \quad \text{se } \varphi(\beta_1) \geq \varphi(\beta_2).$$

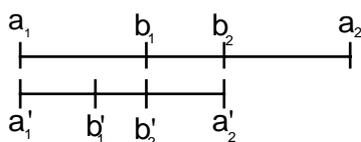


FIGURA 16.1

Nella determinazione di β_1 e β_2 , si può, senza perdita di generalità, imporre la condizione che i due valori siano simmetrici rispetto al valore intermedio dell'intervallo. Imponendo anche la condizione che il nuovo intervallo sia simile al precedente per ciò che riguarda la disposizione dei punti interni β_1 e β_2 si ottiene la seguente relazione:

$$\frac{\alpha_2 - \alpha_1}{\alpha'_2 - \alpha'_1} = \frac{\beta_2 - \alpha_1}{\beta'_2 - \alpha'_1} \implies \frac{\alpha_2 - \alpha_1}{\beta_2 - \alpha_1} = \frac{\beta_2 - \alpha_1}{\beta_1 - \alpha_1} = \frac{\beta_2 - \alpha_1}{\alpha_2 - \beta_2} = \frac{\beta_2 - \alpha_1}{(\alpha_2 - \alpha_1) - (\beta_2 - \alpha_1)}$$

da cui si ottiene

$$\frac{\beta_2 - \alpha_1}{\alpha_2 - \alpha_1} = \frac{\sqrt{5} - 1}{2}$$

cioè il rapporto aureo. Un modo alternativo di procedere consiste nel notare che

$$\alpha_2 - \alpha_1 = (\beta_2 - \alpha_1) + (\alpha_2 - \beta_2)$$

e quindi, indicando con $(\cdot)'$ e $(\cdot)''$ le grandezze ottenute con i valori aggiornati nelle due successive iterazioni rispettivamente, si ha

$$\begin{aligned} \alpha_2 - \alpha_1 &= (\beta_2 - \alpha_1) + (\alpha_2 - \beta_2) = (\alpha_2 - \alpha_1)' + (\beta_1 - \alpha_1) = \\ &= (\alpha_2 - \alpha_1)' + (\beta_2 - \alpha_1)' = (\alpha_2 - \alpha_1)' + (\alpha_2 - \alpha_1)'' \end{aligned}$$

Si ottiene così la classica relazione ricorsiva che genera i numeri di Fibonacci. Adottando questo metodo bisogna calcolare il primo numero di Fibonacci F_k tale che $F_k \geq 2(\alpha_2 - \alpha_1)/\varepsilon$, dove $(\alpha_2 - \alpha_1)$ è l'ampiezza dell'intervallo iniziale e ε è la tolleranza prefissata. Dopodiché si pone

$$\beta_2 := \alpha_1 + (F_{k-1}/F_k)(\alpha_2 - \alpha_1) \quad \text{e} \quad \beta_1 := \alpha_1 + (F_{k-2}/F_k)(\alpha_2 - \alpha_1)$$

e si procede ricorsivamente. I due metodi delineati sono asintoticamente equivalenti. Infatti

$$F_k = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^k - \frac{1}{\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2} \right)^k \implies \lim_{k \rightarrow \infty} \frac{F_k}{F_{k+1}} = \frac{\sqrt{5} - 1}{2}.$$

In altre parole i due metodi tendono a diventare equivalenti al tendere a zero della tolleranza.

16.7 ESEMPIO. Sia $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definita da

$$f(x) := 20(2x_1^2 - x_2)^2 + x_1^2$$

Sia $x = (-1, 1)$ e $h = (0.970843, 0.239714)$. Allora si ha $\varphi(\alpha) = 21 - 166.865\alpha + 416.343\alpha^2 - 310.893\alpha^3 + 71.0701\alpha^4$ il cui grafico è disegnato in figura 16.2. Applicando la ricerca unidimensionale con il rapporto aureo ed errore 0.0001 si ottiene in 19 passi a partire dall'intervallo $[0, 0.5]$ il sottointervallo $[0.28154, 0.281629]$. In figura 16.3 sono mostrati i primi 8 passi con l'indicazione dei successivi intervalli generati. ■

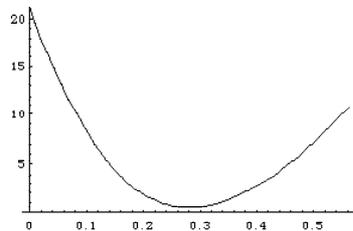


FIGURA 16.2

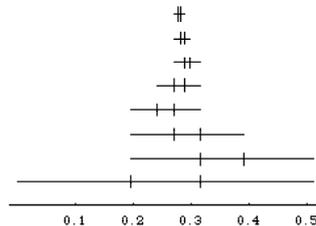


FIGURA 16.3

Un metodo più rapido nel calcolo del passo consiste semplicemente nel trovare un valore di α che soddisfa (16.2) e (16.3). Ad esempio si segue questa strategia nei metodi di Newton e Quasi-Newton. Sia α_0 un valore iniziale. Se α_0 soddisfa le due condizioni allora $\alpha := \alpha_0$. Supponiamo che (16.2) non sia soddisfatta. Allora bisogna cercare un valore inferiore ad α_0 . Questo viene fatto con un'approssimazione quadratica di φ che interpola $\varphi(0)$, $\varphi(\alpha_0)$ e $\varphi'(0) = Df(x)h$ e prendendo il valore $\hat{\alpha}$ che minimizza l'interpolazione. Si ottiene facilmente:

$$\hat{\alpha} = \frac{-\alpha_0^2 \varphi'(0)}{2(\varphi(\alpha_0) - \varphi(0) - \alpha_0 \varphi'(0))}$$

Si pone $\alpha_0 := \max\{\hat{\alpha}, c_1 \alpha_0\}$ (con $c_1 = 0.1$ tipicamente). Se α_0 non soddisfa (16.2) si prosegue ricorsivamente. Se invece α_0 non soddisfa (16.3) (si noti che per il teorema 16.5 non esiste α che non soddisfa né (16.2) né (16.3)) allora bisogna cercare un valore superiore. Questo viene fatto ponendo semplicemente $\alpha_0 := 2\alpha_0$.

Se nell'una o nell'altra procedura non si trova nessun valore che soddisfa sia (16.2) che (16.3), questo avviene perché si sono generati dei valori uno a sinistra e uno a destra dell'intervallo $[\alpha_1, \alpha_2]$ definito dal teorema 16.5. Siano α^- e α^+ questi due valori. Si calcola l'approssimazione quadratica di φ che interpola $\varphi(\alpha^-)$, $\varphi(\alpha^+)$ e $\varphi'(\alpha^-) = Df(x + \alpha^- h)h$ e si prende il valore $\hat{\alpha}$ che minimizza l'interpolazione che è dato da:

$$\hat{\alpha} = \frac{(\alpha^+ - \alpha^-)^2 \varphi'(\alpha^-)}{2(\varphi(\alpha^+) - \varphi(\alpha^-) - (\alpha^+ - \alpha^-)\varphi'(\alpha^-))}$$

poi si pone $\alpha_0 := \min\{\max\{\hat{\alpha}, \alpha^- + c_2(\alpha^+ - \alpha^-)\}, \alpha^+ - c_2(\alpha^+ - \alpha^-)\}$ dove $c_2 = 0.2$ tipicamente (questo corrisponde a prendere $\alpha_0 = \hat{\alpha}$ se $\hat{\alpha}$ sta all'interno di un intervallo un po' più ristretto dell'intervallo $[\alpha^-, \alpha^+]$, altrimenti α_0 è uguale all'estremo corrispondente del nuovo intervallo). Anche questa fase può essere ripetuta più di una volta se necessario. Per la verifica di (16.2) e (16.3) si usano i valori $\beta = 10^{-4}$ (questo significa che non è evidentemente così vincolante la condizione (16.2)) e $\gamma \in [0.7, 0.9]$.

16.3. Metodo del gradiente

Se f è differenziabile una direzione di discesa h da usare in (16.1) è immediatamente disponibile ed è l'opposto del gradiente, cioè $h = -\alpha Df(x)$, con $\alpha > 0$ scalare arbitrario.

Per valutare la complessità computazionale di questo tipo di iterazioni è opportuno riferirsi ad una funzione definita da una forma quadratica positiva definita, cioè $f(x) = xQx > 0$, $\forall x \neq 0$. Infatti la complessità computazionale è legata in questo caso alla velocità di convergenza che dipende essenzialmente dal comportamento asintotico delle iterazioni e quindi è lecito approssimare, nel punto di minimo locale, la funzione data con una forma quadratica definita dall'Hessiano calcolato nel punto di minimo. Quindi $h := -Df(x)/2 = -Qx$ e

$$\alpha := \arg \min (x + \alpha h)Q(x + \alpha h)$$

Da

$$\frac{d(\alpha^2 hQh + 2\alpha hQx + xQx)}{d\alpha} = 2(\alpha hQh + hQx) = 2(\alpha hQh - hh) = 0$$

ricaviamo

$$\alpha = \frac{hh}{hQh} \quad x' := x + \frac{hh}{hQh}h$$

e quindi (tenendo presente che $x^* = 0$ è il minimo di $f(x)$)

$$\frac{\|x' - x^*\|}{\|x - x^*\|} = \frac{\|x + \frac{h}{h} Q h\|}{\|x\|} \quad (16.5)$$

La scelta della norma in (16.5) è arbitraria. Possiamo ad esempio scegliere $\|x\|^2 = x Q x$ così il rapporto espresso in (16.5) corrisponde a $\sqrt{f(x')/f(x)}$ e misura anche il decremento della funzione obiettivo. Quindi:

$$\frac{\|x' - x^*\|^2}{\|x - x^*\|^2} = \frac{(x + \frac{h}{h} Q h) Q (x + \frac{h}{h} Q h)}{x Q x} =$$

(sostituendo $h = -Q x$ e $x Q x = x Q Q^{-1} Q x = h Q^{-1} h$)

$$= \frac{x Q x - \frac{(h h)^2}{h Q h}}{x Q x} = 1 - \frac{h h}{h Q h} \frac{h h}{h Q^{-1} h}$$

Vogliamo ora ottenere una limitazione inferiore di

$$\frac{h h}{h Q h} \frac{h h}{h Q^{-1} h}$$

Possiamo notare che, considerando un cambiamento di base ortogonale sugli autovettori di Q , vale:

$$\min_h \frac{h h}{h Q h} \frac{h h}{h Q^{-1} h} = \min_h \frac{\sum_i \bar{h}_i^2}{\sum_i \bar{h}_i^2 \lambda_i} \frac{\sum_i \bar{h}_i^2}{\sum_i \bar{h}_i^2 / \lambda_i}$$

e, definendo

$$\xi_i := \frac{\bar{h}_i^2}{\sum_i \bar{h}_i^2}$$

otteniamo dei coefficienti di una combinazione convessa sugli autovalori di Q e sui reciproci degli stessi. Si ha:

$$\min_h \frac{h h}{h Q h} \frac{h h}{h Q^{-1} h} = \min_{\xi} \frac{1}{\sum_i \xi_i \lambda_i} \frac{1}{\sum_i \xi_i / \lambda_i}$$

In \mathbb{R}^2 si considerino i punti $(\lambda_i, 1/\lambda_i)$, $i = 1, \dots, n$. Assegnando tutti i possibili valori a ξ otteniamo l'involuppo convesso di questi punti. Il punto dell'involuppo convesso a cui corrisponde il massimo prodotto delle due coordinate del punto stesso sta sulla congiungente $(\lambda_1, 1/\lambda_1)$ con $(\lambda_n, 1/\lambda_n)$ dove λ_1 e λ_n sono rispettivamente il più grande e il più piccolo autovalore di Q . Pertanto basta considerare combinazioni convesse di $(\lambda_1, 1/\lambda_1)$ e $(\lambda_n, 1/\lambda_n)$. Si trova quindi che il valore $\xi = 1/2$ rende massima l'espressione

$$\left(\lambda_1 \xi + \lambda_n (1 - \xi) \right) \left(\frac{1}{\lambda_1} \xi + \frac{1}{\lambda_n} (1 - \xi) \right)$$

con valore

$$\frac{(\lambda_1 + \lambda_n)^2}{4 \lambda_1 \lambda_n}$$

Pertanto si ha

$$\frac{\|x' - x^*\|^2}{\|x - x^*\|^2} \leq 1 - \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2} \implies \frac{\|x' - x^*\|}{\|x - x^*\|} \leq \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \quad (16.6)$$

Si vede quindi che la convergenza è lineare con un basso tasso di convergenza se λ_1 e λ_n non sono molti discosti. Viceversa, per bassi valori di λ_n/λ_1 (matrice Q mal condizionata) non c'è garanzia di convergenza rapida, anzi l'esperienza dimostra che l'iterazione converge con la stessa lentezza di $(\lambda_1 - \lambda_n)/(\lambda_1 + \lambda_n)$.

È importante notare che il metodo del gradiente non si applica solo a funzioni convesse e quadratiche, ma anche a qualsiasi funzione differenziabile. L'analisi della velocità di convergenza fatta precedentemente rimane valida e si riferisce in questo caso all'Hessiano calcolato nel punto di minimo locale.

16.8 ESEMPIO. Si consideri la stessa funzione dell'esempio 16.7, cioè:

$$f(x) := 20(2x_1^2 - x_2)^2 + x_1^2$$

Il minimo è l'origine. Il grafico della funzione è rappresentato in figura 16.4 e le sue linee di livello in figura 16.5. Si tratta di una funzione 'di media difficoltà' per il metodo del gradiente. Il gradiente e l'Hessiano valgono:

$$Df(x) = (2x_1 + 320x_1^3 - 160x_1x_2, -80x_1^2 + 40x_2)$$

$$H(x) = \begin{pmatrix} 2 + 960x_1^2 - 160x_2 & -160x_1 \\ -160x_1 & 40 \end{pmatrix}$$

Nell'origine l'Hessiano è abbastanza mal condizionato essendo $\lambda_n/\lambda_1 = 20$ e la limitazione in (16.6) vale 0.9047. Quindi è lecito aspettarsi una convergenza lenta. Inoltre l'Hessiano è quasi ugualmente mal condizionato sui punti della parabola $x_2 = 2x_1^2$. In questi punti la funzione vale semplicemente x_1^2 . Al di fuori la funzione assume valori immediatamente più elevati a causa del coefficiente più grande che moltiplica il termine $(2x_1^2 - x_2)^2$. Il grafico della funzione assomiglia quindi ad una valle le cui pareti si elevano secondo il termine $10(2x_1^2 - x_2)^2$ e il cui fondo scende più dolcemente secondo il termine x_1^2 . Il metodo del gradiente tende a raggiungere il minimo seguendo il fondovalle, tuttavia il mal condizionamento dell'Hessiano fa sì che l'iterazione proceda a zig-zag.

Si vedano nelle figure 16.4 e 16.5 le prime 47 iterazioni ottenute a partire dal punto $x^0 = (-1, 1)$. Il decremento della funzione obiettivo è rappresentato nelle figure 16.6 (prime 10 iterazioni) e 16.7 (successive). Al 47-mo passo d'iterazione la funzione vale 0.000332912 che è un valore apparentemente 'buono' in quanto poco lontano dal valore ottimo. Tuttavia $x^{47} = (-0.00575427, 0.00156705)$ che è invece più lontano dall'ottimo. Il tasso di convergenza valutato fra x^{46} e x^{47} vale 0.819001. Proseguendo fino alla 70-ma iterazione l'errore sulla funzione è inferiore a 10^{-7} e $x^{70} = (-0.000420061, 0.0000467236)$ che rappresenta un valore accettabile. ■

16.9 ESEMPIO. Si consideri una funzione molto simile a quella dell'esempio precedente, cioè:

$$f(x) := 20(2x_1^2 - x_2)^2 + x_2^2$$

Il minimo è sempre l'origine. Il grafico della funzione è simile a quello dell'esempio precedente. Il gradiente e l'Hessiano valgono:

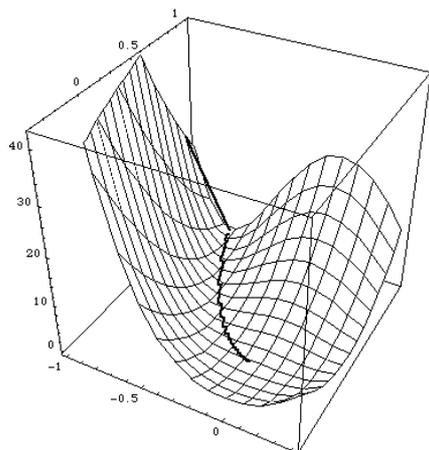


FIGURA 16.4

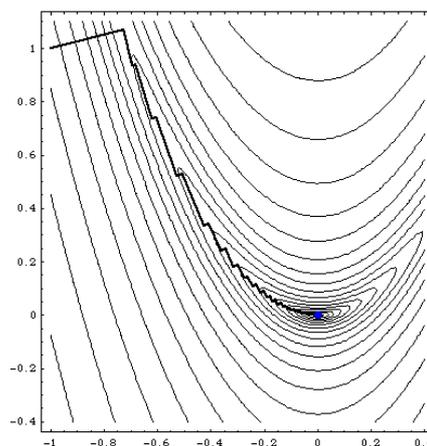


FIGURA 16.5

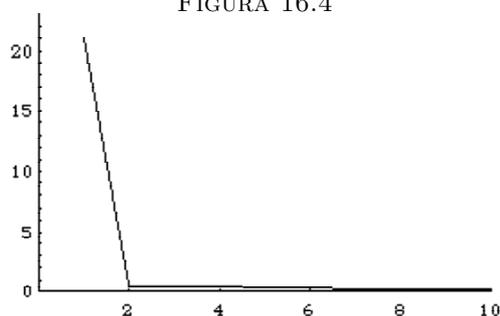


FIGURA 16.6

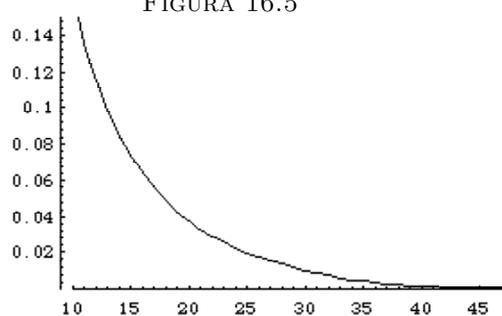


FIGURA 16.7

$$Df(x) = (320x_1^3 - 160x_1x_2, -80x_1^2 + 42x_2), \quad H(x) = \begin{pmatrix} 960x_1^2 - 160x_2 & -160x_1 \\ -160x_1 & 42 \end{pmatrix}$$

L'Hessiano è addirittura singolare nell'origine! Quindi ci si deve aspettare una convergenza sempre più rallentata al tendere verso il minimo. Infatti si ottiene $x^{47} = (-0.1397, 0.0349)$. L'errore sulla funzione è inferiore a 10^{-6} solo alla 227-esima iterazione, però si ha anche $x^{227} = (-0.0215977, 0.000891902)$ e, come si vede, l'errore sulla prima componente è ancora troppo elevato. ■

16.10 ESEMPIO. Si suponga di voler risolvere un sistema lineare di equazioni con il metodo dei minimi quadrati. Sia $Ax = b$ e si voglia minimizzare $f(x) = (Ax - b)^T (Ax - b)$. Si ha $Df(x) = 2A^T(Ax - b)$ e $H(x) = 2A^T A$. Nella seguente esemplificazione numerica la matrice A ha $n = 8$ colonne e $m = 20$ righe ed è stata generata casualmente con componenti intere uniformemente distribuite in $[-10, 10]$. Sia $\bar{x} = \mathbf{1}$ e ε un vettore con componenti reali uniformemente distribuite in $[-2, 2]$. Il vettore b è stato generato come $b := A\mathbf{1} + \varepsilon$. L'ottimo (calcolato risolvendo il sistema lineare $(A^T A)x = A^T b$) è

$$x^* = (1.03607, 0.971916, 1.05297, 0.970883, 0.974246, 1.05493, 0.950354, 0.985431)$$

Partendo da $x^0 = 0$ si ottiene

$$x^{46} = (1.03526, 0.971188, 1.05299, 0.974515, 0.973627, 1.05199, 0.950928, 0.985646)$$

e i valori di f sono rappresentati nelle figure 16.8 e 16.9. Si vedano inoltre in figura 16.10 i valori delle iterate x proiettate sul sottospazio definito dagli autovettori relativi al minimo e massimo autovalore dell'Hessiano calcolato in x^{46} . ■

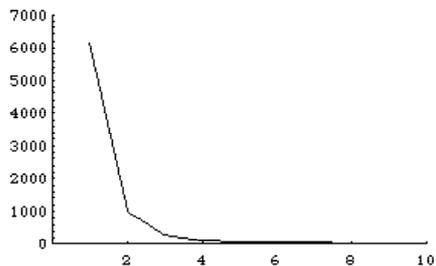


FIGURA 16.8

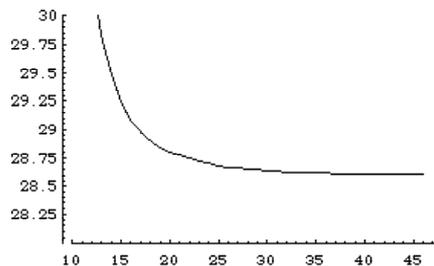


FIGURA 16.9



FIGURA 16.10

16.11 ESEMPIO. Come esempio di convergenza superlineare ottenibile in casi particolari con il metodo del gradiente si supponga di calcolare le radici di un polinomio $p(z)$ a coefficienti reali. Ponendo $z = x + iy$ e $z^* = x - iy$ si ha $|p(z)|^2 = p(z)p(z^*)$. Pertanto $f(x, y) := p(x + iy)p(x - iy)$ è una funzione non negativa a valori e coefficienti reali che si annulla se e soltanto se x e y sono la parte reale e immaginaria di una radice di $p(z)$. Gradiente e Hessiano sono dati da

$$Df(x, y) = (p'(z)p(z^*) + p(z)p'(z^*) \quad i p'(z)p(z^*) - i p(z)p'(z^*))$$

$$H(x, y) = \begin{pmatrix} p''(z)p(z^*) + 2p'(z)p'(z^*) + p(z)p''(z^*) & i p''(z)p(z^*) - i p(z)p''(z^*) \\ i p''(z)p(z^*) - i p(z)p''(z^*) & -p''(z)p(z^*) + 2p'(z)p'(z^*) - p(z)p''(z^*) \end{pmatrix}$$

Si noti che il gradiente si annulla non solo sulle radici di $p(z)$, ma anche su quelle di $p'(z)$ (non ci sono tuttavia altri valori su cui il gradiente si annulla). Sulle radici di $p(z)$ l'Hessiano vale

$$H(x, y) = \begin{pmatrix} 2p'(z)p'(z^*) & 0 \\ 0 & 2p'(z)p'(z^*) \end{pmatrix}$$

e su quelle di $p'(z)$ vale

$$H(x, y) = 2 \begin{pmatrix} \operatorname{Re} p''(z)p(z^*) & \operatorname{Im} p''(z)p(z^*) \\ \operatorname{Im} p''(z)p(z^*) & -\operatorname{Re} p''(z)p(z^*) \end{pmatrix}$$

che non è mai positivo definito (può essere positivo semidefinito, in quanto nullo, solo se $p''(z)p(z^*) = 0$). Quindi gli unici minimi locali sono le radici di $p(z)$. Se la radice non è multipla allora l'Hessiano è perfettamente condizionato e quindi il tasso di convergenza lineare vale zero. Al tendere verso l'ottimo quindi i coefficienti dati da (16.6) tendono a zero e la convergenza è superlineare. Tuttavia una radice multipla è anche radice di $p'(z)$ e l'Hessiano è singolare.

Si consideri ad esempio $p(z) = (z-1)(z-2)(z-1-i)(z-1+i) = z^4 - 5z^3 + 10z^2 - 10z + 4$. Si ottiene

$$\begin{aligned} f(x, y) = & 16 - 80x_1 + 180x_1^2 - 240x_1^3 + 208x_1^4 - 120x_1^5 + 45x_1^6 - 10x_1^7 + x_1^8 + \\ & 20x_2^2 - 80x_1x_2^2 + 152x_1^2x_2^2 - 160x_1^3x_2^2 + 95x_1^4x_2^2 - 30x_1^5x_2^2 + \\ & 4x_1^6x_2^2 + 8x_2^4 - 40x_1x_2^4 + 55x_1^2x_2^4 - 30x_1^3x_2^4 + 6x_1^4x_2^4 + \\ & 5x_2^6 - 10x_1x_2^6 + 4x_1^2x_2^6 + x_2^8 \end{aligned}$$

le cui linee di livello (in scala logaritmica) sono rappresentate in figura 16.11.

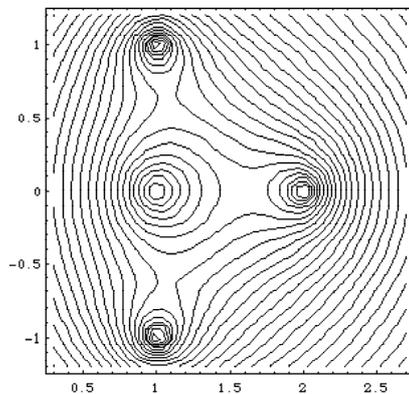


FIGURA 16.11

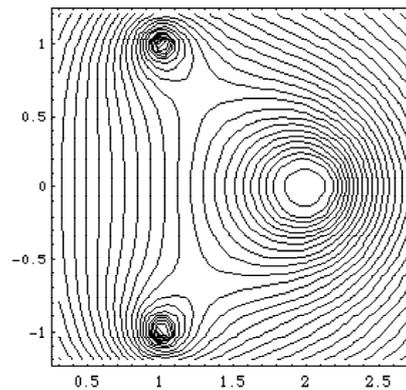


FIGURA 16.12

L'iterazione prodotta dal metodo del gradiente è la seguente

$$x^0 = (0, 1) \quad x^1 = (1.97754, -0.0358536) \quad x^2 = (1.99712, 0.00146851)$$

$$x^3 = (1.99993, 0.000023881) \quad x^4 = (1.99998, 6.72526 \cdot 10^{-6})$$

con una convergenza molto veloce. Sia ora $p(z) = (z-2)^2(z-1-i)(z-1+i) = z^4 - 6z^3 + 14z^2 - 16z + 8$ (linee di livello in figura 16.12). Anche se l'Hessiano è singolare nel punto $(2, 0)$ la convergenza rimane veloce, perché i due autovalori dell'Hessiano tendono insieme a zero. Si ottiene in questo caso

$$x^0 = (0, 1) \quad x^1 = (2.05386, 0.178457) \quad x^2 = (1.9849, 0.00600191)$$

$$x^3 = (1.99996, -0.0000383192) \quad x^4 = (1.99999, -0.0000104068)$$

■

16.4. Metodo di Newton

Il metodo del gradiente corrisponde a minimizzare una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ valutando la sua approssimazione al primo ordine. Se si spinge l'approssimazione fino al secondo ordine si ha il metodo noto come metodo di Newton. Il metodo di Newton è stato originariamente proposto per trovare gli zeri di una funzione $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ valutando l'approssimazione al primo ordine di F . Tuttavia, siccome nel minimo il gradiente $Df(x)$ si annulla, ciò che si fa è essenzialmente la stessa cosa, in quanto si tratta di calcolare gli zeri di $Df : \mathbb{R}^n \rightarrow \mathbb{R}^n$ sfruttando la derivata di Df , cioè l'Hessiano di f . Indicando con \tilde{f} l'approssimazione quadratica di f , abbiamo:

$$\tilde{f}(x+h) = f(x) + Df(x)h + \frac{1}{2}hH(x)h$$

Minimizzando \tilde{f} rispetto ad h si ottiene $Df(x) + H(x)h = 0$ da cui

$$h = -H(x)^{-1}Df(x) \quad (16.7)$$

da usare nello schema iterativo (16.1) con passo d'iterazione $\alpha = 1$. Il valore di x' così ottenuto è l'ottimo se la funzione è quadratica (e quindi \tilde{f} coincide con f). In generale ciò non è vero, per cui bisogna continuare nello schema iterativo. Se x è sufficientemente vicino all'ottimo la direzione (16.7) è di discesa. Infatti in un minimo l'Hessiano è positivo definito e quindi, per continuità, esiste un intorno del minimo su cui tale proprietà viene mantenuta. Siccome una direzione h è di discesa se $Df(x)h < 0$, la positiva definitezza garantisce che la direzione sia di discesa. Le proprietà principali del metodo di Newton sono riassunte nel seguente teorema:

16.12 TEOREMA. *Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ due volte differenziabile in un insieme aperto e convesso $D \subset \mathbb{R}^n$. Sia $x^* \in D$ minimo locale stretto (cioè $H(x^*)$ è positivo definito) ed esista $\rho > 0$ tale per ogni $\{x : \|x - x^*\| \leq \rho\}$ l'Hessiano è di Lipschitz (cioè esiste $\gamma(\rho)$ tale che $\|H(x) - H(y)\| \leq \gamma(\rho)\|x - y\|$ per ogni x, y nell'intorno). Allora esiste un intorno di x^* tale che per ogni x^0 nell'intorno, la successione $\{x^k\}$ generata da*

$$x^{k+1} := x^k - H(x^k)^{-1}Df(x^k) \quad (16.8)$$

converge ad x^* e soddisfa

$$\|x^{k+1} - x^*\| \leq K \|H(x^*)^{-1}\| \|x^k - x^*\|^2$$

DIMOSTRAZIONE.

$$x^{k+1} - x^* = x^k - H(x^k)^{-1}Df(x^k) - x^* + H(x^k)^{-1}Df(x^*) =$$

(siccome $Df(x^*) = 0$ per l'ottimalità)

$$= H(x^k)^{-1}(Df(x^*) - Df(x^k) - H(x^k)(x^* - x^k))$$

e quindi

$$\|x^{k+1} - x^*\| \leq \|H(x^k)^{-1}\| \|Df(x^*) - Df(x^k) - H(x^k)(x^* - x^k)\|$$

Per la continuità di H e l'ipotesi di positiva definitezza, esiste un intorno di x^* di raggio σ ed una costante $\beta(\sigma)$ tale che $\|H(x)^{-1}\| \leq \beta(\sigma)$ per ogni $\{x : \|x - x^*\| \leq \sigma\}$. Inoltre

la continuità di Lipschitz dell'Hessiano implica $\|Df(x^*) - Df(x^k) - H(x^k)(x^* - x^k)\| \leq 1/2 \gamma(\rho) \|x^k - x^*\|^2$ per ogni $\{x : \|x - x^*\| \leq \rho\}$ (si tratta di un classico risultato di analisi). Sia $\mu := \min\{\sigma, \rho\}$. Allora per ogni $\{x : \|x - x^*\| \leq \mu\}$ si ha

$$\|x^{k+1} - x^*\| \leq \frac{1}{2} \beta(\mu) \gamma(\mu) \|x^k - x^*\|^2 \quad (16.9)$$

Sia $\delta := \alpha \min\{\mu; 2(\beta(\mu) \gamma(\mu))^{-1}\}$, con $\alpha < 1$. Allora per ogni $\{x : \|x - x^*\| \leq \delta\}$ si ha

$$\|x^{k+1} - x^*\| \leq \frac{1}{2} \beta(\mu) \gamma(\mu) \|x^k - x^*\| \|x^k - x^*\| \leq \frac{1}{2} \beta(\mu) \gamma(\mu) \delta \|x^k - x^*\| \leq \alpha \|x^k - x^*\| \quad (16.10)$$

Quindi (16.9) prova la velocità di convergenza quadratica e (16.10) prova la convergenza. ■

Il vantaggio della convergenza quadratica del metodo è però bilanciato da due maggiori fattori negativi: al di fuori di un opportuno intorno (peraltro ignoto) del minimo l'iterazione non converge ed è richiesto il calcolo dell'Hessiano ad ogni passo di iterazione, nonché la risoluzione del sistema lineare $H(x^k)h = Df(x^k)$. Sul problema della convergenza si consideri il seguente semplice esempio:

16.13 ESEMPIO. $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x - \ln x$. Il minimo $x^* = 1$ si trova in modo elementare per via analitica. L'iterazione (16.8) diventa

$$x^{k+1} = x^k - (x^k)^2 \left(1 - \frac{1}{x^k}\right) = x^k(2 - x^k)$$

che produce le seguenti iterazioni a partire da due valori vicini:

x^0	2.01	1.99
x^1	-0.0201	0.0199
x^2	-0.040604	0.039404
x^3	-0.0828567	0.0772553
x^4	-0.172579	0.148542
x^5	-0.374941	0.27502
x^6	-0.890462	0.474404
x^7	-2.57385	0.723748
x^8	-11.7724	0.923685
x^9	-162.134	0.994176
x^{10}	-26611.6	0.999966

16.14 ESEMPIO. Si riconsideri la funzione dell'esempio 16.8. Il metodo di Newton produce la seguente iterazione

	x_1^k	x_2^k
x^0	-1.	1.
x^1	-9.87654 10^{-1}	1.95061
x^2	-2.35121 10^{-2}	-1.85803
x^3	-2.33551 10^{-2}	1.09087 10^{-3}
x^4	-9.21417 10^{-8}	-1.09091 10^{-3}
x^5	-7.39599 10^{-9}	-1.42543 10^{-14}

rappresentata anche nelle figure 16.13 e 16.14 (ultime tre iterate). I valori della funzione sono riportati in scala logaritmica (in base 10) in figura 16.15. Come si può notare il valore iniziale x^0 non appartiene all'intorno di convergenza. Tuttavia avviene fortunatamente che x^2 si trovi nell'intorno e da quel momento il valore ottimo viene raggiunto in pochi passi. ■

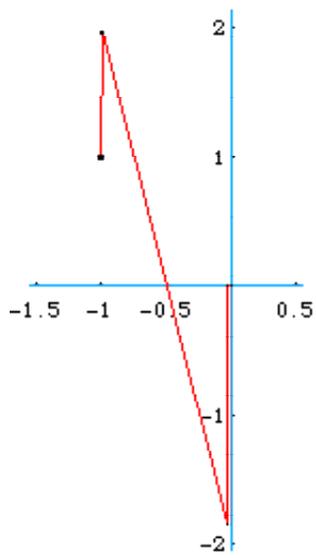


FIGURA 16.13

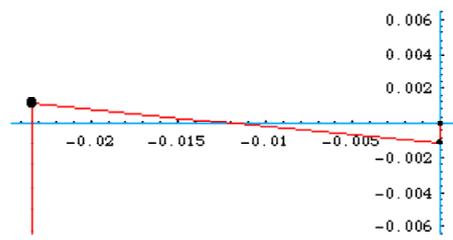


FIGURA 16.14

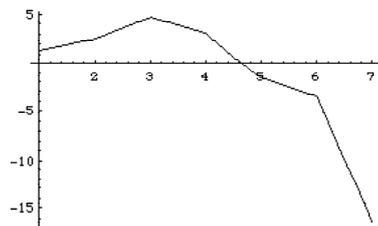


FIGURA 16.15

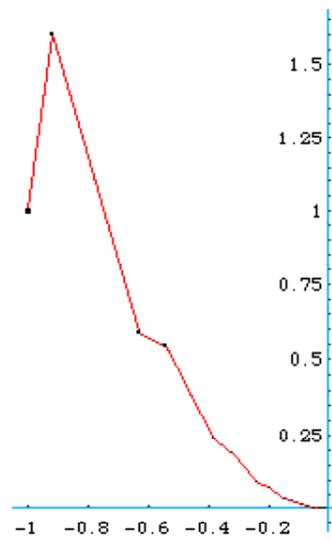


FIGURA 16.16

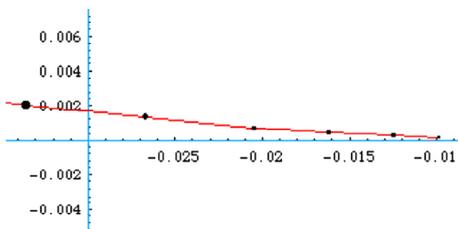


FIGURA 16.17

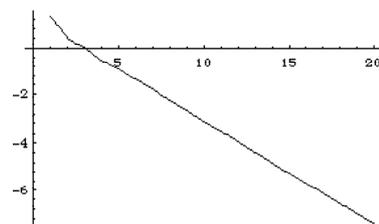


FIGURA 16.18

16.15 ESEMPIO. Si consideri ora l'esempio 16.9 per il quale l'Hessiano nell'ottimo è singolare. Quindi le ipotesi del teorema 16.12 non sono soddisfatte e la convergenza non è quadratica ma semplicemente lineare. In figura 16.16 si vedono le prime 20 iterazioni e in figura 16.17 le ultime cinque. In figura 16.18 sono riportati in scala logaritmica (in base 10) i valori di $f(x^k)$. Dall'andamento lineare del grafico si deduce che la convergenza è lineare. ■

16.5. Metodi quasi-Newton

Considerato il costo di dover calcolare esplicitamente l'Hessiano ad ogni iterazione, si può adottare una strategia di usare al posto dell'Hessiano una sua approssimazione B . I metodi che usano questa approssimazione vengono detti *quasi-Newton* e per essi l'iterazione (16.8) diventa

$$x^{k+1} = x^k - B_k^{-1} Df(x^k) \quad (16.11)$$

Le proprietà di convergenza dell'iterazione (16.11) sono stabilite dal seguente teorema (di facile dimostrazione):

16.16 TEOREMA. *Siano soddisfatte le ipotesi del teorema 16.12. Allora*

$$\| Df(x+h) - Df(x) - Bh \| \leq \frac{1}{2} \| h \|^2 + \| (B - H(x)) h \|^2$$

Se inoltre esiste B_k^{-1} , $\forall k$, allora

$$\frac{\| x^{k+1} - x^* \|}{\| x^k - x^* \|} \leq \| B_k^{-1} \| \left(\frac{1}{2} \| x^k - x^* \|^2 + \| B_k - H(x_k) \|^2 \right)$$

■

Quindi se si riesce ad ottenere un'approssimazione tale che $B_k \rightarrow H(x^*)$ la convergenza di (16.11) è superlineare (si può dimostrare che la più debole condizione $(B_k - H(x^*))(x^k - x^*)/\|x^k - x^*\| \rightarrow 0$ è necessaria e sufficiente alla convergenza superlineare). Si noti che ponendo $B_k = \alpha_k I$ per un opportuno passo α si ottiene il metodo del gradiente.

Un metodo abbastanza comune per approssimare l'Hessiano usa la seguente formula alle differenze finite se si approssima usando il gradiente, dove e_i rappresenta la i -esima colonna di una matrice identica,

$$B_{ij} := \frac{Df_j(x + \delta_i e_i) - Df_j(x)}{2 \delta_i} + \frac{Df_i(x + \delta_j e_j) - Df_i(x)}{2 \delta_j}$$

oppure, direttamente da f

$$B_{ij} := \frac{f(x + \delta_i e_i + \delta_j e_j) - f(x + \delta_i e_i) - f(x + \delta_j e_j) + f(x)}{\delta_i \delta_j}$$

I valori δ_i rappresentano l'approssimazione con cui si calcolano le derivate. Tuttavia non è vero che tanto minori sono i δ tanto migliore è la precisione. Quando δ_i è confrontabile con la precisione di macchina, $x + \delta_i e_i$ può essere uguale a x portando a delle valutazioni inaffidabili di B . Un criterio comunemente adottato consiste nel fissare $\delta_i = \mu^{1/3} x_i$ dove μ è la più piccola quantità per cui 1 e $1 + \mu$ sono rappresentati in virgola mobile da due numeri diversi.

Un altro metodo di approssimazione dell'Hessiano è il metodo della secante esteso da Broyden [1967] a \mathbb{R}^n . Se $Df(x) \in \mathbb{R}$ uno schema semplice e vantaggioso consiste nell'approssimare $H(x)$ (in questo caso un numero rappresentante la pendenza della tangente in x a Df) con la pendenza della secante definita da due iterazioni successive, cioè il rapporto $B(x) := (Df(x) - Df(x')) / (x - x')$ dove x' è la precedente iterata. Si può notare che l'approssimazione lineare

$$Df(x) + B(x)h = Df(x) + \frac{Df(x) - Df(x')}{x - x'} h$$

coincide con $Df(x)$ sia in x che in x' . Il calcolo di $B(x)$ è univoco se $Df(x) \in \mathbb{R}$. Se però $Df(x) \in \mathbb{R}^n$ e imponiamo la sola condizione che $Df(x) + B(x)(x' - x) = Df(x')$ vi sono molti gradi di libertà per la determinazione di $B(x)$. Nel seguito viene esposto un metodo per l'aggiornamento diretto dell'inversa di $B(x)$. Tuttavia tecniche simili si possono usare per aggiornare $B(x)$.

Supponiamo quindi di avere dalle precedenti iterazioni una matrice \hat{B}^{-1} . Vogliamo calcolare la nuova matrice B^{-1} da usare nel punto x in modo da minimizzare $\|B^{-1} - \hat{B}^{-1}\|$ fra tutte le matrici B che sono simmetriche (l'Hessiano è simmetrico) e soddisfano $B(x' - x) = Df(x') - Df(x)$. Per semplicità di notazione si indichi $S := B^{-1} - \hat{B}^{-1}$, $y := x - x'$, $w := Df(x) - Df(x')$, $u := y - \hat{B}^{-1}w$.

La determinazione di B^{-1} può essere effettuata usando il teorema di proiezione in uno spazio di Hilbert \mathcal{S} di matrici simmetriche. In \mathcal{S} sia definito il seguente prodotto interno

$$(S, R) := \text{Tr}(Q S Q^T Q R Q^T) \quad \forall S, R \in \mathcal{S}$$

dove Q è non singolare e per il resto è arbitraria (per il momento, più avanti verrà operata una scelta). Allora dobbiamo risolvere (nell'incognita S ; w e u sono termini noti)

$$\begin{aligned} \min \quad & \|S\| \\ & S w = u \\ & S \subset \mathcal{S} \end{aligned} \tag{16.12}$$

Sia

$$\mathcal{W} : \mathcal{S} \rightarrow \mathbb{R}^n$$

l'operatore che trasforma la matrice simmetrica S nel vettore $S w$. L'insieme

$$M := \{S \in \mathcal{S} : \mathcal{W} S = u\}$$

è una varietà lineare di matrici simmetriche. In base al teorema di proiezione negli spazi di Hilbert, l'elemento di norma minima in M è ortogonale a $\mathcal{N}(\mathcal{W}) \subset \mathcal{S}$. Il complemento ortogonale di $\mathcal{N}(\mathcal{W})$ è uguale all'immagine dell'operatore aggiunto $\mathcal{W}^* : \mathbb{R}^n \rightarrow \mathcal{S}$. Quindi la matrice \hat{S} ottima in (16.12) è data da $\mathcal{W}^* \hat{s}$ con $\hat{s} \in \mathbb{R}^n$ da calcolare. Ricordiamo che l'operatore aggiunto è definito da

$$(s, \mathcal{W} S) = (\mathcal{W}^* s, S) \quad \forall S \in \mathcal{S}, \forall s \in \mathbb{R}^n$$

dove il primo prodotto interno è definito in \mathbb{R}^n (nel modo usuale) e il secondo è definito in \mathcal{S} (nel modo indicato sopra). Quindi cerchiamo un operatore \mathcal{W}^* tale che

$$s^T S w = \text{Tr}(Q (\mathcal{W}^* s) Q^T Q S Q^T) \quad \forall S \in \mathcal{S}, \forall s \in \mathbb{R}^n$$

Notiamo che, dato il dominio e il codominio di \mathcal{W}^* , \mathcal{W}^*s deve essere una matrice simmetrica del tipo $\sum_k s_k W^k$ dove le matrici W^k sono matrici simmetriche, immagini degli elementi della base in \mathbb{R}^n .

Postuliamo che, applicando la definizione di operatore aggiunto, \mathcal{W}^*s si possa esprimere come $r s^T R + R s r^T$, con r vettore arbitrario e R matrice simmetrica arbitraria. Facciamo prima vedere che $r s^T R + R s r^T$ è una combinazione lineare di opportune matrici W^k . L'elemento generico della matrice $r s^T R + R s r^T$ è dato da $\sum_k (r_i s_k R_{kj} + R_{ik} s_k r_j)$ e quindi abbiamo $W_{ij}^k := r_i R_{kj} + R_{ik} r_j$ (simmetrica).

Ora applichiamo la definizione di operatore aggiunto. Tenendo presente che $\text{Tr}(a b^T) = a^T b$, abbiamo

$$\begin{aligned} \text{Tr}(Q(\mathcal{W}^*s)Q^T Q S Q^T) &= \text{Tr}(Q(r s^T R + R s r^T)Q^T Q S Q^T) = \\ r^T Q^T Q S Q^T Q R^T s + s^T R Q^T Q S Q^T Q r &= 2 r^T Q^T Q S Q^T Q R^T s = w^T S s \end{aligned}$$

da cui

$$R = (Q^T Q)^{-1}$$

e

$$2r = (Q^T Q)^{-1}w$$

Allora

$$\mathcal{W}^*s = \frac{1}{2} ((Q^T Q)^{-1}w s^T (Q^T Q)^{-1} + (Q^T Q)^{-1}s w^T (Q^T Q)^{-1})$$

e la matrice $\hat{S} = \mathcal{W}^* \hat{s}$, ottima in (16.12), si ottiene imponendo il vincolo di ammissibilità $S w = u$, cioè $(\mathcal{W}^*s)w = u$ e risolvendo in s . A questo punto possiamo restringere la classe di matrici Q a quelle che soddisfano il vincolo $Q^T Q y = w$ e possiamo riscrivere

$$\mathcal{W}^*s = \frac{1}{2} (y s^T (Q^T Q)^{-1} + (Q^T Q)^{-1} s y^T)$$

e, ponendo $a := (Q^T Q)^{-1}s$, dobbiamo determinare a affinché

$$\frac{1}{2} (y a^T + a y^T) w = u \implies y (a^T w) + a (y^T w) = 2u$$

e siccome, trattandosi di uno scalare, $a^T w = w^T a$, abbiamo

$$(y w^T + (y^T w) I) a = 2u \quad \text{cioè} \quad (y^T w) \left(I + \frac{y w^T}{(y^T w)} \right) a = 2u$$

Applicando la formula

$$(I + c b^T)^{-1} = I - \frac{c b^T}{1 + b^T c}$$

otteniamo

$$a = \frac{2}{y^T w} \left(I - \frac{y w^T}{2 y^T w} \right) u$$

e quindi

$$\hat{S} = \frac{1}{y^T w} \left(y u^T \left(I - \frac{w y^T}{2 y^T w} \right) + \left(I - \frac{y w^T}{2 y^T w} \right) u y^T \right)$$

$$= \frac{y u^T + u y^T}{y^T w} - \frac{y u^T w y^T}{2(y^T w)^2} - \frac{y w^T u y^T}{2(y^T w)^2}$$

e, siccome gli scalari $w^T u = u^T w$ possono commutare con gli altri termini

$$\hat{S} = \frac{y u^T + u y^T}{y^T w} - \frac{(u^T w) y y^T}{(y^T w)^2}$$

Pertanto

$$B^{-1} = \hat{B}^{-1} + \frac{y(y - \hat{B}^{-1}w)^T + (y - \hat{B}^{-1}w)y^T}{y^T w} - \frac{((y - \hat{B}^{-1}w)^T w) y y^T}{(y^T w)^2} \quad (16.13)$$

Il metodo richiede un valore iniziale B_0^{-1} che può venir calcolato con differenze finite. L'iterazione (16.2) con l'aggiornamento (16.13) prende il nome di *Metodo BFGS* (Broyden, Fletcher, Goldfarb, Shanno).

16.6. Convergenza globale

Sia il metodo di Newton che quelli di quasi-Newton non garantiscono la convergenza delle iterazioni a partire da un punto arbitrario. Questa è una caratteristica del metodo del gradiente che però presenta lo svantaggio di una convergenza troppo lenta. Sorge quindi la domanda se sia possibile adottare un metodo che unisca insieme i pregi dei due approcci. Vari metodi sono stati proposti in letteratura e molti di questi rientrano nella classe dei *metodi a regione di confidenza* (*trust region methods*). In generale si tratta di insiemi di regole abbastanza complicate per decidere sia la direzione di spostamento che il passo d'iterazione. Qui ci limitiamo a descrivere una regola semplice, che però presenta garanzia di convergenza unitamente ad una buona velocità di convergenza nelle ultime fasi del processo iterativo.

Ad un generico passo x^k sia $x^k(1) = x^k - \alpha^* Df(x^k)$, con α^* minimo di $\varphi(\alpha) := f(x - \alpha Df(x))$, e $x^k(2) = x^k - B_k^{-1} Df(x^k)$, dove B_k può essere l'Hessiano $H(x^k)$ o una sua approssimazione. Sia $\psi(\beta) := f(x^k(1) + \beta(x^k(2) - x^k(1)))$. Si calcoli ora $\min_{\beta \in [0,1]} \psi(\beta) = \psi(\beta^*)$ e si ponga

$$x^{k+1} := x^k(1) + \beta^*(x^k(2) - x^k(1))$$

Come si vede, si sceglie una direzione che è una combinazione convessa fra l'iterata che sarebbe fornita dal metodo del gradiente e quella fornita dal metodo di Newton o quasi Newton. Se la direzione $-B_k^{-1} Df(x^k)$ non è di discesa il metodo privilegia la direzione del gradiente. Però all'avvicinarsi all'ottimo il valore di β^* tende ad uno e la velocità di convergenza tende ad assumere le caratteristiche del metodo più veloce.

In questo schema sono richieste due ricerche di linea. Sono stati proposti vari metodi più efficaci che non richiedono la valutazione della funzione in un numero così elevato di punti. Tipicamente le funzioni $\varphi(\alpha)$ e $\psi(\beta)$ sono sostituite da approssimazioni quadratiche dalle quali si estrae l'informazione necessaria per calcolare il passo successivo d'iterazione.

16.17 ESEMPIO. Sia

$$f(x) := 1000(2x_1^2 - x_2)^2 + x_1^2$$

Questa funzione è molto mal condizionata. Il metodo del gradiente procede con estrema lentezza. Ad esempio partendo dal punto $x^0 = (-1, 0.5)$ si ottiene $x^{100} = (-0.5397, 0.5830)$, $f(x^{100}) = 0.2915$. Si veda in figura 16.19 (in scala logaritmica) il decremento della funzione $f(x)$. In figura 16.20 sono rappresentati i decrementi relativi al metodo misto gradiente-Newton (decremento più rapido) e al metodo misto gradiente-BFGS (decremento più lento). In ogni caso la differenza di convergenza rispetto al gradiente è notevole. In figura 16.21 si vedono le iterate del metodo misto gradiente-BFGS. ■

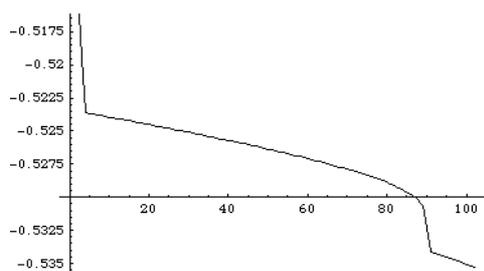


FIGURA 16.19

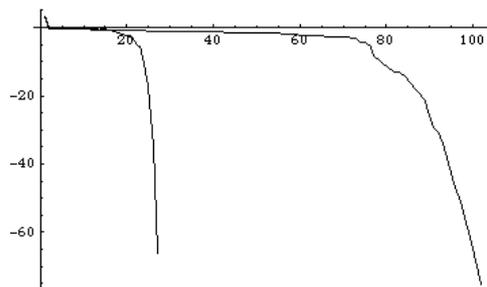


FIGURA 16.20

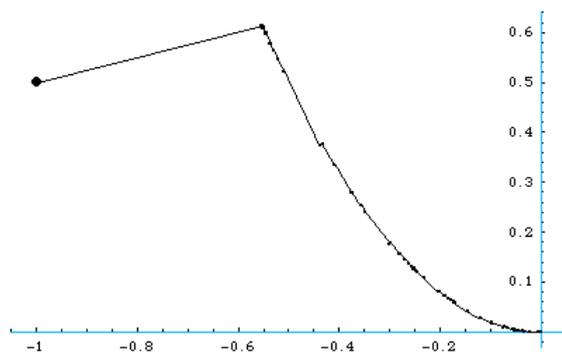


FIGURA 16.21

16.7. Programmazione non lineare vincolata

La programmazione non lineare vincolata distingue a sua volta il caso di vincoli di eguaglianza da quello di vincoli disequaglianza. Le condizioni differenziali di ottimalità necessarie e sufficienti sono già state formulate nella sezione 5.6. Naturalmente i metodi iterativi di risoluzione si basano su queste condizioni.

Presentiamo dapprima il caso con vincoli di eguaglianza. Fra le tipologie possibili di problemi, il caso più semplice è quello in cui i vincoli sono lineari e la funzione obiettivo è quadratica, ovvero:

$$\begin{aligned} \min \quad & \frac{1}{2} x^T Q x + c x \\ & A x = b \end{aligned} \quad (16.14)$$

La condizione necessaria del primo ordine 5.47 e il vincolo forniscono il seguente sistema lineare:

$$\begin{pmatrix} Q & A^T \\ A & \mathbf{0} \end{pmatrix} \begin{pmatrix} x \\ u^T \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} \quad (16.15)$$

(possiamo assumere la regolarità delle soluzioni a causa della linearità dei vincoli). Però le soluzioni di (16.15) non sono necessariamente ottimi di (16.14). La condizione necessaria del secondo ordine 5.48 impone che Q sia positiva semidefinita su $\mathcal{N}(A)$ affinché le soluzioni di (16.15) siano minimi locali. Con tale ipotesi il problema (16.14) diventa convesso e quindi i minimi che si ottengono sono anche globali. Si vede subito che se Q non è positiva

semidefinita su $\mathcal{N}(A)$ ed esistono soluzioni ammissibili allora il problema (16.14) è illimitato. Infatti in questo caso esisterebbe una direzione $h \in \mathcal{N}(A)$ tale che $h^T Q h < 0$ e il problema risulterebbe illimitato lungo h . Il calcolo di (16.15) si può effettuare in vari modi.

Difficoltà possono sorgere se la matrice di (16.15) è singolare oppure se è a grande dimensione.

16.18 ESEMPIO. Si consideri l'istanza elementare:

$$Q := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad c := (c_1 \quad c_2) \quad A := (1 \quad 0)$$

Si può verificare che Q è positiva semidefinita su $\mathcal{N}(A) = \text{lin}(0, 1)$ (è nulla nella fattispecie). Il sistema (16.15) diventa

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ u \end{pmatrix} = \begin{pmatrix} -c_1 \\ -c_2 \\ b \end{pmatrix}$$

che ammette soluzioni (in numero infinito) solo se $b = -c_2$, essendo la matrice singolare. Questo sistema si risolve facilmente, data la dimensione ridotta dell'istanza, ma, se la matrice è invece molto grande, la risoluzione di un sistema lineare singolare può presentare difficoltà. ■

Nel caso generale i metodi si dividono in due classi a seconda se riducano il problema al caso svincolato oppure tengano conto esplicitamente dei vincoli. L'idea di fondo dei primi metodi consiste nello svincolare il problema, ma nel penalizzare ogni scostamento dai valori ammissibili. Ad esempio il problema

$$\min \{f(x) : g_i(x) = 0, i := 1, \dots, m\} \quad (16.16)$$

viene trasformato in

$$\min f(x) + \rho p(g(x)) \quad (16.17)$$

dove p è una funzione cosiddetta di penalità tale che $p(y) = 0$ se $y = 0$, $p(y) > 0$ se $y \neq 0$ e $\rho > 0$ è un opportuno coefficiente. Varie scelte sono possibili per p . Se p viene scelta in modo da essere differenziabile nei punti ammissibili, allora bisogna risolvere il problema (16.17) molte volte per valori crescenti del coefficiente ρ , producendo una successione di punti che tendono all'ottimo del problema.

Per essere più specifici si consideri $p(y) := y^T y$ e quindi $p(g(x)) = \sum_i g_i(x)^2$. Allora le condizioni di primo e secondo ordine del problema svincolato (16.17) diventano

$$Df(x) + 2\rho \sum_i g_i(x) Dg_i(x) = 0 \quad (16.18)$$

$$D^2 f(x) + 2\rho \sum_i (g_i(x) D^2 g_i(x) + Dg_i(x) Dg_i(x)^T) \quad \text{positiva semidefinita}$$

Sia $x(\rho)$ la soluzione di (16.18). Si può dimostrare che, sotto alcune blande ipotesi tecniche, $\lim_{\rho \rightarrow \infty} x(\rho) = \hat{x}$, con \hat{x} minimo locale di (16.16) (Fiacco e McCormick [1968]). Inoltre si ha $\lim_{\rho \rightarrow \infty} 2\rho g_i(x(\rho)) = \hat{u}_i$, con \hat{u}_i variabile duale ottima. In base a questo risultato l'Hessiano della funzione svincolata è costituito da due parti, la prima $D^2 f(x) + 2\rho \sum_i g_i(x) D^2 g_i(x)$ che tende all'Hessiano del Lagrangiano nel minimo di (16.16), e la seconda data da $2\rho Dg^T Dg$

che diventa illimitata al tendere di ρ ad infinito. In altre parole l'Hessiano diventa sempre peggio condizionato all'aumentare del coefficiente di penalità, con tutte le conseguenze computazionali negative che si sono viste.

Una scelta della funzione di penalità che supera questa difficoltà (introducendone altre ovviamente) è data da una funzione non liscia come la norma l_1 , $p(y) := \sum_i |y_i|$. In questo caso non occorre far tendere il coefficiente di penalità ad infinito. Si può dimostrare che esiste un valore di soglia al di là del quale ogni soluzione di (16.17) è soluzione di (16.16). Questa proprietà è dovuta al fatto che la funzione di penalità, per così dire, 'esce' dall'insieme ammissibile già con una pendenza positiva e quando questa supera i valori di derivata direzionale di f fuori dal piano tangente dei vincoli, il minimo deve stare dentro i vincoli. Però questa medesima proprietà introduce delle caratteristiche di non differenziabilità nella funzione da minimizzare che complicano la risoluzione (si veda Coleman e Conn [1982a, 1982b]).

I metodi sequenziali si basano invece sull'uso di un'approssimazione quadratica del problema nei punti dell'iterazione. Sia x^k il punto corrente dell'iterazione, che non assumiamo necessariamente ammissibile, e sia u^k la stima corrente della variabile duale. Definiamo

$$Q := D^2 f(x^k) + \sum_i u_i^k D^2 g_i(x^k) \quad c := Df(x^k) \quad A := Dg(x^k) \quad b := -g(x^k) \quad (16.19)$$

e risolviamo il problema quadratico

$$\begin{aligned} \min \quad & \frac{1}{2} h^T Q h + c h \\ & A h = b \end{aligned} \quad (16.20)$$

risolvendo il sistema lineare

$$\begin{pmatrix} Q & A^T \\ A & \mathbf{0} \end{pmatrix} \begin{pmatrix} h \\ v^T \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} \quad (16.21)$$

Si noti che se x^k è l'ottimo del problema allora la soluzione di (16.21) è $h = 0$ e $v = \hat{u}$. Altrimenti i valori dell'iterazione vengono aggiornati come $x^{k+1} := x^k + h$, e $u^{k+1} := v$.

16.19 ESEMPIO. Sia $f(x) := x_1^2 x_2 - x_2$ e $g(x) := (x_1 - 2)^2 + 4(x_2 - 2)^2 - 1$. Partendo dai valori $x^0 := (1, 1)$ e $u^0 := 1$, si ottiene la successione

	x_1	x_2	u
0	1.	1.	1.
1	0.428571	1.64286	0.5
2	0.449092	2.31314	0.65884
3	0.709919	1.91855	1.05537
4	0.940095	2.06734	1.66824
5	0.987544	1.99132	1.93345
6	0.999763	2.00033	1.99848
7	1.	2.	2.

indicata anche in figura 16.22 insieme con l'insieme ammissibile (ellisse) e le curve di livello (iperboliche). Si può verificare, applicando le condizioni di ottimalità, che $\hat{x} = (1, 2)$ e $\hat{u} = 2$ sono ottimi del problema. ■

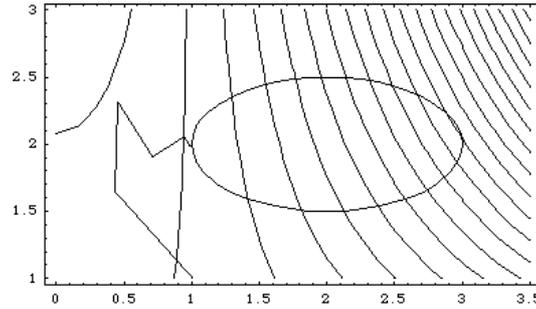


FIGURA 16.22

Nell'esempio precedente non si sono prese speciali precauzioni sulla scelta della lunghezza del passo per ottenere la convergenza. L'implicita scelta della lunghezza del passo uguale ad 1 corrisponde ad una iterazione del metodo di Newton. Questa scelta è certamente pagante nell'intorno dell'ottimo, ma distante dall'ottimo può essere del tutto insoddisfacente. Si tratta inoltre di decidere quale criterio usare nella scelta della lunghezza del passo nella direzione h . È evidente che non basta la sola funzione obiettivo in quanto bisogna anche tenere conto dell'ammissibilità e bisogna altresì tener conto che si minimizza il Lagrangiano. Queste esigenze portano a definire delle cosiddette *funzioni di merito*. Ci limitiamo a citare la funzione di merito detta *Lagrangiano aumentato* e definita da

$$M(x, u, \rho) := f(x) + u g(x) + \frac{1}{2} \rho g(x)^T g(x)$$

L'Hessiano di M è dato da

$$D^2 M = D^2 f(x) + \sum_i (u_i + \rho g_i(x)) D^2 g_i(x) + \rho Dg(x) Dg(x)^T$$

Una generica direzione h può essere decomposta come $h = h^0 + h^1$ dove $h^0 \in \mathcal{N}(Dg(x))$ e $h^1 \in \mathcal{N}(Dg(x))^\perp = \mathcal{R}(Dg(x)^T)$. Siccome, al tendere dell'iterazione verso punti ammissibili, il termine $\sum_i \rho g_i(x) D^2 g_i(x)$ si annulla, il fattore di penalità si riflette solo sul termine $\rho Dg(x)^T Dg(x)$. Per definizione si ha $Dg(x) h^0 = 0$ e $h^1 = Dg(x)^T q$ per un opportuno vettore q . Quindi la forma quadratica (nelle successive espressioni non si indica, per semplicità di notazione, la dipendenza delle varie funzioni dai rispettivi argomenti)

$$(h^1)^T Dg^T Dg h^1 = q^T Dg Dg^T Dg Dg^T q$$

è definita positiva (supponendo rango pieno dello Jacobiano dei vincoli). Allora

$$h^T D^2 M h = (h^0)^T D^2 L h^0 + 2 (h^0)^T D^2 L h^1 + (h^1)^T (D^2 L + \rho Dg^T Dg) h^1$$

Sia: $(h^0)^T D^2 L h^0 \geq \alpha \|h^0\|^2$, con $\alpha > 0$ supponendo verificate le condizioni sufficienti di secondo ordine, $(h^0)^T D^2 L h^1 \geq \beta \|h^0\| \|h^1\|$, β arbitrario, $(h^1)^T D^2 L h^1 \geq \gamma \|h^1\|^2$, γ arbitrario, $(h^1)^T Dg^T Dg h^1 \geq \delta \|h^1\|^2$, $\delta > 0$ per quanto espresso sopra. Siccome

$$\alpha \|h^0\|^2 + 2\beta \|h^0\| \|h^1\| + (\gamma + \rho\delta) \|h^1\|^2 > 0 \quad \text{se} \quad \beta^2 - \alpha(\gamma + \rho\delta) < 0$$

si vede che per un valore di ρ sufficientemente grande $D^2 M$ è definita positiva, garantendo che l'ottimo del problema vincolato sia un minimo del Lagrangiano aumentato svincolato.

16.20 ESEMPIO. Sia $f(x_1, x_2) := x_1^3 + x_1^2 + x_2 - x_2^2$ e $g(x_1, x_2) := x_1^2 - x_2$. In figura 16.23 sono disegnate le linee di livello della funzione obiettivo, l'insieme ammissibile e l'iterazione che verrà successivamente descritta. Si noti che la linea di livello $f(x) = 0.25$ presenta un punto multiplo e per i valori $0.25 < p < 0.4$, la linea di livello $f(x) = p$ si spezza in due curve distinte; infine la linea di livello $f(x) = 0.4$ consiste di un punto e di una linea distinti.

La funzione ha un minimo locale vincolato nell'origine con valore duale ottimo $\hat{u} = 1$. Si ottiene

$$D^2 f(0,0) + \hat{u} D^2 g(0,0) = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$$

e quindi il Lagrangiano è positivo definito solo su $\mathcal{N}(Dg(0,0)) = \text{lin}(1,0)$, ma non su tutto lo spazio. Aggiungendo il termine di penalità si ottiene

$$D^2 M(0,0) = \begin{pmatrix} 4 & 0 \\ 0 & 2\rho - 2 \end{pmatrix}$$

e quindi ponendo $\rho := 2$ si trasforma l'origine in un punto di minimo locale per il Lagrangiano aumentato. Si veda la figura 16.24 dove sono indicate le linee di livello del Lagrangiano aumentato in un intorno dell'origine.

Usando il Lagrangiano aumentato come funzione di merito e inizializzando l'iterazione da $x^0 := (1, 1)$, $u^0 := 1$, si ottiene l'iterazione indicata in figura 16.23 e qui sotto riportata.

	x_1	x_2	u
0	1.	1.	1.
1	0.25	-0.5	1.
2	0.025	0.05	1.1
3	-0.01717	0.03901	1.00109
4	0.00022	-0.00030	1.00061
5	-0.00011	0.00015	1.
6	$0.927 \cdot 10^{-8}$	$-1.236 \cdot 10^{-8}$	1.

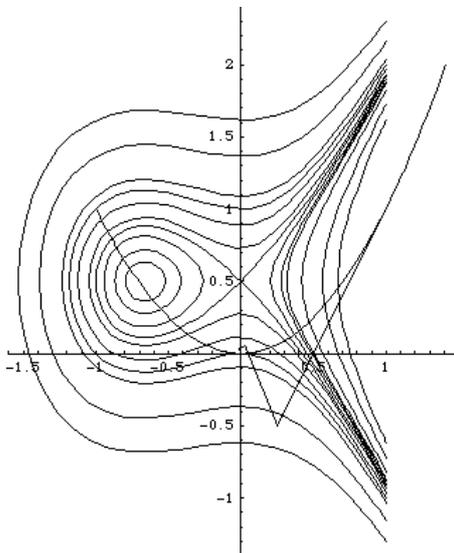


FIGURA 16.23

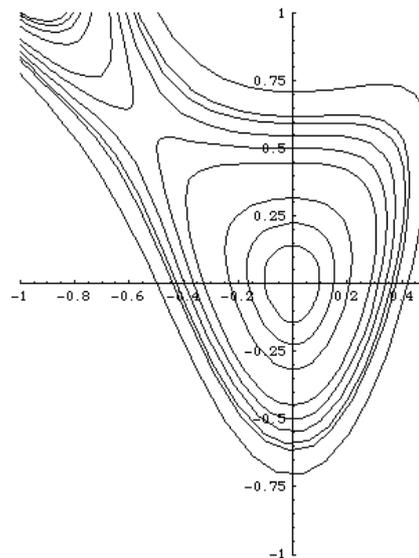


FIGURA 16.24

Se i vincoli sono di disequaglianza bisogna distinguere fra vincoli attivi e vincoli non attivi nel punto corrente d'iterazione. Essendo l'analisi della direzione di discesa di tipo locale i vincoli non attivi possono essere ignorati. Ovviamente l'insieme dei vincoli attivi cambia da iterazione ad iterazione e va correttamente identificato ad ogni passo. Limitando la descrizione dei metodi a quelli sequenziali quadratici, facciamo dapprima riferimento al seguente modello quadratico, analogo di (16.14):

$$\begin{aligned} \min \quad & \frac{1}{2} x^T Q x + c x \\ & A x \leq b \end{aligned} \quad (16.22)$$

La soluzione di (16.22) si ottiene risolvendo il seguente sistema *non* lineare di disequaglianze:

$$\begin{pmatrix} Q & A^T & \mathbf{0} \\ A & \mathbf{0} & I \end{pmatrix} \begin{pmatrix} x \\ u^T \\ s \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} \quad u \geq 0 \quad s \geq 0 \quad u_i s_i = 0 \quad \forall i \quad (16.23)$$

La risoluzione di (16.23), quando Q sia positiva semidefinita si può fare abbastanza facilmente con gli algoritmi ai punti interni che verranno esaminati nel prossimo capitolo (si veda in particolare la sezione 17.3).

Il modello (16.22) viene usato nel caso generale per generare direzioni di discesa con le sostituzioni indicate in (16.19), dove g è limitata ai soli vincoli attivi in x^k . Come nel caso con vincoli di eguaglianza, conviene usare una funzione di merito per controllare il progresso lungo la direzione di discesa. Bisogna modificare la definizione di funzione di penalità data l'asimmetria dei vincoli di disequaglianza. Ad esempio si può usare

$$p(y) := \sum_i \max \{0, (y_i + \varepsilon)\}^2$$

Il termine ε va visto come una tolleranza sulla violazione del vincolo ed è anche utile per evitare discontinuità delle derivate proprio sui punti di frontiera (queste sono state spinte un po' in dentro nella regione ammissibile).

La ricetta per produrre un'iterazione convergente verso dei minimi locali è molto complessa, e, come già sottolineato, richiede un'analisi molto dettagliata del problema da risolvere. Si rinvia alla letteratura citata.