

Introduzione ai metodi di ricerca lineare per problemi di ottimizzazione non lineare non vincolata.

Mehiddin Al-Baali*

Introduzione

Questo capitolo introduce i principali metodi di ottimizzazione per risolvere problemi di ottimizzazione non lineare non vincolata. L'attenzione è focalizzata su tecniche di soluzione che presentano un interesse pratico piuttosto che teorico; l'obiettivo è quello di offrire al lettore la possibilità di utilizzare metodi numerici per risolvere alcuni problemi di ottimizzazione.

Il capitolo contiene la descrizione di alcuni metodi numerici di ottimizzazione e le loro recenti modifiche per risolvere problemi di minimizzazione non vincolata. I metodi che verranno considerati sono il metodo del gradiente, il metodo del gradiente coniugato, il metodo di Newton, il metodo quasi-Newton.

La Sezione 1 presenta alcuni concetti di base relativi all'ottimizzazione vincolata e descrive un metodo di eliminazione delle variabili, che consente di ottenere a partire da un problema di ottimizzazione vincolata la controparte non vincolata. Tuttavia, tutti i metodi sopracitati possono essere utilizzati per affrontare i corrispondenti problemi di ottimizzazione non vincolata, ottenuti applicando metodi di penalità e del Lagrange a problemi di ottimizzazione vincolata.

1 Il Problema di Ottimizzazione

Lo scopo di tale sezione è quello di descrivere alcuni pratici metodi di soluzione per affrontare il seguente problema di ottimizzazione non vincolata:

$$\min_{x \in R^n} f(x). \quad (1)$$

*Department of Mathematics and Statistics, Sultan Qaboos University, Muscat, Oman.
(albaali@squ.edu.om)

In quanto segue, si assume che f (denominata funzione obiettivo) sia una funzione continua e differenziabile di un vettore di n variabili

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (2)$$

che può essere scritto anche come $x = (x_1, x_2, \dots, x_n)^T$, dove T indica il trasposto di un vettore.

È opportuno notare che determinare una soluzione per il problema di massimo $\max_x z(x)$ equivale a risolvere il problema (1) con $f(x) = -z(x)$, dal momento che $\max f(x) = -\min(-f(x))$, questo garantisce che entrambi i problemi abbiano la stessa soluzione x^* con $z(x^*) = -f(x^*)$.

I valori di $f(x^*)$ e $z(x^*)$ rappresentano, rispettivamente, un minimo e un massimo locale di f . Un estremo di f è anche indicato come minimo oppure massimo di f .

Si consideri, ad esempio, la rappresentazione grafica di due funzioni $f(x)$ e $-f(x)$ riportata in Figura 1.

Si osserva che anche se le rappresentazioni grafiche di funzioni con $n \leq 2$ sono utili in alcuni casi, esse vengono generalmente utilizzate a scopo puramente illustrativo.

In quanto segue, vengono riportate alcune definizioni utili per descrivere i metodi usati per risolvere il problema (1).

Definizione 1. Un punto x^* è detto *minimizzatore locale (o relativo)* per una funzione f se la disuguaglianza

$$f(x) \geq f(x^*) \quad (3)$$

è soddisfatta per tutti i valori di x sufficientemente vicini a x^* , cioè in qualche intorno di x^* . In questo caso, il valore $f(x^*)$ rappresenta un minimo locale di f .

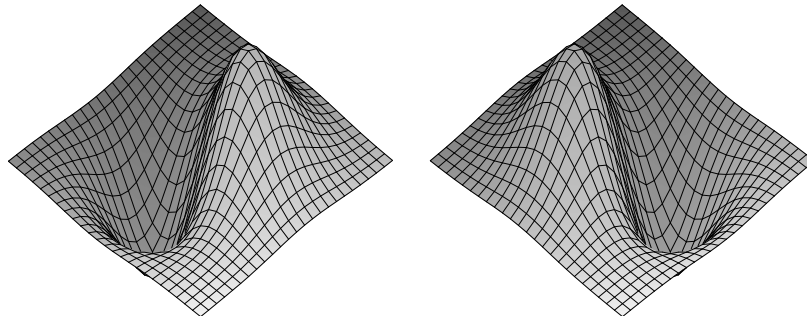


Figura 1: Rappresentazione grafica delle funzioni $f(x) = (x_2 - x_1)e^{-x_1^2 - x_2^2}$ e $-f(x)$

Il punto x^* è chiamato minimizzatore isolato o stretto se la condizione (3) viene sostituita dalla seguente $f(x^*) < f(x)$. Se tali disuguaglianze risultano valide per tutti i valori di x , allora x^* è detto minimizzatore globale o stretto.

Per illustrare questa definizione, è utile considerare il seguente esempio. Un semplice problema di ottimizzazione con una funzione obiettivo in una sola variabile $x \in \mathbb{R}$ può essere rappresentato mediante il modello (1), ponendo

$$f(x) = (x - 2)^2. \quad (4)$$

È chiaro che questa funzione quadratica ha un unico punto di minimo $x^* = 2$, dal momento che la disuguaglianza $f(x) \geq f(2) = 0$ risulta valida per tutti i valori di x . È ovvio che questo minimizzatore locale è anche un minimizzatore globale.

Un'altra funzione obiettivo in una sola variabile, che però non soddisfa la disuguaglianza (3) per tutti i valori di x , è riportata in quanto segue:

$$f(x) = 1 - \frac{1}{c}x(x - 0,3)(x - 0,7)(x - 1,1)(x - 1,5), \quad (5)$$

dove c è una costante positiva.

Per verificare se esiste un minimizzatore x^* , è possibile adottare il seguente procedimento. In particolare, poichè la derivata prima della funzione (5) soddisfa la condizione

$$f'(0) < 0 \quad (6)$$

(chiamata proprietà di discesa, in breve PD), esiste un valore di $x > 0$ (ovvero \hat{x}) tale che $f(\hat{x}) < f(0)$.

È possibile notare che la funzione (5) ha un minimizzatore x^* nell'intervallo $(0; 0,3)$, poichè oltre alla proprietà (6), risulta $f(0) = f(0,3)$ (infatti, $x^* \approx 0,1136$). Inoltre, esiste un altro minimizzatore locale in $(0,7; 1,1)$. Questi risultati possono essere facilmente osservati in Figura 2, che riporta la rappresentazione grafica della funzione, in corrispondenza di $c = 0,3465$.

Nella Sezione 2, verranno analizzati alcuni algoritmi, che soddisfano alcune condizioni standard, per calcolare un minimizzatore oppure una sua approssimazione per una generica funzione.

In \mathbb{R}^2 , la funzione di Rosenbrock

$$f(x) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2 \quad (7)$$

ammette il minimizzatore globale $x^* = (1, 1)^T$, dal momento che $f(1, 1) = 0$ e $f(x) \geq 0$.

Per una funzione generica, come quella riportata in quanto segue:

$$f(x) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2 + x_1^2 \quad (8)$$

non è semplice, generalmente, determinare un minimizzatore locale, anche nel caso in cui la disuguaglianza $f(x) \geq 0$ sia verificata.

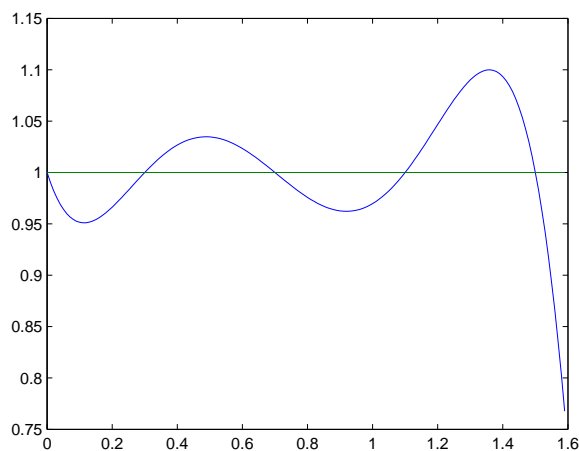


Figura 2: Funzione (5)

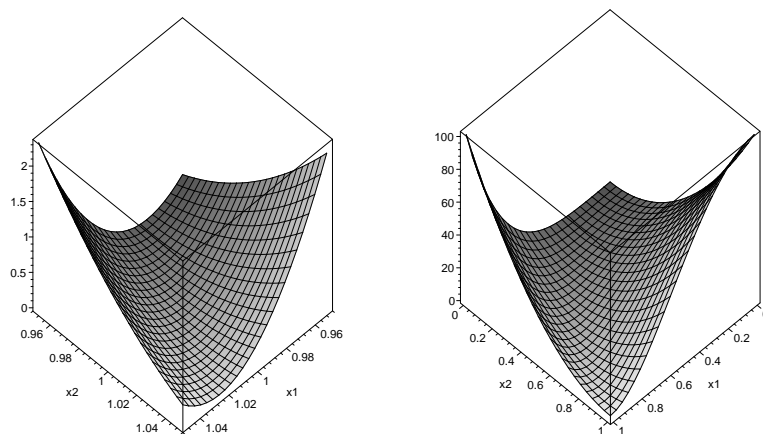


Figura 3: Rappresentazione grafica delle funzioni (7) e (8)

Si considerino, inoltre, i grafici delle funzioni (7) e (8) riportati in Figura 3; non è possibile dall'analisi di tale figura individuare facilmente i minimizzatori delle due funzioni (ovvero $(1, 1)^T$ e $(1/2, 1/4)^T$), nonostante siano state adeguatamente rappresentate.

Sulla base delle considerazioni precedenti, in quanto segue verranno presentati alcuni metodi pratici per determinare un minimizzatore di una funzione. Tali metodi considerano informazioni relative alla derivata prima e, a volte, alla derivata seconda, di $f(x)$ (indicate come vettore gradiente $g(x) = \nabla f(x)$ e matrice Hessiana $G(x) = \nabla^2 f(x)$, rispettivamente).

La Sezione 1.2 descrive queste tecniche e le illustra applicando alcuni metodi analitici a una selezione di problemi semplici da affrontare. Dal momento che questi metodi risultano essere non adeguati per risolvere un generico problema di ottimizzazione, nelle seguenti sezioni verranno considerati diversi metodi numerici. Questi metodi generano iterativamente

una sequenza di minimizzatori approssimati $\{x^{(k)}\}$, a partire da un dato punto iniziale $x^{(1)}$, che sperabilmente converge a un minimizzatore x^* .

Se tale sequenza è costruita in modo tale che il valore $f^{(k)} = f(x^{(k)})$ (similmente, verranno usate in seguito $f^* = f(x^*)$, $g^{(k)} = g(x^{(k)})$, ecc.) si riduce a ogni iterazione, allora x^* è generalmente un minimizzatore, tranne in casi particolari piuttosto rari. La sequenza, inoltre, si dice che converge globalmente se $x^{(1)}$ è un punto qualsiasi, non necessariamente appartenente ad un intorno di x^* . Il tasso di convergenza della sequenza è generalmente studiato considerando alcuni problemi di riferimento. In particolare, la nota funzione di Rosenbrock (7) con punto iniziale $x^{(1)} = (-1, 2, 1)^T$ è generalmente considerata una utile funzione obiettivo a cui fare riferimento per valutare il tasso di convergenza.

È opportuno notare che se sono imposti alcuni limiti sui valori che può assumere x , allora il problema (1) dovrebbe essere soggetto a tali vincoli. In questo caso, si ottiene un problema di ottimizzazione vincolata, che può essere rappresentato matematicamente nel seguente modo.

$$\begin{array}{ll} \min & f(x) \\ \text{soggetto a} & c_i(x) = 0, \quad i = 1, \dots, m' \\ & c_i(x) \geq 0, \quad i = m' + 1, \dots, m, \end{array}$$

dove $c_i(x)$ sono funzioni differenziabili.

Ad esempio, si consideri il semplice problema

$$\begin{array}{ll} \min & f(x) = x_1^2 + x_2^2 \\ \text{soggetto a} & x_1 + 2x_2 = 4. \end{array}$$

Questo problema di ottimizzazione vincolata può essere risolto eliminando una variabile (per esempio, $x_1 = 4 - 2x_2$) per ottenere il corrispondente problema di ottimizzazione non vincolata

$$\min f(x_2) = 4(x_2 - 2)^2 + x_2^2.$$

Per risolvere questo problema quadratico, sulla base della definizione di minimizzatore fornita in precedenza, si riscrive la funzione obiettivo come $f(x_2) = \frac{1}{5}[(5x_2 - 8)^2 + 16]$, che chiaramente ammette il minimizzatore $x_2^* = \frac{8}{5}$ (uno schema più semplice per calcolare il minimizzatore di una funzione verrà presentato in seguito). Sostituendo tale valore nel vincolo $c_1 \equiv x_1 + 2x_2 - 4 = 0$, si ottiene il valore di $x_1^* = \frac{4}{5}$. Quindi, una soluzione per il problema considerato è data da $x^* = (\frac{4}{5}, \frac{8}{5})^T$ con il valore minimo della funzione $f^* = \frac{16}{5}$.

Sebbene il metodo di eliminazione di variabili, appena descritto, rappresenti una strategia di soluzione abbastanza semplice, generalmente non è applicabile a un generico problema di ottimizzazione vincolata. Pertanto, sono stati proposti altri metodi per questo tipo di problema e alcuni di essi

sono basati sulla soluzione di un particolare tipo di problemi non vincolati, che corrispondono al problema vincolato preso in esame, come i metodi di penalità e di Lagrange (si veda ad esempio Fletcher, 1987).

1.1 I Fondamenti dell'Ottimizzazione Non Vincolata

Il vettore gradiente e la matrice Hessiana di una funzione $f(x)$ sono matematicamente definiti mediante le seguenti espressioni:

$$g(x) = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix} \quad (9)$$

e

$$G(x) = \begin{bmatrix} \partial^2 f / \partial x_1^2 & \partial^2 f / \partial x_1 \partial x_2 & \dots & \partial^2 f / \partial x_1 \partial x_n \\ \partial^2 f / \partial x_2 \partial x_1 & \partial^2 f / \partial x_2^2 & \dots & \partial^2 f / \partial x_2 \partial x_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2 f / \partial x_n \partial x_1 & \partial^2 f / \partial x_n \partial x_2 & \dots & \partial^2 f / \partial x_n^2 \end{bmatrix}. \quad (10)$$

Da notare che la matrice $G(x)$ è simmetrica, poichè si è assunto che $\partial^2 f / \partial x_i \partial x_j = \partial^2 f / \partial x_j \partial x_i$.

Esempio 1. Se f è definita dalla funzione di Rosenbrock (7), cioè

$$f(x) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2,$$

segue che

$$g(x) = \begin{bmatrix} -400x_1(x_2 - x_1^2) + 2(x_1 - 1) \\ 200(x_2 - x_1^2) \end{bmatrix} \quad (11)$$

$$G = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}. \quad (12)$$

Sostituendo $x^* = (1, 1)^T$, si ottiene

$$g^* = 0, \quad G^* = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}, \quad (13)$$

dove g^* e G^* indicano $g(x^*)$ e $G(x^*)$, rispettivamente. In questo caso x^* è un punto stazionario, che verrà definito in seguito.

Esercizio. Sia data la funzione

$$f(x) = 2x_1^4 + x_2^2 + 2x_1x_2 + x_1 - x_2 + (1 + x_3)^2.$$

Determinare $g(0)$ e $G(0)$.

Definizione 2. Un vettore \hat{x} è detto punto stazionario di una funzione f se soddisfa la seguente equazione del gradiente:

$$g(x) = 0. \quad (14)$$

Esempio 2. Trovare tutti i punti stazionari per la funzione

$$f(x) = (x_2 - x_1)e^{-x_1^2 - x_2^2}.$$

Soluzione. Il gradiente di f è

$$g(x) = \begin{bmatrix} 2x_1^2 - 2x_1x_2 - 1 \\ 2x_1x_2 - 2x_2^2 + 1 \end{bmatrix} e^{-x_1^2 - x_2^2}.$$

Risolvendo il sistema $g(x) = 0$, si ottengono i punti stazionari $\hat{x} = (1/4, -1/4)^T$ e $\tilde{x} = (-1/4, 1/4)^T$.

È utile osservare che, per molte applicazioni di grande interesse pratico, non è possibile risolvere in maniera esatta il sistema di equazioni non lineari (14). Quindi, nella pratica, la condizione (14) viene sostituita dalla relazione $\|g(x)\| \leq \epsilon$, dove $\|\cdot\|$ indica la norma (definita in seguito), mentre ϵ è una tolleranza (ad esempio, il valore $\epsilon = 10^{-7}$ è considerato prossimo allo zero). Se questa condizione risulta valida per qualche punto \hat{x} , allora esso è considerato un punto stazionario di una funzione f .

Ad esempio, il valore di $\hat{x} = 0,1135784$ è un punto stazionario per la funzione (5) perchè si può dimostrare che $|y'(\hat{x})| = 0,6420 \times 10^{-7}$ (vedi anche la Figura 2), e $\hat{x} = 0,1136$ è inoltre un punto stazionario accettabile approssimato a 4 cifre decimali con $|y'(0,1136)| = 0,00011715728 \leq 0,0002$.

Sono stati, quindi, proposti numerosi metodi numerici di tipo iterativo per risolvere in maniera approssimata tale sistema. Una forma generale di questi metodi è descritta nella Sezione 1.3, assumendo noti alcuni concetti che verranno forniti nelle successive sezioni.

Da notare che il simbolo $\|\cdot\|$ è utilizzato per indicare una delle seguenti norme

$$\|z\|_1 = \sum_{i=1}^n |z_i|, \quad (15)$$

$$\|z\|_2 = \left(\sum_{i=1}^n z_i^2 \right)^{\frac{1}{2}}, \quad (16)$$

$$\|z\|_\infty = \max_{1 \leq i \leq n} |z_i|, \quad (17)$$

dove $z \in \mathbb{R}^n$; tali norme sono dette, rispettivamente, norma ℓ_1 , ℓ_2 (o Euclidea) e ℓ_∞ .

Una volta che i punti stazionari sono stati trovati, è necessario classificarli sulla base di alcuni criteri, che saranno descritti nella Sezione 1.2 e che sono definiti considerando le seguenti definizioni e/o teoremi.

Definizione 3. Una matrice simmetrica G è detta definita positiva, semi-definita positiva, definita negativa o semi-definita negativa se è soddisfatta, rispettivamente, la disuguaglianza $z^T G z > 0$, $z^T G z \geq 0$, $z^T G z < 0$ o $z^T G z \leq 0$, per ogni $z \neq 0$; altrimenti, la matrice è indefinita.

Poichè verificare tali disuguaglianze è solitamente difficile per una generica matrice, si possono ritenere validi i seguenti risultati equivalenti.

Teorema 1. Assumendo che G sia una matrice simmetrica e definita positiva, i seguenti assiomi sono equivalenti:

- (i) $z^T G z > 0 \forall z \neq 0$,
- (ii) tutti gli autovalori di G sono positivi,
- (iii) i fattori di Cholesky $G = LDL^T$, dove L è una matrice triangolare inferiore unitaria e D è una matrice diagonale, esistono per $d_{ii} > 0 \forall i$
- (iv) tutti i minori principali di G sono positivi, ovvero $\Delta_i > 0 \forall i$, dove

$$\Delta_1 = G_{11}, \Delta_2 = \det \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}, \dots, \Delta_n = \det(G).$$

È utile notare che G è simmetrica e definita negativa se la matrice $(-G)$ è definita positiva, ovvero gli assiomi definiti in precedenza (i), (ii) e (iii) con il segno della disuguaglianza $>$ sostituito da $<$ e l'assioma (iv) con $\Delta_i > 0$ sostituito da $(-1)^i \Delta_i > 0$, sono equivalenti.

Esempio 3. L'Hessiana G^* della funzione di Rosenbrock (vedi (13)) è definita positiva per l'assioma (i), perchè

$$z^T G^* z = 802z_1^2 - 800z_1z_2 + 200z_2^2 > 200(2z_1 - z_2)^2 \geq 0.$$

Questo risultato può essere ottenuto semplicemente dall'assioma (iv) con l'osservazione che

$$\Delta_1 = 802 > 0, \quad \Delta_2 = 802(200) - (400)^2 > 0.$$

Esercizio. Dimostrare che $G > 0$ per $G = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$.

Una funzione $f(x)$ è una funzione convessa o strettamente convessa in una data regione se la corrispondente matrice Hessiana $G(x)$ è definita semi-positiva o positiva in quella regione, rispettivamente. Questa proprietà può essere verificata facilmente nel caso in cui l'Hessiana è costante, ovvero se la funzione è quadratica. Più specificatamente, una funzione quadratica è generalmente definita come segue

$$q(x) = \frac{1}{2} x^T G x + b^T x + c, \quad (18)$$

dove G è una matrice simmetrica e costante, b è un vettore costante e c è uno scalare. Si assume che G è definita positiva; quindi, questa funzione ha un unico punto di minimo. Questa funzione verrà usata per dedurre alcuni importanti risultati, considerando che il gradiente della funzione (18) assume la seguente forma:

$$g(x) = Gx + b. \quad (19)$$

Se $x = \hat{x} + \alpha s$, dati \hat{x} e s , segue che

$$q(\hat{x} + \alpha s) = q(\hat{x}) + \alpha s^T \hat{g} + \frac{1}{2} \alpha^2 s^T G s \quad (20)$$

e

$$g(\hat{x} + \alpha s) = \hat{g} + \alpha G s. \quad (21)$$

Dati \hat{x} e s , l'espressione (20) rappresenta una funzione quadratica in una sola variabile α , che ammette il seguente minimizzatore

$$\alpha^* = -\frac{s^T \hat{g}}{s^T G s} = -\frac{s^T (G \hat{x} + b)}{s^T G s} = -\frac{s^T G \hat{x} + s^T b}{s^T G s}. \quad (22)$$

Esempio 4. Determinare il gradiente della seguente funzione:

$$f(x) = x_1^2 + x_1(x_1 + 2x_2) + (x_2 - 1)^2$$

nel punto $\hat{x} = (1, -1)^T$. Trovare, quindi, il minimizzatore di $y(\alpha) = f(\hat{x} + \alpha s)$, dove $s = (-2, 1)^T$.

Soluzione. La funzione è quadratica, perchè può essere scritta nella forma (18) con $G = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$, $b = (0, -2)^T$ e $c = 1$ costanti. Quindi, il suo gradiente può essere ottenuto usando la (19) come

$$g(x) = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} x + \begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 4x_1 + 2x_2 \\ 2x_1 + 2x_2 - 2 \end{pmatrix}$$

che può essere verificata applicando la (9) direttamente a f . Quindi, $\hat{g} = (2, -2)^T$.

Il minimo di $y(\alpha)$ può essere trovato risolvendo l'equazione $y'(\alpha) = 0$, dove $y(\alpha) = f(1 - \alpha, -1 + \alpha)$, o equivalentemente usando la formula (22) come segue

$$\alpha^* = -\frac{s^T \hat{g}}{s^T G s} = -\frac{(-2, 1) \begin{pmatrix} 2 \\ -2 \end{pmatrix}}{(-2, 1) \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{pmatrix} -2 \\ 1 \end{pmatrix}} = \frac{3}{5}.$$

Esercizio. Per ogni matrice costante Q , dimostrare che la funzione $f(x) = \frac{1}{2} x^T Q x + b^T x + c$ è quadratica con $G = \frac{1}{2}(Q + Q^T)$. Dimostrare che la

seguinte funzione è equivalente a quella quadratica descritta sopra:

$$f(x) = \frac{1}{2}x^T \begin{bmatrix} 4 & 4 \\ 0 & 2 \end{bmatrix} x - 2x_2 + 1.$$

Usando la funzione quadratica di cui sopra e lo sviluppo in serie di Taylor per $f(x)$ per un punto \hat{x} , si può scrivere

$$f(x) = \hat{f} + \delta^T \hat{g} + \frac{1}{2} \delta^T \hat{G} \delta + \dots, \quad (23)$$

dove

$$\delta = x - \hat{x}. \quad (24)$$

Pertanto, per valori di $\|\delta\|$ sufficientemente piccoli, $f(x + \delta)$ può essere approssimata da una funzione quadratica, tale approssimazione può essere utile in alcuni casi.

Esempio 5. Usare l'approssimazione quadratica vicino al punto $x^* = (1, 1)^T$, per approssimare la funzione di Rosenbrock (7) in $x^a = (1, 1, 0, 9)^T$, $x^b = (-1, 2, 1)^T$ e $x^c = (1, -1, 2)^T$.

Soluzione. Applicando lo sviluppo in serie di Taylor in x^* e usando la (13) in (23), si ottiene

$$f(x) = f(x^* + \delta) \approx \frac{1}{2} \delta^T G^* \delta = \frac{1}{2} (\delta_1, \delta_2) \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix},$$

dove $\delta = (x_1 - 1, x_2 - 1)^T$. Per approssimare f in x^a, x^b e x^c , si sostituisce δ con $(0, 1, -0, 1)^T$, $(-2, 2, 0)^T$ e $(0, -2, 2)^T$ nell'ultima approssimazione quadratica, ottenendo, rispettivamente, $f^a \approx 9,01$, $f^b \approx 1940,8$ e $f^c \approx 484$.

Confrontando tali approssimazioni con i valori esatti di $f^a = 9,62$, $f^b = 24,2$ e $f^c = 484$, si osserva una accettabile e una cattiva approssimazione per f^a e f^b , rispettivamente, ma si ottiene una approssimazione esatta per f^c (la giustificazione di tale comportamento è lasciata al lettore).

1.2 Ottimizzazione Analitica

In questa sezione, si affronta la risoluzione analitica di un problema di minimizzazione (1) e di un problema di massimizzazione.

Teorema 2. Se x^* è un minimizzatore locale per f (ad esempio, una soluzione del problema (1)), allora $g^* = 0$ e $G^* \geq 0$. Tali condizioni rappresentano le condizioni necessarie del primo e del secondo ordine, rispettivamente.

Inoltre, se x^* è massimizzatore locale per f , allora $g^* = 0$ e $(-G^*) \geq 0$.

La validità di tale risultato è facilmente desumibile osservando la rappresentazione grafica della funzione f (vedi ad esempio le Figure 1 e 2).

Tuttavia, la dimostrazione di tali proprietà si basa sullo sviluppo in serie di Taylor

$$y(\alpha) \equiv f(x^* + \alpha s) = f^* + \alpha s^T g^* + \frac{1}{2} \alpha^2 s^T G^* s + \dots \quad (25)$$

che si ottiene dalla relazione (23) con $\hat{x} = x^*$ e $\delta = \alpha s$ per ogni scalare α e vettore s . Quindi, si può osservare che se $y(0) = f^*$ è un minimo, allora la derivata prima $y'(0) = s^T g^* = 0$ e la derivata seconda $y''(0) = s^T G^* s \geq 0$. Poichè tali espressioni sono valide per ogni valore di s , segue che $g^* = 0$ e $G^* \geq 0$.

Teorema 3. *Se $g^* = 0$ e $G^* > 0$ (condizioni sufficienti del primo e del secondo ordine, rispettivamente), allora x^* è un minimizzatore locale stretto e isolato per f .*

La dimostrazione si basa sullo sviluppo in serie di Taylor introdotto in precedenza (si veda, ad esempio, Fletcher, 1987). Questo risultato conduce al seguente corollario.

Corollary 1. *Sia x^* un punto stazionario, tale che $g^* = 0$. Se inoltre $(-G^*) > 0$, allora x^* è un massimizzatore locale stretto e isolato per f , mentre se nè $G^* \geq 0$ nè $(-G^*) \geq 0$ sono verificate, allora x^* è un punto di sella.*

Da notare che se $g^* = 0$ e $G^* \geq 0$, allora analizzare l'Hessiana G^* non è sufficiente poichè x^* può essere o un minimizzatore oppure un punto di sella. Analogamente, se $g^* = 0$ e $(-G^*) \geq 0$, allora x^* può essere sia un massimizzatore e sia un punto di sella.

Quindi, per determinare un minimizzatore è necessario trovare i punti stazionari risolvendo l'equazione (14) e verificando se la matrice Hessiana è definita positiva in questi punti (vedi la Sezione 1.2 per maggiori dettagli). Tuttavia, poichè generalmente è difficile o impossibile risolvere l'equazione del gradiente (14), sono stati proposti numerosi metodi numerici iterativi. In questo capitolo, si studieranno alcuni dei metodi più popolari e si descriveranno brevemente le loro proprietà teoriche e numeriche. La struttura di base di tali metodi verrà fornita nella Sezione 1.3, che contiene inoltre un algoritmo generale di ricerca lineare (in breve, RL) e alcuni concetti di base.

Esempio 6. *Trovare il minimo per la funzione di Rosenbrock (7).*

Soluzione. Risolvendo $g(x) = 0$ (usando il gradiente (11)), si determina il punto stazionario $x^* = (1, 1)^T$. Quindi, analizzando il valore di G^* dato da (13), segue che $G^* > 0$ (mostrato nell'Esempio 3); $x^* = (1, 1)^T$ è infatti un minimizzatore per f .

Esempio 7. *Trovare e classificare tutti i punti stazionari della funzione*

$$f(x) = x_1^3 + 3x_1x_2^2 - 12x_1. \quad (26)$$

Soluzione. Il gradiente di f è

$$g(x) = \begin{bmatrix} 3x_1^2 + 3x_2^2 - 12 \\ 6x_1x_2 \end{bmatrix}.$$

Per risolvere $g(x) = (g_1, g_2)^T = 0$, si osserva che

$$g_2 = 0 \Rightarrow 6x_1x_2 = 0 \Rightarrow x_1 = 0 \quad \text{o} \quad x_2 = 0.$$

Sostituendo tali valori in $g_1 = 0$, segue rispettivamente che

$$x_1 = 0 \text{ con } \{3x_2^2 - 12 = 0 \Rightarrow x_2 = 2 \quad \text{o} \quad x_2 = -2\}.$$

$$x_2 = 0 \text{ con } \{3x_1^2 - 12 = 0 \Rightarrow x_1 = 2 \quad \text{o} \quad x_1 = -2\}.$$

Si determinano, quindi, i seguenti quattro punti stazionari:

$$\hat{x}^\pm = \begin{pmatrix} 0 \\ \pm 2 \end{pmatrix}, \quad \tilde{x}^\pm = \begin{pmatrix} \pm 2 \\ 0 \end{pmatrix}.$$

Per classificare tali punti, si calcola la matrice Hessiana di f

$$G(x) = \begin{bmatrix} 6x & 6y \\ 6y & 6x \end{bmatrix} = 6 \begin{bmatrix} x & y \\ y & x \end{bmatrix}.$$

Sostituendo i punti stazionari precedentemente determinati in questa espressione, si ha

$$\hat{G}^\pm = 6 \begin{bmatrix} 0 & \pm 2 \\ \pm 2 & 0 \end{bmatrix}, \quad \tilde{G}^\pm = 6 \begin{bmatrix} \pm 2 & 0 \\ 0 & \pm 2 \end{bmatrix}.$$

Considerando l'assioma (iv) del Teorema 1, si calcola

$$\{\hat{\Delta}_1^\pm = 0, \hat{\Delta}_2^\pm < 0\}, \{\tilde{\Delta}_1^+ > 0, \tilde{\Delta}_2^+ > 0\}, \{\tilde{\Delta}_1^- < 0, \tilde{\Delta}_2^- > 0\}.$$

L'assioma (iv) suggerisce che \tilde{x}^+ è un minimizzatore, \tilde{x}^- è un massimizzatore e i rimanenti \hat{x}^\pm sono due punti di sella. Questo risultato è in linea con la rappresentazione grafica della funzione (26), riportata in Figura 4.

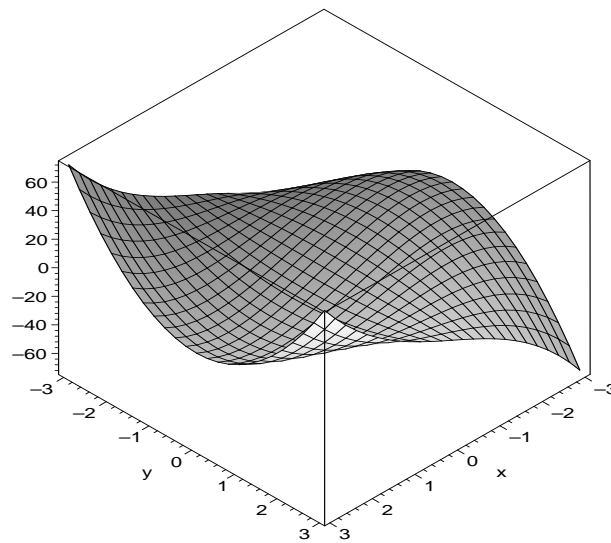


Figura 4: Funzione (26)

Esercizio.

1. Per la funzione,

$$f(x) = 3x_1^2 + x_2^3 - 6x_1x_2 - 9x_2 + 2,$$

verificare quali dei seguenti punti 0 , $(3, 3)^T$, $(-1, -1)^T$ sono di punti di minimo locale.

2. Trovare e classificare tutti i punti stazionari per le due funzioni seguenti:

$$f(x) = (x_2 - x_1)e^{-x_1^2 - x_2^2}.$$

$$f(x) = e^{x_1 - x_2} + e^{x_2 - x_1} + e^{x_1^2} + x_3^2.$$

1.3 Ottimizzazione Numerica

In quanto segue, verranno descritti alcuni metodi numerici per risolvere il problema (1), data una stima iniziale $x^{(1)}$ di un minimizzatore per f . Le due principali strategie per risolvere il problema preso in esame si basano sui metodi di ricerca lineare e sui metodi di trust-region (o metodo della regione di confidenza).

Di seguito, si considera la prima classe di metodi che generano una sequenza di punti $\{x^{(k)}\}$ nel seguente modo: a ogni iterazione k , viene definita una direzione di ricerca $s^{(k)}$, tale che la proprietà di discesa

$$s^{(k)T} g^{(k)} < 0, \tag{27}$$

sia soddisfatta. Questa proprietà, che verrà dimostrata nella Sezione 2, assicura che esiste un valore di $\alpha^{(k)} > 0$ tale che la condizione di decremento del valore della funzione (indicata con C0)

$$f^{(k+1)} < f^{(k)} \quad (28)$$

sia verificata. Per un valore accettabile di $\alpha^{(k)}$, un nuovo punto è dato da

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)}. \quad (29)$$

La procedura ha termine quando una prestabilita condizione di convergenza risulta essere soddisfatta. In particolare, l'algoritmo ha termine quando $x^{(k)}$ è sufficientemente vicino (o quasi vicino) a un minimizzatore locale ovvero quando la seguente condizione è soddisfatta:

$$\|g^{(k)}\| \leq \epsilon, \quad (30)$$

dove $\epsilon > 0$ è una tolleranza (ad esempio, $\epsilon = 10^{-6}$ e in teoria, $\epsilon = 0$). Poichè la convergenza di alcuni metodi non è garantita in un numero finito di iterazioni, si utilizzano i seguenti criteri di terminazione: $\|x^{(k+1)} - x^{(k)}\|$ è sufficientemente piccola, oppure $|f^{(k+1)} - f^{(k)}|$ è sufficientemente piccola, oppure il numero di iterazioni è sufficientemente grande.

Generalmente, i metodi di soluzione sono definiti in modo tale che la sequenza di punti $\{x^{(k)}\}$ converge a un minimizzatore x^* di f (o a una sua approssimazione) con un accettabile tasso di convergenza. Essi si differenziano nel modo in cui vengono determinate la direzione di ricerca $s^{(k)}$ e la lunghezza del passo $\alpha^{(k)}$, in modo tale da soddisfare alcune utili proprietà. La strategia usata per determinare un valore adeguato di $\alpha^{(k)}$ (soluzione del sottoproblema di ricerca lineare) è descritta dettagliatamente nella Sezione 2, mentre i metodi per calcolare $s^{(k)}$ sono illustrati nelle rimanenti sezioni.

Lo schema generale di questo tipo di metodi di ricerca lineare può essere sintetizzato nel seguente modo.

Algoritmo 1. Strategia generale di ricerca lineare

Step 0: Dato un punto iniziale $x^{(1)}$, poni $k := 1$,

Step 1: se la condizione di terminazione (ad esempio, la condizione (30)) è verificata, Stop,

Step 2: definisci una direzione di ricerca $s^{(k)}$ tale che la (27) sia soddisfatta,

Step 3: determina un passo $\alpha^{(k)} > 0$ tale che la (28) sia soddisfatta,

Step 4: calcola una nuova approssimazione del minimizzatore $x^{(k+1)}$ utilizzando la relazione (29),

Step 5: poni $k := k + 1$ e vai allo Step 1.

Anche se le procedure per determinare la direzione di ricerca $s^{(k)}$, utilizzata nello Step 2, e il passo $\alpha^{(k)}$, utilizzato nello Step 3, verranno definite nelle prossime sezioni, in quanto segue viene illustrata l'applicazione dell'algoritmo di ricerca locale ad alcuni esempi, in corrispondenza di alcuni predeterminati valori di $s^{(k)}$ e $\alpha^{(k)}$.

Esempio 8. Risolvere il problema (1), in cui la funzione obiettivo è rappresentata dalla seguente funzione quadratica

$$f(x) = \frac{1}{2}(x_1^2 + x_2^2),$$

considerando come stima iniziale il punto $x^{(1)} = (0, 3162, 0, 9487)^T$, sulla base dei due metodi riportati in quanto segue.

- Ponendo $\alpha^{(k)} = 1$ e utilizzando per determinare la direzione di ricerca $s^{(k)}$ un algoritmo che garantisce il soddisfacimento della proprietà di discesa (27) a ogni iterazione, si ottiene la soluzione del problema dopo 8 iterazioni con la generazione dei punti $x^{(2)} = (0, 0, 8538)^T, \dots, x^{(8)} = (-0, 0001, -0, 0001)^T, x^{(9)} = 0$. Tali punti sono illustrati in Figura 5 che, inoltre, offre un'idea del comportamento dell'algoritmo per ottenere il minimizzatore $x^* = 0$. Si osserva che ogni nuovo punto $x^{(k+1)}$ è più vicino di quello corrente $x^{(k)}$ a x^* , per ogni $k \geq 1$, pertanto la condizione di decremento della funzione (28) è soddisfatta in tutte le iterazioni. Tuttavia, la figura evidenzia la possibilità di ottenere x^* direttamente in un unico step muovendosi lungo la retta che passa attraverso i punti $x^{(1)}$ e x^* . Questa possibilità è illustrata in quanto segue.

- Ponendo $\alpha^{(1)} = 1$ e $s^{(1)} = -\begin{pmatrix} 0, 3162 \\ 0, 9487 \end{pmatrix}$, che rappresenta la cosiddetta migliore direzione di ricerca di Newton, si ottiene il minimizzatore $x^{(2)} = x^*$ in una sola iterazione.

Nonostante l'esempio appena descritto (esempio 8) sottolinei l'utilità del metodo di Newton, quest'ultimo presenta dei limiti, se viene applicato ad alcune tipologie di problemi. Nelle sezioni successive verrà descritto in maniera dettagliata tale metodo, insieme con altre strategie di ricerca locale, e per ognuno di essi si analizzeranno le proprietà teoriche fondamentali. Allo scopo di illustrare le differenze tra i metodi, è utile fare riferimento al seguente semplice esempio.

Esempio 9. Sia data la seguente funzione f

$$f(x) = x_1^2 + 4x_2^2, \quad (31)$$

dove $x^{(1)} = (1, 1)^T$, e siano date le seguenti direzioni di ricerca

$$s^a = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad s^b = \begin{pmatrix} -0, 5 \\ 0 \end{pmatrix}, \quad s^c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad s^d = \begin{pmatrix} -3 \\ 0 \end{pmatrix}.$$

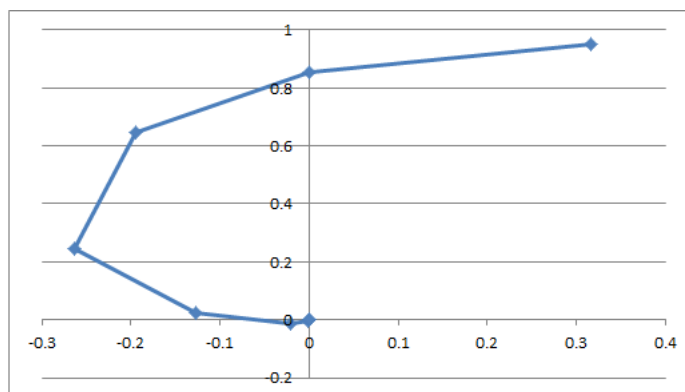


Figura 5: Ricerche lineari sugli assi x_1x_2

Calcolare nuove stime $x^{(k)}$ di $x^* = 0$, utilizzando come direzione di ricerca i vettori $s^{(k)}$ indicati e scegliendo opportunamente i valori di $\alpha^{(k)} = 1, 1/2, 1/4, \dots$

Soluzione. In primo luogo si calcola il gradiente della funzione (31)

$$g(x) = 2 \begin{pmatrix} x \\ 4x \end{pmatrix}.$$

Si eseguono, quindi, le seguenti iterazioni:

$k = 1$: Inizialmente si verifica il soddisfacimento della PD (27) per $k = 1$ ovvero se $s^{aT}g^{(1)} = (0, -1) \begin{pmatrix} 2 \\ 4 \end{pmatrix} = -4 < 0$. Dal momento che tale condizione è verificata, si considera come direzione di ricerca $s^{(1)} = s^a$. Poi si verifica il soddisfacimento della condizione C0 per $k = 1$. In particolare, si calcola $x^+ = x^{(1)} + s^{(1)} = (1, 0)^T$. Dal momento che $f^+ = 1 < f^{(1)} = 5$, segue che la condizione C0 è verificata. Quindi, si sceglie $\alpha^{(1)} = 1$ e si pone $x^{(2)} = (1, 0)^T$ e $f^{(2)} = 1$.

$k = 2$: Analogamente, si pone $s^{(2)} = s^b$, perchè la PD, $s^{(2)T}g^{(2)} < 0$, è verificata. Poichè la condizione C0, $f(x^{(2)} + s^{(2)}) = 0,25 < f^{(2)} = 1$, è soddisfatta, si pone $\alpha^{(2)} = 1$; quindi si ha $x^{(3)} = x^{(2)} + s^{(2)} = (0,5, 0)^T$ e $f^{(3)} = 0,25$.

$k = 3$: Come sopra, si verifica il soddisfacimento della PD per $k = 3$ ovvero se $s^{cT}g^{(3)} = (1, 1) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 > 0$. Poichè la proprietà di discesa non è verificata, non si considera il vettore s^c come direzione di discesa (infatti, se si pone $x^+ = x^c + s^{(3)} = (1,5, 1)^T$, si ha $f^+ = 6,25 > f^{(3)} = 0,25$

ovvero la condizione C0 non è verificata, cioè, x^+ è peggiore del punto precedente).

$k = 3$; (continua): Quindi, si considera s^d e si ripete la procedura. Si nota che la PD $s^{dT} g^{(3)} < 0$ è verificata. Quindi, si considera come direzione di ricerca $s^{(3)} = s^d$. Il valore di $\alpha^{(3)} > 0$ tale che $f^{(4)} < f^{(3)}$, può essere calcolato sfruttando la condizione C0 come illustrato in quanto segue.

Si calcola $x^+ = x^{(3)} + s^{(3)} = (-2, 5, 0)^T$; si nota che $f^+ = 6, 25 > f^{(3)} = 0, 25$, quindi la condizione C0 non è verificata e il valore $\alpha = 1$ viene scartato. Si considera, quindi, un valore più piccolo, ad esempio $\alpha = 0, 5$.

Si calcola $x^+ = x^{(3)} + 0, 5s^{(3)} = (-1, 0)^T$; dal momento che risulta $f^+ = 1 > f^{(3)}$, ne segue che la condizione C0 non è verificata. Quindi, il valore $\alpha = 0, 5$ viene scartato e viene preso in esame un valore più piccolo, ad esempio $\alpha = 0, 25$.

Si calcola $x^+ = x^{(3)} + 0, 25s^{(3)} = (-0, 25, 0)^T$; dal momento che $f^+ = 0, 0625 < f^{(3)}$, segue che la condizione C0 è verificata. Quindi si sceglie $\alpha^{(3)} = 0, 25$ e si pone $x^{(4)} = x^{(3)} + 0, 25s^{(3)} = (-0, 25, 0)^T$ a cui corrisponde $f^{(4)} = 0, 0625$.

Questo esempio mostra chiaramente, in corrispondenza della terza iterazione, che una scelta non corretta di $s^{(k)}$ potrebbe provocare un fallimento dell'algoritmo. Inoltre, mostra che dopo 3 iterazioni si ottiene il punto

$$x^{(2)} + s^{(2)} + 0, 25s^{(4)} = \begin{pmatrix} -0, 25 \\ 0 \end{pmatrix}$$

che è lontano dal minimizzatore $x^* = 0$. Tuttavia, se alla seconda iterazione venisse utilizzato il valore di $\alpha^{(2)} = 0, 5$ al posto del valore $\alpha^{(2)} = 1$ si avrebbe che

$$x^{(2)} + 2s^{(2)} = 0,$$

che fornisce il minimizzatore x^* .

Pertanto, è chiaro che le scelte di $\alpha^{(k)}$ e $s^{(k)}$ influenzano fortemente la capacità dell'Algoritmo 1 di risolvere con successo il problema di ottimizzazione. La Sezione 2 illustra meglio questa situazione e suggerisce alcune strategie utili per trovare valori accettabili di $\alpha^{(k)}$. Le altre sezioni descrivono alcuni metodi e analizzano le proprietà teoriche che vengono utilizzate per definire $s^{(k)}$. In particolare, le Sezioni 3, 4, 5 e 6 considerano, rispettivamente, il metodo di discesa più ripida, il metodo del gradiente coniugato, il metodo di Newton e il metodo quasi-Newton con direzioni di

ricerca definite per esempio come segue:

$$s^{(k)} = -g^{(k)}, \quad (32)$$

$$s^{(k)} = \begin{cases} -g^{(1)} & \text{if } k = 1, \\ -g^{(k)} + \frac{\|g^{(k)}\|^2}{\|g^{(k-1)}\|^2} s^{(k-1)} & \text{if } k > 1, \end{cases} \quad (33)$$

$$s^{(k)} = -G^{(k)-1} g^{(k)}, \quad (34)$$

$$s^{(k)} = -H^{(k)} g^{(k)}, \quad (35)$$

dove

$$H^{(k)} = \begin{cases} I & \text{if } k = 1, \\ \left\{ H - \frac{H\gamma\gamma^T H}{\gamma^T H \gamma} + \frac{\delta\delta^T}{\delta^T \gamma} \right\}^{(k-1)} & \text{if } k > 1, \end{cases} \quad (36)$$

In particolare, $H^{(k)}$ approssima $G^{(k)-1}$, I è la matrice identità e

$$\delta^{(k-1)} = x^{(k)} - x^{(k-1)}, \quad \gamma^{(k-1)} = g^{(k)} - g^{(k-1)}. \quad (37)$$

È utile notare che la (33) e la (35) definiscono le ben note formule FR e DFP che appartengono alle classi dei metodi del gradiente coniugato e quasi-Newton, rispettivamente.

2 Il sottoproblema di ricerca lineare

In questa sezione, si affronta la determinazione della lunghezza del passo $\alpha^{(k)}$, necessaria nello Step 3 dell'Algoritmo 1, che è definita idealmente risolvendo il seguente sottoproblema di ricerca lineare

$$\min_{\alpha > 0} y(\alpha), \quad (38)$$

dove

$$y(\alpha) = f(x^{(k)} + \alpha s^{(k)}). \quad (39)$$

Indicando

$$x^+ = x^{(k)} + \alpha s^{(k)} \quad (40)$$

e $s = s^{(k)}$, e usando la regola della catena, si ottiene

$$\begin{aligned} y'(\alpha) &= \frac{d}{d\alpha} f(x^+) \\ &= \frac{\partial f}{\partial x_1^+} \frac{dx_1^+}{d\alpha} + \cdots + \frac{\partial f}{\partial x_n^+} \frac{dx_n^+}{d\alpha} \\ &= g_1^+ s_1 + \cdots + g_n^+ s_n \\ &= s^{(k)T} g^+ \\ &= s^{(k)T} g(x^{(k)} + \alpha s^{(k)}). \end{aligned} \quad (41)$$

$$= s^{(k)T} g(x^{(k)} + \alpha s^{(k)}). \quad (42)$$

Quindi, la proprietà di discesa (27) può essere scritta come

$$y'(0) < 0 \quad (43)$$

che comporta l'esistenza di una soluzione $\alpha^* > 0$ per il problema (38). Pertanto, si cerca di determinare un valore di $\alpha > 0$ che soddisfa la condizione di decremento della funzione

$$C0: \quad y(\alpha) < y(0) \quad (44)$$

che equivale a (28) quando $\alpha = \alpha^{(k)}$, risolvendo il problema (38) nei due modi seguenti:

- in maniera esatta scegliendo $\alpha^{(k)} = \alpha^*$ (riferito a una ricerca lineare esatta) o
- in maniera non esatta scegliendo $\alpha^{(k)} \approx \alpha^*$ (per esempio, $\hat{\alpha}$, riferito a una ricerca lineare inesatta o approssimata).

Il primo caso richiede che venga risolta l'equazione

$$y'(\alpha) = 0 \quad (45)$$

esattamente, mentre il secondo caso richiede di trovare un valore di $\alpha > 0$ tale che la condizione C0 sia soddisfatta (in seguito saranno prese in considerazione ulteriori condizioni). Per illustrare tali strategie, si considera il semplice esempio riportato di seguito.

Esempio 10. *Risolvere il problema (38) se*

$$y(\alpha) = \alpha^3 - \alpha + 1$$

in due modi: (i) in maniera esatta e (ii) in maniera non esatta sfruttando la condizione di decremento C0.

Soluzione. Poichè la PD $y'(0) = -1 < 0$ è verificata, i due casi possono essere applicati con successo, notando che $y(0) = 1$.

Caso (i) Risolvendo l'equazione $y'(\alpha) = 0$, per $\alpha > 0$, si ottiene la radice positiva $\alpha^* = 1/\sqrt{3} \approx 0,5774$ che è accettabile, perchè il minimo $y(\alpha^*) = 1 - 2/3\sqrt{3} \approx 0,6151 < 1$.

Caso (ii). In questo caso, la condizione C0 può essere scritta come $y(\alpha) < 1$. Si pone $\alpha = 1$ e si ottiene $y(1) = 1$. Poichè tale scelta non soddisfa la condizione C0, si considera $\alpha = 1/2$ e si ottiene $y(1/2) = 0,6250 < 1$. Poichè $\alpha = 1/2$ soddisfa la condizione C0, esso rappresenta una soluzione approssimata di α^* .

Si consideri ora una funzione con un minimizzatore α^* non noto con certezza. L'obiettivo che si vuole raggiungere consiste nel determinare una approssimazione del minimizzatore a 4 cifre decimali.

Esempio 11. Sia $y(\alpha)$ definita dalla relazione (5) in cui x e f sono sostituiti da α e y , rispettivamente; segue che

$$y(\alpha) = 1 - \frac{1}{c}\alpha(\alpha - 0,3)(\alpha - 0,7)(\alpha - 1,1)(\alpha - 1,5), \quad (46)$$

dove c è una costante positiva (si sceglie $c = 0,3465$). È possibile notare che $y(0) = 1$ e $y'(0) = -1$, quindi la PD è soddisfatta, pertanto esiste un minimizzatore esatto per y (vedi inoltre le Figure 2 e 6).

Nonostante sia possibile ottenere un minimizzatore esatto α^* ($\approx 0,1136$ oppure $0,9194$), sono richiesti accurati calcoli numerici. Infatti, questi minimizzatori, che soddisfano la condizione $|y'(\alpha^*)| \leq 0,0001$, possono essere trovati applicando due volte il metodo di Newton-Raphson alla (45), e considerando due punti iniziali sufficientemente vicini a α^* (per esempio, $0,1$ e $0,9$ che possono essere dedotti analizzando la figura).

Tuttavia, è più semplice usare la ricerca lineare non esatta per trovare un valore di α che soddisfa la condizione C0. Per esempio, sia $\alpha = 1$; dal momento che $y(1) \approx 0,9697 < y(0)$, si ha che, in base alla condizione C0, $\alpha = 1$ rappresenta una approssimazione accettabile di α^* .

Esercizio. Trovare una soluzione approssimata per i seguenti problemi:

$$\min y(\alpha) = \alpha^4 - \alpha^2 - \alpha + 1$$

$$\min y(\alpha) = 100\alpha^4 + \alpha^2 - 2\alpha + 1.$$

In genere è difficile o impossibile risolvere il problema (38). Quindi, viene considerata la possibilità di trovare una soluzione approssimata del problema (38) che soddisfa la condizione C0 e altre condizioni standard.

Poichè la condizione C0 garantisce un decremento molto limitato di y (o f), un metodo di ottimizzazione che si base su tale condizione potrebbe fallire nel risolvere il problema (1). Pertanto, al fine di superare tale limite, generalmente si considera la condizione di sufficiente decremento della funzione, riportata in quanto segue

$$C1 : \quad y(\alpha) \leq y(0) + \rho\alpha y'(0), \quad (47)$$

dove $\rho \in (0, 1/2)$. L'equazione associata alla condizione C1 individua una retta (indicata, nel seguito, come ρ -retta). La condizione C1 consente di determinare un intervallo accettabile di valori di α , che contiene α^* , un minimizzatore di $y(\alpha)$ se esso esiste.

Per mostrare l'utilità del limite superiore su ρ , è utile fare riferimento alla funzione quadratica (4), nella quale si sostituisce x con α e si scrive la condizione C1. Si ottiene, quindi, $(\alpha - 2)^2 \leq 4 - 4\rho\alpha$. Tale disuguaglianza non è soddisfatta in corrispondenza del minimizzatore $\alpha = 2$ e in corrispondenza di ogni altro valore di $\rho > 1/2$.

Generalmente, si pone $\rho = 10^{-4}$ (valore usato in seguito se non diversamente specificato). Tuttavia, considerando la funzione (46) con valori di $\rho = 1/4$ e $\rho = 1/40$, che possono essere sostituiti nella condizione C1 (47), si ottiene

$$\text{C1a: } y(\alpha) \leq 1 - \frac{1}{4}\alpha \quad \text{e} \quad \text{C1b: } y(\alpha) \leq 1 - \frac{1}{40}\alpha,$$

rispettivamente. Le rappresentazioni grafiche della funzione (46) e delle due ρ -rette in C1a e C1b, sono riportate nelle Figure 6 e 7, rispettivamente. La condizione C1a conduce ad accettabili valori di α negli intervalli $(0; 0,1686)$ e $(1,6314; \infty)$, mentre C1b produce i seguenti intervalli $(0; 0,2794)$, $(0,7779; 1,0220)$ e $(1,5206; \infty)$. È chiaro che un valore più piccolo di ρ consente di ottenere valori più appropriati di α .

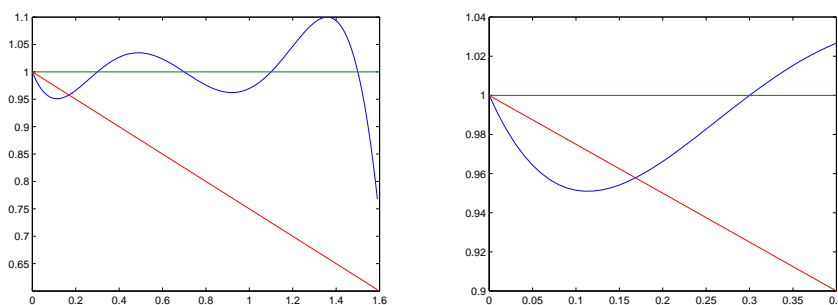


Figura 6: Funzione (46) e ρ -retta in (47) con $\rho = 0,25$

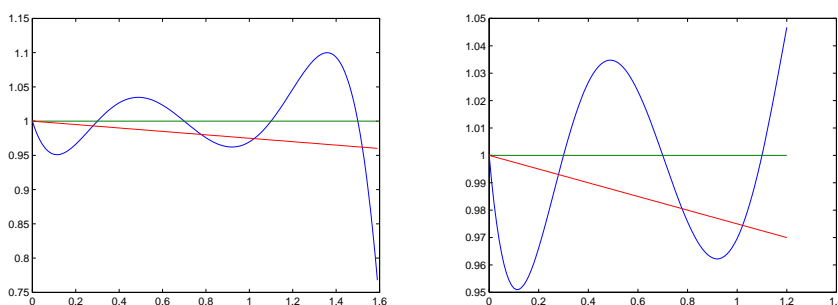


Figura 7: Rappresentazione grafica della funzione (46) e della ρ -retta, definita dalla condizione (47), con $\rho = 0,025$

Sono stati proposti diversi algoritmi per determinare un valore di α che soddisfi la condizione C1. In questa trattazione, si farà riferimento all'approccio descritto in quanto segue.

Algoritmo 2. Risolvere il sottoproblema di ricerca lineare per la condizione C1

Step 0: Dato $y(0)$, $y'(0) < 0$ e $\rho < 0,5$, sia $b_1 = \infty$ e poni $j = 1$,

Step 1: scegli $\alpha = \alpha_j \in (0, b_j)$,
 se C1 è verificata allora l'algoritmo ha termine,
 altrimenti poni $b_{j+1} = \alpha_j$, $j = j + 1$ e vai allo Step 1.

Il valore iniziale di $\alpha_1 > 0$ può essere stimato basandosi sui valori disponibili di $y(0)$, $y'(0)$ e su una possibile stima di $y(\alpha)$ (vedi la (62)). Tuttavia, in quanto segue si pone $\alpha_1 = 1$.

Per $j > 1$, il valore di α_j può essere scelto come minimizzatore della funzione quadratica che interpola i dati disponibili $y(0)$, $y'(0)$ e $y(\alpha_j)$ (vedi ad esempio Fletcher, 1987). Tuttavia, per semplicità di trattazione, si considerano i valori di $\alpha_j = \alpha_1, \alpha_1/2, \alpha_1/4, \alpha_1/8, \dots$, nell'ordine evidenziato (cioè, si sceglie il valore più grande di $\alpha_{j+1} = \frac{\alpha_1}{2^j}$, $j \geq 0$) tale che C1 sia soddisfatta. Questa procedura (indicata come ricerca lineare con backtracking) può essere illustrata per la funzione (46) e la condizione (47) come segue.

Esempio 12. Assumendo che il valore di $y(\alpha)$ venga definito secondo la condizione (46), applicare la ricerca lineare con backtracking per trovare un valore di α che soddisfa le condizioni C1a e C1b.

Soluzione.

- Considerando C1b e $\alpha = 1$, si trova $y(1) \approx 0,9697 < 1 - 1/40 = 0,9750$. Quindi C1b è verificata e $\alpha = 1$ soddisfa questa condizione.
- Adesso si vuole determinare un valore di α che soddisfa C1a.

1. si considera $\alpha = 1$ e si sostituisce tale valore in C1a, si ottiene che $y(1) = 0,9697 \not\leq 1 - 0,25$. Quindi, la C1a non è verificata e
2. si considera $\alpha = 1/2$ e si determina $y(1/2) = 1,0346 > 0,875$. Quindi, la C1a non è ancora verificata e
3. si pone $\alpha = 1/4$ e si determina $y(1/4) = 0,9828 > 0,9375$ e allora
4. si considera $\alpha = 1/8$ affinché C1a diventi $y(1/8) = 0,9513 < 0,96875$. Quindi C1a è verificata per $\alpha = 1/8$.

Dal momento che i valori di α che soddisfano la condizione C1 possono variare notevolmente, al fine di limitare tale fenomeno viene imposto anche il soddisfacimento della seguente condizione:

$$\text{C2: } \quad y'(\alpha) \geq \sigma y'(0), \quad (48)$$

dove $\sigma \in (\rho, 1)$ e la retta associata a tale condizione ha una pendenza negativa (indicata con σ -pendenza). Il test sulla pendenza insieme con la condizione C1 (condizione di Wolfe-Powell, indicata con WP) consente di

determinare un intervallo di valori accettabili di α , secondo quanto evidenziato nell'algoritmo 3, riportato in quanto segue. In pratica, generalmente si sceglie $\sigma = 0,9$.

Per illustrare meglio tale procedura, si considera la funzione (46), C1b determinata secondo la C1, scegliendo $\rho = 0,025$, e C2b calcolata in base alla C2 con $\sigma = 0,25$, si ottiene

$$\text{C1b: } y(\alpha) \leq 1 - \frac{1}{40}\alpha \quad \text{e} \quad \text{C2b: } y'(\alpha) \geq -\frac{1}{4}, \quad (49)$$

rispettivamente. Rappresentando la retta associata alla C2b (cioè, la σ -retta) in Figura 7, si ottiene la Figura 8. Da notare che i valori accettabili di α appartengono agli intervalli $(0,0749; 0,2794)$ e $(0,7779; 1,0220)$. L'ultimo intervallo è lo stesso di quello ottenuto per C1, in quanto è possibile osservare che la pendenza $y'(\alpha) = \sigma y'(0) = -0,25$ in corrispondenza di $\alpha = 0,7452$ con $y(0,7452) = 0,9884 > 0,9806 = f(0,7779)$.

Generalmente, si sceglie un valore di α sufficientemente vicino al minimizzatore α^* , imponendo il soddisfacimento delle condizioni forti di Wolfe-Powell ovvero della condizione C1 e

$$\text{C2s: } |y'(\alpha)| \leq -\sigma y'(0), \quad (50)$$

che implica la C2. Le rette associate alla C2s hanno pendenze negative e positive (indicate come σ -pendenza e $-\sigma$ -pendenza, rispettivamente).

In corrispondenza di un valore sufficientemente piccolo di σ , i valori di α che soddisfano la condizione C2s sono sufficientemente prossimi al minimizzatore α^* e in corrispondenza di $\sigma = 0$ si ottiene la condizione di ricerca lineare esatta (45).

Al fine di rappresentare le condizioni introdotte, si considera il caso in cui $\sigma = 0,25$ e si rappresentano graficamente le rette associate alla condizione C2s (cioè, le rette σ e $-\sigma$) in Figura 8, ottenendo la Figura 9. I valori accettabili di α appartengono agli intervalli $(0,0749; 0,1770)$ e $(0,7779; 1,0221)$. È possibile notare che l'ultimo intervallo è lo stesso ottenuto utilizzando la condizione C1, per le considerazioni introdotte in precedenza e $y'(\alpha) = -\sigma y'(0) = 0,25$ in corrispondenza di $\alpha = 1,0256$ con $y(1,0256) = 0,9753 > 0,9744 = f(1,0221)$.

Si noti che la ricerca lineare con backtracking non necessariamente produce un valore di α che soddisfa le condizioni C2 oppure C2s. Pertanto, sono stati proposti alcuni algoritmi che consentono di calcolare valori accettabili per queste condizioni (si veda ad esempio Fletcher, 1987).

In quanto segue, verranno presentati alcuni semplici algoritmi per determinare il valori di α . In particolare, modificando opportunamente l'Algoritmo 2 si ottiene l'Algoritmo 3, riportato in quanto segue, che consente di determinare un valore di α tale da soddisfare le condizioni WP, C1 e C2.

Algoritmo 3. *Procedura per risolvere il sottoproblema di ricerca lineare per le condizioni C1 e C2*

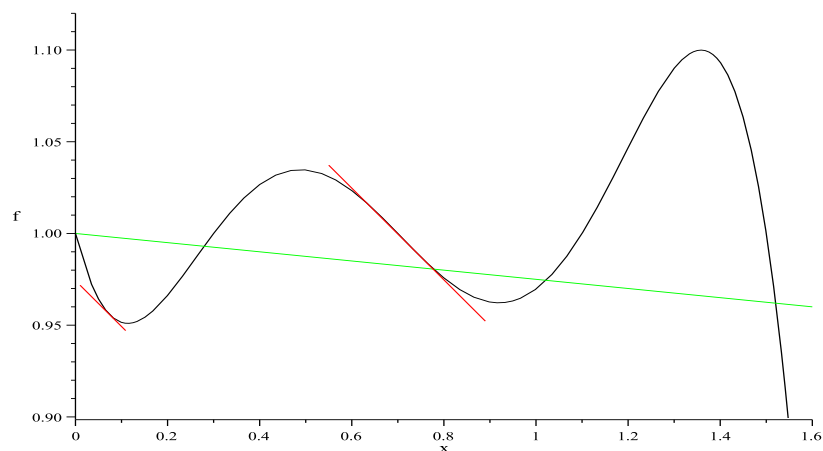


Figura 8: Funzione (46), ρ -retta con $\rho = 0,025$ e σ -pendenza con $\sigma = 0,25$

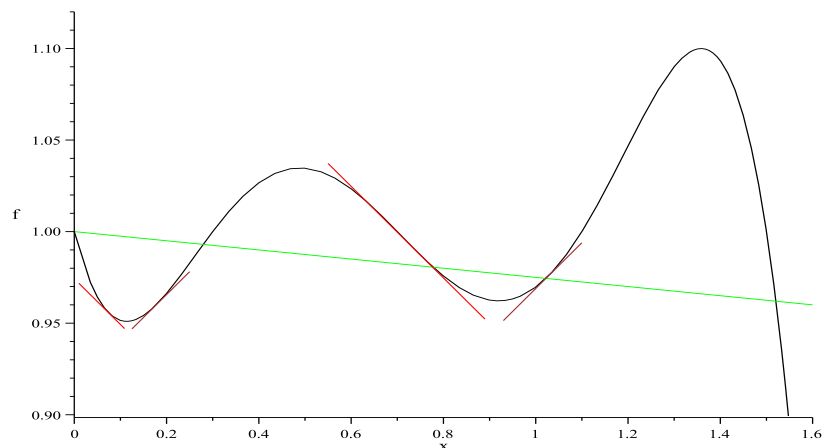


Figura 9: Funzione (46), ρ -retta con $\rho = 0,025$ e pendenze σ e $-\sigma$ con $\sigma = 0,25$

Step 0: Dati $y(0), y'(0) < 0$, $\rho < 0,5$ e $\sigma \in (\rho, 1)$, poni $a_1 = 0$, calcola b_1 in base alla condizione (51) e poni $j = 1$,

Step 1: scegli $\alpha = \alpha_j \in (a_j, b_j)$,
se $C1$ è soddisfatta allora vai allo Step 2,
altrimenti poni $b_{j+1} = \alpha_j$ e vai allo Step 3,

Step 2: calcola $y'(\alpha) = y'(\alpha_j)$ e verifica il soddisfacimento della condizione $C2$,
se $C2$ è soddisfatta allora l'algoritmo ha termine con un valore di α che soddisfa entrambe le condizioni $C1$ e $C2$,
altrimenti poni $a_{j+1} = \alpha_j$, poichè $y'(\alpha_j) < 0$,

Step 3: poni $j = j + 1$, nota che $y'(a_j) < 0$, e vai allo Step 1.

L'ideale sarebbe porre $b_1 = \infty$, come nell'Algoritmo 2, ma in questo caso l'Algoritmo 3 non ha termine se la disuguaglianza $y'(\alpha_j) < 0$ è soddisfatta per tutti gli indici j (ad esempio, si consideri la funzione (46) e si scelga $\alpha_1 = 2$). Quindi, si assume che la funzione $y(\alpha)$ sia limitata inferiormente da qualche valore \bar{f} (per esempio, $\bar{f} = -10^{99}$ se non è disponibile una stima per tale valore). Questa scelta garantisce che la condizione C1 sia soddisfatta per uguaglianza, per $\alpha = b_1$ e $y(b_1) = \bar{f}$, può essere, quindi, risolta per b_1 ottenendo:

$$b_1 = \frac{\bar{f} - y(0)}{\rho y'(0)} \quad (51)$$

(vedi Fletcher, 1987). In tal modo, l'Algoritmo 3 ha termine in corrispondenza dello Step 1, se $y(\alpha)$ non ammette un minimizzatore nell'intervallo $(0, b_1)$, altrimenti, ha termine nello Step 2 con un valore di α che soddisfa le condizioni C1 e C2. È inoltre interessante notare che, nel primo caso, il valore scelto per $\alpha_j \in (a_j, b_j)$ non dovrebbe essere molto diverso da a_j (per esempio, $\alpha_j \leq 2a_j$).

In quanto segue, viene evidenziato in che modo lo Step 2 nell'Algoritmo 3 può essere modificato per determinare un valore di α tale che le condizioni WP forti, C1 e C2s siano verificate.

Algoritmo 4. *Procedura per risolvere il sottoproblema di ricerca lineare per le condizioni C1 e C2s*

Step 0: *Dati $y(0)$, $y'(0) < 0$, $\rho < 0,5$ e $\sigma \in (\rho, 1)$, poni $a_1 = 0$, calcola b_1 secondo la (51) e poni $j = 1$,*

Step 1: *scegli $\alpha = \alpha_j \in (a_j, b_j)$,
se C1 è verificata allora vai allo Step 2
altrimenti poni $b_{j+1} = \alpha_j$ e vai allo Step 3,*

Step 2: *calcola $y'(\alpha) = y'(\alpha_j)$ e verifica il soddisfacimento della C2s,
se C2s è verificata allora termina è stato determinato un valore di α
che soddisfa C1 e C2s
altrimenti se $y'(\alpha_j) > 0$, cioè C2 è verificata, allora poni $b_{j+1} = \alpha_j$,
altrimenti, cioè $y'(\alpha_j) < 0$, allora poni $a_{j+1} = \alpha_j$,*

Step 3: *poni $j = j + 1$, si noti che $y'(a_j) < 0$, e vai allo Step 1.*

Al-Baali e Fletcher (vedi ad esempio Fletcher, 1987) forniscono alcuni algoritmi più generali e dimostrano la loro terminazione in un numero finito di operazioni. Inoltre, suggeriscono alcuni schemi per ottenere un valore adeguato per $\alpha_j \in (a_j, b_j)$, per $j > 1$, allo scopo di trovare un minimizzatore di una funzione quadratica, che interpola i valori di $y(a_j), y'(a_j)$ e $y(b_j)$, oppure di una funzione cubica che interpola questi valori e quello relativo a $y'(b_j)$. Tuttavia, per semplicità, si sceglie $\alpha_j = \frac{1}{2}(a_j + b_j)$, sebbene il numero di valutazioni della funzione potrebbe essere incrementato per ottenere un valore di α accettabile, come evidenziato in precedenza.

Esempio 13. *Assumendo che $y(\alpha)$ sia stata determinata in base alla condizione (46), trovare un valore di α che soddisfa le condizioni WP, C1 e C2s, con $\rho = 0,025$ e (i) $\sigma = 0,25$ e (ii) $\sigma = 0,1$.*

Soluzione. Caso (i): Utilizzando tali valori nelle condizioni C1 e C2, si determinano i valori relativi alla relazioni C1b e C2b, rispettivamente, specificati nella (49). Considerando la soluzione dell'Esempio 12, si osserva che $\alpha = 1$ soddisfa la C1b. Per valutare il soddisfacimento della condizione C2, si calcola $y'(1) = 0,1890$ e si sostituisce in C2s, si ottiene $|y'(1)| = |0,1890| \leq 0,25$. Quindi, $\alpha = 1$ è accettabile per le condizioni WP forti nel caso (i).

Caso(ii): C1 non cambia, mentre se si sostituisce $\sigma = 0,1$ e $y'(0) = -1$ in C2s, si ottengono le condizioni

$$\text{C1b : } y(\alpha) \leq 1 - 0,025\alpha, \quad \text{C2sc : } 0,1 \geq y'(\alpha) \geq -0,1. \quad (52)$$

Come nel caso (i), si osserva che $\alpha = 1$ soddisfa C1b. Tuttavia, C2sc diventa $|y'(1)| = |0,1890| \not\leq 0,1$.

Quindi il valore $\alpha = 1$ non soddisfa C2sc e si deve cercare un altro valore come indicato in quanto segue. Poichè $y'(1) > 0$ e $y'(0) < 0$, si sceglie un valore di $\alpha \in (0,1)$ (dopo aver scelto un valore, si controlla sempre il soddisfacimento di C1b prima di validare C2s) come segue. Analogamente a quanto fatto per il caso (i) dell'Esempio12,

1. si sceglie $\alpha = 1/2$ e si determina $y(1/2) = 1,0346 > 1 - 0,025(\frac{1}{2}) = 0,9875$. La C1b non è verificata, quindi
2. si sceglie $\alpha = 1/4$ e si valuta $y(1/4) = 0,9828 \leq 1 - 0,025(\frac{1}{4}) = 0,9938$. La C1b è verificata per $\alpha = 1/4$, che quindi rappresenta un valore è accettabile per C1b.
3. Sostituendo questo valore in C2sc, si ottiene $y'(1/4) = 0,3484 > 0,1$, quindi la condizione C2s non risulta essere verificata.

Quindi il valore di $\alpha = 1/4$ non soddisfa la C2sc e si determina un altro valore. In particolare, poichè $y'(1/4) > 0$ e $y'(0) < 0$, si sceglie un valore di $\alpha \in (0,1/4)$ in modo tale che C1b sia soddisfatta.

Quindi si sceglie $\alpha = 1/8$ e si ottiene $y(1/8) = 0,9513 \leq 1 - 0,025(\frac{1}{8}) = 0,9969$; pertanto il valore $\alpha = 1/8$ soddisfa la C1b. Sostituendo tale valore in C2sc, si ottiene $|y'(1/8)| = |0,0587| < 0,1$. Pertanto, $\alpha = 1/8 = 0,125$ soddisfa le condizioni WP forti, C1 e C2s.

È opportuno notare che nonostante altri valori di α , che soddisfano le condizioni C1b e C2sc, siano presenti in prossimità del primo valore scelto per α , ovvero $\alpha = 1$ (vedi la Figura 9), i valori disponibili di $y'(0)$ e $y'(1)$ non consentono di individuarli. Infatti, se si sceglie $\alpha = 0,9$, si determina $y(0,9) = 0,9626$ e $y'(0,9) = -0,0415$ che soddisfano entrambe le condizioni C1b e C2sc nel caso (ii).

Esercizio.

1. Assumendo che $y(\alpha)$ sia dato da (46), trovare un valore accettabile di α per le condizioni C1 e C2s con $\rho = 0,25$ e $\sigma = 0,9$, rispettivamente.
2. Sia $y(\alpha)$ dato da (4) con x sostituito da α . Applicare l'Algoritmo 4 con $\rho = 10^{-7}$, $\sigma = 0,01$ e $\bar{f} = 0$. Usare $\alpha_1 = 0,6$ e per $j \geq 1$, $\alpha_{j+1} = 2a_j$ se $y'(a_j) < 0$ e $b_j \gg a_j$, altrimenti $\alpha_{j+1} = \frac{1}{2}(a_j + b_j)$.
3. Usare la ricerca lineare con backtracking per trovare una soluzione approssimata per il problema

$$\min y(\alpha) = \alpha^4 - 1,6\alpha^3 + 2,5\alpha^2 - 6\alpha + 1.$$

4. Determina in maniera non esatta una soluzione per il problema

$$\min y(\alpha) = 100\alpha^4 + (\alpha - 1)^2$$

che soddisfa le condizioni WP.

Per applicare le strategie di soluzione del sottoproblema di ricerca lineare, in corrispondenza del problema di ottimizzazione (1), introdotte in precedenza, si usano le relazioni (39) per y e (42) per $y'(\alpha)$ per un qualsiasi scalare α ; di conseguenza si ricava che

$$y(0) = f^{(k)}, \quad y'(0) = s^{(k)T} g^{(k)}. \quad (53)$$

$$y(\alpha^{(k)}) = f^{(k+1)}, \quad y'(\alpha^{(k)}) = s^{(k)T} g^{(k+1)}. \quad (54)$$

Quindi, la proprietà di discesa (27) e la condizione di ricerca lineare esatta (indicata con RLE) possono essere scritte, rispettivamente, come

$$y'(0) < 0 \quad \Leftrightarrow \quad s^{(k)T} g^{(k)} < 0, \quad (55)$$

$$y'(\alpha^{(k)}) = 0 \quad \Leftrightarrow \quad s^{(k)T} g^{(k+1)} = 0. \quad (56)$$

Inoltre, le condizioni di ricerca lineare definite in precedenza (28), (47), (48) e (50) (indicate con C0, C1, C2 e C2s) possono essere scritte, rispettivamente, nel seguente modo.

$$y(\alpha^{(k)}) < y(0) \quad \Leftrightarrow \quad f^{(k+1)} < f^{(k)} \quad (57)$$

$$y(\alpha^{(k)}) \leq y(0) + \rho\alpha^{(k)}y'(0) \quad \Leftrightarrow \quad f^{(k+1)} \leq f^{(k)} + \rho\alpha^{(k)}s^{(k)T}g^{(k)} \quad (58)$$

$$y'(\alpha^{(k)}) \geq \sigma y'(0) \quad \Leftrightarrow \quad s^{(k)T}g^{(k+1)} \geq \sigma s^{(k)T}g^{(k)} \quad (59)$$

$$|y'(\alpha^{(k)})| \leq -\sigma y'(0) \quad \Leftrightarrow \quad |s^{(k)T}g^{(k+1)}| \leq -\sigma s^{(k)T}g^{(k)}, \quad (60)$$

dove $\rho \in (0, 1/2)$ e $\sigma \in (\rho, 1)$.

È importante notare che se la funzione è quadratica e assume la forma (18), allora la soluzione esatta del sottoproblema di ricerca lineare $\min_{\alpha} q(x^{(k)} + \alpha s^{(k)})$ è data da

$$\alpha^{(k)} = -\frac{s^{(k)T}g^{(k)}}{s^{(k)T}Gs^{(k)}} \quad (61)$$

che è un valore positivo, poichè G è definita positiva, e soddisfa la condizione (56). È, inoltre, uguale a 1 se $Gs^{(k)} = -g^{(k)}$ (direzione di Newton). Questo risultato si ricava dalla condizione (22), nella quale s e \hat{g} sono sostituiti da $s^{(k)}$ e $g^{(k)}$, rispettivamente. Alternativamente, può essere determinato ponendo $x^+ = x + \alpha s$ e scrivendo (56) come

$$\begin{aligned} 0 = s^T g^+ &= s^T (Gx^+ + b) \\ &= s^T (Gx + \alpha Gs + b) \\ &= s^T (g + \alpha Gs) \\ &= g^T s + \alpha s^T Gs. \end{aligned}$$

Risolvendo tale equazione per α , ne consegue che $\alpha^* = -\frac{g^T s}{s^T Gs}$ che equivale a (61).

Nel caso di una funzione generica, un valore di α che soddisfa le condizioni C1, C1 e C2 o C1 e C2s può essere determinato, applicando gli algoritmi 2, 3 o 4, rispettivamente, a ogni iterazione k di un metodo di ottimizzazione che genera una sequenza di punti tali che $f^{(k+1)} < f^{(k)}$. In questo caso, Fletcher (1987) usa i valori di queste funzioni e $y'(0) = s^{(k)T} g^{(k)} < 0$ per definire

$$\alpha_1 = \min \left(1, -2 \frac{\Delta f^{(k)}}{s^{(k)T} g^{(k)}} \right), \quad (62)$$

dove $\Delta f^{(k)} = f^{(k)} - f^{(k+1)}$, per $k > 1$, and $\Delta f^{(1)} = f^{(1)}$. Tuttavia, per semplicità, in quanto segue si pone $\alpha_1 = 1$.

Di seguito viene riportato un semplice esempio per illustrare il ruolo del sottoproblema di ricerca lineare nel problema di ottimizzazione (1).

Esempio 14. *Sia f una funzione quadratica*

$$f(x) = 2x_1^2 + x_2^2 + 2x_1x_2 + x_1 - x_2, \quad (63)$$

dove $x^{(1)} = (-1, 1)^T$. *Mostrare che la direzione di ricerca $s^{(1)} = (1, 1)^T$ è accettabile e usare tale direzione per trovare un nuovo punto $x^{(2)}$, usando (i) la ricerca lineare esatta, (ii) la ricerca lineare non esatta basata su C1 con $\rho = 10^{-4}$, (iii) le condizioni WP con ρ definito come in (ii) e $\sigma = 0,9$ e (iv) le condizioni forti WP con ρ e σ definiti come in (iii).*

Soluzione. Si noti che

$$g(x) = \begin{pmatrix} 4x_1 + 2x_2 + 1 \\ 2x_1 + 2x_2 - 1 \end{pmatrix}.$$

Rimuovendo l'apice ⁽¹⁾, si ha che

$$s^T g = (1 \ 1) \begin{pmatrix} -1 \\ -1 \end{pmatrix} = -2 < 0.$$

Quindi, la PD è soddisfatta e s è una direzione di ricerca accettabile. Per trovare un nuovo punto, si pone

$$x^+ = x + \alpha s = \begin{pmatrix} -1 + \alpha \\ 1 + \alpha \end{pmatrix},$$

ne segue che

$$g^+ = \begin{pmatrix} -1 + 6\alpha \\ 4\alpha - 1 \end{pmatrix}.$$

In quanto segue vengono descritti due modi diversi per determinare un nuovo punto.

Caso (i): Ricerca lineare esatta. Risolvendo l'equazione della ricerca lineare

$$s^T g^+ = 10\alpha - 2 = 0$$

esattamente, si ottiene $\alpha^* = 1/5$. Quindi, si pone $\alpha^{(1)} = 1/5$ e x^+ diventa

$$x^{(2)} = \begin{pmatrix} -1 + 1/5 \\ 1 + 1/5 \end{pmatrix} = \begin{pmatrix} -0,8 \\ 1,2 \end{pmatrix}.$$

(Si noti che $f^{(2)} = -1,2 < f^{(1)} = -1$ e $g^{(2)} = (0,2, -0,2)^T$.)

Caso (ii): Si considera una ricerca lineare non esatta basata sulla condizione C1. È necessario trovare $\alpha^{(1)}$ che soddisfa la condizione (58) con $k = 1$ e $\rho = 10^{-4}$, ovvero

$$C1 : \quad f^+ \leq -1 - 0,0002\alpha,$$

è soddisfatta. Quindi, si procede nel modo seguente.

1. Si pone $\alpha = 1$, allora $x^+ = x + s = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$, $f^+ = 2$

2. Poichè $f^+ > f = -1$, si pone $\alpha = 1/2$,

$$x^+ = x + 0,5s = \begin{pmatrix} -0,5 \\ 1,5 \end{pmatrix}, \quad f^+ = -0,75$$

3. Poichè $f^+ > -1$, si pone $\alpha = 1/4$,

$$x^+ = x + 0,25s = \begin{pmatrix} -0,75 \\ 1,25 \end{pmatrix}, \quad f^+ = -1,1875$$

4. Poichè $-1,1875 < -1$, C0 è soddisfatta. Quindi sostituendo in C1, segue che

$$-1,1875 \leq -1 - 0,0002\left(\frac{1}{4}\right) = -1,0001.$$

Poichè è soddisfatta, si pone $\alpha^{(1)} = 1/4$ e x^+ è accettabile, si definisce

$$x^{(2)} = \begin{pmatrix} -0,75 \\ 1,25 \end{pmatrix}, \quad f^{(2)} = -1,1875.$$

(Si noti che $g^{(2)} = (0,5, 0)^T$, quindi $x^{(2)}$ non è un minimizzatore per f .)

Caso(iii): Si considera una ricerca lineare non esatta basata sulle condizioni WP C1 e C2. Usando il valore accettabile per C1, dato in (ii), ovvero $x^+ = (-0,75, 1,25)^T$ per $\alpha = 1/4$, si determina $g^+ = (0,5, 0)^T$. Sostituendo tale valore, $s = (1,1)^T$ e $g = (-1,-1)^T$ in (59), si ottiene $s^T g^+ = 0,5 \geq -1,8 = \sigma s^T g$. Quindi, la C2 è soddisfatta e $\alpha^{(2)}$ e $x^{(2)}$ sono i valori determinati nel caso (ii).

Caso (iv): Si considera una ricerca lineare non esatta basata sulle condizioni WP, C1 e C2s. Come nel caso (iii), ma considerando la (60), si ottiene $|s^T g^+| = 0,5 \leq 1,8 = -\sigma s^T g$. Quindi, la C2s è soddisfatta e $\alpha^{(2)}$ e $x^{(2)}$ sono i valori determinati nel caso (ii).

Osservazioni sull'esempio precedente.

- Il valore di $x^{(2)}$ può essere trovato considerando $y(\alpha) = f^+$, ovvero sostituendo $x^+ = (-1 + \alpha, 1 + \alpha)^T$ in (63) ottenendo

$$y(\alpha) = 2(-1 + \alpha)^2 + (1 + \alpha)^2 + 2(\alpha^2 - 1) - 2.$$

Di conseguenza,

- quando si applica la ricerca lineare esatta, si risolve $y'(\alpha) = 0$ e si ottiene come sopra $\alpha^* = 1/5$
- quando si applica la ricerca lineare non esatta, la condizione C1 può essere scritta come

$$y(\alpha) \leq -1 - 0,0002\alpha.$$

Si valuta, quindi, questa condizione scegliendo $\alpha = 1, 1/2, 1/4$. Si determina che $\alpha = 1/4$ è il valore accettabile, come nel caso precedente.

- Da notare che la funzione è quadratica, perchè può essere scritta nella forma (18), cioè $f(x) = \frac{1}{2}x^T Gx + b^T x$, dove $G = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$ e $b = (1, -1)^T$. Quindi, sostituendo la matrice Hessiana, $s^{(1)} = (1, 1)^T$ e $g^{(1)} = (-1, -1)^T$ in (61), per $k = 1$, segue che $\alpha^{(1)} = \frac{1}{5}$ come nel caso precedente.

Esercizio. Sia f una funzione di Rosenbrock (7) e sia $x^{(1)} = 0$. Trovare un nuovo punto $x^{(2)}$ basato sulla condizione C0, applicando una ricerca lineare non esatta e usando come direzione di ricerca (i) $s^{(1)} = (2, 2)^T$ e (ii) $s^{(1)} = (1, 0)^T$.

3 Metodo della Discesa più Ripida o Metodo del Gradiente

In questa sezione, viene presentato il metodo della discesa più ripida (o metodo del gradiente) in cui la direzione di ricerca in ogni iterazione è determinata

come segue:

$$s^{(k)} = -g^{(k)}. \quad (64)$$

Tale direzione è accettabile, in quanto

$$g^{(k)T} s^{(k)} = -\|g^{(k)}\|_2^2 < 0$$

, quindi, la proprietà di discesa (27) è soddisfatta in tutte le iterazioni e, di conseguenza, il sottoproblema di ricerca lineare (38)-(39) può essere risolto per valori di $\alpha^{(k)} > 0$.

La direzione di ricerca sopra indicata è ottenuta nel seguente modo. Usando lo sviluppo in serie di Taylor (23), in cui δ viene sostituito da una direzione di ricerca non nota s e ignorando il terzo termine quadratico, si ottiene l'approssimazione lineare di f ; in tal modo il problema di ottimizzazione (1) diventa

$$\min_s f(x) = f(x^{(k)} + s) \approx f^{(k)} + g^{(k)T} s, \quad s = x - x^{(k)}.$$

Dal momento che questo problema non ha soluzione, si normalizza $g^{(k)}$ e s , ottenendo

$$\min_s f(x) = f(x^{(k)} + s) \approx f^{(k)} + \frac{g^{(k)T} s}{\|g^{(k)}\| \|s\|}, \quad s = x - x^{(k)}.$$

Poichè per $\cos(g^{(k)}, s) = \frac{g^{(k)T} s}{\|g^{(k)}\| \|s\|}$, il più piccolo valore possibile è dato da $\cos(g^{(k)}, s) = -1$. Risolvendo questa equazione per s , si ottiene la direzione (64), motivo per cui il metodo è detto metodo della discesa più ripida (DR). Nonostante questo metodo sia globalmente convergente, nella pratica la convergenza risulta molto lenta. Una caratteristica di questo metodo con ricerca lineare esatta è data da

$$g^{(k)T} g^{(k+1)} = 0,$$

per ogni k (cioè, i gradienti sono ortogonali).

Si nota che la direzione di ricerca $s^{(1)} = (1, 1)^T$ usata nell'Esempio 14 è uguale alla direzione DR $s^{(1)} = -g^{(1)}$. Quindi, tale esempio illustra una iterazione del metodo DR con le diverse condizioni relative alla ricerca lineare. Tuttavia, nell'esempio seguente vengono presentate nel dettaglio le iterazioni del metodo DR.

Esempio 15. *Applicare il metodo DR con ricerca lineare esatta per approssimare il minimizzatore della funzione*

$$f(x) = 2x_1^2 + x_2^2 + 2x_1x_2 + x_1 - x_2.$$

Partendo da $x^{(1)} = 0$, terminare il metodo quando $\|g^{(k)}\|_\infty \leq 0,01$.

Soluzione. Prima di tutto, si calcola il gradiente della funzione nel seguente modo

$$g(x) = \begin{pmatrix} 4x_1 + 2x_2 + 1 \\ 2x_1 + 2x_2 - 1 \end{pmatrix}.$$

Dal momento che è necessario applicare la ricerca lineare esatta, la funzione è quadratica e la direzione di ricerca è data da $s^{(k)} = -g^{(k)}$, si procede calcolando il passo in base alla condizione (61), che può essere scritta come segue

$$\alpha^{(k)} = \frac{g^{(k)T}g^{(k)}}{g^{(k)T} \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} g^{(k)}}.$$

Sostituendo $x^{(1)} = 0$, si ottiene $g^{(1)} = (1, -1)^T$ e quindi $s^{(1)} = (-1, 1)^T$. Per $k = 1$, si ha

$$\alpha^{(1)} = \frac{(1, -1) \begin{pmatrix} 1 \\ -1 \end{pmatrix}}{(1, -1) \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}} = \frac{1 + 1}{4 - 4 + 2} = 1.$$

quindi $x^{(2)} = x^{(1)} + \alpha^{(1)}s^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $g^{(2)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$.

Poichè $\|g^{(2)}\|_\infty = 1 > 0,01$, si opera come nell'iterazione precedente e si ottiene

$$\alpha^{(2)} = 0,2, \quad x^{(3)} = \begin{pmatrix} -0,8 \\ 1,2 \end{pmatrix}, \quad g^{(3)} = \begin{pmatrix} 0,2 \\ -0,2 \end{pmatrix}.$$

Poichè $\|g^{(k)}\|_\infty > 0,01$, per $k \leq 6$, si reitera fino a che

$$\alpha^{(6)} = 0,2, \quad x^{(7)} = \begin{pmatrix} -0,9920 \\ 1,4880 \end{pmatrix}, \quad g^{(7)} = \begin{pmatrix} 0,0080 \\ -0,0080 \end{pmatrix}.$$

Poichè $\|g^{(7)}\|_\infty = 0,008 < 0,01$, l'algoritmo ha termine con $x^{(7)}$ e $f^{(7)} = -1,2499$ che approssimano il minimizzatore e il minimo di f , rispettivamente.

Si nota che a ogni iterazione $k = 1, 2$, $g^{(k)T}g^{(k+1)} = 0$.

Esempio 16. Risolvere il problema riportato nell'esempio 15, considerando la ricerca lineare non esatta con la condizione C1.

Soluzione. Per ogni k , si ha che $s^{(k)T}g^{(k)} = -g^{(k)T}g^{(k)}$, quindi, la condizione (58) può essere scritta come

$$C1: \quad f^{(k+1)} \leq f^{(k)} - 0,0001\alpha(g^{(k)T}g^{(k)}).$$

Per trovare una soluzione, si sceglie $\alpha = 1, 1/2, 1/4, \dots$. Dall'esempio precedente, si ha $x^{(1)} = 0$ a cui corrisponde $f^{(1)} = 0$ e si trova $g^{(1)} = (1, -1)^T$.

Per $k = 1$, si pone $\alpha = 1$, quindi $x^+ = x + s = (-1, 1)^T$, $f^+ = -1$. Sostituendo in C1, si ottiene $-1 \leq -0,0001(2)$. Dal momento che la condizione risulta soddisfatta, si pone

$$\alpha^{(1)} = 1, x^{(2)} = x^{(1)} + \alpha^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, g^{(2)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, f^{(2)} = -1.$$

Dal momento che $\|g^{(2)}\|_\infty = 1 > 0,01$, si opera come nell'iterazione precedente e si osserva che per $k = 2$ C1 non è soddisfatta per $\alpha = 1, 1/2$, ma per $\alpha = 1/4$, si ottiene

$$C1 : -1,1875 \leq -1 - 0,0001\left(\frac{1}{4}\right)(2) = -1,0001.$$

Dal momento che la condizione risulta soddisfatta, si pone

$$\alpha^{(2)} = 0,25, x^{(3)} = \begin{pmatrix} -0,75 \\ 1,25 \end{pmatrix}, g^{(3)} = \begin{pmatrix} 0,5 \\ 0 \end{pmatrix}, f^{(3)} = -1,1875.$$

L'esecuzione delle altre iterazioni viene lasciata al lettore per esercizio.

Esercizio. Applicare il metodo di discesa più ripida alla funzione quadratica

$$f(x) = u^2, u = x_1^2 + (x_1 - x_2)^2 + x_3^2,$$

partendo dal punto iniziale $x = (0, 0, 1)^T$ e usando la ricerca lineare esatta (i); la la ricerca lineare non esatta (ii).

L'esempio precedente mostra che il metodo del gradiente è lento (per la motivazione, vedere ad esempio Fletcher, 1987, e Nocedal e Wright, 1999). Per esaminare meglio tale comportamento, si considera l'esempio riportato in quanto segue.

Esempio 17. Applicare il metodo DR con la ricerca lineare con backtracking alla funzione di Rosenbrock (7), partendo dal punto iniziale $x^{(1)} = (-1, 2, 1)^T$. Eseguire la prima iterazione e terminare l'esecuzione quando la disuguaglianza $\|g^{(k)}\|_2 \leq 10^{-7}$ risulta verificata.

Soluzione. La funzione di Rosenbrock e il suo gradiente sono dati, rispettivamente, da

$$f(x) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2,$$

e

$$g(x) = \begin{bmatrix} -400x_1(x_2 - x_1^2) + 2(x_1 - 1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

((7) e (11)).

Sostituendo il valore si $x^{(1)}$, si ottiene $f^{(1)} = 24,2$.

Per $k = 1$, si ha

$$g^{(1)} = \begin{pmatrix} -215,6 \\ -88 \end{pmatrix}, s^{(1)} = -g^{(1)} = \begin{pmatrix} 215,6 \\ 88 \end{pmatrix}, g^{(1)T} s^{(1)} \approx -5,4227 \times 10^4 < 0$$

(l'ultimo PD è atteso). Per trovare $\alpha^{(1)}$, si omette l'indicazione dell'apice $^{(1)}$ e si pone

$$x^+ = x + \alpha s = \begin{pmatrix} -1, 2 + 215, 6\alpha \\ 1 + 88\alpha \end{pmatrix}.$$

Si sceglie $\alpha = 1$, si ottiene $f^+ \approx 10^{11} > f^{(1)}$. Quindi, similmente, si reitera la procedura

scegliendo $\alpha = 1/2, 1/4, \dots, 1/2^9 = 1/512$, si ottiene $f^+ \approx 10^{10}, 10^8, 10^7, 10^6, 10^4, 543, 211, 149, 35$, rispettivamente, che sono $> f^{(1)}$, fino a che per $\alpha = \frac{1}{2^{10}} = \frac{1}{1024}$, si trova $f^+ = 5, 1011 < f^{(1)}$. Sostituendo in C1, si ottiene

$$5, 1011 \leq 24, 2 + 0, 0001\left(\frac{1}{1024}\right)(-5, 4227 \times 10^4) \approx 24, 1962.$$

Poichè la condizione è verificata, si pone

$$\alpha^{(1)} = \frac{1}{1024} \approx 9, 7656 \times 10^{-4}, \quad x^{(2)} = \begin{pmatrix} -0, 9895 \\ 1, 0859 \end{pmatrix}, \quad f^{(2)} = 5, 1011.$$

Da notare che il costo computazionale necessario per determinare tale punto ha richiesto 10 valutazioni della funzione (che potrebbero essere ridotte se fosse usata l'interpolazione).

Sono state eseguite altre iterazioni dell'algorithmo considerando $s^{(k)} = -g^{(k)}$, per $k \geq 2$, e usando l'implementazione in MATLAB dello stesso ed è stata usata come condizione di terminazione $\|g^{(k)}\|_2 \leq 10^{-7}$. Il problema è stato risolto con un costo computazionale pari a 16596 ricerche lineari e 165088 valutazioni della funzione.

Questo esempio mostra chiaramente che il metodo del gradiente è molto lento e che potrebbe non riuscire a trovare una soluzione in un numero finito di iterazioni. Pertanto, sono stati considerati altri metodi, illustrati nelle seguenti sezioni.

4 Metodo del Gradiente Coniugato

Nel metodo del gradiente coniugato (GC), la direzione di ricerca è calcolata sulla base delle seguenti condizioni:

$$s^{(k+1)} = \begin{cases} -g^{(1)} & \text{if } k = 1, \\ -g^{(k+1)} + \beta^{(k)} s^{(k)} & \text{if } k \geq 1, \end{cases} \quad (65)$$

dove $\beta^{(k)}$ è un parametro. In particolare, il metodo è definito da una delle seguenti formule:

$$\beta^{(k)} = \frac{\|g^{(k+1)}\|^2}{\|g^{(k)}\|^2} \quad (66)$$

$$\beta^{(k)} = \frac{\gamma^{(k)T} g^{(k+1)}}{\|g^{(k)}\|^2}, \quad (67)$$

dove $\|\cdot\|$ indica $\|\cdot\|_2$ e

$$\gamma^{(k)} = g^{(k+1)} - g^{(k)}. \quad (68)$$

Queste formule sono note come formule di Fletcher-Reeves (FR) e Polak-Ribière (PR), rispettivamente.

Il nome del metodo è dovuto al fatto che le direzioni di ricerca possono essere viste come una versione modificata della direzione del gradiente, che si ottiene ponendo $\beta^{(k)} = 0$, e soddisfano la proprietà coniugata (69) (definita sotto) e la condizione di ricerca lineare esatta (56) ovvero

$$s^{(k)T} g^{(k+1)} = 0.$$

Questa condizione garantisce il soddisfacimento della proprietà di discesa (27) per ogni scelta di $\beta^{(k)}$, come evidenziato in quanto segue.

Per $k = 1$, si nota che $g^{(1)T} s^{(1)} = -\|g^{(1)}\|^2 < 0$ e per $k \geq 1$, segue che

$$\begin{aligned} g^{(k+1)T} s^{(k+1)} &= -g^{(k+1)T} g^{(k+1)} + \beta^{(k)} g^{(k+1)T} s^{(k)} \\ &= -\|g^{(k+1)}\|_2^2 < 0. \end{aligned}$$

Pertanto, esiste una soluzione per il sottoproblema di ricerca lineare (38)-(39) per qualche $\alpha^{(k)} > 0$. In quanto segue, verranno illustrate importanti proprietà dei metodi del gradiente coniugato.

Teorema 4. *Si assume che la funzione obiettivo f sia una funzione quadratica del tipo (18), con matrice Hessiana G simmetrica e definita positiva. La classe di metodi del gradiente coniugato con ricerche lineari esatte termina in un punto stazionario dopo al massimo n iterazioni. Inoltre, per ogni $i \neq j$ valgono le seguenti proprietà:*

$$s^{(i)T} G s^{(j)} = 0, \quad (69)$$

$$g^{(i)T} g^{(j)} = 0, \quad (70)$$

$$s^{(i)T} g^{(i)} = -g^{(i)T} g^{(i)}. \quad (71)$$

Queste proprietà garantiscono, rispettivamente, che le direzioni di ricerca siano coniugate rispetto a G , i gradienti siano ortogonali e la proprietà di discesa sia soddisfatta.

Per la dimostrazione di tale risultati teorici il lettore può fare riferimento ad esempio al contributo di Fletcher, 1987.

Esempio 18. *Applicare il metodo FR con ricerca lineare esatta alla seguente funzione*

$$f(x) = 2x_1^2 + x_2^2 + 2x_1x_2 + x_1 - x_2,$$

considerando come punto iniziale $x^{(1)} = 0$. Si ottiene lo stesso risultato se viene applicato il metodo PR.

Soluzione. Dal momento che il metodo FR sceglie la direzione di ricerca come $s^{(1)} = -g^{(1)}$, esattamente a quanto viene fatto dal metodo DR, si ottiene (vedi Esempio 15)

$$g^{(1)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, s^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \alpha^{(1)} = 1, x^{(2)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, g^{(2)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

Quindi, la formula FR assume la seguente forma

$$\beta_1 = \frac{\|g^{(2)}\|^2}{\|g^{(1)}\|^2} = \frac{(-1)^2 + (-1)^2}{(1)^2 + (-1)^2} = 1,$$

$$s^{(2)} = -g^{(2)} + \beta_1 s^{(1)} = -\begin{pmatrix} -1 \\ -1 \end{pmatrix} + (1) \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

Dal momento che la funzione obiettivo è quadratica, applicando la ricerca lineare esatta si ottiene (dalla formula (61)) che

$$\alpha^{(2)} = -\frac{s^{(2)T} g^{(2)}}{s^{(2)T} G s^{(2)}} = \frac{(0, 2) \begin{pmatrix} -1 \\ -1 \end{pmatrix}}{(0, 2) \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix}} = \frac{1}{4}.$$

Quindi,

$$x^{(3)} = x^{(2)} + \alpha^{(2)} s^{(2)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1,5 \end{pmatrix}.$$

Dal momento che $g^{(3)} = 0$, allora $x^* = (-1, 1,5)^T$. Quindi, il metodo converge alla soluzione in 2 iterazioni. Tale risultato conferma quanto sintetizzato nel teorema introdotto precedentemente. Inoltre, si nota che le direzioni di ricerca $s^{(1)}$ e $s^{(2)}$ sono coniugate rispetto a G .

Anche il metodo PR con ricerca lineare esatta richiede al massimo 2 iterazioni, come evidenziato nel seguito. I valori di $x^{(2)}$ e di $g^{(2)}$ sono uguali a quelli calcolati in precedenza per il metodo FR. Applicando la formula PR si ottiene

$$\beta_1 = \frac{(g^{(2)} - g^{(1)})^T g^{(2)}}{\|g^{(1)}\|^2} = \frac{(-2)(-1) + 0}{(1)^2 + (-1)^2} = 1.$$

Quindi, operando in modo analogo a quanto fatto per l'applicazione del metodo FR, è possibile verificare facilmente che il metodo termina con $x^* = x^{(3)} = (-1, 1,5)^T$.

Confrontando il numero di iterazioni richieste nell'esempio precedente con quelle eseguite nel caso dell'Esempio 15, è evidente che le prestazioni computazionali dei metodi del gradiente coniugato sono migliori rispetto a quelle del metodo DR, nel caso in cui viene applicata la ricerca lineare esatta. Questo comportamento viene generalmente osservato anche quando si utilizza la ricerca lineare non esatta, come illustrato nell'esempio riportato in quanto segue. È utile evidenziare che quando si applica la ricerca lineare

non esatta, si registra un deterioramento delle prestazioni del metodo del gradiente coniugato; tuttavia, questo comportamento non è osservato nel caso dell'esercizio seguente.

Esercizio. Risolvere l'Esercizio 18, usando la ricerca lineare con backtracking e dimostrare che la procedura per risolvere il problema quadratico è identica, tranne che il valore di $\alpha^{(2)} = 1/4$ è ottenuto dalla ricerca lineare con backtracking piuttosto che dalla ricerca lineare esatta.

Nel caso in cui viene utilizzata una ricerca lineare non esatta oppure la funzione obiettivo non è una funzione quadratica, le proprietà evidenziate non sono di solito soddisfatte, tranne nei casi in cui la proprietà di discesa è garantita grazie a una oculata scelta del parametro $\beta^{(k)}$ e, inoltre, risultano soddisfatte adeguate condizioni sulla ricerca lineare. In particolare, per una generica funzione, se vengono applicate le condizioni forti di Wolfe-Powell (58) e (60) con $0 < \rho < \sigma < 1/2$ (generalmente vengono usati i valori $\rho = 10^{-4}$ e $\sigma = 0,1$), allora le direzioni FR soddisfano la PD e il corrispondente metodo converge globalmente. Questo primo risultato sulla convergenza è dovuto a Al-Baali (vedi Fletcher, 1987, e Nocedal e Wright, 1999), mentre il metodo PR può, in casi rari, ciclare indefinitivamente senza convergere.

Tuttavia, il metodo PR, che prevede una fase di rinizializzazione in corrispondenza di un certo numero di iterazioni spesso converge molto più rapidamente del metodo FR. Infatti, per il metodo PR e per i metodi del gradiente coniugato, la PD viene garantita utilizzando, quando si ritiene necessario, come direzione di ricerca quella della discesa più ripida. Questa strategia è illustrata in quanto segue.

Ad ogni iterazione, una volta calcolato $s^{(k+1)}$, si verifica se la disuguaglianza $s^{(k+1)T}g^{(k+1)} < 0$ è soddisfatta; in caso negativo, si pone $s^{(k+1)} = -g^{(k+1)}$ (cioè, si sceglie $\beta^{(k)} = 0$). Tuttavia, sono state proposte altre scelte per il parametro $\beta^{(k)}$ che garantiscono il soddisfacimento della PD e garantiscono la convergenza. In particolare, si considera

$$\beta^{(k)} = \max\left(\frac{\gamma^{(k)T}g^{(k+1)}}{\|g^{(k)}\|^2}, 0\right) \quad (72)$$

(denominato PR⁺) che è stato inizialmente proposto da Powell e considerato da Gilbert e Nocedal, i quali hanno dimostrato che il metodo corrispondente che utilizza le condizioni forti di Wolfe-Powell converge globalmente e generalmente presenta delle buone prestazioni nelle applicazioni pratiche (vedi ad esempio Nocedal e Wright, 1999).

Esempio 19. Applicare i metodi FR e PR con ricerca lineare non esatta con backtracking alla funzione di Rosenbrock (7), partendo dalla stima iniziale $x^{(1)} = (-1, 2, 1)^T$ e terminando l'algoritmo quando $\|g^{(k)}\|_2 \leq 10^{-7}$. Presentare i risultati dettagliati delle prime due iterazioni del metodo FR.

Soluzione. La funzione di Rosenbrock e il suo gradiente sono dati da

$$f(x) = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2,$$

e

$$g(x) = \begin{bmatrix} -400x_1(x_2 - x_1^2) + 2(x_1 - 1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

(cf (7) e (11)), rispettivamente.

Per $k = 1$, il metodo FR determina la direzione DR $s^{(1)} = -g^{(1)}$. Quindi, sfruttando il risultato nell'Esempio 17, è facile verificare che

$$g^{(1)} = \begin{pmatrix} -215,6 \\ -88 \end{pmatrix}, x^{(2)} = \begin{pmatrix} -0,9895 \\ 1,0859 \end{pmatrix}, g^{(2)} = \begin{pmatrix} 38,3380 \\ 21,3840 \end{pmatrix}$$

e $f^{(2)} = 5,1011$. Quindi, applicando la formula di FR formula si ottiene

$$\beta_1 = \frac{\|g^{(2)}\|^2}{\|g^{(1)}\|^2} = 0,0355$$

e

$$s^{(2)} = - \begin{pmatrix} 38,3380 \\ 21,3840 \end{pmatrix} + 0,0355 \begin{pmatrix} 215,6 \\ 88 \end{pmatrix} = \begin{pmatrix} -30,6842 \\ -18,2600 \end{pmatrix}.$$

Nonostante sia possibile verificare che la PD sia verificata, si calcola il valore di $s^{(2)T}g^{(2)} = -1566,8 < 0$, perchè verrà usato successivamente. Si prosegue come evidenziato nel seguito. Per semplificare la notazione, nelle relazioni usate per determinare $\alpha^{(2)}$, si rimuove l'apice (2) e si impone

$$x^+ = x + \alpha s = \begin{pmatrix} -0,9895 - 30,6842\alpha \\ 1,0859 - 18,2600\alpha \end{pmatrix}.$$

Sostituendo $\alpha = 1, 1/2, \dots, 1/2^8 = 1/256$, si ottengono i corrispondenti valori di $f^+ > f^{(2)}$, si reitera la procedura fino al valore $\alpha = \frac{1}{2^9} = \frac{1}{512}$, quando si ottiene $f^+ = 4,4610 < f^{(2)} = 5,1011$. Sostituendo in C1, si ha

$$4,4610 \leq 5,1011 + 0,0001\left(\frac{1}{512}\right)(-1566,8) \approx 5,1008.$$

Poichè la condizione è verificata, si pone

$$\alpha^{(2)} = \frac{1}{512} \approx 0,0020, x^{(3)} = \begin{pmatrix} -1,0494 \\ 1,0502 \end{pmatrix}, f^{(3)} = 4,4610.$$

Da notare che il costo computazionale per trovare tale punto è pari a 9 valutazioni della funzione (che potrebbe essere ridotto utilizzando l'interpolazione).

Utilizzando l'implementazione in MATLAB di questo algoritmo, il problema è stato risolto considerando come direzione di ricerca $s^{(k)}$ definita in base al metodo FR per $k \geq 3$ e terminando l'esecuzione quando

Tabella 1: Applicazioni alla funzione di Rosenbrock (7)

Metodo	nls	nfe
SD	16596	165088
FR	365	4592
PR	1805	46475
PR ⁺	2239	56461

$\|g^{(k)}\|_2 \leq 10^{-7}$ è risultata soddisfatta. Il costo computazionale richiesto per risolvere tale problema viene mostrato in Tabella 4, dove nls and nfe indicano, rispettivamente, il numero di ricerche lineari e di valutazioni delle funzioni che sono state eseguite.

La procedura è stata rieseguita, considerando i metodi PR e PR+ al posto della strategia FR. I risultati ottenuti sono riportati in tabella 4. Questa tabella mostra chiaramente che i metodi GC presentano delle prestazioni notevolmente migliori rispetto al metodo DR, nonostante essi convergano più lentamente alla soluzione. Anche se i risultati ottenuti evidenziano che FR funziona notevolmente meglio di entrambi i metodi PR e PR⁺, questi ultimi due metodi con le condizioni di Wolfe-Powell sono di solito da preferire per cercare la soluzione di un problema generico.

Nonostante questa lenta convergenza, vale la pena notare che i metodi GC sono utili per risolvere problemi su larga scala (n è grande), perchè sono caratterizzati da un uso limitato di memoria. Tuttavia, per un valore sufficientemente piccolo di N , i metodi di Newton e quasi-Newton che vengono presentati, rispettivamente, nelle Sezioni 5 e 6 sono da preferire.

5 Il Metodo di Newton e le sue varianti

In questa sezione, si considera la direzione di ricerca di Newton $s^{(k)}$ definita come segue

$$s^{(k)} = -G^{(k)-1}g^{(k)}, \quad (73)$$

dove $G^{(k)}$ e $g^{(k)}$ sono, rispettivamente, l'Hessiana e il gradiente di f calcolati in $x^{(k)}$. Tale direzione soddisfa la proprietà di discesa (27) se $g^{(k)T}s^{(k)} = -g^{(k)T}G^{(k)-1}g^{(k)} < 0$ che è verificata se $G^{(k)}$ è definita positiva. Di conseguenza, il sottoproblema di ricerca lineare (38)-(39) può essere risolto.

La direzione di ricerca precedente può essere ricavata nel seguente modo. Usando lo sviluppo in serie di Taylor (23) con $\hat{x} = x^{(0)}$, segue che

$$f(x) = f(x^{(0)}) + g(x^{(0)})^T\delta + \frac{1}{2}\delta^T G(x^{(0)})\delta + \dots$$

dove $\delta = x - x^{(0)}$. Quindi, il gradiente nel punto stazionario $x = x^*$ soddisfa la condizione

$$\begin{aligned} 0 = g(x^*) &= g^{(0)} + G^{(0)}(x^* - x^{(0)}) + \dots \\ &\approx g^{(0)} + G^{(0)}(x^* - x^{(0)}) \end{aligned}$$

(che si ricava da (20)-(21) con $\hat{x} = x^{(0)}$ e $\alpha s = x^* - x^{(0)}$). Questa espressione è una relazione approssimata con valori esatti. Quindi, usando un valore approssimato $x^{(1)} \approx x^*$, segue che

$$0 = g^{(0)} + G^{(0)}(x^{(1)} - x^{(0)})$$

l'equazione precedente può essere risolta rispetto a $x^{(1)}$, ottenendo $x^{(1)} = x^{(0)} - G^{(0)^{-1}}g^{(0)}$. Ora sostituendo le vecchie e le nuove stime $x^{(0)}$ e $x^{(1)}$ di x^* con $x^{(k)}$ e $x^{(k+1)}$, rispettivamente, si definisce un metodo iterativo definito sulla base della seguente condizione

$$x^{(k+1)} = x^{(k)} + \delta^{(k)}, \quad \delta^{(k)} = -G^{(k)^{-1}}g^{(k)}.$$

Se si pone $s^{(k)} = \delta^{(k)}$, si definisce il metodo di Newton senza ricerca lineare (cioè, $\alpha^{(k)} = 1$) che produce $x^{(k+1)} = x^{(k)} + s^{(k)}$, dove $s^{(k)}$ è dato da (73). Questa direzione di ricerca è generalmente calcolata risolvendo il seguente sistema di equazioni lineari $G^{(k)}s = -g^{(k)}$ per s . Il metodo di Newton con ricerca lineare è definito analogamente a quanto fatto sopra, ma si pone $\delta^{(k)} = \alpha^{(k)}s^{(k)}$ (cioè, da (29) con $s^{(k)}$ definito come (73)).

È utile osservare che l'applicazione di questo metodo a una funzione quadratica richiede solo una iterazione, poichè nessuna approssimazione è usata nello sviluppo in serie di Taylor di cui sopra. Per una funzione generale, il metodo converge in modo quadratico se $x^{(1)}$ è sufficientemente vicino a x^* (vedi Fletcher, 1987). Queste proprietà sono illustrate negli esempi riportati nel seguito.

Esempio 20. *Si vuole risolvere il problema quadratico dell'Esempio 18, applicando il metodo di Newton senza ricerca lineare.*

Soluzione. In corrispondenza del punto iniziale $x^{(1)} = 0$, si ha che $g^{(1)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ e l'Hessiana $G^{(1)} = G = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$ è costante. Quindi l'esecuzione del metodo di Newton comporta la determinazione della direzione di ricerca

$$\begin{aligned} s^{(1)} &= -G^{(1)^{-1}}g^{(1)} \\ &= -\frac{1}{4} \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1,5 \end{pmatrix} \end{aligned}$$

quindi $x^{(2)} = x^{(1)} + s^{(1)} = (-1, 1,5)^T$ che è uguale a x^* .

Esercizio. Risolvere il problema precedente a partire dalla stima iniziale $x^{(1)} = (1, 2)^T$.

Esempio 21. *Risolvere il problema dell'Esempio 19, applicando il metodo di Newton senza ricerca locale e con ricerca locale con backtracking (indicate con NT0 e NT1, rispettivamente).*

Soluzione. Usando le espressioni del gradiente e dell'Hessiana di f , date da (11) e (12) rispettivamente, si calcola $g^{(k)}$ e $G^{(k)}$. Quindi, la direzione di ricerca (73) soddisfa la PD se $G^{(k)}$ è definita positiva, che accade fortunatamente in tutte le iterazioni di entrambi i metodi. Un nuovo punto $x^{(k+1)}$ è dato da (29) con passo $\alpha^{(k)} = 1$ per il metodo NT0, mentre per il metodo NT1 i valori del passo sono determinati applicando l'Algoritmo 2 ad ogni iterazione del metodo. I due metodi sono stati eseguiti, partendo dal valore standard di $x^{(1)}$, fino al soddisfacimento della seguente condizione di terminazione $\|g^{(k)}\|_2 \leq 10^{-7}$.

Si osserva che entrambi i metodi hanno determinato la soluzione del problema e il numero di ricerche lineari (nls) e valutazioni della funzione (nfe) richieste per risolvere il problema sono $nls/nfe = 6/7$ per NT0 e $nls/nfe = 21/29$ per NT1. Tali valori sono notevolmente più limitati rispetto a quelli ottenuti per i metodi GC, riportati in Tabella 4.

Nonostante NT0 converga alla soluzione del problema sostanzialmente più velocemente di NT1, esso potrebbe fallire nella risoluzione di altri problemi perchè la condizione di decremento della funzione (57) non è garantita. Inoltre, anche se NT1 assicura il soddisfacimento della condizione di decremento della funzione obiettivo (58) e consente di ottenere la migliore prestazione del metodo di Newton, tale metodo potrebbe fallire quando il punto iniziale è sufficientemente lontano da un punto stazionario x^* come mostrato nel seguente esempio.

Esempio 22. *Applicare il metodo di Newton senza ricerca locale al seguente problema*

$$f(x) = x_1^4 + x_1x_2 + (1 + x_2)^2,$$

considerando come stima iniziale (i) $x^{(1)} = (.75, -1, 25)^T$ e (ii) $x^{(1)} = 0$.

Soluzione. Per il caso (i), applicando il metodo di Newton alla funzione precedente, si determina la soluzione in 5 iterazioni con $f^{(5)} = -0,5824$. Tuttavia, per il caso (ii), il metodo fallisce alla prima iterazione, perchè $G^{(1)}$ non è definita positiva.

Una motivazione per questo comportamento è che $x^{(1)}$ nel caso (i) è sufficientemente vicino al minimizzatore $x^* = (0,6959, -1,3479)^T$, mentre questa condizione non è soddisfatta nel caso (ii).

Quindi lo svantaggio del metodo di Newton è che $x^{(1)}$ deve essere scelto sufficientemente vicino a x^* , tale condizione è difficile da garantire nei casi di grande interesse pratico. Nonostante il calcolo dell'Hessiana $G^{(k)}$ di f possa sembrare un ulteriore svantaggio di tale metodo, tale matrice può essere calcolata mediante tecniche di differenziazione automatica oppure usando diversi software disponibili (es, Maple).

Per ovviare a queste difficoltà, si considerano i metodi quasi-Newton, descritti nella successiva sezione. Tuttavia, lo svantaggio descritto è stato superato anche con l'applicazione di diverse tecniche di aggiornamento. Ad esempio, Fletcher (1987), prima di calcolare la direzione di ricerca (73), sostituisce l'Hessiana $G^{(k)}$ con

$$\hat{G}^{(k)} = G^{(k)} + \nu I,$$

dove ν rappresenta il minimo numero intero $\nu = 0, 1, 2, \dots$ tale che la matrice $\hat{G}^{(k)}$ sia definita positiva.

Esempio 23. Partendo dalla stima iniziale $x^{(1)} = 0$, applicare una iterazione del metodo di Newton modificato alla seguente funzione

$$f(x) = x_1^4 - 3x_1x_2 + (2 + x_2)^2.$$

Soluzione. In corrispondenza di $x^{(1)} = 0$, si nota che

$$\hat{G}^{(1)}(\nu) = G^{(1)} + \nu I = \begin{bmatrix} \nu & -3 \\ -3 & 2 + \nu \end{bmatrix}$$

che è indefinita per $\nu \leq 2$ e definita positiva per $\nu = 3$. Quindi, si considera l'ultimo valore di ν per definire la direzione di ricerca come

$$s^{(1)} = -[\hat{G}^{(1)}(3)]^{-1}g^{(1)} = -\frac{1}{6} \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix} \begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \end{pmatrix}.$$

Applicando la ricerca locale con backtracking si determina $\alpha^{(1)} = 1/2$. Quindi, $x^{(2)} = (-1, -1)^T$ e $f^{(2)} = -1$.

Esercizio. Partendo dalla stima iniziale $x^{(1)} = 0$, applicare una iterazione del metodo di Newton modificato alla funzione

$$f(x) = x_1^4 - x_1^2 + x_1x_2 + x_2^2 + x_2.$$

6 Metodi Quasi-Newton

La classe dei metodi quasi-Newton è simile al metodo Newton con ricerca lineare, l'unica differenza risiede nel fatto che l'inversa della Hessiana $G^{(k)-1}$ viene approssimata con una matrice simmetrica $H^{(k)}$ che soddisfa la seguente condizione, detta condizione quasi-Newton (o equazione secante)

$$H^{(k+1)}\gamma^{(k)} = \delta^{(k)}, \quad (74)$$

dove

$$\delta^{(k)} = x^{(k+1)} - x^{(k)}, \quad \gamma^{(k)} = g^{(k+1)} - g^{(k)}. \quad (75)$$

Quindi, la direzione di ricerca è definita come segue

$$s^{(k)} = -H^{(k)}g^{(k)} \quad (76)$$

e soddisfa la proprietà di discesa (27) se $H^{(k)}$ è definita positiva.

La matrice iniziale $H^{(1)}$ è scelta in modo da essere una matrice definita positiva (in genere, $H^{(1)} = I$) e, a ogni iterazione, la corrente matrice $H^{(k)}$ definita positiva viene aggiornata ad una nuova approssimazione $H^{(k+1)}$. La prima formula di approssimazione proposta è nota in letteratura come aggiornamento inverso DFP. Essa è definita mediante la relazione (36) o, in modo equivalente, da

$$H_{DFP}^{(k+1)} = \left\{ H - \frac{H\gamma\gamma^T H}{\gamma^T H\gamma} + \frac{\delta\delta^T}{\delta^T\gamma} \right\}^{(k)}. \quad (77)$$

Altre due formule di aggiornamento inverso, note in letteratura, sono la BFGS e la SR1, definite rispettivamente come

$$H_{BFGS}^{(k+1)} = \left\{ H + \left(1 + \frac{\gamma^T H\gamma}{\delta^T\gamma} \right) \frac{\delta\delta^T}{\delta^T\gamma} - \frac{\delta\gamma^T H + H\gamma\delta^T}{\delta^T\gamma} \right\}^{(k)} \quad (78)$$

e

$$H_{SR1}^{(k+1)} = \left\{ H + \frac{(\delta - H\gamma)(\delta - H\gamma)^T}{(\delta - H\gamma)^T\gamma} \right\}^{(k)}. \quad (79)$$

Si osserva che questi aggiornamenti sono simmetrici e soddisfano la condizione quasi-Newton (74). La ragione per imporre tale condizione può essere giustificata sulla base delle seguenti relazioni. Applicando l'espansione dello sviluppo in serie di Taylor a $g^{(k+1)}$ nel punto $x^{(k)}$, si ha che

$$\begin{aligned} \gamma^{(k)} &= g(x^{(k)} + \delta^{(k)}) - g^{(k)} \\ &= g^{(k)} + G^{(k)}\delta^{(k)} + \dots - g^{(k)} \\ &\approx G^{(k)}\delta^{(k)}. \end{aligned}$$

Approssimando $G^{(k)-1}$ con $H^{(k+1)}$, questa relazione di approssimazione porta alla condizione (74).

È possibile mostrare che gli aggiornamenti BFGS e DFP garantiscono che le approssimazioni dell'Hessiana siano definite positive se la condizione di curvatura

$$\delta^{(k)T}\gamma^{(k)} > 0 \quad (80)$$

è verificata, mentre per l'aggiornamento SR1 deve essere verificata la seguente condizione:

$$\delta^{(k)T}\gamma^{(k)} > \min(\delta^{(k)T}B^{(k)}\delta^{(k)}, \gamma^{(k)T}H^{(k)}\gamma^{(k)}). \quad (81)$$

Poichè l'aggiornamento SR1 presenta, in alcuni casi, prestazioni migliori rispetto all'aggiornamento BFGS, è conveniente considerare il seguente aggiornamento combinato BFGS/SR1

$$H_{sw}^{(k+1)} = \begin{cases} H_{SR1}^{(k+1)} & \text{se } \delta^{(k)T}\gamma^{(k)} > \gamma^{(k)T}H^{(k)}\gamma^{(k)}, \\ H_{BFGS}^{(k+1)} & \text{altrimenti.} \end{cases} \quad (82)$$

Esempio 24. Risolvere il problema quadratico dell'Esempio 18, applicando il metodo DFP con $H^{(1)} = I$ e ricerca lineare esatta.

Soluzione. Poichè $H^{(1)} = I$, la direzione di ricerca (76) conduce alla direzione di ricerca più ripida $s^{(1)} = -g^{(1)}$. Quindi, come nell'Esempio 15 e (18), si ottiene

Per $k = 1$, $x^{(1)} = 0$,

$$g^{(1)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, s^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \alpha^{(1)} = 1, x^{(2)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, g^{(2)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

Per calcolare $H^{(2)}$ con (77), al fine di semplificare la notazione, si elimina l'apice $^{(1)}$, si sostituisce $^{(2)}$ con $+$ e si calcola

$$\delta = x^+ - x = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \gamma = g^+ - g = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

Dal momento che $\delta^T \gamma = 2 > 0$, la matrice DFP aggiornata è definita positiva e viene determinata come segue. Si calcola $\gamma^T H \gamma = 4$ e si sostituisce in (77), ottenendo

$$H^{(2)} = I - \frac{1}{4} \begin{pmatrix} -2 \\ 0 \end{pmatrix} (-2 \ 0) + \frac{1}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix} (-1 \ 1) = \begin{bmatrix} 0,5 & -0,5 \\ -0,5 & 1,5 \end{bmatrix}.$$

che è definita positiva.

Per $k = 2$,

$$s^{(2)} = -H^{(2)} g^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \alpha^{(2)} = 0,5, x^{(3)} = \begin{pmatrix} -1 \\ 1,5 \end{pmatrix}, g^{(3)} = 0.$$

Quindi, $x^* = x^{(3)}$. Quindi, il metodo DFP con ricerca lineare esatta converge alla soluzione in $n = 2$ iterazioni. Si può, inoltre, notare che le direzioni di ricerca $s^{(1)}$ e $s^{(2)}$ sono coniugate rispetto a G e che $H^{(2)} = G^{-1}$. Questo comportamento è giustificato (anche per BFGS) da alcune proprietà riportate in quanto segue. Da notare che la condizione di curvatura (80) è verificata, dal momento che il soddisfacimento dell'equazione relativa alla ricerca lineare esatta (56) impone $\delta^{(k)T} \gamma^{(k)} = -\delta^{(k)T} g^{(k)} > 0$ (che è illustrato anche nell'esempio riportato in precedenza). Tuttavia, l'applicazione della ricerca lineare esatta non necessariamente garantisce la condizione (81) come mostrato nell'esempio seguente.

Esercizio. Risolvere il problema dell'ultimo esempio, applicando la regola di aggiornamento BFGS.

Esempio 25. Risolvere il problema dell'ultimo esempio, applicando la regola di aggiornamento SR1.

Soluzione. Poichè la prima iterazione è uguale a quella eseguita per DFP, si opera come segue.

Per $k = 1$, si ottiene

$$x^{(2)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, g^{(2)} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \delta^{(1)} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \gamma^{(1)} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

Per calcolare $H^{(2)}$ sfruttando la (79), al fine di semplificare la notazione, si elimina l'apice (1) , si sostituisce (2) con $+$ e si calcola

$$\delta - H\gamma = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, (\delta - H\gamma)^T \gamma = -2.$$

Quindi, sostituendo in (79), si ottiene la matrice modificata

$$H^+ = I - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1 \ 1) = \begin{bmatrix} 0,5 & -0,5 \\ -0,5 & 0,5 \end{bmatrix}.$$

Poichè tale matrice è indefinita, il metodo SR1 non permette di risolvere il problema preso in considerazione.

Tale comportamento del metodo era prevedibile dal momento che $\delta^T \gamma = 2$, $\delta^T B \delta = 2$, $\gamma^T H \gamma = 4$ e quindi la condizione (81) non è soddisfatta per $k = 1$. Con la scelta di tali valori, il metodo dell'aggiornamento combinato coincide con il metodo BFGS.

Poichè, in pratica, la seconda condizione di ricerca lineare (59) implica la (80), ma non necessariamente la (81) anche quando viene usata la ricerca lineare esatta, i metodi BFGS, DFP e l'aggiornamento combinato BFGS/SR1 con le condizioni di WP (58) e (59) sono da preferire al metodo SR1, mentre generalmente quest'ultimo aggiornamento è preferibile se la tecnica della trust region (che non verrà considerata in questo capitolo) è utilizzata al posto della tecnica della ricerca lineare. Quindi, se la condizione (81) non è verificata, l'aggiornamento SR1 non viene di solito effettuato (ie, $H^{(k+1)} = H^{(k)}$). Analogamente, questa strategia, indicata come tecnica dello skipping, potrebbe essere utilizzata per altri metodi di aggiornamento che possono produrre approssimazioni indefinite dell'Hessiana (anche per i metodi BFGS e DFP quando viene usata solo la condizione (58)). L'aggiornamento con la tecnica dello skipping è illustrato nel seguente esempio.

Esempio 26. *Risolvere il problema riportato nell'ultimo esempio, applicando la regola di aggiornamento SR1 con la tecnica dello skipping.*

Soluzione. Dal precedente esempio si ricava che l'aggiornamento SR1 di $H^{(1)}$ produce una matrice indefinita. Quindi l'aggiornamento non viene effettuato, si pone $H^{(2)} = H^{(1)}$ e si passa alla successiva iterazione

Per $k = 2$, si ottiene $s^{(2)} = -H^{(2)}g^{(2)} = (1, 1)^T$. Quindi, applicando la tecnica di ricerca lineare, si ricava

$$\alpha^{(2)} = 0,25, x^{(3)} = \begin{pmatrix} -0,75 \\ 1,25 \end{pmatrix}, g^{(3)} = \begin{pmatrix} 0,5 \\ 0 \end{pmatrix}.$$

Pertanto $H^{(2)}$ può essere modificata applicando la regola di aggiornamento SR1 e si ottiene

$$H^{(3)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{2,625} \begin{pmatrix} -1,25 \\ -0,75 \end{pmatrix} (-1,25 \quad -,75) = \begin{bmatrix} 0,4048 & -0,3571 \\ -0,3571 & 0,7857 \end{bmatrix}.$$

Poichè la matrice è definita positiva, si passa all'iterazione $k = 3$ e si ottiene

$$x^{(4)} = \begin{pmatrix} -0,9524 \\ 1,4286 \end{pmatrix}, \quad g^{(4)} = \begin{pmatrix} 0,0476 \\ -0,0476 \end{pmatrix}, \quad H^{(4)} = \begin{bmatrix} 0,5 & -1 \\ -1 & 1 \end{bmatrix}$$

che è uguale a G^{-1} .

Per $k = 4$, si ha $x^{(5)} = x^* = (1, -1, 25)^T$, $f^* = -1,25$. Quindi, il problema è stato risolto con $nfe = 7$.

Nonostante l'applicazione della regola di aggiornamento SR1, con la tecnica dello skipping, ha consentito di risolvere il problema, si è registrato un aumento del costo computazionale. Da notare che l'onere computazionale dopo lo skipping rimane uguale a quello dei metodi BFGS e DFP con ricerche lineari esatte. Infatti, tali regole di aggiornamento appartengono alla famiglia delle regole di aggiornamento di Broyden che godono delle seguenti proprietà.

- Se vengono usate ricerche lineari esatte, allora
 - i metodi della famiglia Broyden generano la stessa sequenza di punti $\{x^{(k)}\}$,
 - per una funzione quadratica $f(x)$ con Hessiana G definita positiva, i metodi della famiglia Broyden
 - * terminano in al massimo n iterazioni,
 - * $H^{(n+1)} = G^{-1}$,
 - * le condizioni quasi-Newton, cioè, $H^{(k+1)}\gamma^{(j)} = \delta^{(j)}$, $j \leq k$, risultano essere soddisfatte,
 - * quando $B^{(1)} = I$, i vettori $\{\delta^{(k)}\}$ e $\{\gamma^{(k)}\}$ sono coniugati.
- Se vengono utilizzate ricerche lineari non esatte, allora
 - per una funzione quadratica $f(x)$ con Hessiana G definita positiva, si ha $H^{(n+1)} = G^{-1}$ solo se viene applicato un metodo SR1 con la sequenza $\{\delta^{(k)}\}$ uniformemente indipendente, che garantisce anche la terminazione in al massimo n iterazioni,
 - per una funzione uniformemente convessa $f(x)$, il metodo BFGS con ricerca lineare esatta, che soddisfa le condizioni di Wolfe-Powell (58) e (59) converge globalmente e superlinearmente. Questa proprietà è stata estesa ad una famiglia di metodi di Broyden ristretta, che non include i metodi DFP e SR1,

- per una funzione generica $f(x)$, è generalmente conveniente usare il metodo BFGS con ricerca lineare esatta che soddisfa le condizioni di Wolfe-Powell (58) e (59).

Per illustrare l'ultima proprietà, vengono presentati i risultati ottenuti sulla base del seguente esperimento.

Esempio 27. *Applicare i metodi di Newton, BFGS, DFP, SR1 e BFGS/SR1 con ricerca lineare non esatta con backtracking alla funzione di Rosenbrock (7), considerando come stima iniziale il punto standard $x^{(1)} = (-1, 2, 1)^T$ e fermandosi quando $\|g^{(k)}\|_2 \leq 10^{-7}$. Applicare le regole di aggiornamento, considerando anche la tecnica dello skipping.*

Soluzione. Come evidenziato nell'Esempio 17, la prima iterazione dei metodi è uguale a quella del metodo di discesa, poichè si pone $H^{(1)} = I$. L'applicazione dei vari metodi ha evidenziato che essi, con la sola eccezione del metodo SR1 senza la tecnica dello skipping, sono in grado di risolvere il problema di Rosenbrock. Per ogni metodo, si riporta nella Tabella 2 il numero di ricerche lineari (nls) e di valutazioni della funzione (nfe) richieste per risolvere il problema. Sulla base dell'analisi di questi risultati si possono fare le seguenti considerazioni. Nonostante il metodo di Newton presenti le migliori prestazioni, tale metodo potrebbe fallire nel risolvere alcuni problemi come evidenziato precedentemente. Anche il metodo SR1, con la tecnica dello skipping, ha determinato la soluzione del problema, anche se presenta le peggiori prestazioni. Quindi, è chiaro che il metodo BFGS è il più efficiente ed è leggermente migliore di BFGS/SR1.

Poichè il metodo con backtracking non garantisce che la matrice sia definita positiva, la sperimentazione è stata ripetuta, considerando il punto iniziale $(-12, 1)^T$. In questo caso è stato osservato che tutti i metodi falliscono, eccetto il metodo combinato BFGS/SR1. Tuttavia, quando è stato utilizzato il metodo con lo skipping, si è notato che BFGS e BFGS/SR1 presentano delle buone prestazioni, mentre il metodo SR1 converge più lentamente e la convergenza del metodo DFP è ancora più lenta. Pertanto, è chiaro che dovrebbero essere utilizzate le condizioni di Wolfe-Powell (58) e (59). Infatti, generalmente, il metodo BFGS con le condizioni forti di Wolfe-Powell (58) e (60) è il più adatto a risolvere problemi di tipo generale, anche se presenta delle difficoltà nella soluzione di alcuni tipi di problemi mal condizionati.

Riferimenti bibliografici:

1. R. Fletcher, "Practical Methods of Optimization, 2nd Edition, 1987, Chichester, Wiley.
2. J.E. Dennis and R.B. Schnabel, "Numerical Methods for Unconstrained Optimization and Nonlinear Equation, 2nd Edition, 1996, SIAM.

Tabella 2: Applicazioni alla funzione di Rosenbrock (7)

Metodo	nls	nfe
Newton	21	29
BFGS	34	54
DFP	49	69
SR1	3 fallito	70 fallito
SR1/skip	64	88
BFGS/SR1	35	60

3. J. Nocedal and S.J. Wright, "Numerical Optimization, 1999, Springer-Verlag.
4. P.E. Gill, W. Murray and M.H. Wright, "Practical Optimization, 1981, London, Academic Press.
5. A.R. Conn, N.I.M. Gould and P.L. Toint, "Trust-Region Methods, 2000, SIAM.
6. M.S. Bazaraa, H.D. Sherali and C.M. Shetty, "Nonlinear Programming, Theory and Algorithms, 1993, Chichester, Wiley.
7. E.M.L. Beale, Introduction to Optimization, 1988, Chichester, Wiley.
8. B.D. Bunday, "Basic Optimization Methods, 1984, London, Edward Arnold Ltd.
9. E.K.P. Chong and S.H. Zak, "An Introduction to Optimization, 1996, Wiley.
10. A.V. Fiacco and G.P. McCormick, "Classics in Applied Mathematics, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, 1990, Philadelphia, SIAM.
11. P.E. Frandsen, K. Jonasson, H.B. Nielsen and O. Tingleff, "Unconstrained Optimization, 2004, 3rd Edition, Informatics and Mathematical Modelling Technical University of Denmark.
12. D.G. Luenberger, Optimization by Vector Space Methods, 2nd Edition, 1997, Wiley.
13. M. Minoux, "Mathematical Programming: Theory and Algorithms, 1986, Chichester, Wiley.
14. S.G. Nash and A. Sofer, "Linear and Nonlinear Programming, 1996, McGraw Hill.
15. A.L. Peressini, F.E. Sullivan and J.J. Uhl, Jr., "The Mathematics of Nonlinear Programming, 1988, Springer-Verlag.

16. W. Sun and Y. Yuan, "Optimization Theory and Methods: Nonlinear Programming, 2006, Springer.
17. G.R. Walsh, "Methods of optimization, 1985, Chichester, Wiley.
18. M. Al-Baali, A. Jamal-Eldeen and M. Soubh, "Numerical Analysis (2), 1993, Damascus University Publications (in Arabico).