I METODI DI DISCESA

1. Generalità sui metodi di discesa

Per la risoluzione di un sistema lineare $Ax = b$ con matrice A reale, simmetrica e definita positiva, un'altra famiglia di metodi iterativi è data dai così detti *metodi di discesa*.

Nota: Ricordiamo che dati due vettori colonna $x, y \in R^n$ *con la notazione <x,y> si intende il prodotto scalare* $x^T y$.

Il risultato teorico alla base di questi metodi è il seguente:

Teorema 1:

Sia $A \in R^{nxn}$, *matrice simmetrica e definita positiva,* $b, x \in R^n$*, allora la soluzione del sistema lineare*

$$Ax = b \qquad (1)$$

coincide con il punto di minimo della seguente funzione quadratica

$$F(x) = \frac{1}{2} < Ax, x > - < b, x > = \frac{1}{2} x^T A x - b^T x = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j - \sum_{i=1}^{n} b_i x_i \qquad (2)$$

ove la forma quadratica

$$Q(x) = <Ax,x> = x^T Ax \text{ è positiva per } x \neq 0.$$

Il teorema afferma quindi che risolvere il sistema (1) equivale a minimizzare la funzione (2).

Dimostrazione:

Se consideriamo il sistema (1) e definiamo il vettore residuo

$$r = Ax-b$$

se x* è la soluzione del sistema (1) , allora $r = Ax^* - b = 0$.

Ora consideriamo la funzione quadratica:

$$F(x) = \frac{1}{2}x^T A x - b^T x = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j - \sum_{i=1}^{n} b_i x_i$$

e cerchiamo il suo punto di minimo.

A questo scopo calcoliamo il gradiente di F ed uguagliamolo a zero:

$$\nabla f = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \\ \\ \frac{\partial F}{\partial x_n} \end{bmatrix}$$

Sviluppiamo $F(x)$

$F(x) = \frac{1}{2}(a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1i}x_1x_i + \cdots + a_{1n}x_1x_n +$

$+a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2i}x_2x_i + \cdots + a_{2n}x_2x_n +$

$\ldots$

$+a_{i1}x_ix_1 + a_{i2}x_ix_2 + \cdots + a_{ii}x_i^2 + \cdots + a_{in}x_ix_n +$

$\ldots$

$+ a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{ni}\, x_nx_i \;\; + \cdots + a_{nn}x_n^2) - (b_1x_1 + b_2x_2 + \cdots +$

$b_nx_n) +$

$\frac{\partial F}{\partial x_1} = \frac{1}{2}(2\, a_{11}x_1 + a_{12}x_2 + a_{1i}x_i + \cdots + a_{1n}x_n + a_{21}x_2 + a_{i1}x_i + \cdots a_{n1}x_n) - b_1$

Per ipotesi la matrice A è simmetrica quindi $a_{ij} = a_{ji}$, quindi la precedente relazione si può semplificare in:

$$\frac{\partial F}{\partial x_1} = \frac{1}{2}(2\, a_{11}x_1 + 2a_{12}x_2 + 2a_{1i}x_i + \cdots + 2a_{1n}x_n) - b_1$$

cioè

$$\frac{\partial F}{\partial x_1} = a_{11}x_1 + a_{12}x_2 + a_{1i}x_i + \cdots + a_{1n}x_n - b_1$$

Analogamente si ricava che

$$\frac{\partial F}{\partial x_i} = a_{i1}x_1 + a_{i2}x_2 + a_{ii}x_i + \cdots + a_{in}x_n - b_i \qquad \text{i=1,…,n}$$

che si può scrivere come

$$\frac{\partial F}{\partial x_i} = \sum_{j=1}^{n} a_{ij}x_j - b_i = 0 \quad i = 1, \ldots, n$$

e, in termini vettoriali,

$$\nabla F = Ax\text{-}b$$

Poiché abbiamo definito r=Ax-b, risulta che il vettore che annulla il gradiente coincide con la soluzione del sistema lineare, che rende nullo il residuo.

$$r = \nabla F = Ax^{*} - b = 0.$$

Verifichiamo che il punto che annulla il gradiente è effettivamente un punto di minimo. A tale scopo calcoliamo la matrice Hessiana H di F:

$$H_F(x,y) = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \cdots & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} & & & \frac{\partial^2 F}{\partial x_2 \partial x_n} \\ & & \cdots & & \\ & & & \cdots & \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \frac{\partial^2 F}{\partial x_n \partial x_2} & & & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix}$$

Osserviamo che, poiché

$$\frac{\partial F}{\partial x_i} = a_{i1}x_1 + a_{i2}x_2 + a_{ii}x_i + \cdots a_{ij}x_j + .. + a_{in}x_n - b_i$$

si ha

$$\frac{\partial F}{\partial x_i \partial x_j} = a_{ij} \quad \text{e quindi} \quad H_F(x,y) = A$$

Risulta

$$H_F(x,y) = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & a_{2n} \\ \dots & & & & \\ \dots & & & & \\ a_{n1} & a_{n2} & \dots & \dots & a_{nn} \end{bmatrix}$$

La matrice Hessiana coincide con la matrice A che ha determinate maggiore di zero, ed elemento $a_{11} > 0$ essendo definita positiva.

Vale infatti il Criterio di Sylvester per le matrici simmetriche e definite positive.

Il criterio di Sylvester è un teorema che fornisce una condizione necessaria e sufficiente affinché una matrice simmetrica sia definita positiva. Stabilisce che una matrice simmetrica è definita positiva se e solo se i determinanti di tutte le sottomatrici principali di testa $A_k, k = 1, .., n$ sono **positivi**. Poiché la sottomatrice principale di testa di ordine 1 è l'elemento a_{11} e la sottomatrice principale di testa di ordine n coincide con la matrice A, segue che $a_{11} > 0$ ed il determinante dell'Hessiano è positivo. Quindi il punto che annulla il gradiente è il minimo della forma quadratica.

Quindi il vettore x^* che minimizza la funzione $F(x)$ coincide con la soluzione del sistema lineare (1); viceversa la soluzione del sistema (1) con matrice A simmetrica definita positiva minimizza la corrispondente funzione quadratica (2).

Esempio n=2

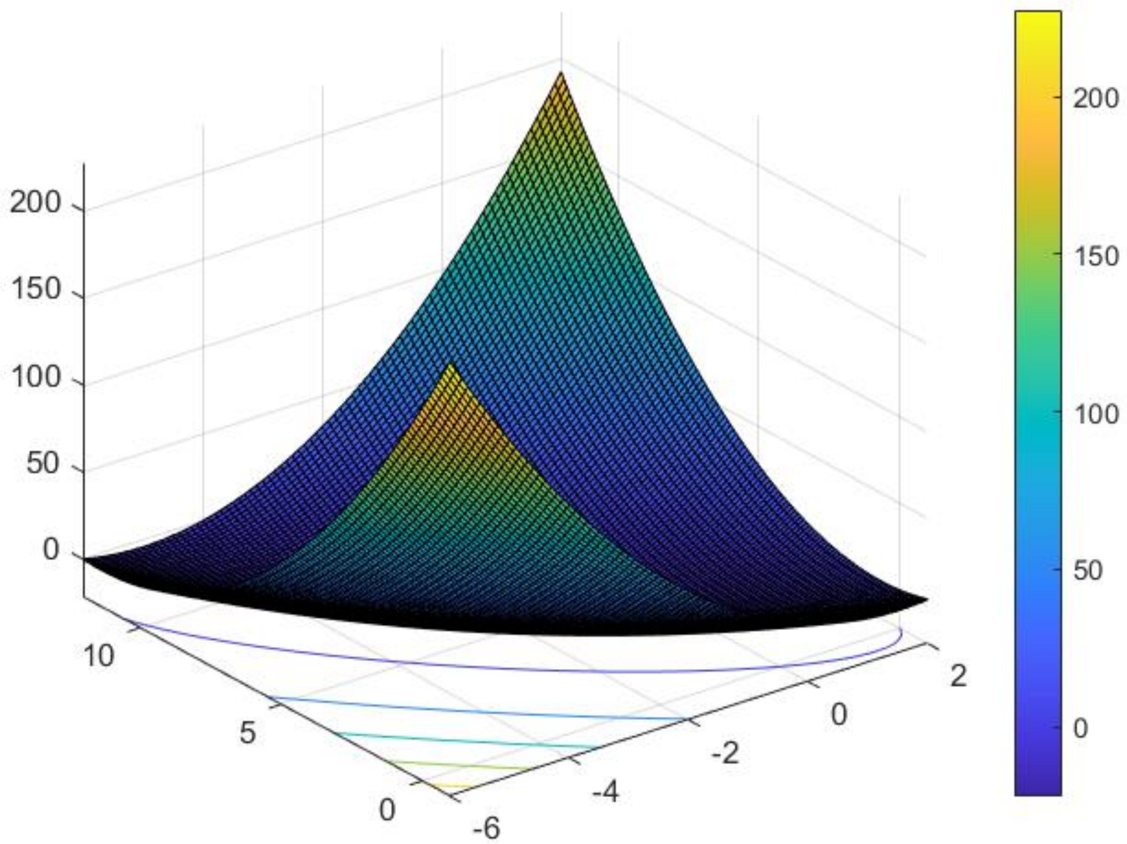
$$A = \begin{bmatrix} 8 & 4 \\ 4 & 3 \end{bmatrix} \quad \mathrm{b} = \begin{bmatrix} 8 \\ 10 \end{bmatrix} \quad x = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$$

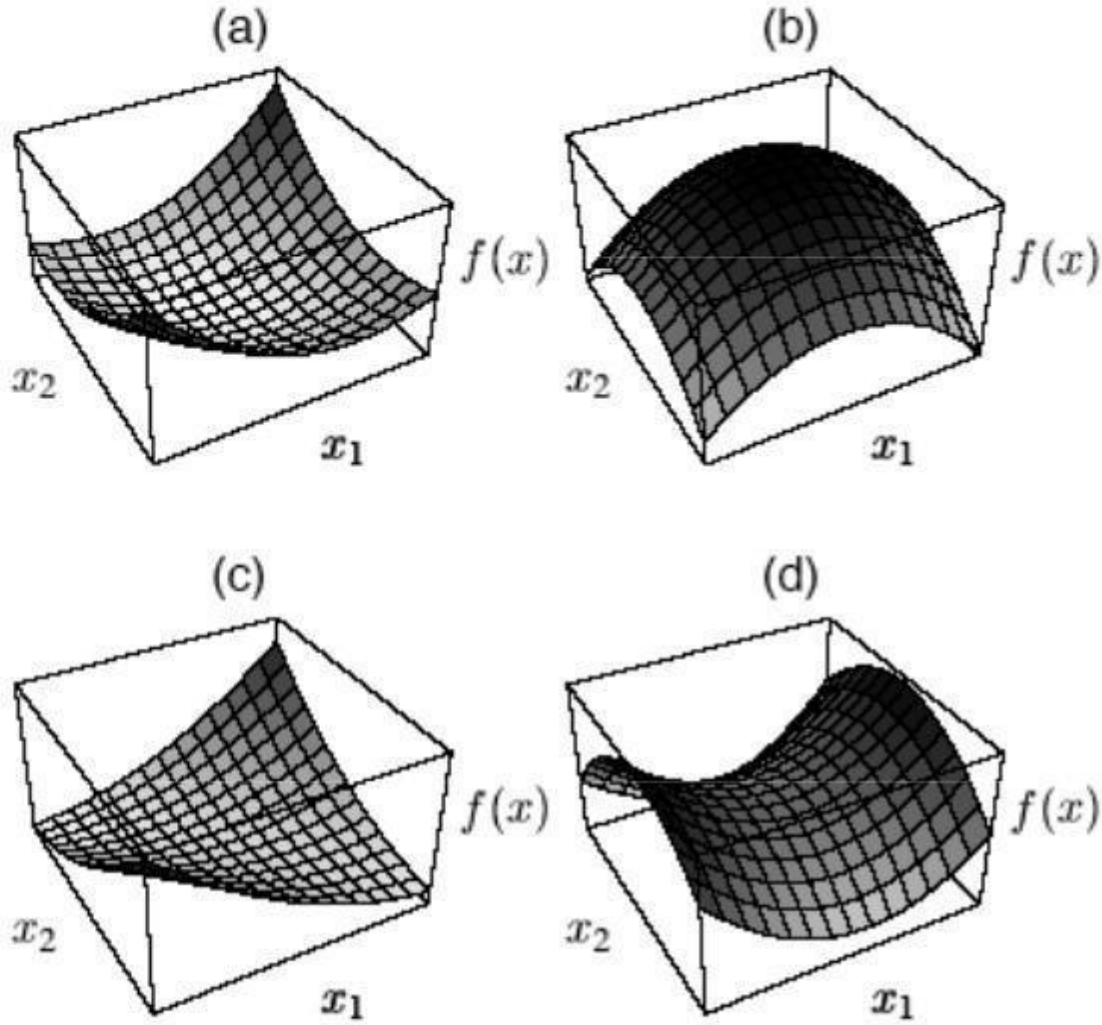
La matrice A è simmetrica e definita positiva.

Vale il seguente risultato: Forma quadratica associata ad una matrice simmetrica e definita positiva è strettamente convessa, quindi se ammette minimo, esso è unico

$$F(x_1, x_2) = \frac{1}{2}(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2) - (b_1x_1 + b_2x_2)$$

Grafico della forma quadratica associata al sistema lineare Ax=b





Se la matrice *A*, della forma quadratica è *definita positiva* allora il minimo corrisponde alla soluzione del sistema ; mentre se *A* è *definita negativa*, (b) in questo caso il paraboloide è rovesciato ed il suo massimo corrisponde alla soluzione del sistema; se invece *A* è *semidefinita positiva , (c)* allora vuol dire che almeno un suo autovalore è nullo e sarà perciò singolare, le soluzioni del sistema sono più di una e corrispondono all'insieme dei punti della linea sul fondo del paraboloide; infine se A è *non definita*, (d) la soluzione è un cosiddetto punto di sella.

Il risultato del teorema 1 ci permette di affermare che per la risoluzione di sistemi lineari con matrice simmetrica definita positiva in generale possono essere usati i metodi per determinare il minimo di una funzione quadratica, noti come **metodi di discesa**.

Questi metodi iterativi consistono nel determinare, a partire da un vettore *x al passo k, che indicheremo con* $x^{(k)}$, un vettore direzione *p al passo k*, $p^{(k)}$, opportuno e nel correggere $x^{(k)}$ in questa direzione in modo che il valore della funzione quadratica F nel nuovo iterato $x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)}$ diminuisca, cioè

$$F(x^{(k)} + \alpha^{(k)} p^{(k)}) < F(x^{(k)}).$$

Perché ciò avvenga il parametro $\alpha^{(k)}$ e la direzione $p^{(k)}$ devono essere scelti in modo opportuno. I differenti metodi di discesa sono caratterizzati dalla scelta della direzione di discesa $p^{(k)}$ fra le direzioni di discesa ammissibili.

La determinazione del parametro $\alpha^{(k)}$ che permette di rendere minima la F nella direzione $p^{(k)}$ differenzia i metodi suddetti quando vengono impiegati per la soluzione del sistema lineare $Ax = b$ (quindi in cui A sia nota) da quando vengono impiegati per determinare il minimo di una funzione qualsiasi (in cui quindi la matrice A non è nota).

Un **generico algoritmo di discesa per la minimizzazione**

1. Parti con qualche $x^{(0)}$, $k = 0$
2. Determina la direzione di discesa $p^{(k)}$
3. Scegli lo step-size $\alpha^{(k)}$ tale che
$$F(x^k + \alpha^{(k)} p^{(k)}) < F(x^{(k)})$$
4. Aggiorna l'iterato
$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)}$$
5. Incrementa il contatore k=k+1

Fino a convergenza

2. Scelta dello step-size

Nel caso in cui la matrice A sia nota si vede infatti che

$$F\left(x^{(k)}+\alpha^{(k)}p^{(k)}\right)=\frac{1}{2}<A\left(x^{(k)}+\alpha^{(k)}p^{(k)}\right),\left(x^{(k)}+\alpha^{(k)}p^{(k)}\right)>-$$

$$<b,\left(x^{(k)}+\alpha^{(k)}p^{(k)}\right)>$$

$$=\tfrac{1}{2}<Ax^{(k)},x^{(k)}>+\alpha^{(k)}<Ax^{(k)},p^{(k)}>+\tfrac{1}{2}(\alpha^{(k)})^2<Ap^{(k)},p^{(k)}>+$$

$$-<b,x^{(k)}>-\alpha^{(k)}<b,p^{(k)}>$$

$$=F(x^{(k)})+\frac{1}{2}(\alpha^{(k)})^2<Ap^{(k)},p^{(k)}>+\alpha^{(k)}<r^{(k)},p^{(k)}>, \qquad r^{(k)}=Ax^{(k)}-b,$$

è una **funzione quadratica** in t.

Per determinarne il valore di t che rende minima F nella direzione $p^{(k)}$, cioè:

$$arg\ \min_{\alpha^{(k)}} F(x^{(k)}+\alpha^{(k)}\,p^{(k)})$$

basta considerare

$$\frac{dF}{d\alpha^{(k)}}=\alpha^{(k)}<Ap^{(k)},p^{(k)}>+<r^{(k)},p^{(k)}>$$

e uguagliarla a 0; si ottiene per il parametro$\alpha^{(k)}$ il valore

$$\alpha^{(k)}=-\frac{<r^{(k)},p^{(k)}>}{<Ap^{(k)},p^{(k)}>}=-\frac{(r^{(k)})^Tp^{(k)}}{(p^{(k)})^TAp^{(k)}} \qquad (3)$$

Questo significa che, una volta determinata la direzione $p^{(k)}$, ovvero scelto il metodo di minimizzazione, la relazione (3) ci fornisce il valore da assegnare allo stepsize $\alpha^{(k)}$ per ottenere il minimo valore possibile della F lungo la direzione $p^{(k)}$,.

Osservazione: il valore di $\alpha^{(k)}$fornito dalla (3) ci fornisce un minimo per F in quanto se effettuiamo la derivata seconda in t otteniamo

$$\frac{d^2F}{dt^2}=<Ap^{(k)},p^{(k)}>$$

che è una quantità positiva per ogni direzione di $p^{(k)}$, essendo la matrice A per ipotesi simmetrica e definita positiva.

Nel caso in cui il metodo venga applicato per determinare il minimo di una funzione e quindi la matrice *A* non sia nota, il valore di t ottimale si ottiene risolvendo un problema di minimo monodimensionale per la funzione *F*(t).

Nel punto di minimo ottenuto usando la formula (3) per il valore del parametro t abbiamo il seguente risultato:

Teorema:

Nel punto di minimo $x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)}$, *ottenuto muovendosi lungo la direzione* $p^{(k)}$ *con* $\alpha^{(k)}$ *dato dalla (3), il vettore residuo* $r^{(k+1)} = Ax^{(k+1)} - b$ *risulta ortogonale alla direzione* $p^{(k)}$ *, cioè*

$$< r^{(k+1)}, p^{(k)} >= 0 \qquad (*)$$

Dimostrazione:

Essendo

$$r^{(k+1)} = Ax^{(k+1)} - b = A\left(x^{(k)} + \alpha^{(k)} p^{(k)}\right) - b = r^{(k)} + \alpha^{(k)} A\, p^{(k)}$$

Si ottiene

$$< r^{(k+1)}, p^{(k)} >=< r^{(k)}, p^{(k)} > +\alpha^{(k)} < A\, p^{(k)}, p^{(k)} >$$

che risulta 0 scegliendo $\alpha^{(k)}$ secondo la (3).

$$< r^{(k+1)}, p^{(k)} >=< r^{(k)}, p^{(k)} > - \frac{< r^{(k)}, p^{(k)} >}{< Ap^{(k)}, p^{(k)} >} < A\, p^{(k)}, p^{(k)} > = 0$$

3. Condizioni di ammissibilità per la direzione di discesa

Per quanto riguarda la scelta delle direzioni di discesa $p^{(k)}$ se si considera la relazione

$$\alpha^{(k)} = -\frac{<r^{(k)},p^{(k)}>}{<Ap^{(k)},p^{(k)}>}$$

si può affermare che $p^{(k)}$ non deve essere ortogonale al residuo $r^{(k)}$, o, in modo equivalente, (visto che il gradiente della forma quadratica calcolato in $x^{(k)}$ è uguale al residuo del sistema lineare valutato in $x^{(k)}$) non deve essere ortogonale al gradiente $\nabla F(x^{(k)})$ di F perché questo porterebbe a $t^{(k)}$=0.

Se inoltre consideriamo lo sviluppo in serie di Taylor della $F\left(x^{(k)} + \alpha^{(k)}p^{(k)}\right)$, in un intorno di $x^{(k)}$, cioè

$$F\left(x^{(k+1)}\right) = F\left(x^{(k)} + \alpha^{(k)}p^{(k)}\right) = F\left(x^{(k)}\right) + \alpha^{(k)}\nabla F(x^{(\dot{k})})^T p^{(k)} + \cdots$$

che si può scrivere come

$$\left(x^{(k+1)}\right) = F\left(x^{(k)} + \alpha^{(k)}p^{(k)}\right) = F\left(x^{(k)}\right) + \alpha^{(k)} < \nabla F(x^{(k)}), p^{(k)} > \quad + \cdots$$

e richiediamo che si abbia

$$F(x^{(k)} + \alpha^{(k)}p^{(k)}) < F(x^{(k)}) \quad per\ \alpha^{(k)} > 0 \ , \qquad (4)$$

allora la direzione $p^{(k)}$ deve soddisfare la seguente condizione

$$< \nabla F(x^{(k)}), p^{(k)} > \quad < 0$$

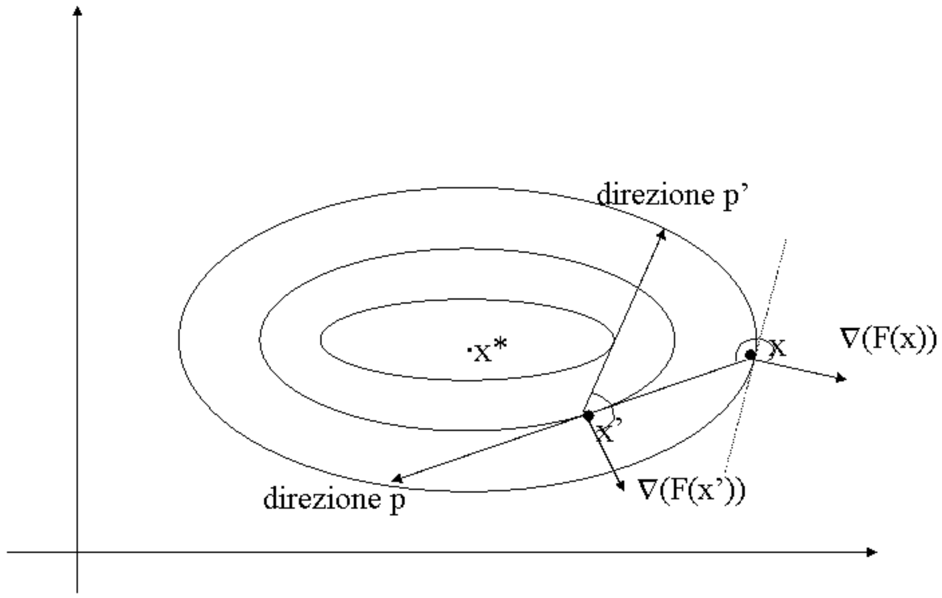
che rappresenta la condizione di **direzione ammissibile**.

Questa condizione ci dice che l'angolo fra la nuova direzione di discesa e il gradiente $\nabla F(x^{(k)})$ deve avere coseno negativo (ricordiamo che $< \nabla F(x^{(k)}), p^{(k)} >= ||\nabla F(x^{(k)}|| \cdot ||p^{(k)}|| \cos\theta$ dove l'angolo θ è l'angolo fra i due vettori) cioè l'angolo θ deve essere maggiore di π/2 e minore di 3π/2 .

Poiché la condizione di ammissibilità per la direzione di discesa è data in funzione del gradiente della F(x) i metodi di discesa vengono anche chiamati metodi del gradiente.

<u>Es:</u> Interpretazione geometrica dei metodi di discesa

Nel caso n = 2 la funzione $F(x)$=cost è rappresentata da ellissi concentriche il cui centro coincide con il minimo della funzione quadratica $F(x)$ e costituisce la soluzione del problema. Il seguente grafico mostra quanto affermato dalla condizione di ammissibilità per la direzione di discesa.



4. Metodo della Discesa più Ripida (Steepest Descent)

Il metodo di Discesa più Ripida (Steepest Descent) è caratterizzato dalla scelta, ad ogni passo k, della direzione $p^{(k)}$ come l'antigradiente della F calcolato nell'iterato k -esimo, ovvero

$$p^{(k)} = -\nabla F(x^{(k)}) = -Ax^{(k)} + b = -r^{(k)}. \tag{5}$$

Poiché il gradiente è la direzione di massima crescita, questo significa che ad ogni passo il vettore $p^{(k)}$ essendo l'antigradiente di F coincide con la direzione di massima decrescita.

In questo caso la (3) diventa

$$\alpha^{(k)} = -\frac{< r^{(k)}, p^{(k)} >}{< Ap^{(k)}, p^{(k)} >} = \frac{< r^{(k)}, r^{(k)} >}{< Ar^{(k)}, r^{(k)} >}$$

e

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)}.$$

Algoritmo steepest descent (Algoritmo del gradiente)

1. Parti con qualche $x^{(0)}$, $k = 0$
2. Calcola la direzione di discesa più ripida
$$p^{(k)} = -\nabla f(x^{(k)})$$
3. Scegli lo stepsize $\alpha^{(k)}$ tale che
$$F\left(x^k + \alpha^{(k)} p^{(k)}\right) < F\left(x^{(k)}\right)$$
$$\alpha^{(k)} = \frac{< r^{(k)}, r^{(k)} >}{< Ar^{(k)}, r^{(k)} >}$$
4. Aggiorna l'iterato
$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)}$$
$$r^{(k+1)} = r^{(k)} + \alpha^{(k)} A\, p^{(k)}$$
5. Incrementa il contatore k=k+1

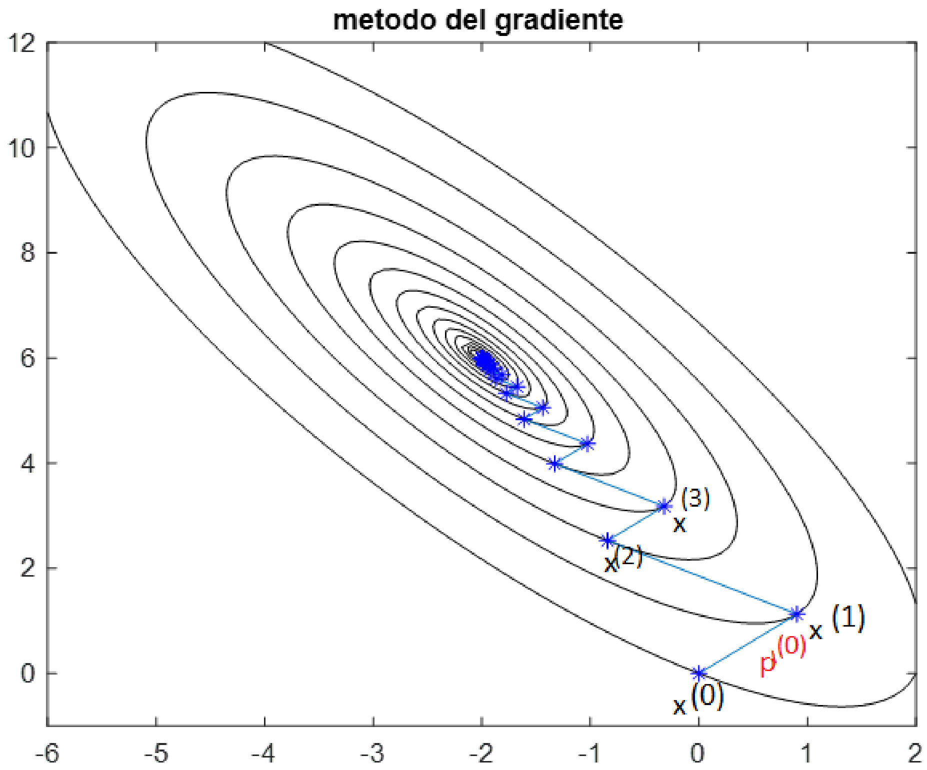
Fino a convergenza

Si considera che il procedimento iterativo ha raggiunto la convergenza quando $\frac{||r^{(k+1)}||_2}{||b||} \leq tolleranza$

Esempio

$$A = \begin{bmatrix} 8 & 4 \\ 4 & 3 \end{bmatrix} \quad \mathrm{b} = \begin{bmatrix} 8 \\ 10 \end{bmatrix} \quad x = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$$

Iterato iniziale $x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$



La direzione $p^{(0)}$ è opposta alla direzione del gradiente alla curva di livello $\mathrm{F}(x^{(0)}) = cost$ nel punto $x^{(0)}$ (Ricordiamo che il gradiente $\nabla F\left(x^{(0)}\right) = r^{(0)}$ è perpendicolare alla tangente alla curva di livello nel punto $x^{(0)}$.

L'iterato $x^{(1)}$ si trova a partire da $x^{(0)}$ nella direzione di $p^{(0)}$ nella posizione individuata da $\alpha^{(1)}$, $x^{(1)} = x^{(0)} + \alpha^{(1)} p^{(0)}$.

Dal grafico si **nota carattere a zigzag** del metodo del gradiente, dovuto al fatto che il gradiente di una iterata è ortogonale al gradiente di quello precedente. Si noterà che, nonostante la convergenza dell'algoritmo, quest'ultimo è relativamente lento a causa di questo avanzamento a zig zag.

Velocità di Convergenza

La velocità di convergenza di un metodo iterativo si può misurare considerando di quanto si è ridotto l'errore iniziale alla k-esima iterazione .

Per misurare l'errore si definisce la norma indotta dalla matrice simmetrica definita positiva A su x come

$$\| x \|_A^2 = x^T A x.$$

Per il **metodo del gradiente Steepest Descent** vale la seguente relazione

$$\| x^{(k)} - x^* \|_A \leq \left(\frac{K(A)-1}{K(A)+1} \right)^k) \cdot \| x^{(0)} - x^* \|_A$$

Pertanto, definendo l'errore al passo k

$$e_A^{(k)} = \| x^{(k)} - x^* \|_A$$

si ha

$$e_A^{(k)} \leq \left(\frac{K(A)-1}{K(A)+1} \right)^k) \cdot e_A^{(0)}$$

dove K(A) è l'indice di condizionamento di A, dato da $K(A)=\|A\|\cdot\|A^{-1}\|=\lambda_{max}/\lambda_{min}$.

Tanto più K(A) è alto tanto più il rapporto $\left(\frac{K(A)-1}{K(A)+1} \right) \approx 1$ e quindi tanto più è lenta la convergenza.

Poiché la funzione quadratica F(x) data dalla (2) assegnata la F(x)=cost rappresenta l'espressione di un iperellissoide con eccentricità legata dal rapporto $\frac{\lambda_{\max}}{\lambda_{\min}}$, possiamo dire che ad una matrice A mal condizionata corrisponde un' iperellissoide molto allungato, mentre ad un K(A) piccolo corrisponde un iperellissoide più arrotondato.