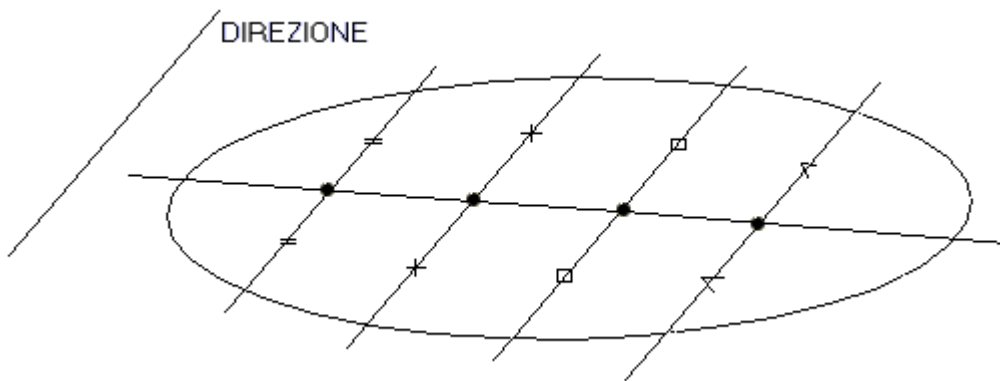


# I METODI DI DISCESA

## Il gradiente coniugato

### Definizione di direzioni coniugate

Data una conica ed una direzione, tutti i punti medi delle corde parallele alla direzione sono allineati e formano una direzione che si dice coniugata alla direzione data.



### 1. Metodo del Gradiente Coniugato

In questo metodo la scelta della direzione di discesa  $p^{(k)}$  tiene conto non solo del gradiente della  $F(x^{(k)})$  cioè di  $r^{(k)}$ , ma anche della direzione di discesa  $p^{(k-1)}$ . In particolare nel metodo del Gradiente Coniugato, al generico passo  $k$ , partendo dal punto  $x^{(k)}$  che è stato ottenuto muovendosi lungo la direzione  $p^{(k-1)}$ , e in cui è stato calcolato il residuo  $r^{(k)}$  (ortogonale a  $p^{(k-1)}$ ), sceglie la nuova direzione di discesa come quella appartenente al piano  $\pi_k$  passante per  $x^{(k)}$  e individuato dai due vettori ortogonali  $r^{(k)}$  e  $p^{(k-1)}$ .

Più precisamente si ha

$$p^{(k)} = -r^{(k)} + \gamma_k p^{(k-1)}, \quad k = 1, 2, \dots \quad (6)$$

Il piano  $\pi_k$  interseca la funzione  $F(x)$  con un'ellisse passante per  $x^{(k)}$ . Poiché  $x^{(k)}$  era stato scelto come punto di minimo nella direzione  $p^{(k-1)}$ , in  $x^{(k)}$  l'ellisse risulta tangente alla direzione di discesa  $p^{(k)}$ .

Poiché il punto di minimo nel piano  $\pi_k$  coincide con il centro dell'ellisse, il parametro  $\gamma_k$  sarà scelto in modo che la direzione  $p^{(k)}$  punti verso il centro dell'ellisse, cioè sia il **coniugato**, rispetto all'ellisse, di  $p^{(k-1)}$ . Ciò significa che deve soddisfare la seguente relazione:

$$\langle Ap^{(k)}, p^{(k-1)} \rangle = \langle p^{(k)}, Ap^{(k-1)} \rangle = 0. \quad (7)$$

Sostituendo l'espressione  $p^{(k)} = -r^{(k)} + \gamma_k p^{(k-1)}$  nella seconda delle (7) si ottiene

$$\gamma_k = \frac{\langle r^{(k)}, Ap^{(k-1)} \rangle}{\langle p^{(k-1)}, Ap^{(k-1)} \rangle}. \quad (7\text{-bis})$$

Utilizzando tale valore nella (6) si ottiene la nuova direzione  $p^{(k)}$  e il nuovo punto  $x^{(k)}$  viene calcolato come punto di minimo nella direzione  $p^{(k)}$ , cioè

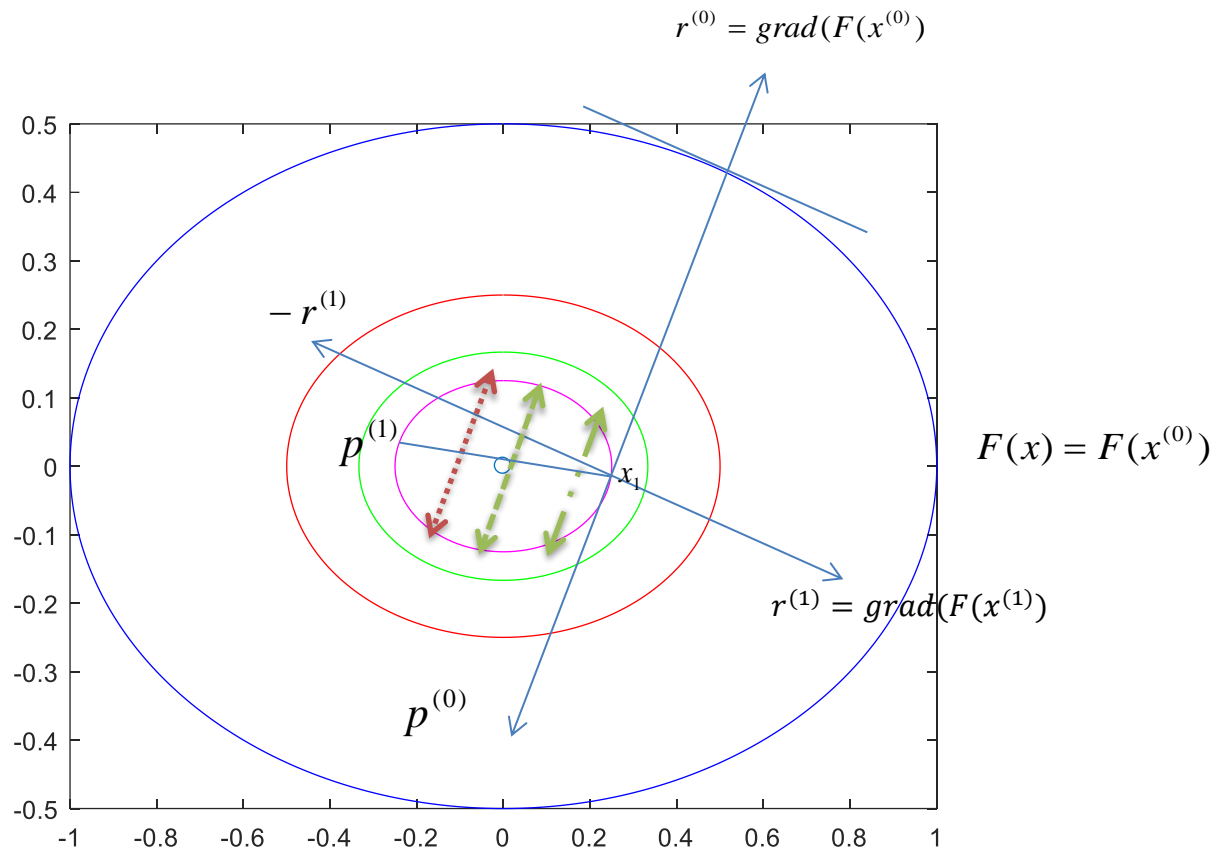
$$x^{(k+1)} = x^{(k)} + \lambda_k p^{(k)} \quad (8)$$

con

$$\lambda_k = -\frac{\langle r^{(k)}, p^{(k)} \rangle}{\langle Ap^{(k)}, p^{(k)} \rangle} > \quad k = 1, 2, \dots \quad (9)$$

Si noti che  $\lambda_k$  in (8) gioca il ruolo di  $\alpha_k$  nelle formule del paragrafo precedente e quindi la (9) rappresenta il valore che assume  $\lambda_k$  nel punto di minimo nella direzione  $p^{(k)}$ .

Le relazioni (6), (7-bis), (8) e (9) definiscono sostanzialmente il metodo del gradiente coniugato.



La direzione  $p^{(1)}$  parte da  $x^{(1)}$ , appartiene al piano individuato da  $-r^{(1)}$  e  $p^{(0)}$  ed è coniugata rispetto a  $p^{(0)}$ , cioè è la direzione congiungente i punti medi delle corde parallele alla direzione  $p^{(0)}$  e passa quindi per il centro delle ellissi che corrisponde al minimo del funzionale.

Quindi il metodo del gradiente coniugato, nel caso  $n=2$ , raggiunge la soluzione in 2 passi.

Ci avvaliamo ora di alcuni risultati che permettono di semplificarne le formule riducendone la complessità computazionale.

La prima semplificazione si ottiene osservando che per il residuo è possibile definire una formula ricorsiva che lo aggiorna utilizzando una quantità che è necessaria anche per calcolare altre grandezze, cioè

$$r^{(k)} = Ax^{(k)} - b = Ax^{(k-1)} - b + \lambda_k Ap^{(k)}$$

cioè

$$r^{(k)} = r^{(k-1)} + \lambda_k Ap^{(k)}.$$

(10)

La seconda semplificazione segue dal seguente

**Teorema:**

*Nel metodo del gradiente coniugato le direzioni di discesa  $p^{(k)}$ , con  $k=0,1,\dots$ , formano un sistema di direzioni coniugate, mentre i vettori residui  $r^{(k)}$ , con  $k=0,1,\dots$ , formano un sistema ortogonale, cioè*

$$\langle r^{(k)}, r^{(j)} \rangle = 0 \quad k \neq j, \quad j=0,1,\dots,k-1 \quad (11)$$

$$\langle Ap^{(k)}, p^{(j)} \rangle = 0 \quad k \neq j, \quad j=0,1,\dots,k-1. \quad (11\text{bis})$$

Ciò significa che la direzione  $p^{(k)}$  è coniugata non solo a  $p^{(k-1)}$  ma a tutte le precedenti direzioni di discesa e che il residuo  $r^{(k)}$  è ortogonale a tutti i precedenti residui:

**Osservazione:** Poiché in  $\mathbb{R}^n$  non si possono avere più di  $n$  vettori che costituiscono un sistema ortogonale, in linea teorica questa classe di metodi appartiene ai metodi diretti poiché viene costruita una successione  $\{x^{(k)}\}_{k=0,1,\dots}$  di vettori tali che

$$x^{(k)} = x^* = A^{-1}b \quad \text{quando } k = n-1.$$

In pratica, però, a causa degli errori di arrotondamento, il metodo non termina al passo  $k=n-1$  e viene quindi utilizzato come metodo iterativo.

In molti casi comunque si verifica che il numero di iterazioni che occorrono per raggiungere la precisione richiesta è di gran lunga inferiore alla dimensione del sistema e questo rende il metodo molto utile per problemi di grosse dimensioni.

Utilizziamo ora il fatto che i residui in due passi successivi sono ortogonali per ottenere un'ulteriore proprietà di ortogonalità. Infatti sostituendo nella relazione che esprime il

risultato generale, (cioè che residuo ad ogni passo è ortogonale alla direzione del passo precedente),

$$\langle r^{(k+1)}, p^{(k)} \rangle = 0 \quad (12)$$

l'espressione di  $p^{(k)}$  data dalla (6) si ottiene

$$\langle r^{(k+1)}, -r^{(k)} + \gamma_k p^{(k-1)} \rangle = 0$$

$$0 = -\langle r^{(k+1)}, r^{(k)} \rangle + \gamma_k \langle r^{(k+1)}, p^{(k-1)} \rangle$$

utilizzando la proprietà di ortogonalità (11), segue:

$$\langle r^{(k+1)}, p^{(k-1)} \rangle = 0.$$

Si può verificare che in generale:.

$$\langle r^{(k+1)}, p^{(j)} \rangle = 0 \quad j < k + 1$$

E' inoltre possibile trovare una nuova espressione semplificata per il parametro  $\lambda_k$  dato dalla (9).

Infatti poiché

$$\langle r^{(k)}, p^{(k)} \rangle = \langle r^{(k)}, -r^{(k)} + \gamma_k p^{(k-1)} \rangle = -\langle r^{(k)}, r^{(k)} \rangle + \gamma_k \underbrace{\langle r^{(k)}, p^{(k-1)} \rangle}_{=0} = -\langle r^{(k)}, r^{(k)} \rangle$$

$\langle r^{(k)}, p^{(k-1)} \rangle = 0$  perché il residuo ad ogni passo è ortogonale alla direzione al passo precedente

si ottiene

$$\lambda_k = \frac{\langle r^{(k)}, r^{(k)} \rangle}{\langle Ap^{(k)}, p^{(k)} \rangle} \quad k = 1, 2, \dots$$

(13)

Da questa formula si vede che se il residuo non è nullo  $\lambda_k$  è sempre positivo.

Utilizzando ora la formula ricorrente (10) per il residuo e la nuova espressione (13) di  $\lambda_k$  è possibile trovare una formula computazionalmente più efficiente per  $\gamma_k$ .

e quindi l'espressione di  $\gamma_k$  diviene

$$\gamma_k = \frac{\langle r^{(k)}, r^{(k)} \rangle}{\langle r^{(k-1)}, r^{(k-1)} \rangle} . \quad (14)$$

In definitiva l'algoritmo del **Gradiente Coniugato** può essere schematizzato come segue:

```

Scelto  $x^{(0)}$  arbitrario, si calcola  $r^{(0)} = Ax^{(0)} - b$ , si prende
 $p^{(0)} = -r^{(0)}$ 
 $k=0$ ;
while arresto  $\geq \varepsilon$ 

     $\lambda_k = \frac{\langle r^{(k)}, r^{(k)} \rangle}{\langle Ap^{(k)}, p^{(k)} \rangle}$ 

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

     $r^{(k+1)} = r^{(k)} + \lambda_k Ap^{(k)}$ 

    arresto =  $\| r^{(k+1)} \|_2^2$ 

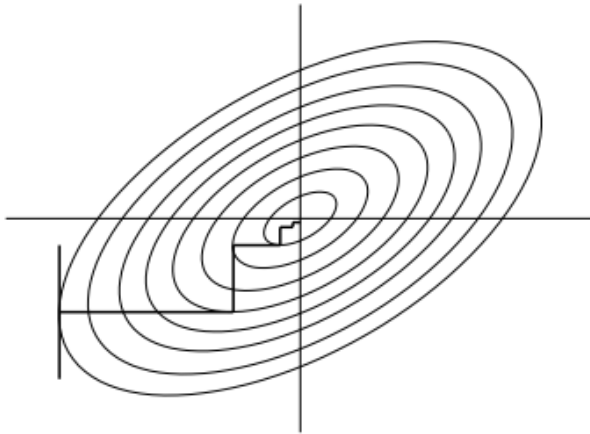
     $\gamma_{k+1} = \frac{\langle r^{(k+1)}, r^{(k+1)} \rangle}{\langle r^{(k)}, r^{(k)} \rangle}$ 

     $p^{(k+1)} = -r^{(k+1)} + \gamma_{k+1} p^{(k)}$ 

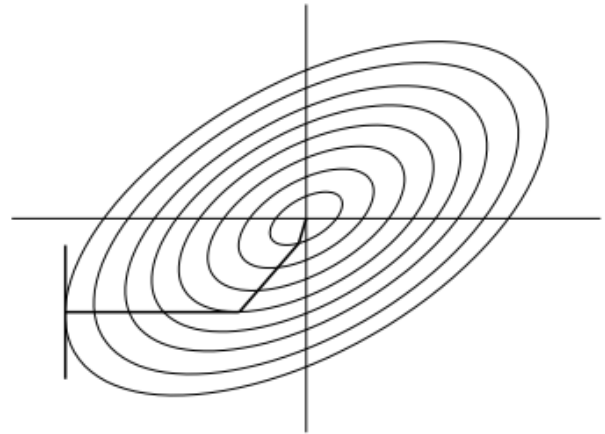
     $k=k+1$ 
end while

```

**Osservazione:** l'algoritmo del gradiente coniugato così ottimizzato necessita di un'unica moltiplicazione matrice per vettore per ogni iterazione.



**Steepest descent**



**Gradiente Coniugato**

### Velocità di Convergenza

Per il **metodo del gradiente coniugato** vale la seguente relazione

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

cioè

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

che mostra come la convergenza di questo metodo, pur rimanendo sempre legata all'indice di condizionamento di  $A$  sia più veloce di quella del metodo di Steepest Descent a parità di valori di  $K(A)$ .

Comunque, se la matrice  $A$  è molto mal condizionata può accadere che siano necessari molti passi di iterazione per ottenere la convergenza.

Poiché l'obiettivo dei metodi iterativi è quello di ottenere una buona approssimazione della soluzione del sistema  $Ax = b$  con, mediamente, poche iterazioni, sono state studiate tecniche di preconditionamento che trasformano il problema originale in un problema equivalente ma meglio condizionato.

**Osservazione:**

Poiché la funzione quadratica  $F(x)$  data dalla (2) assegnata la  $F(x)=\text{cost}$  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.