

Sistemi Lineari

Metodi iterativi

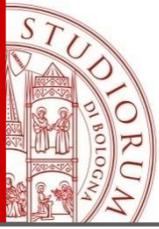
Sistemi Lineari

$$\begin{array}{c}
 \begin{matrix}
 \text{matrice} \\
 \text{coefficienti}
 \end{matrix}
 \left[\begin{array}{cccc}
 a_{11} & a_{12} & \cdots & a_{1n} \\
 a_{21} & a_{22} & \cdots & a_{2n} \\
 \vdots & \vdots & \ddots & \vdots \\
 a_{m1} & a_{m2} & \cdots & a_{mn}
 \end{array} \right]
 \begin{matrix}
 \left(\begin{array}{c}
 x_1 \\
 x_2 \\
 \vdots \\
 x_n
 \end{array} \right) \\
 \mathbf{x}
 \end{matrix}
 =
 \begin{matrix}
 \left(\begin{array}{c}
 b_1 \\
 b_2 \\
 \vdots \\
 b_m
 \end{array} \right) \\
 \mathbf{b} \\
 \text{vettore} \\
 \text{termini noti}
 \end{matrix}
 \end{array}$$

$\mathbf{A} \mathbf{x} = \mathbf{b}$

vettore incognite

$$\sum_{j=1}^n a_{ij} x_j = b_i \quad 1 \leq i \leq m$$



Metodi Numerici per risolvere sistemi lineari $m=n$

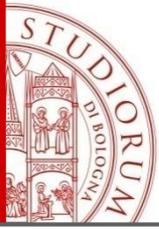
METODI DIRETTI

L'esatta soluzione viene costruita, in assenza di errori di arrotondamento nei dati e nei calcoli, in un numero finito di passi.

METODI ITERATIVI

La soluzione è ottenuta come limite di una successione di soluzioni di problemi lineari più semplici.

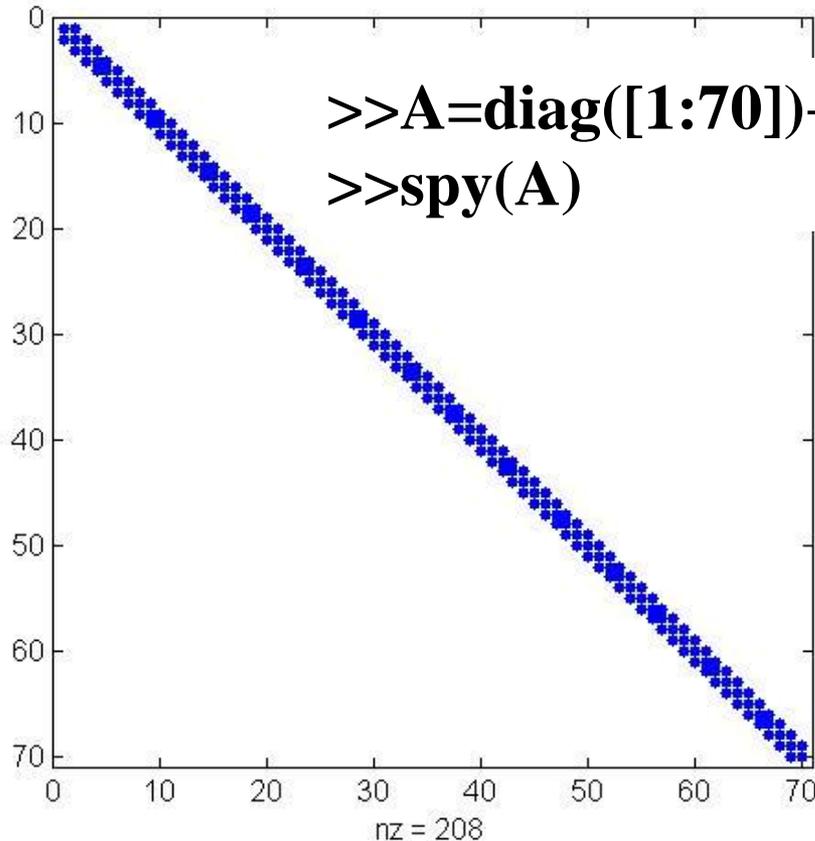
- La matrice A non viene modificata durante il calcolo.
- Anche in assenza di errori di arrotondamento si deve comunque operare un troncamento del procedimento risolutivo commettendo un errore
- Non è necessario formare esplicitamente la matrice A



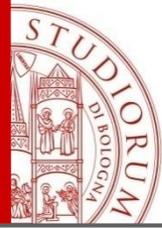
Matrici sparse

Una matrice è ritenuta **sparsa** quando il numero degli elementi che possono essere diversi dallo zero è di ordine $O(n)$

Esempio: matrice tridiagonale

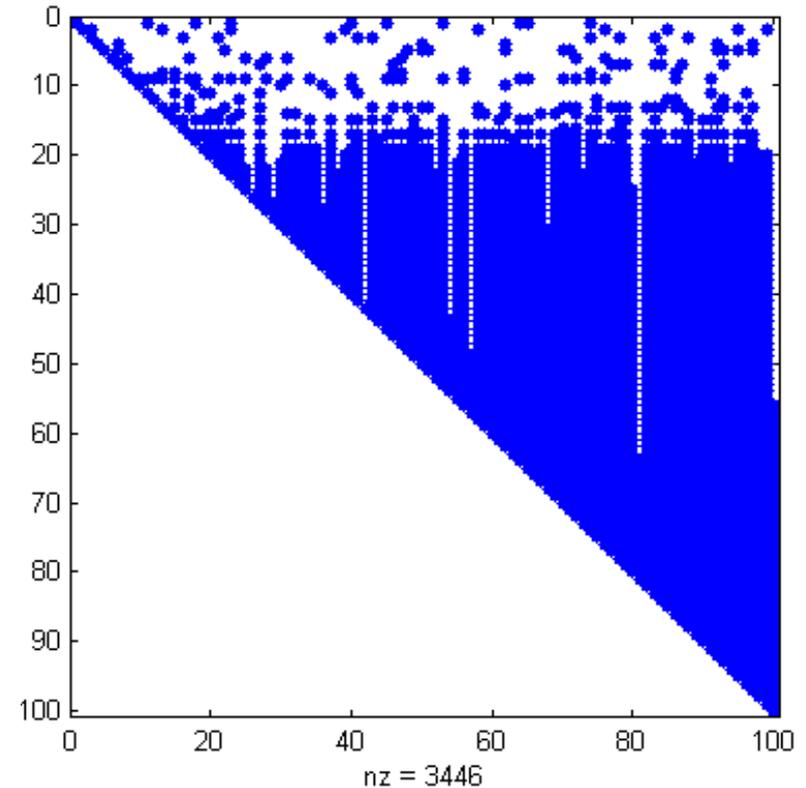
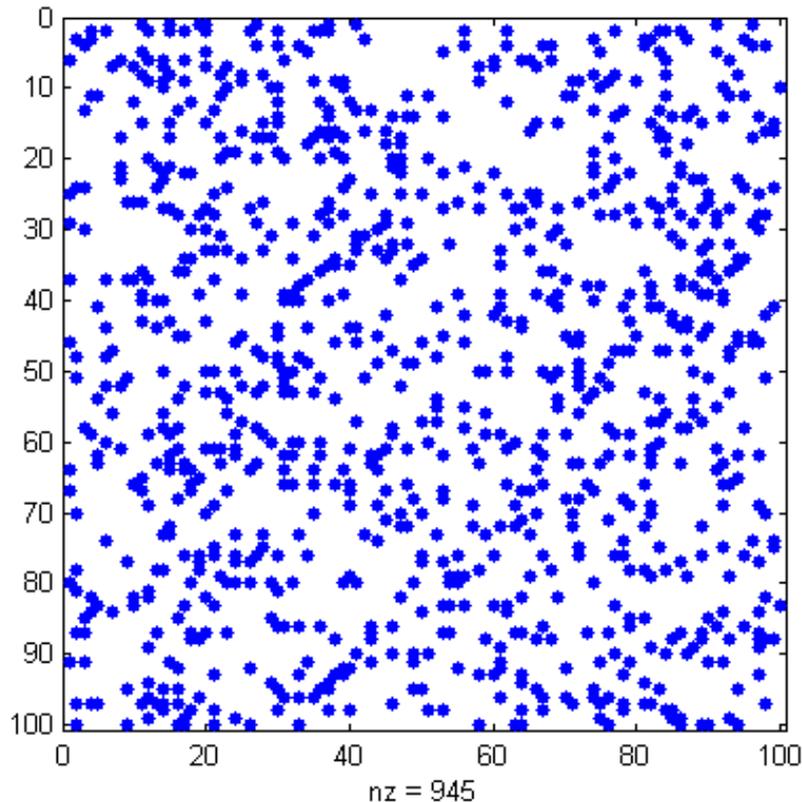


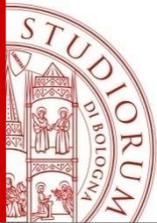
```
>>A=diag([1:70])+diag([1:69],-1)+diag([1:69],1);  
>>spy(A)
```



Matrici sparse, effetto fill-in

```
>> A=sprand(100,100,0.1)
>> [L,U,P]=lu(A);
>> spy(A);
>> spy(U);
```





Metodi iterativi

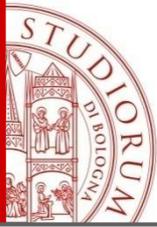
$$\mathbf{A} \mathbf{x} = \mathbf{b} \quad \det \mathbf{A} \neq \mathbf{0}$$

Partendo da un vettore iniziale $\mathbf{x}^{(0)}$ viene generata una successione “infinita” di vettori

$$\mathbf{x}^{(0)} \rightarrow \mathbf{x}^{(1)} \rightarrow \mathbf{x}^{(2)} \rightarrow \dots$$

che “*converge*” verso la soluzione cercata

- Ogni singolo passo richiede una quantità di calcoli paragonabile alla quantità necessaria nella moltiplicazione della matrice \mathbf{A} con un vettore. $O(n^2)$, $\mathbf{A} \in R^{n \times n}$
- Adatti per matrici di **grandi dimensioni, sparse, strutturate**



Costo Computazionale

- Numero di operazioni del metodo di Gauss,
$$n_1 = O(n^3/3)$$
- Metodo iterativo con **N** iterazioni,
$$n_2 = O(Nn^2)$$
 in generale.
- **Esempio** Se $n=10^5$, allora $n_1 \approx 3 \times 10^{14}$ e $n_2 \approx 10^{10}N$.
Tutto dipende del **numero di iterazioni**, però se questo numero non è troppo alto, per esempio $N = 3 \times 10^2$, allora si ottiene $n_2 \approx 3 \times 10^{12}$, quindi $n_1/n_2 = 100$: il tempo di calcolo del metodo diretto è cento volte più alto di quello del metodo iterativo.



Metodi iterativi: decomposizione

$$\mathbf{A} \mathbf{x} = \mathbf{b}$$

$$\det \mathbf{A} \neq \mathbf{0}$$

Decomponiamo

$$\mathbf{A} = \mathbf{N} - \mathbf{P}$$

$$\mathbf{N} \mathbf{x} = \mathbf{P} \mathbf{x} + \mathbf{b}$$

Partendo da un generico vettore $\mathbf{x}^{(0)}$ definiamo

$$\mathbf{N} \mathbf{x}^{(k)} = \mathbf{P} \mathbf{x}^{(k-1)} + \mathbf{b} \quad k=1,2,\dots$$

Calcoliamo la successione

$$\mathbf{x}^{(k)} = \mathbf{N}^{-1} \mathbf{P} \mathbf{x}^{(k-1)} + \mathbf{N}^{-1} \mathbf{b}$$

$$\mathbf{M} = \mathbf{N}^{-1} \mathbf{P}$$

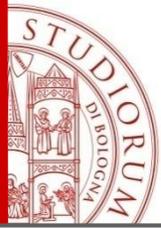
matrice di
iterazione

Fino a convergenza

Condizioni

$$\det \mathbf{N} \neq \mathbf{0}$$

$\mathbf{N} \mathbf{y} = \mathbf{z}$ sistema di “facile” soluzione



Studio della Convergenza

$$\mathbf{N} \mathbf{x}^{(k)} = \mathbf{P} \mathbf{x}^{(k-1)} + \mathbf{b}, \quad \mathbf{x}^{(k)} \text{ sol calcolata la passo } k$$

$$\mathbf{N} \mathbf{x} = \mathbf{P} \mathbf{x} + \mathbf{b}, \quad \mathbf{x} \text{ soluzione esatta di } \mathbf{A} \mathbf{x} = \mathbf{b}$$

sottraendo

$$\mathbf{N} (\mathbf{x}^{(k)} - \mathbf{x}) = \mathbf{P} (\mathbf{x}^{(k-1)} - \mathbf{x})$$

$$\mathbf{x}^{(k)} - \mathbf{x} = \mathbf{N}^{-1} \mathbf{P} (\mathbf{x}^{(k-1)} - \mathbf{x})$$

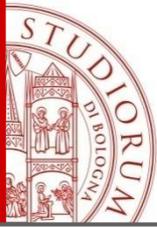
**Vettore
errore**

$$\mathbf{e}^{(k)} = \mathbf{M} \mathbf{e}^{(k-1)}$$

$$\mathbf{e}^{(k)} = \mathbf{M} (\mathbf{M} \mathbf{e}^{(k-2)}) = \mathbf{M}^2 \mathbf{e}^{(k-2)}$$

.....

$$\mathbf{e}^{(k)} = \mathbf{M}^k \mathbf{e}^{(0)} \quad \text{Vettore errore iniziale}$$



Metodi iterativi: convergenza

$$\mathbf{e}^{(k)} = \mathbf{M}^k \mathbf{e}^{(0)}$$

$$\mathbf{e}^{(0)} = \mathbf{x}^{(0)} - \mathbf{x}$$

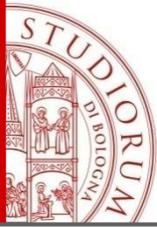
Condizione necessaria e sufficiente per la convergenza

$$\lim_{k \rightarrow \infty} \|\mathbf{e}^{(k)}\| = 0 \quad \text{per ogni } \mathbf{e}^{(0)} \quad \text{è che} \quad \lim_{k \rightarrow \infty} \mathbf{M}^k = \mathbf{0}$$

Se $\lim_{k \rightarrow \infty} \mathbf{M}^k = \mathbf{0}$ allora la matrice \mathbf{M} si dice convergente

Teorema.

La matrice di iterazione \mathbf{M} è convergente se e solo se tutti i suoi autovalori sono in valore assoluto minori di 1, cioè: $\rho(\mathbf{M}) < 1$



Metodi iterativi: convergenza

$$\rho(M) < 1$$

$$M = N^{-1} P$$

$$\rho(M) = \max_{1 \leq i \leq n} |\lambda_i(M)|$$

raggio spettrale

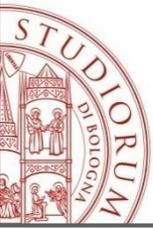
Condizione di difficile verifica

Lemma. Per ogni norma naturale $\|\cdot\|$ e per ogni matrice M

$$\rho(M) \leq \|M\|$$

Condizione sufficiente:

Teorema. Se esiste una norma matriciale indotta $\|\cdot\|$ per cui $\|M\| < 1$ il metodo iterativo con matrice di iterazione M è convergente.



Starting - Stopping

x_0 Se è disponibile: una buona stima della soluzione
altrimenti: zero

Criterio di Arresto

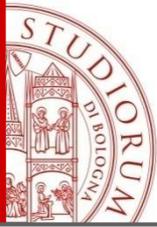
Definito il vettore residuo al passo k : $r_k = b - Ax_k$

si utilizza il criterio:

$$\|r_k\| < \varepsilon \|b\|,$$

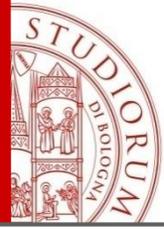
In questo modo, se la matrice è ben condizionata, a residui piccoli corrisponderanno errori piccoli

$$\frac{\|x - x_k\|}{\|x\|} \leq \frac{\|A^{-1}\| \|r_k\|}{\|x\|} \leq \kappa(A) \frac{\|r_k\|}{\|b\|} \leq \varepsilon \kappa(A) \quad \text{con } x_0=0$$



Criteri di arresto: osservazioni

- Deve comunque essere previsto un **numero massimo di iterazioni**
- Può accadere che un metodo iterativo la cui matrice di iterazione M è tale che $\rho(M) < 1$, per gli effetti indotti dagli errori di arrotondamento non converga in pratica. Questo accade, in particolare, quando la **matrice A è fortemente mal condizionata e $\rho(M)$ molto vicino a 1.**

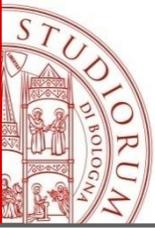


Metodi iterativi/Costruzione

Strategia generale per costruire metodi iterativi: splitting

$$\boxed{\mathbf{A} = \mathbf{N} - \mathbf{P}} \quad \longleftrightarrow \quad \boxed{\mathbf{A} = \mathbf{D} + \mathbf{L} + \mathbf{U}}$$

$$\mathbf{A} = \begin{pmatrix} & & \mathbf{U} \\ & \mathbf{D} & \\ \mathbf{L} & & \end{pmatrix}$$



Metodo di Gauss-Seidel

Carl Friedrich Gauss (1777-1855), Philipp Ludwig von Seidel (1821-1896)

Il metodo di Gauss-Seidel corrisponde alla scelta

$$N=D+L, \quad P=-U$$

Da $Nx^{(k)} = Px^{(k-1)} + b$ si ottiene

$$x^{(k)} = D^{-1} (b - Lx^{(k)} - Ux^{(k-1)})$$

$$N = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \dots & 0 \\ a_{31} & a_{32} & a_{33} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix} \quad P = \begin{bmatrix} 0 & -a_{12} & -a_{13} & \dots & -a_{1n} \\ 0 & 0 & -a_{23} & \dots & -a_{2n} \\ 0 & 0 & 0 & \dots & -a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

Metodo di Gauss-Seidel

$$x^{(k)} = D^{-1} (b - Lx^{(k)} - Ux^{(k-1)})$$

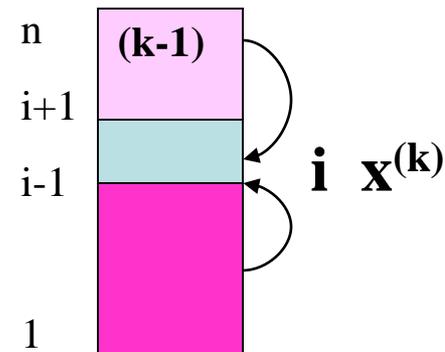
Per $k=1,2,\dots,\text{itermax}$

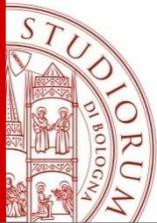
Per $i=1,2,\dots,n$

$$x_i^{(k)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^n a_{ij} x_j^{(k-1)} \right]$$

Se il criterio di arresto scelto e' verificato: stop

Occupazione di memoria:
1 solo vettore x , b e matrice A
(opportunamente memorizzata)





Convergenza

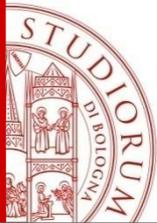
Matrice di iterazione $M = N^{-1} P = -(D+L)^{-1} U$

La condizione **NECESSARIA E SUFFICIENTE** per la convergenza sulla matrice di iterazione

$$\rho(M) < 1$$

è difficile da calcolare in generale.

Per le applicazioni è importante stabilire delle condizioni **SUFFICIENTI** per la convergenza del metodo iterativo; in altre parole è importante individuare delle classi di matrici per le quali un particolare metodo converge



Convergenza

Definizione

Una matrice A è detta a predominanza diagonale stretta se i suoi elementi verificano la condizione

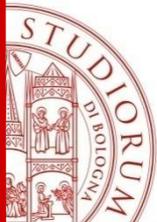
$$|a_{ii}| > \sum_{i \neq j} |a_{ij}| \quad \forall i = 1, 2, \dots, n$$

Teorema.

Se la matrice A è a **predominanza diagonale in senso stretto** (per righe o per colonne), allora il metodo di Gauss-Seidel è **convergente**.

Teorema.

Se la matrice A è **simmetrica /hermitiana e definita positiva**, allora il metodo di Gauss-Seidel è **convergente**.



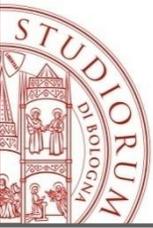
Metodi di rilassamento

Obiettivo: accelerare la convergenza dei metodi iterativi

$$x_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \omega x_i^{(k + \frac{1}{2})} \quad k = 1, 2, \dots$$

**Componente
calcolata al passo
k-esimo**

**Componente
calcolata a partire
da $x^{(k)}$ mediante
uno dei metodi
iterativi**



Metodi di rilassamento

(Successive Over Relaxation)

Partendo dal metodo di Gauss Seidel

$$x^{(k+1)} = (1 - \omega)x^{(k)} + \omega \left[(b - Lx^{(k+1)} - Ux^{(k)})D^{-1} \right]$$

$\omega=1$ processo normale: Gauss-Seidel

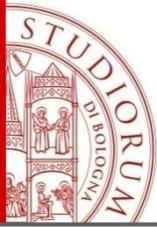
$\omega>1$ **SOR** (Successive Over Relaxation), sovrarilassamento

$\omega<1$ sottorilassamento

$$x^{(k+1)} = \underbrace{(D + \omega L)^{-1} [(1 - \omega)D - \omega U]}_{\mathbf{H}(\omega)} x^{(k)} + \omega(D + \omega L)^{-1} b$$

$\mathbf{H}(\omega)$ matrice di iterazione

Condizione di convergenza: **$\rho(\mathbf{H}(\omega)) < 1$**



Metodi di rilassamento (SOR)

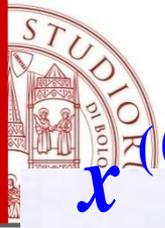
Per $k=1,2,\dots,\text{itermax}$

Per $i=1,2,\dots,n$

$$x_i^{(GS)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right]$$
$$x_i^{(k+1)} = (1 - \omega) x_i^{(k)} + \omega x_i^{(GS)}$$

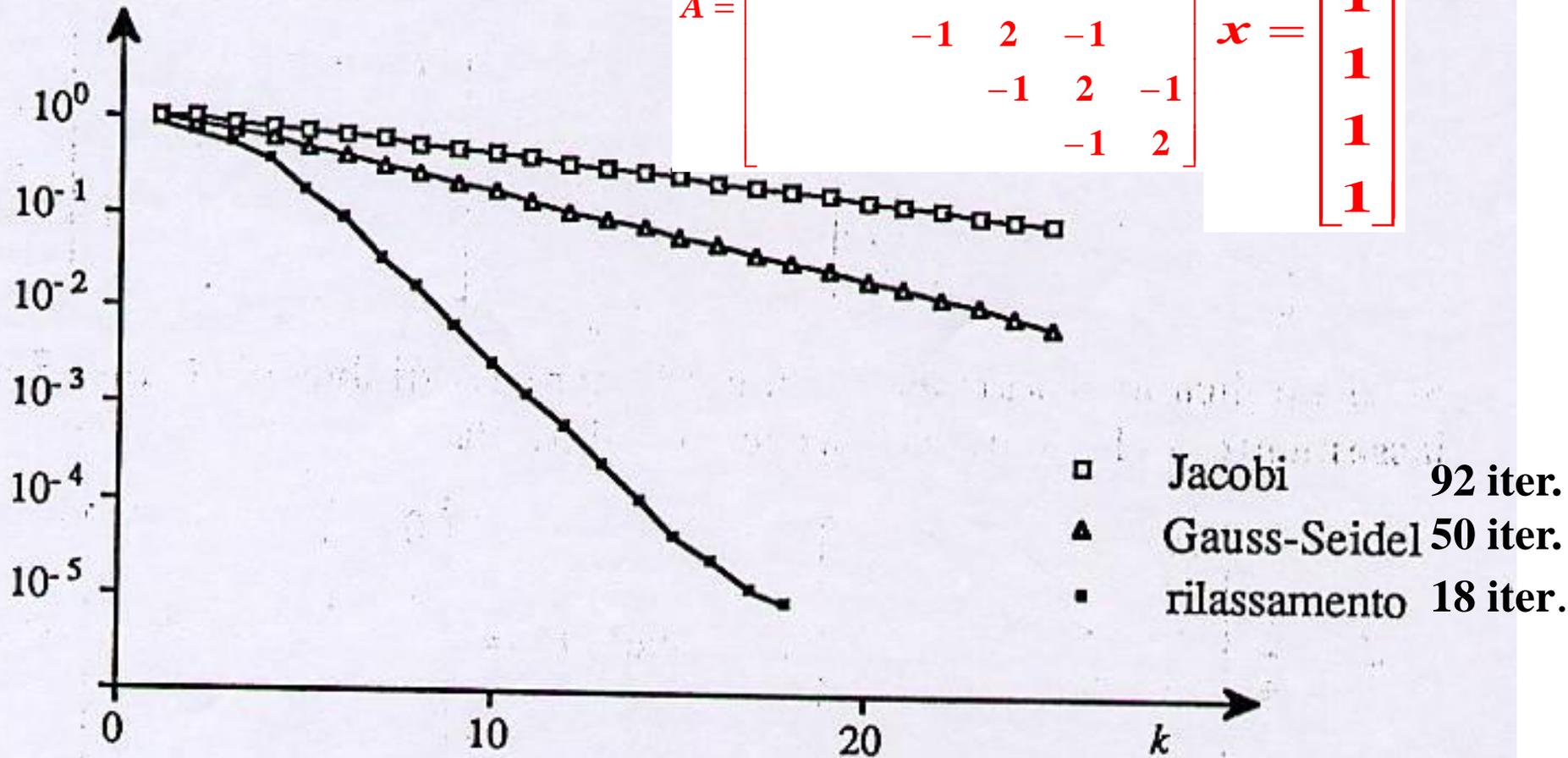
Se il criterio di arresto scelto e' verificato: stop

ESEMPIO:

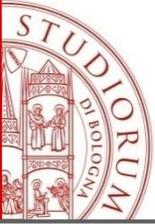


$$x^{(0)} = 0 \quad \varepsilon = 10^{-5}$$

$$A = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & -1 & 2 & -1 & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$



- Andamento degli errori dei metodi iterativi di Jacobi, Gauss-Seidel e rilassamento.



Metodi di rilassamento: convergenza

Teorema Kahan. Per ogni matrice di iterazione di un metodo di rilassamento $H(\omega)$, risulta

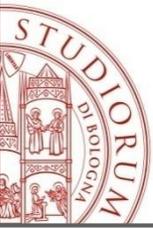
$$\rho(H(\omega)) \geq |\omega - 1|$$

Quindi condizione **necessaria** per la convergenza è che

$$|\omega - 1| < 1$$

e se ω è reale, allora una condizione **NECESSARIA** è

$$0 < \omega < 2$$

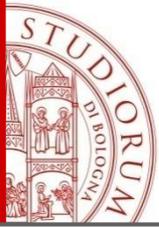


Metodi di rilassamento

La condizione $0 < \omega < 2$ del teorema di Kahan risulta anche sufficiente per la convergenza dei metodi di rilassamento SOR se la matrice A è simmetrica definita positiva.

Teorema Ostrowski-Reich.

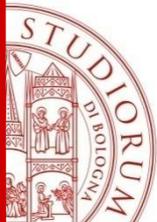
Se A è simmetrica definita positiva (SPD),
il metodo SOR converge se e solo se
 ω è un numero reale tale che $0 < \omega < 2$



Esempio di SOR

$$\left\{ \begin{array}{l} 4X_1 + 2X_2 = 2 \\ 2X_1 + 10X_2 + 4X_3 = 6 \\ 4X_2 + 5X_3 = 5 \end{array} \right.$$

Soluzione: $(X_1, X_2, X_3) = (0.41379, 0.17241, 0.86206)$

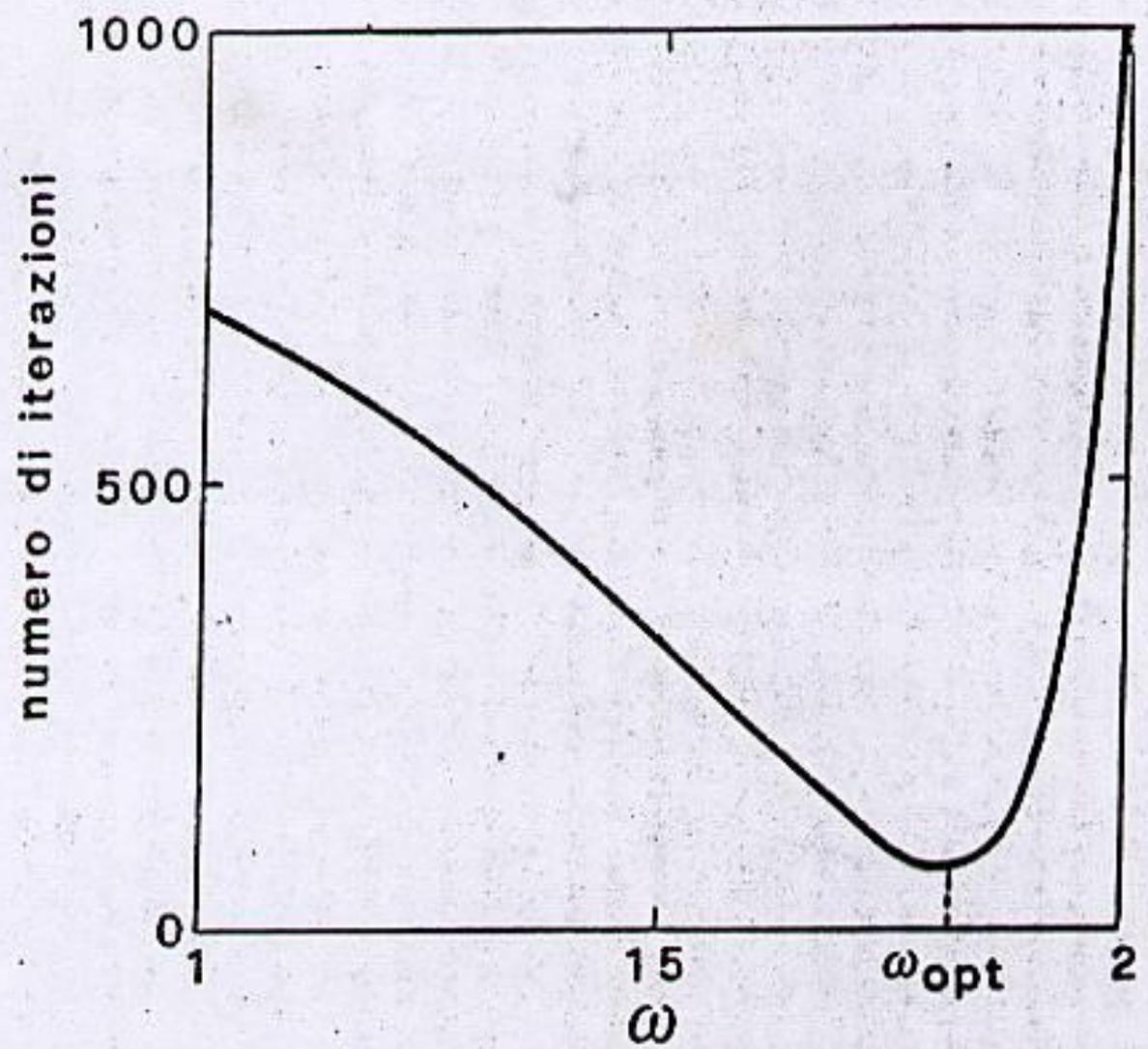


Effetto del parametro ω

Esempio con il metodo SOR con $n_{\max}=50$ e $\text{tol}=0.000001$

$$\frac{\|x_k - x_{k-1}\|_{\infty}}{\|x_k\|_{\infty}} \leq \text{tol}$$

ω	Numero iterazioni	ω	Numero Iterazioni
0,7	33	1,25	12
0,8	27	1,3	14
0,9	22	1,4	17
1	17	1,5	22
1,1	13	1,6	30
1,15	10	1,7	43
1,175	10		
1,2	10		



– Andamento tipico del numero di iterazioni richiesto dallo schema SOR per raggiungere una prefissata tolleranza in funzione di ω .

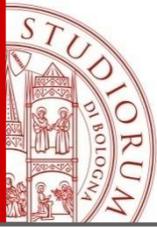
Metodi di discesa:

- Metodo del Gradiente (Steepest Descent)
- Metodo del Gradiente Coniugato (Conjugate Gradient)

Problema

$$Ax=b$$

Ip: A matrice Sim. Def.Pos. SPD (o SND)



Principio dei metodi di discesa

Forma quadratica:

$$f(x) = \frac{1}{2} x^T A x - x^T b + c$$

A matrice
b, x vettori
c scalare

$$\min_{x \in \mathbb{R}^n} f(x) \quad x : \nabla f(x) = \mathbf{0}$$

$$\nabla f(x) = Ax - b = \mathbf{0} \quad \text{e} \quad H(x) = A \text{ def. pos.}$$

Trovare il punto critico equivale a risolvere

$$A x = b$$

Se **A** è simmetrica e definita positiva allora questo punto è anche un minimo di $f(x)$.

Esempio: bivariato

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

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ -8 \end{bmatrix} \quad c = 0$$

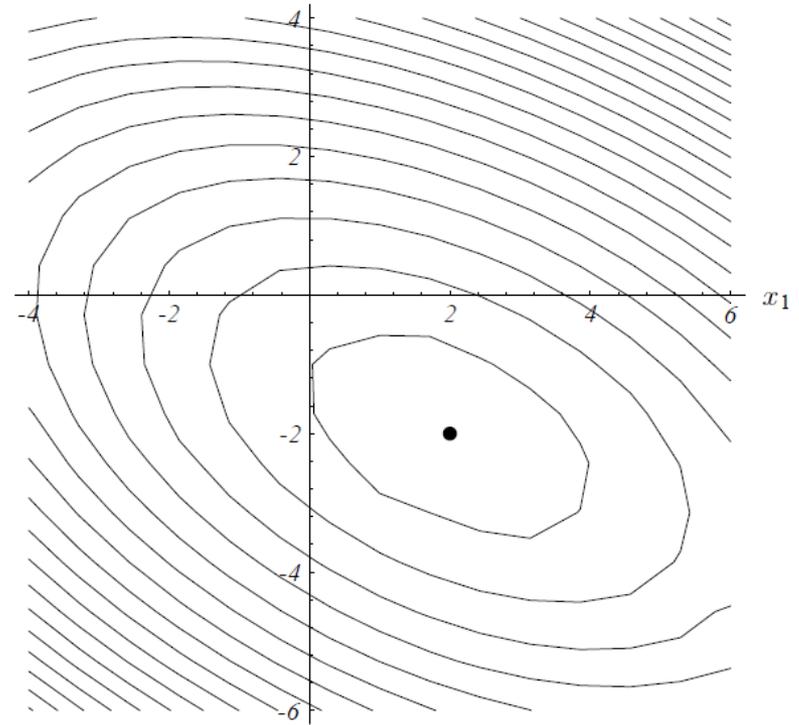
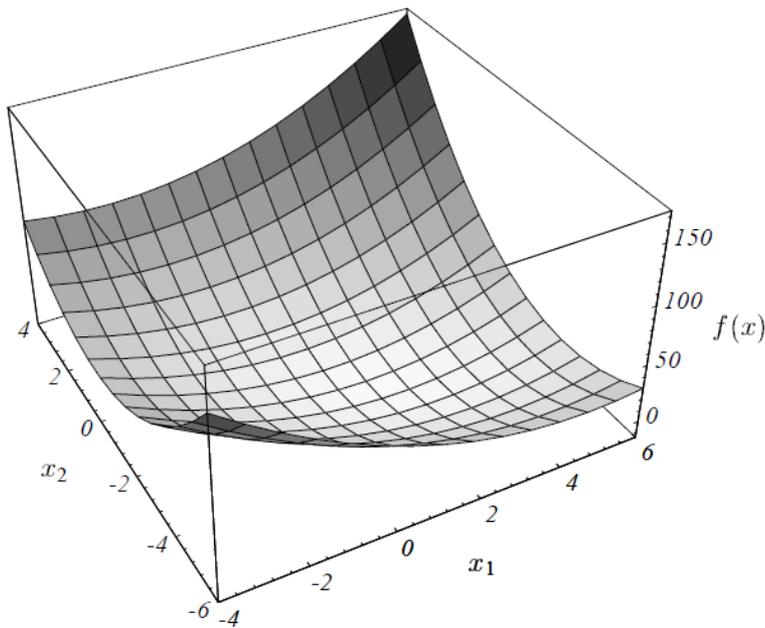
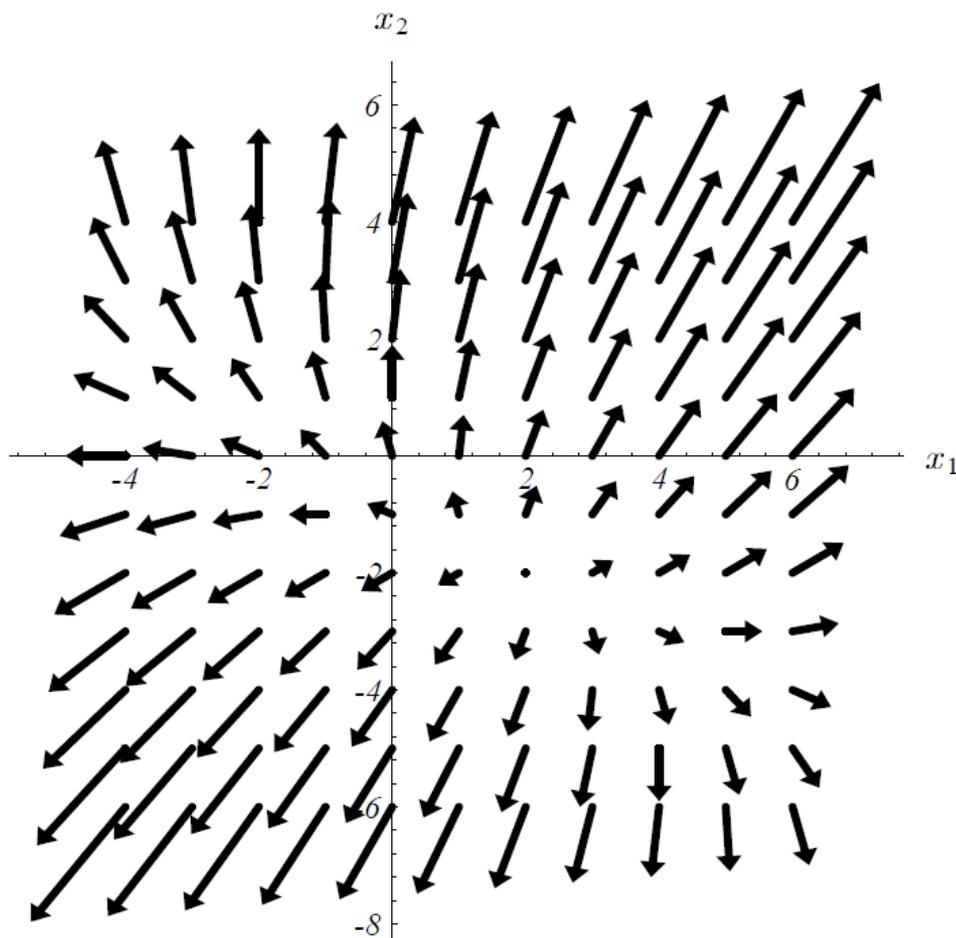


Grafico di $f(x)$
Il punto di minimo è la soluzione
di $Ax=b$

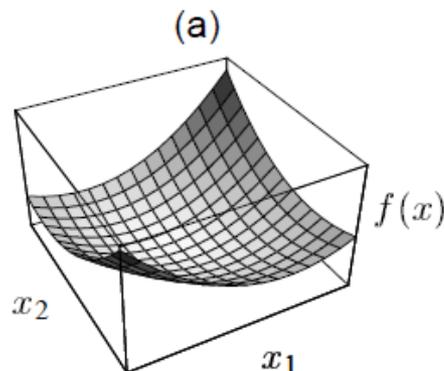
Curve di livello,
su ciascuna ellisse $f(x)=\text{cost.}$



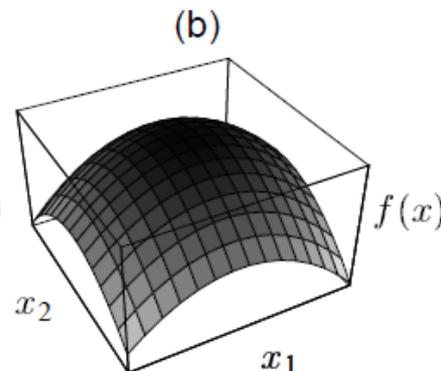
Gradiente della forma quadratica; la direzione dei vettori gradienti è verso la massima crescita di $F(\mathbf{x})$

A simmetrica

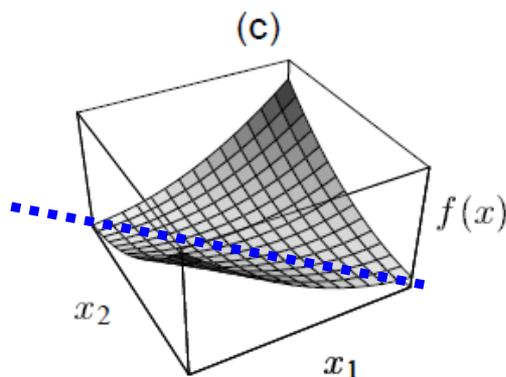
**Definita
positiva**



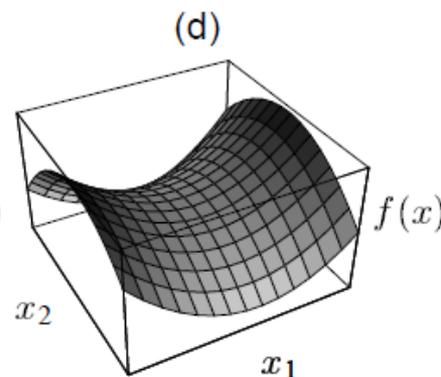
**Definita
negativa**



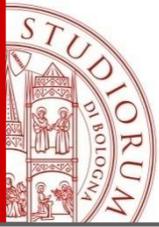
Singolare
soluzioni



Indefinita



Se A non è simmetrica si trova la sol. del sistema $\frac{1}{2}(A^T + A)x = b$



Metodi di discesa

Scelta di p_k non nulla direzione di discesa ed uno scalare $\alpha > 0$ tale che posto:

$$x_{k+1} = x_k + \alpha p_k$$

si abbia:

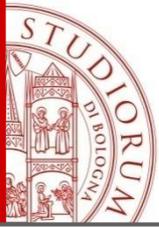
$$f(x_{k+1}) \leq f(x_k), \quad \text{vale} = \Leftrightarrow x_{k+1} = x_k = x$$

**Condizione
per direzione di discesa ammissibile:**

$$\langle \nabla f(x_k), p_k \rangle < 0$$

Formula di Taylor $f(x_{k+1}) = f(x_k) + \alpha_k \langle \nabla f(x_k), p_k \rangle + o(\alpha^2)$

Se $\langle \nabla f(x_k), p_k \rangle < 0$ allora $f(x_{k+1}) \leq f(x_k)$



Metodo del gradiente

SCELTA DELLA DIREZIONE DI DISCESA:

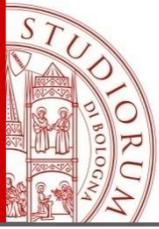
Strategia più semplice per calcolare il minimo di $f(x)$:
Poiché il gradiente di f è direzione di massima crescita,

seguire localmente la direzione di massima pendenza

$$p_k = -\nabla f(x_k)$$

$$-\nabla f(x) = -\left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]^T = b - Ax = r(x)$$

$r(x)$ residuo del sistema lineare



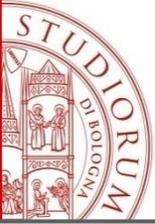
Metodo del gradiente

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{r}_k$$

Come scegliere α_k ?

$$f(\mathbf{x}_{k+1}) = \min_{\alpha_k} f(\mathbf{x}_k + \alpha_k \mathbf{r}_k)$$



Calcoliamo l' α ottimale

$$f(x_{k+1}) = \min_{\alpha_k} f(x_k + \alpha_k r_k)$$

$$f(x_k + \alpha_k r_k) = \frac{1}{2} (x_k + \alpha_k r_k)^T A (x_k + \alpha_k r_k) - (x_k + \alpha_k r_k)^T b$$

$$\frac{\partial f}{\partial \alpha} = (x_k + \alpha_k r_k)^T A r_k - r_k^T b = 0$$

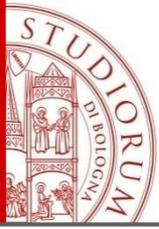
$$x_k^T A r_k + \alpha_k r_k^T A r_k - r_k^T b = 0$$

$$r_k^T (A x_k - b) + \alpha_k r_k^T A r_k = 0$$

$$-r_k^T r_k + \alpha_k r_k^T A r_k = 0$$



$$\alpha_k = \frac{r_k^T r_k}{r_k^T A r_k}$$

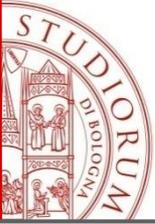


Calcolo residuo

Per calcolare il residuo conviene (dal punto di vista della propagazione degli errori) utilizzare la rappresentazione:

$$\begin{aligned}r_{k+1} &= b - Ax_{k+1} \\ &= b - A(x_k + \alpha_k p_k) \\ &= r_k - \alpha_k A p_k\end{aligned}$$

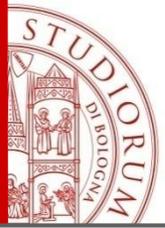
Per il metodo dei gradienti: $(p_k = r_k)$



Algoritmo del gradiente

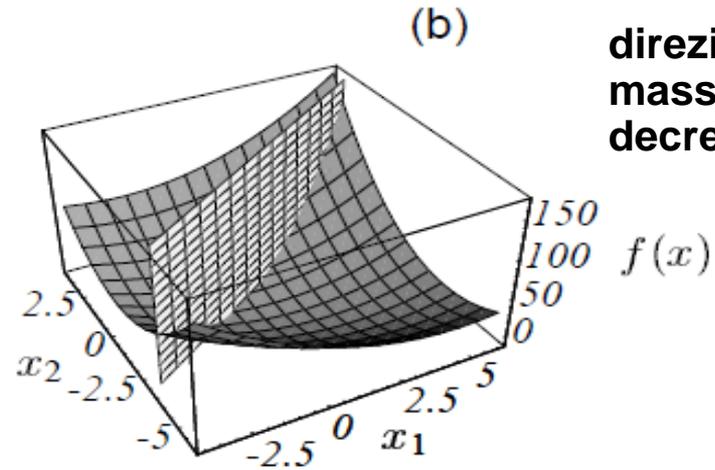
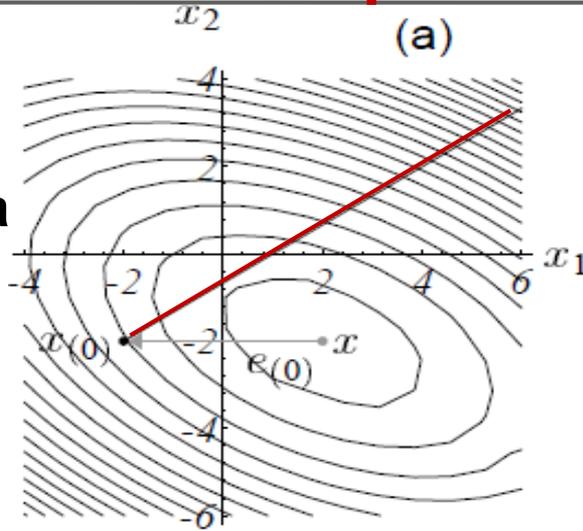
```

$$p_0 = r_0 = b - Ax_0$$
for  $k = 0, 1, \dots,$   
  if  $r_k = 0$  then  $x = x_k$  stop  
  else  $w_k = Ar_k$   
    
$$\alpha_k = \frac{r_k^T r_k}{\langle w_k, r_k \rangle}$$
  
    
$$x_{k+1} = x_k + \alpha_k r_k$$
  
    
$$r_{k+1} = r_k - \alpha_k w_k$$
  
  endif  
endfor
```

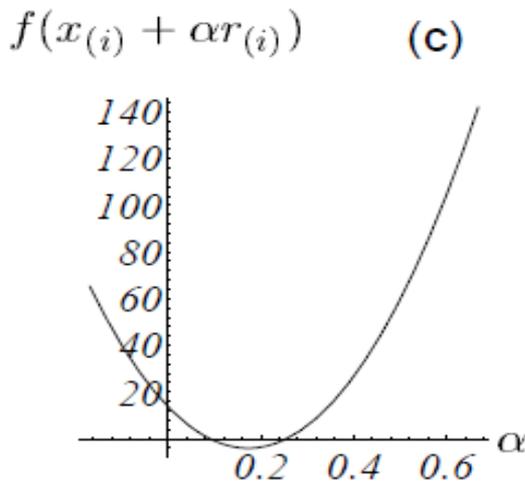


Metodo del gradiente o della più ripida discesa

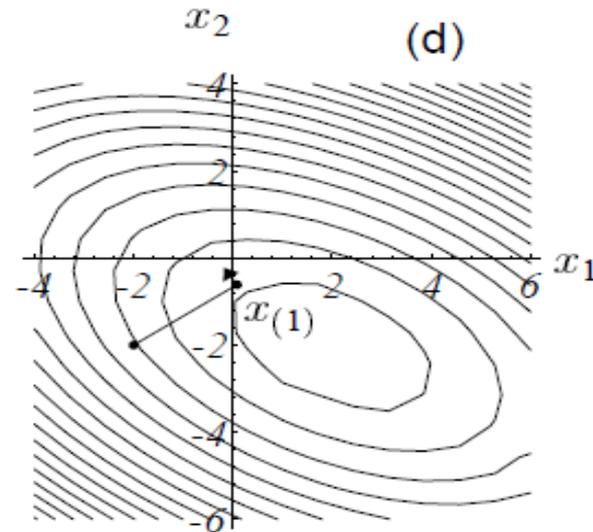
Partenza da $x_0 = [-2, -2]$



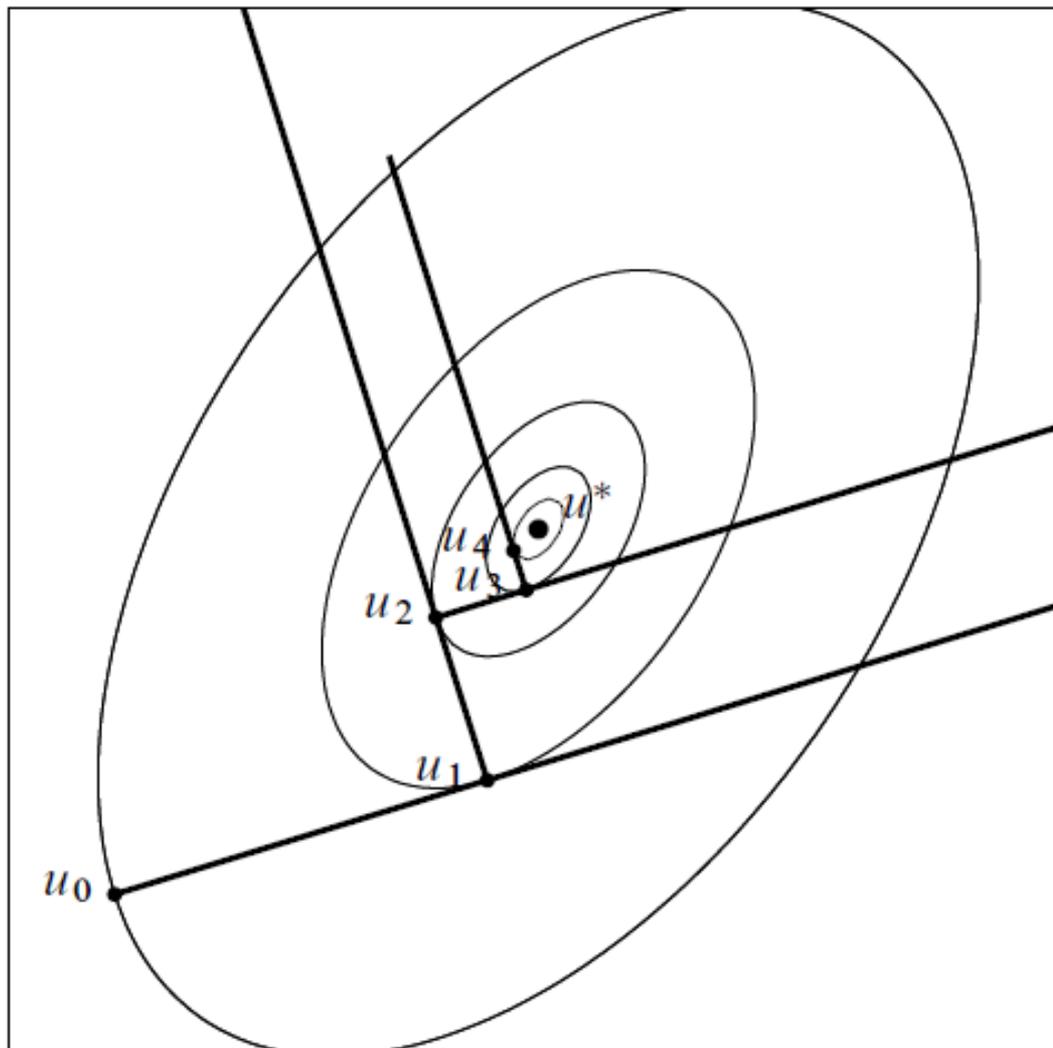
direzione di massima decrescita



parabola intersezione delle 2 superfici in (b)

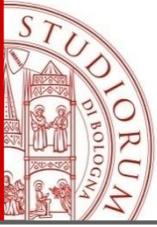


Il gradiente nel punto di minimo della intersezione è ortogonale al gradiente al passo precedente

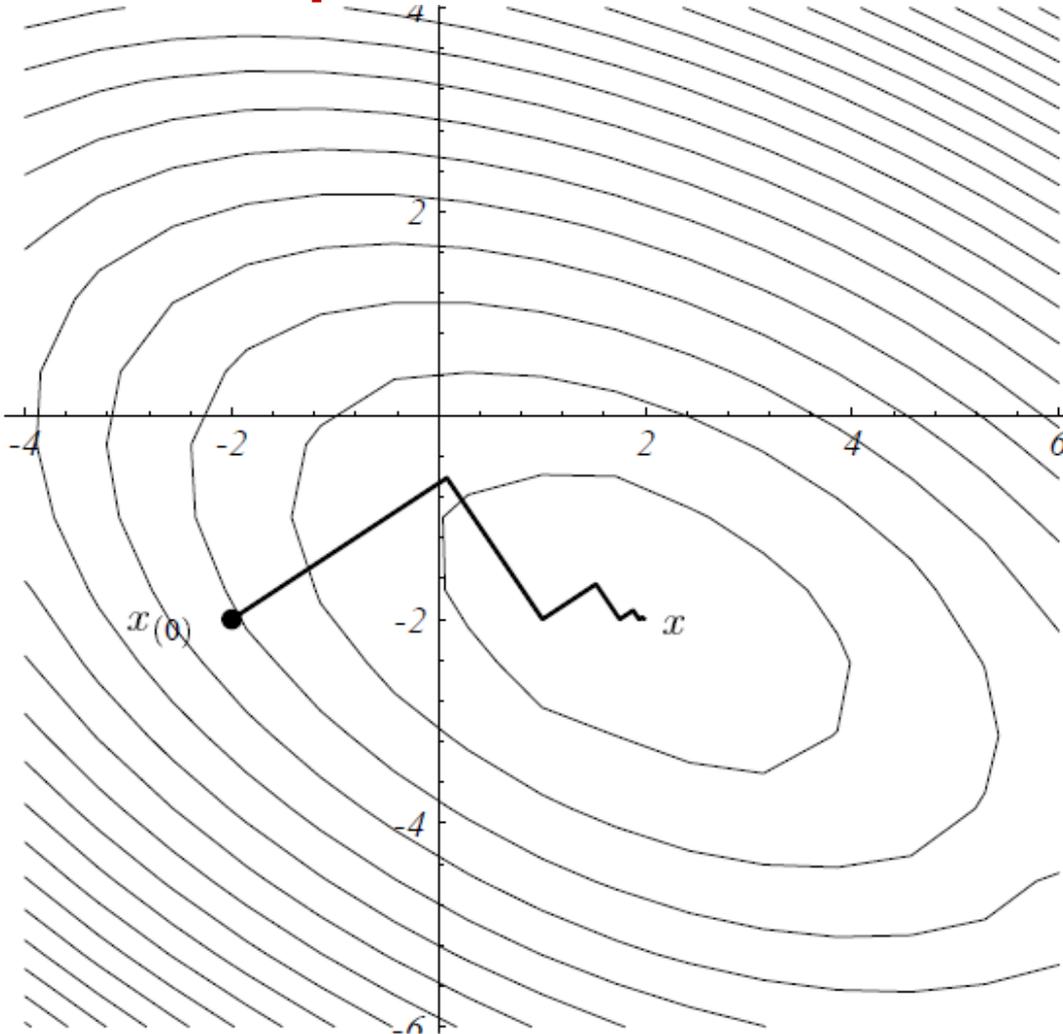


Se la matrice è SPD allora le isolinee sono sempre ellissi.

La direzione di discesa è sempre ortogonale alla precedente direzione di discesa, in due dimensioni significa alternare tra due diverse direzioni ortogonali di discesa

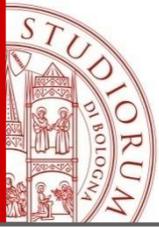


Metodo del gradiente: comportamento a 'zig zag'



**Partenza da $[-2, -2]$
Soluzione $[2, -2]$**

**Il gradiente nel
punto di minimo
è ortogonale al
gradiente del
passo precedente...**



Convergenza del Metodo del gradiente

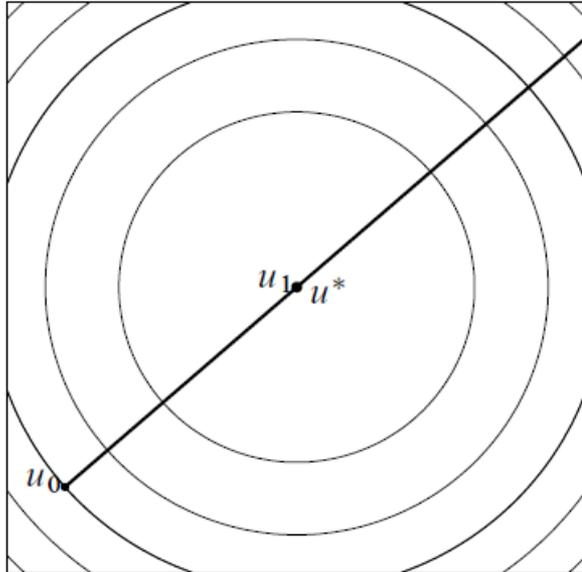
$$e_k = x_k - x \quad \text{Errore}$$

$$\|e_k\|_A = e_k^T A e_k \quad \text{Norma energia}$$

$$\|e_k\|_A \leq \left(\frac{k_2(A) - 1}{k_2(A) + 1} \right)^k \|e_0\|_A$$

Più la matrice è malcondizionata (l'indice di condizionamento è grande) più la convergenza è lenta.

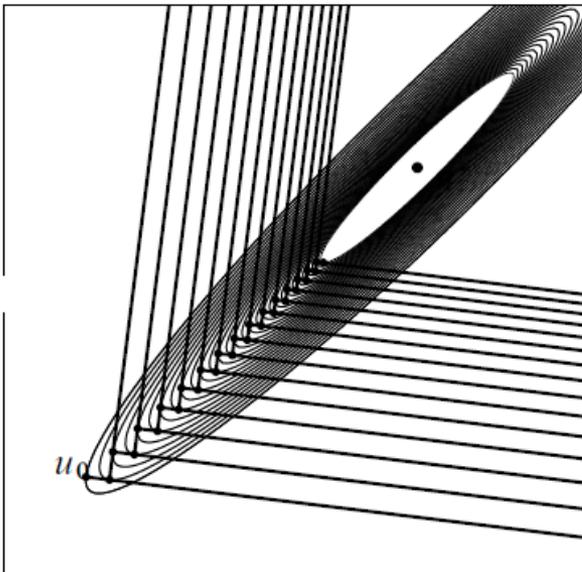
Esempio: convergenza del Metodo del gradiente



**Convergenza
in un solo
passo**

$$K(A) = \frac{\lambda_{\max}}{\lambda_{\min}}$$

Quando tutti gli autovalori di A sono uguali, gli ellissoidi sono sferici, $K(A)=1$. Non importa il punto di partenza, il residuo punta al centro, il metodo converge in una iterazione.

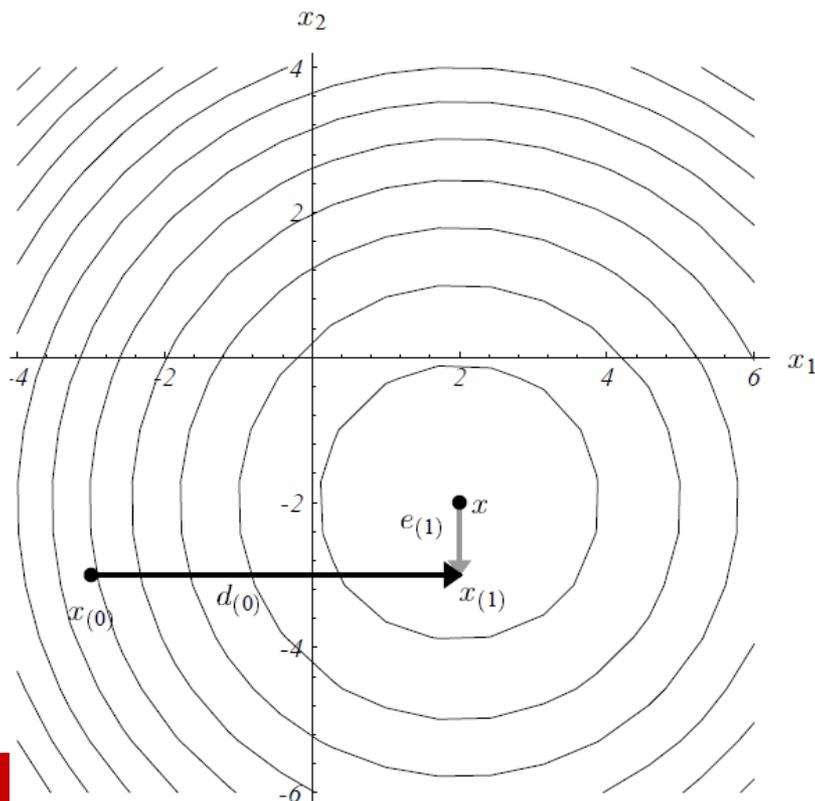


**Convergenza
lenta**

Direzioni coniugate

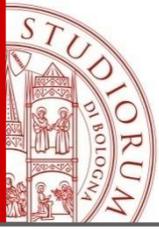
Nel metodo del gradiente o della più ripida discesa una direzione può essere riusata. Per evitarlo, prendiamo n direzioni di ricerca ortogonali p_0, p_1, \dots, p_{n-1} .

In ogni direzione facciamo un solo passo e della giusta lunghezza per allinearci con x . Dopo n passi abbiamo x .



Usiamo gli assi del sistema coord. come direzioni di ricerca. Il primo passo (orizzontale) corregge la coordinata x_1 , e con il secondo passo (verticale) si trova la soluzione.

Nota: e_1 è ortogonale a d_0



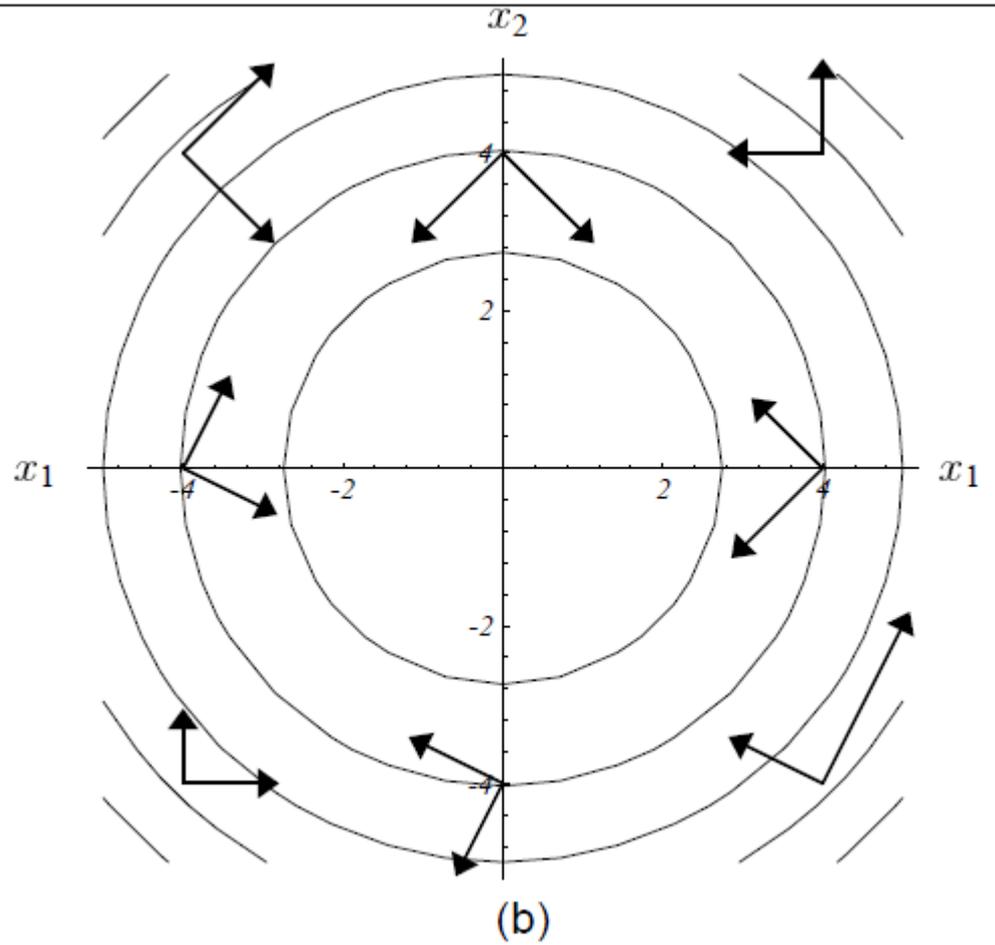
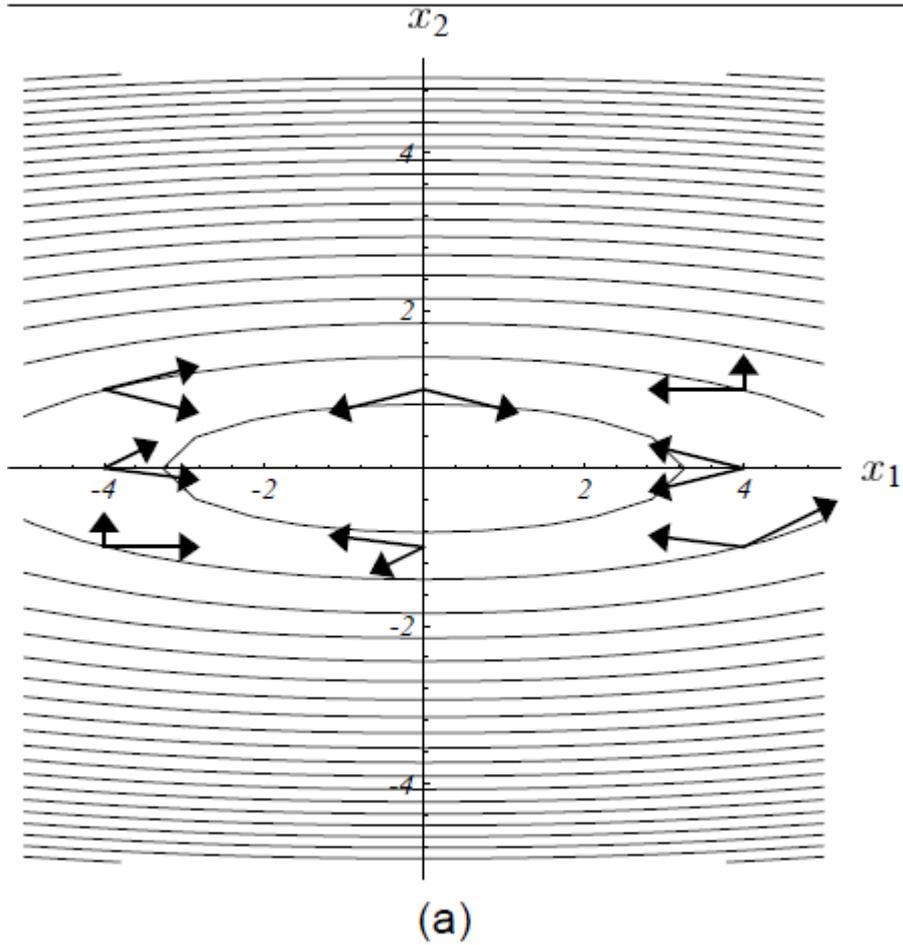
Direzioni coniugate

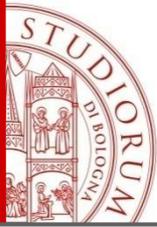
$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$$

Definiamo 2 vettori \mathbf{p}_k e \mathbf{p}_j **A-ortogonali** o **coniugati** se

$$\mathbf{p}_k^T \mathbf{A} \mathbf{p}_j = \mathbf{0}$$

Tale scelta consentirà di avere la soluzione con un numero di iterazioni non superiore all'ordine della matrice .





Scelta direzioni di discesa

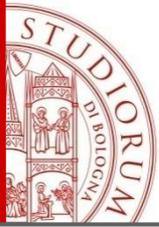
$$p_k = \begin{cases} r_0 & \text{se } k = 0 \\ r_k + \beta_k p_{k-1} & \text{se } k \geq 1 \end{cases}$$

Fattore di correzione

β_k è tale che $p_k^T A p_{k-1} = 0$

$$p_k^T A p_{k-1} = r_k^T A p_{k-1} + \beta_k p_{k-1}^T A p_{k-1} = \mathbf{0}$$

$$\beta_k = -\frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}$$



Scelta di α ottimale

Sostituendo r_k con p_k nell'espressione per alfa ottimo si ha per il metodo dei gradienti coniugati:

$$\alpha_k = \frac{p_k^T r_k}{p_k^T A p_k}$$

La direzione p_k è una direzione di decrescita della forma $f(x)$, si ha infatti:

$$p_k^T \nabla f(x) = -p_k^T r_k = -(r_k^T r_k + \beta_k \underbrace{p_{k-1}^T r_k}_{=0})^T = -r_k^T r_k < 0$$

se α ottimo

Algoritmo CG Hestenes-Stiefel (1952)



```

$$p_0 = r_0 = b - Ax_0$$
for  $k = 0, 1, \dots, n-1$   
  if  $r_k = 0$  then  $x = x_k$  stop  
  else  $w_k = Ap_k$   
    
$$\alpha_k = \frac{p_k^T r_k}{\langle w_k, p_k \rangle}$$
  
    
$$x_{k+1} = x_k + \alpha_k p_k$$
  
    
$$r_{k+1} = r_k - \alpha_k w_k$$
  
    
$$\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$$
  
    
$$p_{k+1} = r_{k+1} + \beta_{k+1} p_k$$
  
  endif  
endfor
```



Convergenza

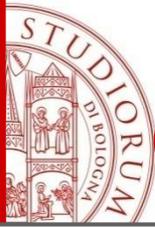
Gradiente coniugato

CG converge in n iterazioni, in teoria.

Per effetto degli errori di arrotondamento il residuo perde in precisione ed eventuali cancellazione numeriche nei vettori direzione di ricerca fanno perdere la A -ortogonalità.

Il primo problema è risolto ricalcolando di tanto in tanto il residuo come

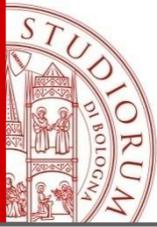
$$\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$$



Convergenza Gradiente coniugato

$$\|e_k\|_A \leq 2 \left[\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \right] \|e_0\|_A$$

La convergenza è funzione dell'indice di condizionamento

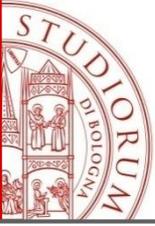


Complessità Gradiente Coniugato

Il maggior onere è dato dal calcolo del prodotto matrice-vettore, se il numero di iterazioni è pari a n (dimensione della matrice) si ha allora un costo complessivo di

$$\text{Complessità } O(n^3)$$

Il metodo diventa interessante per risolvere sistemi lineari con matrice sparsa e/o quando il numero di iterazioni necessarie è decisamente inferiore ad n . (Si vedano a questo proposito le tecniche di preconditionamento)



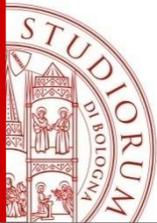
Starting - Stopping

x_0 Se è disponibile: una buona stima della soluzione
altrimenti: zero

Criterio di Arresto

Di solito $\|r_k\| < \varepsilon \|r_0\|$

$$\frac{\|x - x_k\|}{\|x\|} \leq \frac{\|A^{-1}\| \|r_k\|}{\|x\|} \leq \kappa(A) \frac{\|r_k\|}{\|b\|} \leq \varepsilon \kappa(A) \quad \text{con } x_0=0$$



Precondizionamento

$$\mathbf{A} \mathbf{x} = \mathbf{b}$$

Precondizionatore

Sinistro

$$\mathbf{M}^{-1} \mathbf{A} \mathbf{x} = \mathbf{M}^{-1} \mathbf{b}$$

Destro

$$\mathbf{A} \mathbf{M}^{-1} \mathbf{y} = \mathbf{b}$$

$$\mathbf{y} = \mathbf{M} \mathbf{x}$$

Centrato

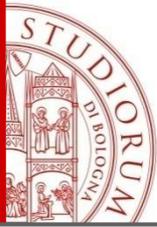
$$\mathbf{M}_L^{-1} \mathbf{A} \mathbf{M}_R^{-1} \mathbf{y} = \mathbf{M}_L^{-1} \mathbf{b}$$

$$\mathbf{y} = \mathbf{M}_R \mathbf{x}$$

In generale non è possibile individuare a priori il precondizionatore ottimale

M è un buon precondizionatore per A se:

- $\mathbf{M}^{-1} \mathbf{A}$ *assomiglia* ad un matrice normale (ideale: identità) ed i suoi autovalori sono contenuti in una regione sufficientemente piccola del piano complesso
- Basso costo computazionale
- Poca memoria



Precondizionatori

- Precondizionatori diagonali

$$M = \text{diag}(A)$$

- Fattorizzazioni LU incomplete (ILU) e fattorizzazioni incomplete di Cholesky (IC)

$$M = L_{in} U_{in}$$

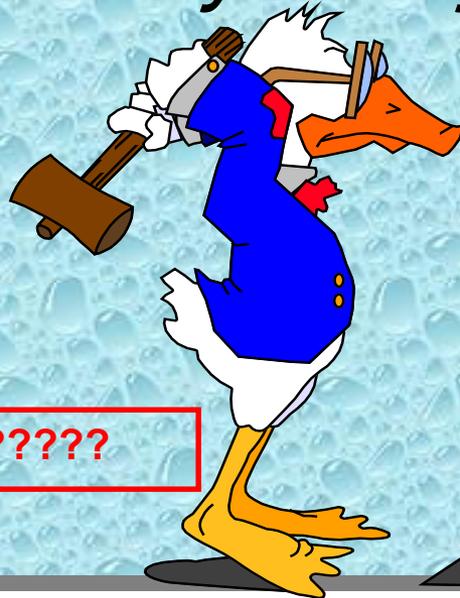
L_{in}, U_{in} approssimazioni della fattorizzazione LU di A

ILU(0),IC(0) dove “0” indica che non è stato introdotto fill-in durante il processo di fattorizzazione.

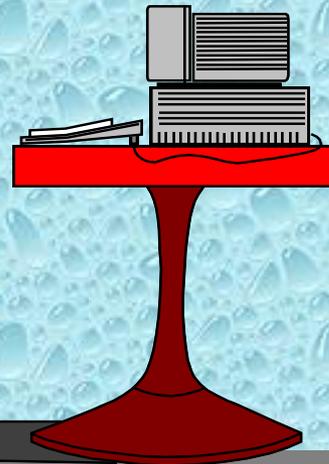
ILU(p),IC(p) p intero positivo, livello di fill-in. L'accuratezza è aumentata accettando che elementi non nulli vengano costruiti laddove A presenta elementi nulli.

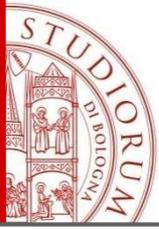
That is all, folks...

Thank you for your patience!



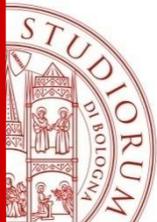
?????





Quale metodo scelgo? Se la Matrice dei coefficienti è ..

- **Densa e/o piccola**
 - Metodo di Gauss
- **Sparsa di Grandi dimensioni (o non disponibile)**
 - Metodi iterativi
- **Matrice a banda (es tridiagonale)**
 - Thomas o Gauss rivisitato
- **Simmetrica def. Pos. densa e/o piccole dimensioni**
 - Cholesky
- **Simmetrica def. Pos. Grandi dimensioni e sparsa**
 - gradienti coniugati
- **Simmetrica indefinita grandi dimensioni e sparsa**
 - Minres
- **Non simmetrica grande e sparsa**
 - GMRES, BiCGstab



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Serena Morigi

Dipartimento di Matematica

serena.morigi@unibo.it

<http://www.dm.unibo.it/~morigi>