METODI ITERATIVI di tipo GRADIENTE

Dati: $A \in \mathbb{R}^{n\times n}$ sdp, $\underline{b} \in \mathbb{R}^{n\times 1}$ 24/10/2023

? $\underline{x}$: $A\underline{x} = \underline{b}$

voglio costruire $\{\underline{x}^{(k)}\}_{k\geq 0}$: $\lim_{k\to\infty} \underline{x}^{(k)} = \underline{x}$

Def $\Phi(\underline{x}) = \frac{1}{2}\, \underline{x}^T A \underline{x} - \underline{x}^T \underline{b} \in \mathbb{R}$

$\Phi : \mathbb{R}^{n\times 1} \longrightarrow \mathbb{R}$

se A è sdp $\Rightarrow$ $\Phi(\underline{x})$ è convessa

Φ ha ! pto di min assoluto $\tilde{\underline{x}}$ e stazionario

cioè $\nabla\Phi(\tilde{\underline{x}}) = \underline{0}$

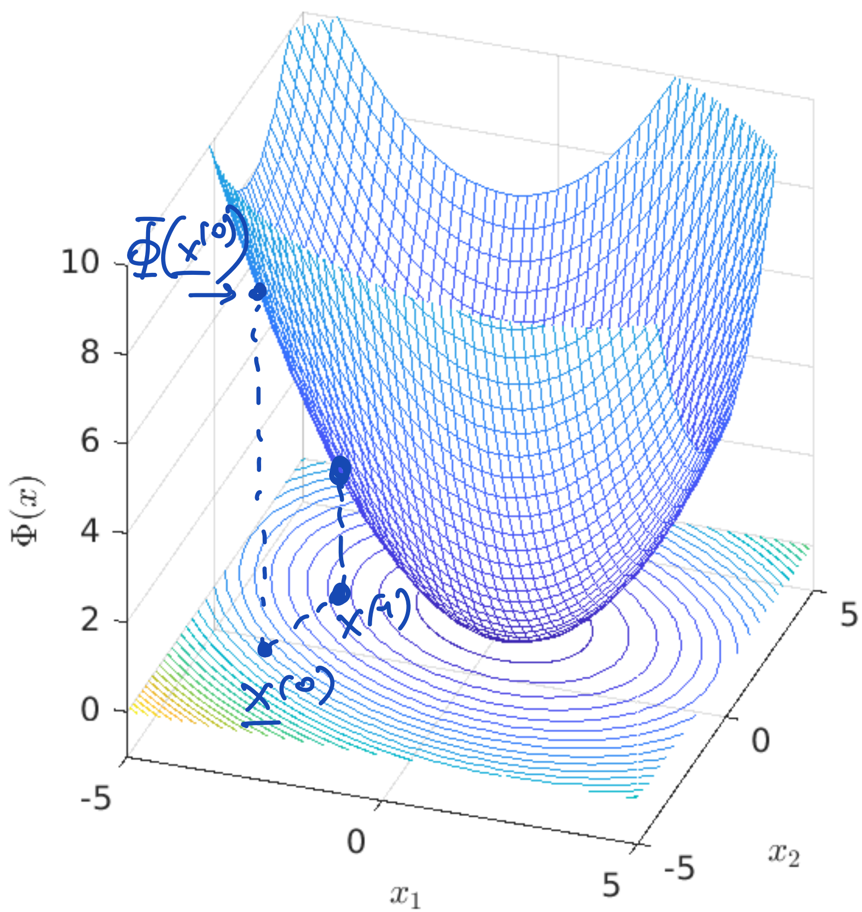
$$\nabla\Phi(\underline{x}) = \begin{bmatrix} \frac{\partial\Phi(\underline{x})}{\partial x_1} \\ \frac{\partial\Phi(\underline{x})}{\partial x_2} \\ \vdots \\ \frac{\partial\Phi(\underline{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{n\times 1}$$

$\nabla\Phi(\underline{x}) = A\underline{x} - \underline{b}$

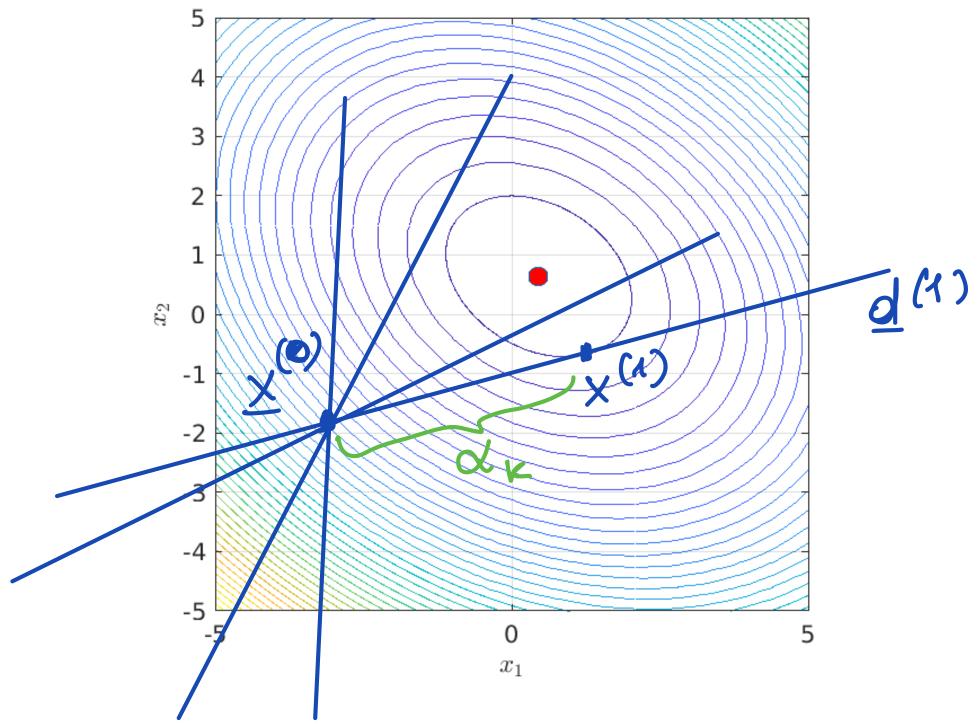
$$\Rightarrow \quad \nabla\Phi(\tilde{\underline{x}}) = 0 = A\tilde{\underline{x}} - \underline{b}$$

cioè il punto di min $\tilde{\underline{x}}$ è esatt la sol del sist lin $A\underline{x} = \underline{b}$

$n=2$



$\Phi(x^{(0)})$
10
8
6
4
2
0
$\Phi(x)$
$x^{(1)}$
$x^{(0)}$
-5
0
5
x_1
-5
0
5
x_2



5
4
3
2
1
0
-1
-2
-3
-4
-5
x_2
-5
0
5
x_1
$x^{(0)}$
$x^{(1)}$
$d^{(1)}$
α

$\underline{x}^{(0)}$ dato in $\mathbb{R}^{n\times 1}$

per $k \geq 0$ si costruisce

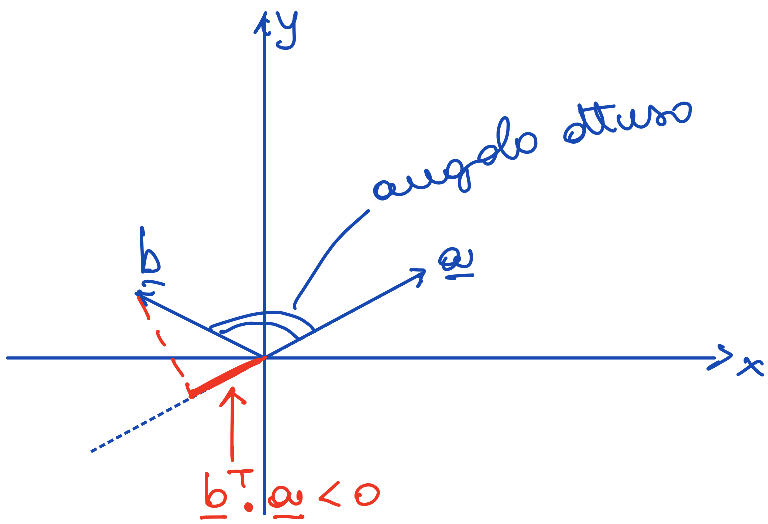
$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \underbrace{\alpha_k}_{\text{step size}} \underline{d}^{(k)}$$

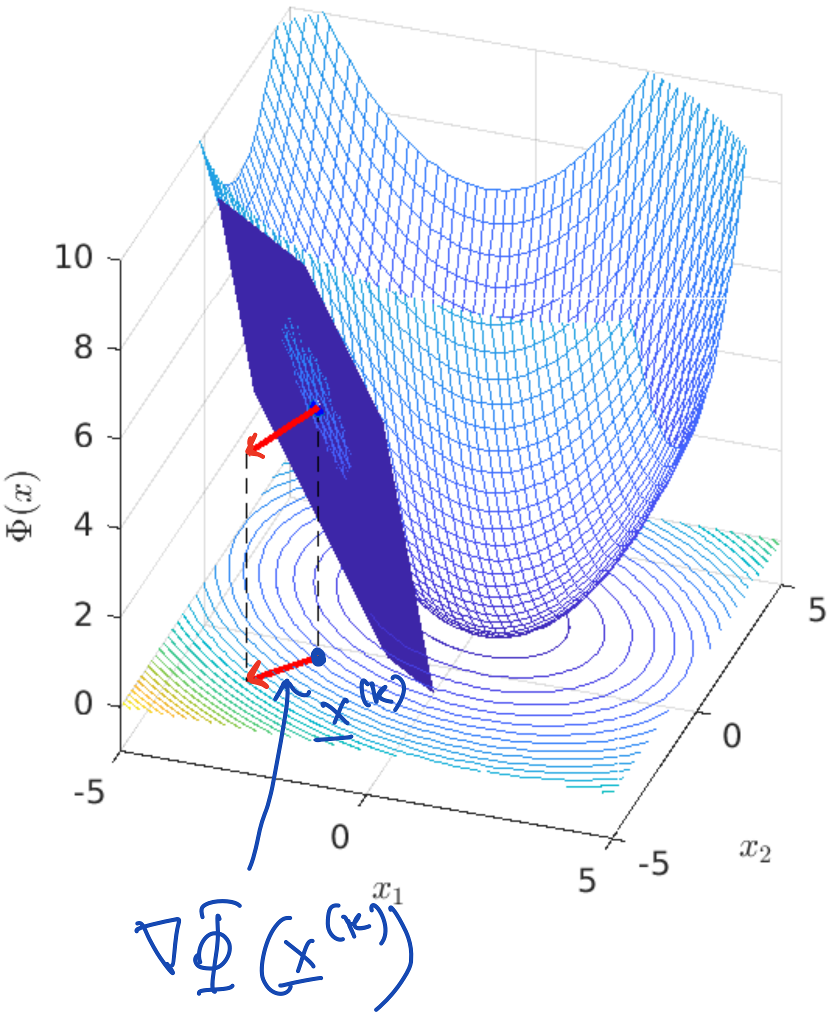
$\uparrow$ direzione di discesa

Def Dico che $\underline{d} \in \mathbb{R}^{n\times 1}$ è una direzione di discesa al passo k se:

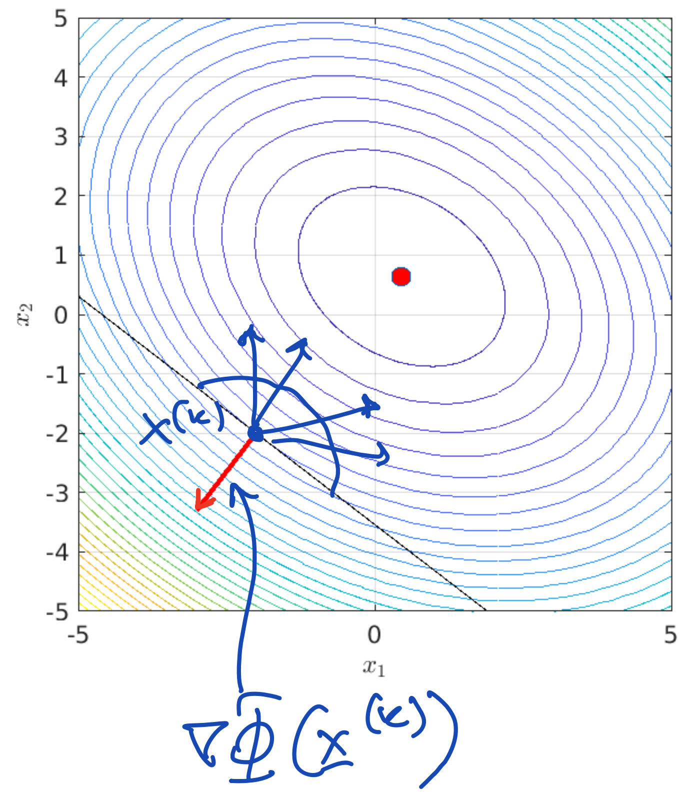
$\underline{d}^T \cdot \nabla \underline{\Phi}(\underline{x}^{(k)}) < 0$ quando $\nabla \underline{\Phi}(\underline{x}^{(k)}) \neq 0$

o $\underline{d} = \underline{0}$ quando $\nabla \underline{\Phi}(\underline{x}^{(k)}) = \underline{0}$

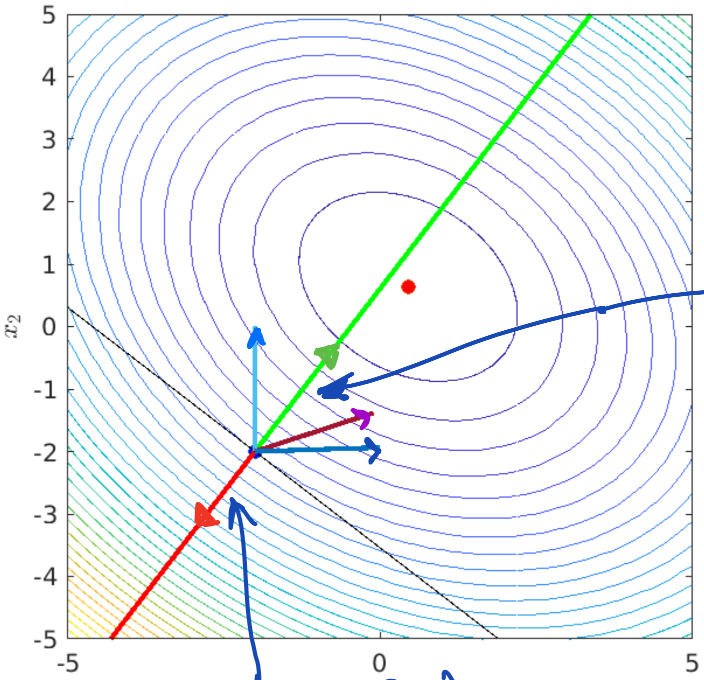




10
8
6
4
2
0
$\Phi(x)$
-5
0
5
x_1
5
0
-5
x_2
$x^{(k)}$
$\nabla\Phi(x^{(k)})$



5
4
3
2
1
0
-1
-2
-3
-4
-5
x_2
-5
0
5
x_1
$x^{(k)}$
$\nabla\Phi(x^{(k)})$



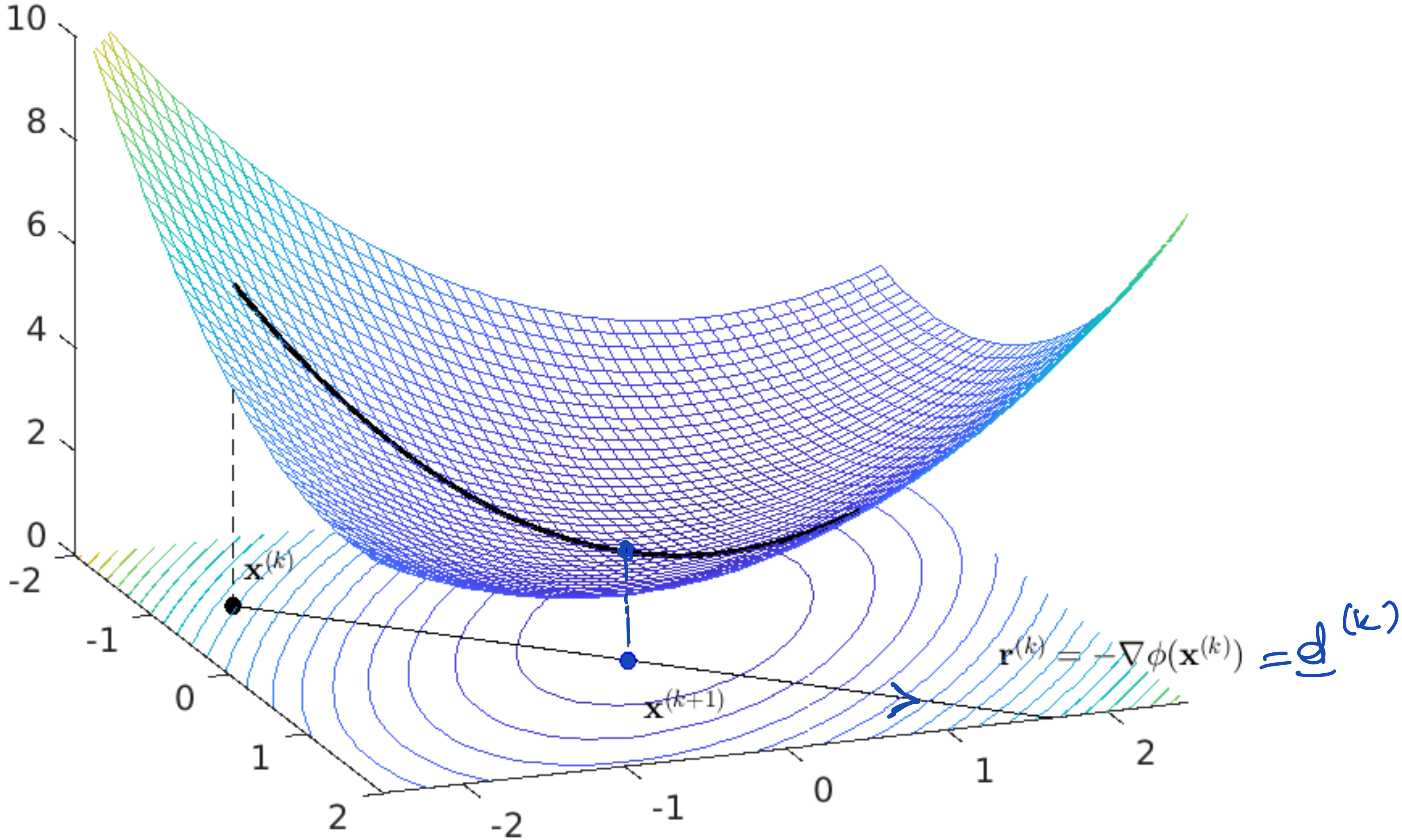
$\nabla \underline{\Phi}(\underline{x}^{(k)})$

Il metodo del gradiente sceglie come direz di discesa:

$$\underline{d}^{(k)} = -\underbrace{\nabla \underline{\Phi}(\underline{x}^{(k)})}_{A\underline{x}^{(k)}-\underline{b}} = \underbrace{\underline{b} - A\underline{x}^{(k)}}_{\substack{\text{residuo} \\ \text{al passo } k \\ \underline{r}^{(k)}}}$$

Ora devo definire α_k

direzione di discesa del gradiente e sezione di Phi



$\Phi(\underline{x})\Big|_{\underline{d}^{(k)}}$ è una parabola, scelgo α_k : $\underline{x}^{(k+1)}$ sia il pto di minimo della parabola

$\underline{x} = \underline{x}^{(k)} + \alpha\, \underline{d}^{(k)}$ $\in$ retta che passa per $\underline{x}^{(k)}$ con direzione dettata da $\underline{d}^{(k)}$

$$\Phi(\underline{x})\Big|_{\underline{d}^{(k)}} = \Phi\left(\underline{x}^{(k)} + \alpha\, \underline{d}^{(k)}\right) = \varphi(\alpha)$$

calcolare $\min_{\alpha} \varphi(\alpha)$

$\equiv\ ?\,\alpha : \varphi'(\alpha) = 0$

trovo $\alpha = \alpha_k = \dfrac{\left(\underline{d}^{(k)}\right)^T \cdot \underline{r}^{(k)}}{\underbrace{\left(\underline{d}^{(k)}\right)^T \cdot A \cdot \underline{d}^{(k)}}_{\neq 0}} \in \mathbb{R}$ $\neq 0$ perché A è sdp

Test d'arresto

ci si ferma quando $\dfrac{\|\underline{r}^{(k)}\|}{\|\underline{b}\|} < \varepsilon$

con ε dato, tolleranza

Algoritmo del metodo del gradiente

Input: A, $\underline{b}$, $\underline{x}^{(0)}$, ε, kmax

Output: $\underline{x}$, res, k, resvec

res $= \|\underline{r}^{(k)}\| / \|\underline{b}\|$ all'ultima iter.

k ← iterazioni svolte

resvec ← vettore dei valori $\dfrac{\|\underline{r}^{(k)}\|}{\|\underline{b}\|}$ $\forall k$

$$\underline{r}^{(0)} = \underline{b} - A\underline{x}^{(0)}, \quad \text{res} = \|\underline{r}^{(0)}\| / \|\underline{b}\|$$

$$\underline{d}^{(0)} = \underline{r}^{(0)}, \quad k = 0$$

while $k \leq$ kmax & res $> \varepsilon$

$$\alpha_k = \frac{(\underline{d}^{(k)})^T \underline{r}^{(k)}}{(\underline{d}^{(k)})^T A \underline{d}^{(k)}}$$

$A\underline{d}^{(k)} \sim 2n^2$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha_k \cdot \underline{d}^{(k)}$$

$$\underline{r}^{(k+1)} = \underline{b} - A\underline{x}^{(k+1)} = \underline{b} - A\left(\underline{x}^{(k)} + \alpha_k \underline{d}^{(k)}\right) = \underbrace{\underline{b} - A\underline{x}^{(k)}} - \alpha_k A\underline{d}^{(k)} = \underline{r}^{(k)} - \alpha_k A\underline{d}^{(k)}$$

$$\underline{d}^{(k+1)} = \underline{r}^{(k+1)}$$

$$\text{res} = \|\underline{r}^{(k+1)}\| / \|\underline{b}\|$$

$$k = k+1$$

$$\underline{x} = \underline{x}^{(k)}$$

1 iteraz costa $\sim 2n^2$ operaz elem

= 1 prod mat - vett.

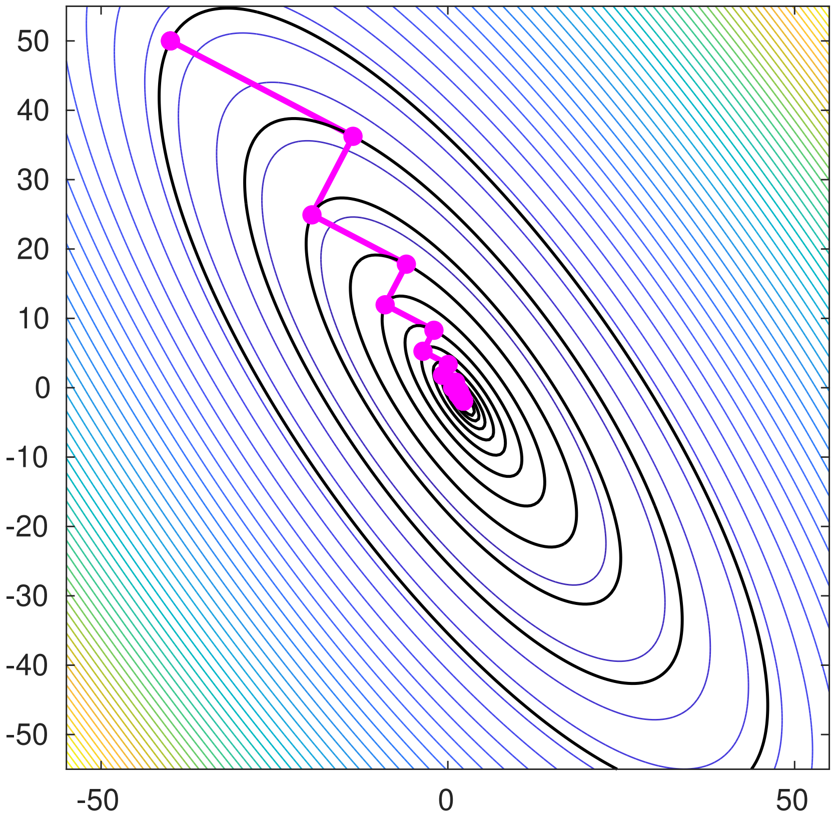
Limiti del Metodo del Gradiente

31/10/2023

Dati

$$A = \begin{bmatrix} 10 & 6 \\ 6 & 6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 11 \\ 2 \end{bmatrix}$$

risolviamo $A\mathbf{x} = \mathbf{b}$ con il metodo del gradiente, posto $\mathbf{x}_0 = [-40; 50]$, tol$=10^{-10}$, kmax$=100$.



Si ha convergenza in 78 iterazioni: TROPPE!!

Le direzioni di discesa sono a 2 a 2 ortogonali.

Vorrei convergere seguendo una via più diretta.

METODO DEL GRADIENTE CONIUGATO

$\underline{d}^{(0)} = \underline{r}^{(0)}$, α_K come nel metodo del grad.

Scelgo $\underline{d}^{(k)}$:

1) $\left(\underline{d}^{(k+1)}\right)^T A \, \underline{d}^{(j)} = 0 \quad \forall j = 0, \dots, K$

(poichè ~~A~~ A è sdp ⇒ è non sing, questa condiz ⇒ che le $\underline{d}^{(k)}$ formino una base di $\mathbb{R}^n$)

2) $\left(\underline{r}^{(k+1)}\right)^T \underline{d}^{(j)} = 0 \quad \forall j = 0, \dots, K$

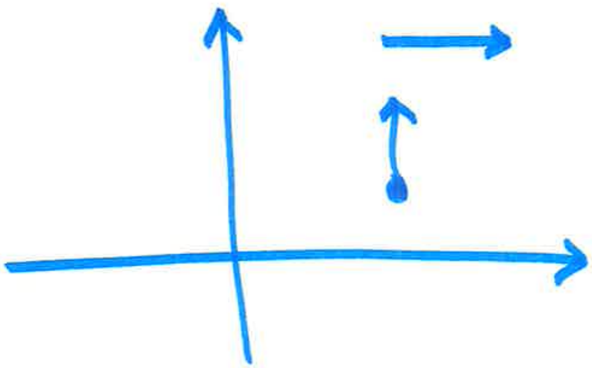
$\left(\mathcal{B} = \{\underline{d}^{(0)}, \underline{d}^{(1)}, \dots, \underline{d}^{(n-1)}\}\right.$ è una base per $\mathbb{R}^n$

quando ~~k=~~ $k = n-1 \Rightarrow \underline{r}^{(n)}$ è ortogonale alla base $\mathcal{B}$

ma $\underline{r}^{(n)} \in \mathbb{R}^n$

$$\Rightarrow \underline{r}^{(n)} = \underline{0}$$

$$\underline{b} - A\underline{x}^{(n)} = 0$$



cioè $\underline{x}^{(n)} \equiv \underline{x}$

la sol esatta del sist $A\underline{x} = \underline{b}$

Il metodo del GC arriva a terminazione (alla sol esatta) al più in n iterazioni (dove n = dimens. sistema)

Si costruisce $\mathbb{R} \ni \beta_k = \dfrac{(A\underline{d}^{(k)})^T \underline{r}^{(k+1)}}{(\underline{d}^{(k)})^T A \underline{d}^{(k)}}$

definisco: $\underline{d}^{(k+1)} = \underline{r}^{(k+1)} - \beta_k \underline{d}^{(k)}$

Algoritmo del GC

Input: A sdp, $\underline{b}$, $\underline{x}^{(0)}$, ε, kmax

Output: $\underline{x}$, res, K, resvec

$\underline{x}$ sol calcolata

res $\|\underline{r}^{(k)}\| / \|\underline{b}\|$

K n° it svolte

resvec vettore dei valori $\|r^{(k)}\| / \|b\|$ $\forall k$

$$\underline{r}^{(0)} = \underline{b} - A\underline{x}^{(0)}; \quad \underline{d}^{(0)} = \underline{r}^{(0)}; \quad res = \|\underline{r}^{(0)}\| / \|\underline{b}\|; \quad k = 0$$

while (res > ε & $k \leq$ kmax)

$$\alpha_k = \frac{(\underline{d}^{(k)})^T \underline{r}^{(k)}}{(\underline{d}^{(k)})^T A \underline{d}^{(k)}} \quad A\underline{d}^{(k)} \rightarrow \underline{v}$$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha_k \underline{d}^{(k)}$$

$$\underline{r}^{(k+1)} = \underline{r}^{(k)} - \alpha_k \boxed{A\underline{d}^{(k)}} \quad \underline{v}$$

$$\beta_k = \frac{(A\underline{d}^{(k)})^T \underline{r}^{(k+1)}}{(\underline{d}^{(k)})^T \boxed{A\underline{d}^{(k)}}} \quad \underline{v}$$

costo comput: 1 prod mat x vett $\forall$ iterazione

$$\underline{d}^{(k+1)} = \underline{r}^{(k+1)} - \beta_k \, \underline{d}^{(k)}$$

ricordarci di costruire resvec

$$res = \|\underline{r}^{(k+1)}\| / \|\underline{b}\|$$

$$k = k+1$$

end

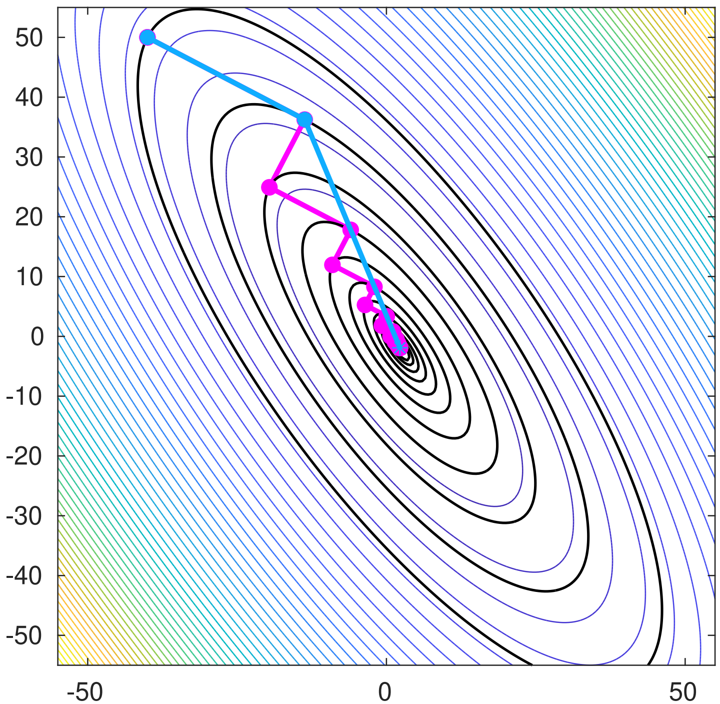
$$\underline{x} = \underline{x}^{(k)}$$

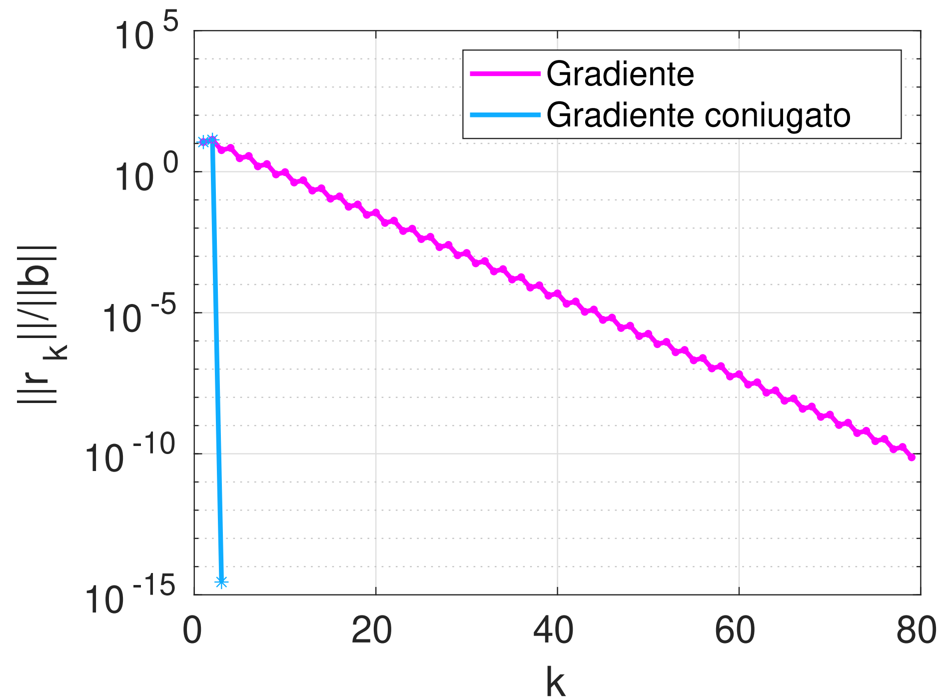
Confronto Gradiente – Gradiente Coniugato

Dati

$$A = \begin{bmatrix} 10 & 6 \\ 6 & 6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 11 \\ 2 \end{bmatrix}$$

con $\mathbf{x}_0 = [-40; 50]$, tol$=10^{-10}$, kmax$=100$.





Gradiente converge in 78 it (err $\simeq 10^{-10}$)

Gradiente Coniugato converge in 2 it (err $\simeq 10^{-15}$)

Teorema di conv. del metodo del gradiente

Se A è sdp, il metodo del grad converge $\forall \underline{x}^{(0)} \in \mathbb{R}^n$, inoltre

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

dove $\|\underline{v}\|_A = \sqrt{\underline{v}^T A \underline{v}}$.

$\rho = \dfrac{K(A)-1}{K(A)+1}$ fattore di riduz dell'errore ad ogni passo

$\rho < 1$

Se ρ è prossimo a 1

sto riducendo l'err molto poco ad ogni passo.

$\rho \sim 1 \equiv K(A) \gg 1$ $\quad$ ($K(A)$ = condiz di A)

Se $K(A) \gg 1 \Rightarrow$ G impiega molte iterazioni per convergere

A mal condizionata

Teorema di convergenza per Grad. Coniugato

Sia A sdp, GC converge $\forall \underline{x}^{(0)} \in \mathbb{R}^n$ e si ha

$$\|\underline{x} - \underline{x}^{(k)}\|_A \leq \frac{2 \cdot c^k}{1 + c^{2k}} \|\underline{x} - \underline{x}^{(0)}\|_A$$

dove $c = \dfrac{\sqrt{K(A)} - 1}{\sqrt{K(A)} + 1}$

Il condiz di A influisce ancora nella converg del metodo, ma meno rispetto al met del grad perché qui ho $\sqrt{K(A)}$ invece di $K(A)$.

#it per arrivare a conv $\sim \sqrt{K(A)}$ per il GC

Se A non è sdp si possono applicare varianti del metodo del grad. coniugato:

BCG bi conjugate gradient
CGS conjugate gradient square
BiCGStab bi conjugate gradient stabilized

metodi di Krylov per matrici generiche