

## 4 Approssimazione nel senso dei minimi quadrati

In questo Capitolo si affrontano due problemi che richiedono di determinare un vettore che rende *minimo* il valore di una *somma di quadrati*: il problema di determinare la soluzione di un sistema di equazioni lineari *nel senso dei minimi quadrati* ed il problema di determinare la funzione che meglio approssima dati assegnati *nel senso dei minimi quadrati*.

### 4.0.16 Definizione (Soluzione di un sistema nel senso dei minimi quadrati)

Siano  $A \in \mathbb{R}^{n \times k}$  e  $b \in \mathbb{R}^n$ . Un vettore  $x^* \in \mathbb{R}^k$  è una *soluzione del sistema*  $Ax = b$  *nel senso dei minimi quadrati* se verifica una delle tre proprietà *equivalenti*:

(A) Per ogni  $y \in \mathbb{R}^k$  si ha:  $\|Ax^* - b\|_2 \leq \|Ay - b\|_2$ ;

(B) Per ogni  $y \in \mathbb{R}^k$  si ha:  $\|Ax^* - b\|_2^2 \leq \|Ay - b\|_2^2$ ;

(C) Il vettore  $x^*$  è un *minimo assoluto* della funzione  $n: \mathbb{R}^k \rightarrow \mathbb{R}$ , *norma del residuo*, definita da:

$$n(x) = \|Ax - b\|_2$$

Si osservi che se  $x^*$  è soluzione di  $Ax = b$  allora è anche soluzione di  $Ax = b$  nel senso dei minimi quadrati. Infatti:  $n(x^*) = 0$  e per ogni  $x \in \mathbb{R}^k$  si ha  $n(x) \geq 0$ .

### 4.0.17 Definizione (Funzione che meglio approssima dati assegnati nel senso dei minimi quadrati)

Siano  $I$  un intervallo non degenere,  $F$  un sottospazio vettoriale di *dimensione finita* dello spazio delle funzioni continue da  $I$  in  $\mathbb{R}$ ,  $x_0, \dots, x_k$  numeri reali in  $I$  e  $y_0, \dots, y_k$  numeri reali. Un elemento  $f^*$  di  $F$  è una *funzione che meglio approssima i dati*  $(x_0, y_0), \dots, (x_k, y_k)$  *nel senso dei minimi quadrati* se: Per ogni  $f \in F$  si ha:

$$(f^*(x_0) - y_0)^2 + \dots + (f^*(x_k) - y_k)^2 \leq (f(x_0) - y_0)^2 + \dots + (f(x_k) - y_k)^2$$

ovvero se  $f^*$  è un *minimo assoluto* della funzione  $SQ: F \rightarrow \mathbb{R}$ , *scarto quadratico*, definita da:

$$SQ(f) = (f(x_0) - y_0)^2 + \dots + (f(x_k) - y_k)^2$$

Interpretando i dati come coordinate di punti in un piano cartesiano, lo scarto quadratico ha un semplice *significato geometrico*. Il valore  $SQ(f)$  è somma di  $k + 1$  addendi. Il  $j$ -esimo addendo è il quadrato della lunghezza del segmento individuato, sulla retta verticale di ascissa  $x_j$ , dal valore  $f(x_j)$  e dall'ordinata del  $j$ -esimo dato  $y_j$ . Il valore  $SQ(f)$  è una *misura dello scostamento del grafico di  $f(x)$  dai dati*.

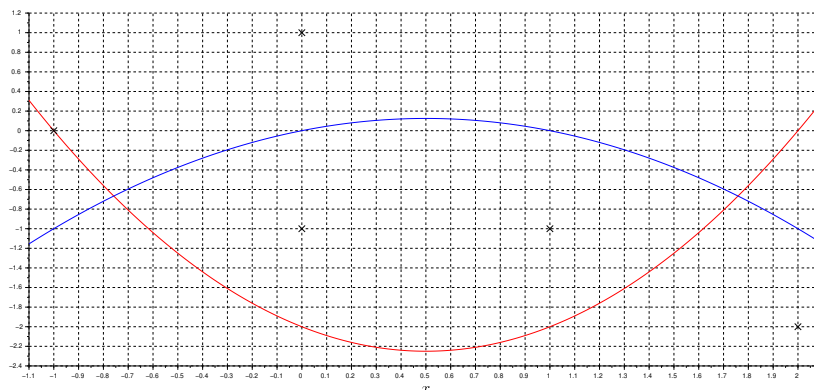
Si osservi che, contrariamente a quanto richiesto nei problemi di interpolazione, i numeri reali  $x_0, \dots, x_k$  *non* devono necessariamente essere distinti.

---

### Esercizi

---

E0 Nella figura seguente sono rappresentati i dati ( $\times$ ), il grafico della funzione  $f$  (in rosso) e quello della funzione  $g$  (in blu). Determinare  $SQ(f)$ ,  $SQ(g)$  e decidere quale funzione tra  $f$  e  $g$  meglio approssima i dati nel senso dei minimi quadrati.



E1 ★ Dimostrare che  $SQ(f) = 0$  se e solo se  $f$  interpola i dati.

## 4.1 Migliore approssimazione in spazi con prodotto scalare

Alla ricerca delle soluzioni di un sistema nel senso dei minimi quadrati e delle funzioni che meglio approssimano i dati nel senso dei minimi quadrati premettiamo la nozione di migliore approssimazione in uno spazio con prodotto scalare e la sua ricerca.

Sia  $V$  uno spazio vettoriale su  $\mathbb{R}$  con *prodotto scalare*. Indicato con  $a \cdot b$  il prodotto scalare di  $a$  e  $b$ , si indica con  $\|a\|$  la norma di  $a$  indotta dal prodotto scalare:

$$\|a\| = \sqrt{a \cdot a}$$

Siano poi  $W$  un sottospazio vettoriale di  $V$  di *dimensione finita* e  $v$  un elemento di  $V$ .

### 4.1.1 Definizione (Migliore approssimazione in uno spazio con prodotto scalare)

Un elemento  $w^*$  di  $W$  è una *migliore approssimazione di  $v$  in  $W$*  se verifica una delle due proprietà *equivalenti*:

(A) Per ogni  $w \in W$  si ha:  $\|v - w^*\| \leq \|v - w\|$ ;

(B) Il vettore  $w^*$  è un *minimo assoluto* della funzione  $d: W \rightarrow \mathbb{R}$ , *distanza da  $v$* , definita da:

$$d(w) = \|v - w\|$$

Come vedremo, la nozione di migliore approssimazione in uno spazio con prodotto scalare è strettamente connessa a quella di *proiezione ortogonale*:

### 4.1.2 Definizione (Proiezione ortogonale)

Un elemento  $w^*$  di  $W$  è *proiezione ortogonale di  $v$  su  $W$*  se il vettore  $v - w^*$  è ortogonale a  $W$ , ovvero se:

$$\text{Per ogni } w \in W \text{ si ha: } (v - w^*) \cdot w = 0 \quad (*)$$

Sia  $w_1, \dots, w_j$  un insieme di *generatori* di  $W$ . La condizione (\*) è *equivalente* a:

$$(v - w^*) \cdot w_i = 0 \quad i = 1, \dots, j$$

Allora una combinazione lineare  $a_1 w_1 + \dots + a_j w_j$  è proiezione ortogonale di  $v$  su  $W$  se e solo se per  $i = 1, \dots, j$  si ha:  $(v - (a_1 w_1 + \dots + a_j w_j)) \cdot w_i = 0$ , ovvero se e solo se:

$$\begin{aligned} v \cdot w_1 &= (a_1 w_1 + \dots + a_j w_j) \cdot w_1 = a_1 (w_1 \cdot w_1) + \dots + a_j (w_j \cdot w_1) \\ &\vdots \\ v \cdot w_j &= (a_1 w_1 + \dots + a_j w_j) \cdot w_j = a_1 (w_1 \cdot w_j) + \dots + a_j (w_j \cdot w_j) \end{aligned} \quad (**)$$

Posto:

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_j \end{bmatrix}, \quad F = \begin{bmatrix} w_1 \cdot w_1 & \cdots & w_j \cdot w_1 \\ \vdots & & \vdots \\ w_1 \cdot w_j & \cdots & w_j \cdot w_j \end{bmatrix}, \quad c = \begin{bmatrix} v \cdot w_1 \\ \vdots \\ v \cdot w_j \end{bmatrix}$$

la condizione (\*\*) si riformula: *una combinazione lineare  $a_1 w_1 + \dots + a_j w_j$  è proiezione ortogonale di  $v$  su  $W$  se e solo se la colonna  $a$  è soluzione del sistema  $Fx = c$  delle equazioni normali.*

### 4.1.3 Teorema (di esistenza ed unicità della proiezione ortogonale)

Esiste un solo elemento di  $W$  proiezione ortogonale di  $v$  su  $W$ .

*Dimostrazione.* Sia  $w_1, \dots, w_r$  una *base ortonormale* di  $W$  (sicuramente esistente...). Allora la matrice  $F$  del sistema delle equazioni normali è la matrice identica ed il sistema ha una sola soluzione:

$$a = \begin{bmatrix} v \cdot w_1 \\ \vdots \\ v \cdot w_r \end{bmatrix}$$

L'unico elemento di  $W$  proiezione ortogonale di  $v$  su  $W$  è:

$$w^* = v \cdot w_1 w_1 + \cdots + v \cdot w_r w_r$$

Si osservi che il Teorema appena dimostrato prova che *le equazioni normali  $Fx = c$  hanno sempre almeno una soluzione*. Infatti: *le componenti di ciascuna soluzione delle equazioni normali sono coordinate, rispetto ai generatori di  $W$ , della proiezione ortogonale*. L'esistenza della proiezione ortogonale  $w^*$  garantisce l'esistenza di almeno una combinazione lineare dei generatori di  $W$  tale che  $a_1 w_1 + \cdots + a_j w_j = w^*$ . Inoltre, se  $w_1, \dots, w_j$  sono generatori *linearmente dipendenti* (e quindi *non* una base) di  $W$  allora esistono *infinite* combinazioni lineari che generano lo stesso elemento  $w^*$ . In tal caso *il sistema  $Fx = c$  ha infinite soluzioni* (in particolare: la matrice  $F$  non è invertibile). Se, invece, gli elementi sono generatori *linearmente indipendenti* (e quindi *una base*) di  $W$  allora esiste *una sola* combinazione lineare che genera l'elemento  $w^*$ . In tal caso *il sistema  $Fx = c$  ha una sola soluzione* (in particolare: la matrice  $F$  è invertibile).

#### 4.1.4 Teorema (di esistenza ed unicità della migliore approssimazione)

La proiezione ortogonale  $w^*$  di  $v$  su  $W$  è l'*unica* migliore approssimazione di  $v$  in  $W$ , ovvero:  $w^*$  è l'*elemento di  $W$  più vicino a  $v$* .

*Dimostrazione.* Per ogni  $w \in W$  si ha:

$$\|v - w\|^2 = \|(v - w^*) + (w^* - w)\|^2$$

Il primo addendo, per definizione, è ortogonale a  $W$  ed il secondo è un elemento di  $W$ . Allora, per il Teorema di Pitagora:

$$\|(v - w^*) + (w^* - w)\|^2 = \|v - w^*\|^2 + \|w^* - w\|^2$$

e quindi:

$$\|v - w\|^2 = \|v - w^*\|^2 + \|w^* - w\|^2$$

Poiché  $\|w^* - w\|^2 \geq 0$  e si ha  $\|w^* - w\|^2 = 0$  se e solo se  $w = w^*$ , allora:

- (a)  $\|v - w\|^2 \geq \|v - w^*\|^2$  ovvero  $\|v - w\| \geq \|v - w^*\|$ :  $w^*$  è una migliore approssimazione;
- (b)  $\|v - w\|^2 = \|v - w^*\|^2$ , ovvero  $\|v - w\| = \|v - w^*\|$ , se e solo se  $w = w^*$ :  $w^*$  è l'*unica* migliore approssimazione.

#### 4.1.5 Esercizio

(1) Siano  $V = \mathbb{R}^4$  con prodotto scalare canonico,

$$W = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\} \quad \text{e} \quad v = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

Determinare la migliore approssimazione di  $v$  in  $W$ .

*Soluzione:* Occorre determinare la proiezione ortogonale  $w^*$  di  $v$  su  $W$ . Poiché si ha *un solo* generatore di  $W$ , le equazioni normali si riducono ad *una* equazione in *una* incognita. Si ha:

$$F = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = 2 \quad , \quad c = v \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = 1$$

e le equazioni normali sono:  $2x = 1$ . Si ottiene l'unica soluzione  $a = \frac{1}{2}$  da cui:

$$w^* = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Per il Teorema di esistenza ed unicità della migliore approssimazione in uno spazio con prodotto scalare, la migliore approssimazione di  $v$  in  $W$  è  $w^*$ .

(2) Siano  $V = \mathbb{R}^4$  con prodotto scalare canonico,

$$W = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix} \right\} \quad \text{e} \quad v = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

Determinare la migliore approssimazione di  $v$  in  $W$ .

*Soluzione:* Come nel caso precedente, occorre determinare la proiezione ortogonale  $w^*$  di  $v$  su  $W$ . Poiché si hanno *due* generatori di  $W$ , le equazioni normali sono un sistema di *due* equazioni in *due* incognite. Poiché i generatori sono *dipendenti*, il sistema ha *infinite* soluzioni. Si ha:

$$F = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

e le equazioni normali sono:  $Fx = c$ . L'insieme delle soluzioni è:

$$S = \left\{ \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \lambda \in \mathbb{R} \right\}$$

Tutti gli elementi di  $S$  individuano lo *stesso* elemento di  $W$ :

$$\left(\frac{1}{2} + 2\lambda\right) \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \lambda \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = w^*$$

Si ottiene, come giusto, *lo stesso elemento di  $W$*  ottenuto nel punto precedente. Per il Teorema di esistenza ed unicità della migliore approssimazione in uno spazio con prodotto scalare, la migliore approssimazione di  $v$  in  $W$  è  $w^*$ .

(3) Siano  $I = [0, 2\pi]$  e  $V = C(I)$  con prodotto scalare definito da:

$$f \cdot g = \frac{1}{\pi} \int_0^{2\pi} f(t)g(t) dt$$

Siano infine:

$$W = \text{span} \{ 1, \cos t, \sin t \} \quad \text{e} \quad v = t^2$$

Determinare la migliore approssimazione di  $v$  in  $W$ .

*Soluzione:* Come nel caso precedente, occorre determinare la proiezione ortogonale  $w^*$  di  $v$  su  $W$ . Si osservi che in questo esempio lo spazio vettoriale  $V$  ha *dimensione infinita*. Poiché si hanno *tre* generatori di  $W$ , le equazioni normali sono un sistema di *tre* equazioni in *tre* incognite. Poiché i generatori sono *indipendenti*, il sistema ha *una* soluzione. Si ha:

$$\begin{array}{lll} 1 \cdot 1 = \frac{1}{\pi} \int_0^{2\pi} dt = 2 & & v \cdot 1 = \frac{8}{3} \pi^2 \\ 1 \cdot \cos t = 0 & \cos t \cdot \cos t = 1 & v \cdot \cos t = 4 \\ 1 \cdot \sin t = 0 & \cos t \cdot \sin t = 0 \quad \sin t \cdot \sin t = 1 & v \cdot \sin t = -4\pi \end{array}$$

e le equazioni normali sono:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} \frac{8}{3} \pi^2 \\ 4 \\ -4\pi \end{bmatrix}$$

L'unica soluzione è:

$$a = \begin{bmatrix} \frac{4}{3} \pi^2 \\ 4 \\ -4\pi \end{bmatrix}$$

da cui:

$$w^* = \frac{4}{3} \pi^2 + 4 \cos t - 4\pi \sin t$$

Per il Teorema di esistenza ed unicità della migliore approssimazione in uno spazio con prodotto scalare, la migliore approssimazione di  $v$  in  $W$  è  $w^*$ .

L'elemento determinato è il *minimo assoluto su  $W$*  della funzione  $d : C(I) \rightarrow \mathbb{R}$  definita da:

$$d(f) = \|v - f\| = \left[ \frac{1}{\pi} \int_0^{2\pi} (t^2 - f(t))^2 dt \right]^{\frac{1}{2}}$$

– Osservazione: *Serie di Fourier*

Sia  $g \in C(I)$ . Si ricordi che, introdotto in  $C(I)$  il prodotto scalare definito nel punto (3) dell'Esercizio, e posto:

$$a_0 = g \cdot 1 \quad , \quad a_k = g \cdot \cos kt \quad , \quad b_k = g \cdot \sin kt \quad k = 1, 2, \dots$$

si chiama *serie di Fourier generata dalla funzione  $g$*  la serie:

$$\frac{1}{2} a_0 + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt)$$

ovvero, introdotte le *somme parziali*:

$$s_0(t) = \frac{1}{2} a_0 \quad , \quad s_j(t) = \frac{1}{2} a_0 + \sum_{k=1}^j (a_k \cos kt + b_k \sin kt) \quad j = 1, 2, \dots$$

la *successione*:  $s_0(t), s_1(t), \dots$

Posto:

$$W_k = \text{span} \{ 1, \cos t, \sin t, \dots, \cos kt, \sin kt \} \quad k = 0, 1, 2, \dots$$

la migliore approssimazione  $w_k^*$  di  $g$  in  $W_k$ , ovvero il minimo assoluto *su  $W_k$*  della funzione  $d : C(I) \rightarrow \mathbb{R}$  definita da:

$$d(f) = \|g - f\| = \left[ \frac{1}{\pi} \int_0^{2\pi} (g(t) - f(t))^2 dt \right]^{\frac{1}{2}}$$

è  $s_k$ . Poiché  $W_0 \subset W_1 \subset W_2 \subset \dots$ , la successione  $d(s_0), d(s_1), d(s_2), \dots$  è *non crescente*, dunque *convergente*. Un risultato classico dell'Analisi Matematica mostra che  $\lim d(s_k) = 0$ , risultato che si esprime anche con l'asserto:

$$g(t) = \frac{1}{2} a_0 + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt) \quad \text{nel senso della convergenza in media}$$

Si osservi che la sola continuità di  $g(t)$  *non assicura* la validità *puntuale* dell'uguaglianza in tutti i punti dell'intervallo  $I$ . Si consideri ad esempio il caso  $I = [0, 2\pi]$  e  $g(t) = t^2$ . Per ogni  $j$  si ha  $s_j(0) = s_j(2\pi)$  perciò se  $\lim s_j(0) = g(0)$  allora  $\lim s_j(2\pi) = g(0)$  Ma  $g(0) \neq g(2\pi)$ , dunque la convergenza puntuale della serie *non può sussistere* sull'intero intervallo  $I$ .

Lo studio della convergenza della successione  $s_0(t), s_1(t), \dots$  — ovvero del *significato* dell'uguaglianza precedente — è argomento vasto e non elementare.<sup>41</sup>

## Esercizi

**E2** ★ Sia  $W = \text{span}\{w_1, \dots, w_j\}$  un sottospazio vettoriale di  $V$ , spazio vettoriale su  $\mathbb{R}$  con prodotto scalare. Dimostrare che  $v \in V$  è ortogonale a  $W$  se e solo se:

$$v \cdot w_i = 0 \quad i = 1, \dots, j$$

<sup>41</sup>Si veda, ad esempio: J.E. Marsden: *Elementary Classical Analysis*, Capitolo 10.

E3 ★ Sia  $V$  uno spazio vettoriale su  $\mathbb{R}$  con prodotto scalare. Dimostrare il Teorema di Pitagora: Siano  $a$  e  $b$  due elementi di  $V$ . Allora:

$$a \cdot b = 0 \quad \Rightarrow \quad \|a + b\|^2 = \|a\|^2 + \|b\|^2$$

E4 Si consideri un sistema di riferimento cartesiano nello spazio. Siano poi  $\pi$  il piano di equazione  $3x_1 - x_2 = 0$  e  $P$  il punto di coordinate  $(4, 1, 0)$ .

(a) Verificare che  $P \notin \pi$ .

Posto  $V = \mathbb{R}^3$  con prodotto scalare canonico,  $W = \{x \in \mathbb{R}^3 : 3x_1 - x_2 = 0\}$  e:

$$v = \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}$$

(b) Determinare una base di  $W$ ;

(c) Determinare la migliore approssimazione  $w^*$  di  $v$  in  $W$ ;

(d) Il punto  $P^*$  di coordinate le componenti di  $w^*$  è il punto di  $\pi$  più vicino a  $P$ . Determinare la distanza di  $P$  da  $\pi$ .

## 4.2 Calcolo delle soluzioni di un sistema nel senso dei minimi quadrati

Siano  $A \in \mathbb{R}^{n \times k}$  di colonne  $a_1, \dots, a_k \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^n$  e si consideri  $\mathbb{R}^n$  con prodotto scalare canonico. Allora:

- Le soluzioni di  $Ax = b$  nel senso dei minimi quadrati sono le *coordinate* rispetto ad  $a_1, \dots, a_k$  della migliore approssimazione di  $b$  in  $\text{span}\{a_1, \dots, a_k\}$ , ovvero della proiezione ortogonale di  $b$  su  $\text{span}\{a_1, \dots, a_k\}$ .

*Infatti:* La migliore approssimazione di  $b$  in  $\text{span}\{a_1, \dots, a_k\}$  è il minimo assoluto  $y^*$  della funzione  $d : \text{span}\{a_1, \dots, a_k\} \rightarrow \mathbb{R}$  definita da:

$$d(y) = \|b - y\| = \|y - b\|$$

Allora, posto  $y^* = Ax^*$ , per ogni  $x \in \mathbb{R}^k$  si ha  $Ax \in \text{span}\{a_1, \dots, a_k\}$  e quindi:

$$\|Ax - b\|_2 = \|Ax - b\| \geq \|y^* - b\| = \|Ax^* - b\|_2$$

Dunque, per definizione,  $x^*$  è una soluzione di  $Ax = b$  nel senso dei minimi quadrati. Le componenti di  $x^*$  sono coordinate rispetto ad  $a_1, \dots, a_k$  della migliore approssimazione  $y^*$ .

- Le coordinate rispetto ad  $a_1, \dots, a_k$  della migliore approssimazione di  $b$  in  $\text{span}\{a_1, \dots, a_k\}$ , ovvero della proiezione ortogonale di  $b$  su  $\text{span}\{a_1, \dots, a_k\}$ , sono le *soluzioni* del sistema delle equazioni normali definite dai generatori  $a_1, \dots, a_k$ . Ricordando che per ogni  $a, b \in \mathbb{R}^n$  si ha  $a \cdot b = b^T a$ , il sistema delle equazioni normali definite dai generatori  $a_1, \dots, a_k$  si scrive:

$$A^T Ax = A^T b$$

ovvero:

### 4.2.1 Teorema

Le soluzioni di  $Ax = b$  nel senso dei minimi quadrati sono le soluzioni del sistema delle equazioni normali:

$$A^T Ax = A^T b$$

### 4.2.2 Osservazione

Si consideri il sistema delle equazioni normali relative ad  $Ax = b$ . Come già osservato, il sistema ha sempre *almeno una* soluzione. Si ha inoltre:

- (i) La matrice  $A^T A \in \mathbb{R}^{k \times k}$  delle equazioni normali è *simmetrica* e *semidefinita positiva*.<sup>42</sup> La matrice è poi *definita positiva*, in particolare *invertibile*, se e solo se le colonne di  $A$  sono *linearmente indipendenti*.

*Infatti:* Per ogni  $x \in \mathbb{R}^k$  si ha:

$$A^T A x \cdot x = x^T A^T A x = (Ax)^T A x = Ax \cdot Ax = \|Ax\|^2 \geq 0$$

dunque la matrice  $A^T A$  è semidefinita positiva. Inoltre, si ha:

$$A^T A x \cdot x = \|Ax\|^2 = 0 \quad \text{se e solo se} \quad Ax = 0$$

e quindi  $A^T A$  è definita positiva se e solo se la condizione  $Ax = 0$  è equivalente a  $x = 0$ , ovvero se e solo se le colonne di  $A$  sono linearmente indipendenti.

- (ii) Sia  $S(A, b) \subset \mathbb{R}^k$  l'insieme delle soluzioni di  $Ax = b$  nel senso dei minimi quadrati. Sussistono i risultati seguenti:<sup>43</sup>

- Esiste *un solo* elemento di minima norma in  $S(A, b)$ .
- La funzione da  $\mathbb{R}^n$  in  $\mathbb{R}^k$  che associa a  $b$  l'elemento di minima norma in  $S(A, b)$  è un'*applicazione lineare*. La matrice  $k \times n$  che la definisce si chiama *pseudoinversa* di  $A$  e si indica con  $A^+$ .

Se  $n \geq k$  e le colonne di  $A$  sono linearmente indipendenti allora  $S(A, b)$  ha un solo elemento:

$$x^* = (A^T A)^{-1} A^T b$$

Dunque  $x^*$  è l'elemento di minima norma in  $S(A, b)$  e la matrice pseudoinversa di  $A$  è:

$$A^+ = (A^T A)^{-1} A^T$$

Se, inoltre,  $n = k$  allora  $A^+ = A^{-1}$ .

La ricerca delle soluzioni del sistema  $Ax = b$  nel senso dei minimi quadrati è dunque ricondotta alla costruzione e soluzione delle equazioni normali. Un procedimento *numericamente* preferibile alla determinazione della soluzione delle equazioni normali si ottiene estendendo la nozione di fattorizzazione QR al caso di matrici *non quadrate*.

#### 4.2.3 Definizione (fattorizzazione QR, caso non quadrato)

Sia  $A \in \mathbb{R}^{n \times k}$  con  $n \geq k$ . La coppia  $U, T$  è una *fattorizzazione QR* di  $A$  se:

- $A = UT$
- il fattore sinistro  $U$  è una matrice  $n \times k$  ad elementi reali con *colonne ortonormali* rispetto al prodotto scalare canonico in  $\mathbb{R}^n$ ;
- il fattore destro  $T$  è una matrice  $k \times k$  ad elementi reali *triangolare superiore*.

La ricerca di una fattorizzazione QR può essere effettuata, se le colonne di  $A$  sono linearmente indipendenti, con la procedura GS definita nel Paragrafo 2.7 del Capitolo 2. Esistono procedure per la ricerca di una fattorizzazione QR *più generali* di GS e ad essa preferibili da un punto di vista numerico. La funzione predefinita `qr` di *Scilab* realizza una di queste ultime.

- `qr`

Questa *funzione predefinita* restituisce una coppia di matrici che approssima una fattorizzazione QR di una matrice assegnata. Precisamente, se  $A$  è una matrice  $n \times k$  con  $n \geq k$  e:

$$[U, T] = \text{qr}(A, 'e')$$

allora la coppia  $U, T$  *approssima* una fattorizzazione QR di  $A$ . Come già osservato le colonne di  $A$  *possono* essere linearmente dipendenti.

<sup>42</sup>Si ricordi che una matrice simmetrica  $M \in \mathbb{R}^{n \times n}$  è *semidefinita positiva* se per ogni  $x \in \mathbb{R}^n$  si ha  $Mx \cdot x \geq 0$ . Se, inoltre,  $Mx \cdot x > 0$  per *tutti* gli  $x \neq 0$ , la matrice è *definita positiva*. Se  $x \neq 0$  e  $Mx = 0$  allora  $Mx \cdot x = 0$  ed  $M$  non è definita positiva, ovvero: Se  $M$  è definita positiva allora  $Mx = 0$  se e solo se  $x = 0$ , ovvero  $M$  è invertibile.

<sup>43</sup>Si veda, ad esempio: L. Aceto e M. Ciampa: *Complementi di Algebra e Fondamenti di Geometria, Capitolo 4, Decomposizione ai valori singolari* (<http://pagine.dm.unipi.it/~a008363/x-appunti.php>).

#### 4.2.4 Esempio

Sia:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Una fattorizzazione QR si determina utilizzando opportunamente la procedura GS. Procedendo come nell'Esempio 2.30:

– *Primo passo*

Si cercano  $\Omega = (\omega_1, \omega_2) \in \mathbb{R}^{3 \times 2}$  a colonne ortogonali e  $\Theta \in \mathbb{R}^{2 \times 2}$  triangolare superiore con uno sulla diagonale tali che  $\Omega\Theta = A$ , ovvero tali che, dette  $a_1, a_2$  le colonne di  $A$ :

$$\omega_1 = a_1 \quad , \quad \omega_1\theta_{12} + \omega_2 = a_2$$

Se esistono matrici siffatte allora, *necessariamente*:

$$\omega_1 = a_1 \quad , \quad \theta_{12} = \frac{\omega_1 \cdot a_2}{\omega_1 \cdot \omega_1} = 1$$

e quindi:

$$\omega_2 = a_2 - \omega_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Dunque:

$$\Omega = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad \Theta = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

– *Secondo passo*

Si *normalizzano* le colonne di  $\Omega$  lasciando inalterato il risultato del prodotto. Posto:

$$\Delta = \text{diag}(\|\omega_1\|, \|\omega_2\|) = \text{diag}(\sqrt{2}, 1)$$

si pone:

$$U = \Omega\Delta^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad T = \Delta\Theta = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & 1 \end{bmatrix}$$

Se  $A$  è una matrice a colonne linearmente indipendenti e  $(U, T)$  è una fattorizzazione QR di  $A$  allora:

$$A^+ = (A^T A)^{-1} A^T = (T^T T)^{-1} T^T U^T = T^{-1} (T^T)^{-1} T^T U^T = T^{-1} U^T$$

Per la matrice in esame si ha allora:

$$A^+ = \begin{bmatrix} \frac{1}{\sqrt{2}} & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Siano  $A$  a colonne linearmente indipendenti e  $U, T$  una fattorizzazione QR di  $A$ . Si ha:

*Il sistema delle equazioni normali  $A^T A x = A^T b$  è equivalente al sistema  $T x = U^T b$*

Infatti, tenuto conto che  $U^T U = I$ , si ricava:

$$A^T A = T^T T \quad \text{e} \quad A^T b = T^T U^T b$$

dunque il sistema delle equazioni normali si riscrive:

$$T^T T x = T^T U^T b$$



L'asserto si ottiene considerando che se le colonne di  $A$  sono linearmente indipendenti allora la matrice  $T$ , e quindi  $T^T$ , è invertibile. Infatti, ragionando per assurdo: Se  $Ty = 0$  per qualche  $y \neq 0$  allora  $Ay = UTy = 0$  per qualche  $y \neq 0$ , ovvero: le colonne di  $A$  sono linearmente dipendenti.

Si ha, inoltre:<sup>44</sup>

$$c_2(A^T A) = (c_2(T))^2$$

ovvero:

*Le proprietà di condizionamento di  $T$  sono (quasi sempre) migliori di quelle di  $A^T A$*

Dunque: un procedimento per la ricerca delle soluzioni del sistema  $Ax = b$  nel senso dei minimi quadrati *numericamente preferibile* alla costruzione e soluzione delle equazioni normali  $A^T Ax = b$  è quello di *calcolare una coppia  $U, T$  fattorizzazione QR di  $A$  e poi risolvere il sistema  $Tx = U^T b$ .*

#### 4.2.5 Osservazione (backslash)

*Scilab* ha una funzione predefinita per la ricerca delle soluzioni di un sistema di equazioni lineari: `backslash`.

– `backslash`

Questa *funzione predefinita* restituisce un vettore che approssima una soluzione o una soluzione nel senso dei minimi quadrati del sistema di equazioni lineari descritto dai dati di ingresso. Precisamente, detta  $u$  la precisione di macchina, dati  $A$  matrice  $n \times k$  e  $b$  colonna ad  $n$  componenti, `backslash(A,b)` o, più usualmente, `A\b`, restituisce la colonna a  $k$  componenti così determinata:

Se  $n = k$  allora:

- `[S,D,P] = EGPP(A)`;
- Se  $\det D = 0$  allora: `rcond = 0`; altrimenti: `rcond =` una stima di  $c_1(A)^{-1}$ ;
- Se `rcond > 20u` allora: `A\b = SI(D,SA(S,Pb))`;

Se  $n \neq k$  oppure `rcond ≤ 20u` allora:

- `A\b` = una colonna che approssima una soluzione di  $Ax = b$  nel senso dei minimi quadrati determinata utilizzando opportunamente una fattorizzazione QR di  $A$ .

#### 4.2.6 Esempio

Siano:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

La matrice  $A$  non è invertibile ma  $b$  è uguale alla prima colonna di  $A$  e il sistema ha infinite soluzioni:

$$S(A,b) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \lambda \in \mathbb{R}$$

Le soluzioni di  $Ax = b$  nel senso dei minimi quadrati coincidono con le soluzioni.

In *Scilab* si ha:

```
-->A = [1,1;1,1]; b = [1,1]';
```

```
-->x = A \ b
```

```
Warning :
```

```
matrix is close to singular or badly scaled. rcond = 0.0000D+00
computing least squares solution. (see lsq).
```

```
x =
```

```
1.
```

```
0.
```

<sup>44</sup>Si veda, ad esempio: M. Ciampa: *Calcolo Numerico* (<http://pagine.dm.unipi.it/~a008363/x-appunti.php>), Osservazione 5.21 punto (c).

Dopo aver avvisato l'utilizzatore che la stima di  $c_1(A)^{-1}$  è inferiore a  $20u \approx 2 \cdot 10^{-15}$  (e quindi  $c_1(A)$  è maggiore di  $(20u)^{-1} \approx 4 \cdot 10^{14}$ ), *Scilab* assegna ad  $x$  un valore che *approssima* una delle soluzioni di  $Ax = b$  nel senso dei minimi quadrati:

```
-->x == [1,0]
ans =
```

```
F
T
```

```
-->format(25)
```

```
-->x
x =
```

```
0.9999999999999998889777
0.
```

Si osservi che l'elemento di  $S(A, b)$  approssimato da *Scilab* non è quello di minima norma.

### 4.3 Calcolo delle funzioni che meglio approssimano dati assegnati nel senso dei minimi quadrati

Siano  $I$  un intervallo non degenere,  $F$  un sottospazio vettoriale di *dimensione finita* dello spazio delle funzioni continue da  $I$  in  $\mathbb{R}$ ,  $x_0, \dots, x_k$  numeri reali in  $I$  e  $y_0, \dots, y_k$  numeri reali. Ricordiamo che un elemento  $f^*$  di  $F$  è una *funzione che meglio approssima i dati*  $(x_0, y_0), \dots, (x_k, y_k)$  nel senso dei minimi quadrati se: Per ogni  $f \in F$  si ha:

$$(f^*(x_0) - y_0)^2 + \dots + (f^*(x_k) - y_k)^2 \leq (f(x_0) - y_0)^2 + \dots + (f(x_k) - y_k)^2$$

ovvero se  $f^*$  è un *minimo assoluto* della funzione *scarto quadratico*  $SQ : F \rightarrow \mathbb{R}$  definita da:

$$SQ(f) = (f(x_0) - y_0)^2 + \dots + (f(x_k) - y_k)^2$$

Sia  $f_1, \dots, f_j$  una *base* di  $F$ . Il problema si traduce allora nella ricerca di  $a_1, \dots, a_j \in \mathbb{R}$  tali che  $a_1 f_1(x) + \dots + a_j f_j(x)$  sia un minimo assoluto della funzione  $SQ$ . Poiché per ogni  $f \in F$  si ha:

$$SQ(f) = (f(x_0) - y_0)^2 + \dots + (f(x_k) - y_k)^2 = \left\| \begin{bmatrix} f(x_0) - y_0 \\ \vdots \\ f(x_k) - y_k \end{bmatrix} \right\|_2^2$$

allora:

$$SQ(a_1 f_1(x) + \dots + a_j f_j(x)) = \left\| \begin{bmatrix} a_1 f_1(x_0) + \dots + a_j f_j(x_0) - y_0 \\ \vdots \\ a_1 f_1(x_k) + \dots + a_j f_j(x_k) - y_k \end{bmatrix} \right\|_2^2$$

Posto:

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_j \end{bmatrix}, \quad A = \begin{bmatrix} f_1(x_0) & \dots & f_j(x_0) \\ \vdots & & \vdots \\ f_1(x_k) & \dots & f_j(x_k) \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} y_0 \\ \vdots \\ y_k \end{bmatrix}$$

si riscrive:

$$SQ(a_1 f_1(x) + \dots + a_j f_j(x)) = \|Aa - b\|_2^2$$

Osservato che  $A$  e  $b$  sono la matrice e colonna del sistema di equazioni che traduce le condizioni di *interpolazione* dei dati con elementi di  $F$  e ricordata la definizione di soluzione del sistema  $Ax = b$  nel senso dei minimi quadrati si deduce che: *Le coordinate delle funzioni che meglio approssimano i dati nel senso dei minimi quadrati sono le componenti delle soluzioni nel senso dei minimi quadrati del sistema che traduce le condizioni di interpolazione dei dati.*

### 4.3.1 Esercizio

Determinare gli elementi di  $P_1(\mathbb{R})$  che meglio approssimano i dati:

$$(0, 1) \quad , \quad (0, 2) \quad , \quad (1, 1) \quad , \quad (2, 0)$$

nel senso dei minimi quadrati.

*Soluzione.* Sia  $1, x$  una base di  $P_1(\mathbb{R})$ . Il sistema che traduce le condizioni di interpolazione dei dati con un elemento di  $P_1(\mathbb{R})$  è  $Ax = b$  con:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

I coefficienti che individuano gli elementi di  $P_1(\mathbb{R})$  che meglio approssimano i dati sono le soluzioni del sistema nel senso dei minimi quadrati. Poiché le colonne di  $A$  sono linearmente indipendenti il sistema ha *una sola* soluzione nel senso dei minimi quadrati. Il sistema delle equazioni normali è:

$$\begin{bmatrix} 4 & 3 \\ 3 & 5 \end{bmatrix} x = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

da cui:

$$x^* = \begin{bmatrix} \frac{17}{11} \\ -\frac{8}{11} \end{bmatrix}$$

e l'elemento cercato è:

$$p^*(x) = \frac{17}{11} - \frac{8}{11}x$$

Si osservi che, detta  $b^*$  la proiezione ortogonale di  $b$  sul sottospazio di  $\mathbb{R}^4$  generato dalle colonne di  $A$  si ha:  $b^* = Ax^*$  e quindi:

$$\|Ax^* - b^*\| = 0$$

Dunque: *l'elemento  $p^*$  migliore approssimazione dei dati  $(x_i, b_{i+1}), i = 0, \dots, 3$  interpola i dati  $(x_i, b_{i+1}^*), i = 0, \dots, 3$ .*

Le nozioni di soluzione di un sistema di equazioni lineari nel senso dei minimi quadrati e di funzione che meglio approssima i dati nel senso dei minimi quadrati possono essere estese modificando le funzioni  $n$  e SQ con l'introduzione di un coefficiente positivo, detto *peso*, per ciascun addendo. L'esempio seguente illustra queste estensioni.

### 4.3.2 Esercizio

Sia  $F$  un sottospazio vettoriale di dimensione due dello spazio delle funzioni continue da  $I$  in  $\mathbb{R}$  e  $p_0, p_1, p_2$  numeri reali positivi. Determinare gli elementi di  $F$  minimi assoluti della funzione SQ:  $F \rightarrow \mathbb{R}$  definita da:

$$\text{SQ}(f) = p_0 (f(x_0) - y_0)^2 + p_1 (f(x_1) - y_1)^2 + p_2 (f(x_2) - y_2)^2$$

*Soluzione.* Per ogni  $f \in F$  si ha:

$$\text{SQ}(f) = p_0 (f(x_0) - y_0)^2 + p_1 (f(x_1) - y_1)^2 + p_2 (f(x_2) - y_2)^2 = \left\| \begin{bmatrix} \sqrt{p_0} (f(x_0) - y_0) \\ \sqrt{p_1} (f(x_1) - y_1) \\ \sqrt{p_2} (f(x_2) - y_2) \end{bmatrix} \right\|_2^2$$

Procedendo come sopra si ottiene che, detta  $f_1, f_2$  una base di  $F$  e posto:

$$a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad , \quad \Delta = \text{diag}(\sqrt{p_0}, \sqrt{p_1}, \sqrt{p_2}) \quad , \quad A = \begin{bmatrix} f_1(x_0) & f_2(x_0) \\ f_1(x_1) & f_2(x_1) \\ f_1(x_2) & f_2(x_2) \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}$$

si riscrive:

$$\text{SQ}(a_1 f_1(x) + a_2 f_2(x)) = \|\Delta Aa - \Delta b\|_2^2$$

Si deduce che: *Le coordinate delle funzioni che meglio approssimano i dati nel senso dei minimi quadrati, con pesi  $p_0, p_1, p_2$ , sono le componenti delle soluzioni nel senso dei minimi quadrati del sistema  $\Delta Ax = \Delta b$ . Queste ultime sono le soluzioni del sistema delle equazioni normali:*

$$A^T \Delta^2 Ax = A^T \Delta^2 b$$

### Esercizi

E5 Siano:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 2 & -1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Determinare una fattorizzazione QR di  $A$  utilizzando la procedura GS ed utilizzarla per calcolare le soluzioni del sistema  $Ax = b$  nel senso dei minimi quadrati.

E6 ♠ Verificare che in *Scilab* dopo gli assegnamenti:

```
-->A = [1,1,0;1,1,1]';
```

```
-->[U,T] = qr(A,'e');
```

la coppia  $(U, T)$  è un'approssimazione della fattorizzazione QR di  $A$ :

$$U = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 \\ 0 & -1 \end{bmatrix}, \quad T = \begin{bmatrix} -\sqrt{2} & -\sqrt{2} \\ 0 & -1 \end{bmatrix}$$

Confrontare la fattorizzazione con quella ottenuta nell'Esempio 4.2.4.

E7 Siano  $A$  e  $A^+$  come nell'Esempio 4.2.4. Determinare  $A^+A$  e  $AA^+$ .

Il primo risultato giustifica il termine "pseudoinversa" utilizzato per la matrice  $A^+$ . Dimostrare che per ogni  $B \in \mathbb{R}^{n \times k}$  a colonne linearmente indipendenti si ha:

$$B^+B = I$$

E8 ♠ Utilizzare la funzione `backslash` per determinare gli elementi di  $P_1(\mathbb{R})$  che meglio approssimano i dati:

$$(0, 1), \quad (0, 2), \quad (1, 1), \quad (2, 0)$$

nel senso dei minimi quadrati e disegnare, su uno stesso piano cartesiano, i dati ed il grafico dell'elemento ottenuto.

E9 Sia  $F : \mathbb{R}^3 \rightarrow \mathbb{R}$  definita da:

$$F(x) = x_1^2 + 2x_2^2 + 3x_3^2 + (x_1 - x_2 + x_3 - 1)^2 + (-x_1 - 4x_2 + 2)^2$$

Determinare una matrice  $A \in \mathbb{R}^{n \times 3}$  ed una colonna  $b \in \mathbb{R}^n$  tali che per ogni  $x \in \mathbb{R}^3$  si abbia:

$$F(x) = \|Ax - b\|_2^2$$

Determinare poi il minimo assoluto di  $F$ .

E10 Assegnato un sistema di riferimento cartesiano in un piano  $\pi$ , siano  $c_1, \dots, c_j \in \mathbb{R}^2$  i vettori delle coordinate di  $j$  punti distinti di  $\pi$ . Si consideri la funzione  $\lambda : \mathbb{R}^2 \rightarrow \mathbb{R}$  definita da:

$$\lambda(x) = \|x - c_1\|^2 + \dots + \|x - c_j\|^2$$

\* Dare un'interpretazione geometrica della funzione  $\lambda$ .

\* Determinare  $A \in \mathbb{R}^{2j \times 2}$  e  $b \in \mathbb{R}^{2j}$  tali che per ogni  $x \in \mathbb{R}^2$  si abbia:

$$\lambda(x) = \|Ax - b\|_2^2$$

\* Siano:

$$c_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad c_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad c_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Determinare il minimo assoluto di  $\lambda$ .

*E11* Assegnato un sistema di riferimento cartesiano in un piano  $\pi$ , siano  $c_1, \dots, c_j \in \mathbb{R}^2$  i vettori delle coordinate di  $j$  punti distinti di  $\pi$ . Assegnati  $k_1, \dots, k_j$  numeri reali *positivi*, si consideri la funzione EP :  $\mathbb{R}^2 \rightarrow \mathbb{R}$  definita da:

$$\text{EP}(x) = \frac{1}{2} k_1 \|x - c_1\|^2 + \dots + \frac{1}{2} k_j \|x - c_j\|^2$$

\* Dare un'interpretazione meccanica della funzione EP.

\* Determinare  $A \in \mathbb{R}^{2j \times 2}$  e  $b \in \mathbb{R}^{2j}$  tali che per ogni  $x \in \mathbb{R}^2$  si abbia:

$$\text{EP}(x) = \frac{1}{2} \|Ax - b\|_2^2$$

\* Siano:

$$c_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad c_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad c_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

e  $k_1 = k_2 = 1, k_3 = 4$ . Determinare il minimo assoluto di EP.

---