Lezione 33

In questa lezione si conclude lo studio dei problemi di *approssimazione nel senso dei minimi quadrati*: si estende la nozione di *fattorizzazione QR* a matrici non quadrate per descrivere un procedimento di ricerca delle soluzioni di un sistema nel senso dei minimi quadrati, si descrive la funzione predefinita `backslash` di *Scilab* e, infine, si determinano le funzioni che meglio approssimano i dati 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. Sappiamo che le soluzioni di $Ax = b$ nel senso dei minimi quadrati sono le colonne delle *coordinate* rispetto ad $a_1, \dots, a_k$ della migliore approssimazione di b in $\operatorname{span}\{a_1, \dots, a_k\}$, ovvero della proiezione ortogonale di b su $\operatorname{span}\{a_1, \dots, a_k\}$, e che queste ultime sono *le soluzioni* del sistema delle equazioni normali:

$$A^\mathsf{T} A x = A^\mathsf{T} b$$

- *Osservazione.*

  (A) La matrice $A^\mathsf{T} A \in \mathbb{R}^{k\times k}$ delle equazioni normali è *simmetrica* e *semidefinita positiva.*[1]
  La matrice $A^\mathsf{T} A$ è *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:

  $$A^\mathsf{T} A x \cdot x = x^\mathsf{T} A^\mathsf{T} A x = (Ax)^\mathsf{T} A x = Ax \cdot Ax = \| Ax \|^2 \geqslant 0$$

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

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

  e quindi $A^\mathsf{T} A$ 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.

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

  * Esiste *un solo* elemento di minima norma in $S(A, b)$.
  * La funzione che associa a b l'elemento di minima norma in $S(A, b)$ è un'*applicazione lineare.* La matrice che la definisce si chiama *pseudoinversa di A* e si indica con A^+.

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

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

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

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

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

La ricerca delle soluzioni del sistema $Ax = b$ nel senso dei minimi quadrati è dunque ricondotto alla costruzione e soluzione delle equazioni normali. *Se le colonne di A sono linearmente indipendenti*, 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.*

- *Definizione* (fattorizzazione QR, caso non quadrato).
  Sia $A \in \mathbb{R}^{n\times k}$ con $n \geqslant k$. La coppia U, T è una *fattorizzazione QR* di A se:

  * U è una matrice $n \times k$ ad elementi reali con *colonne ortonormali* rispetto al prodotto scalare canonico in $\mathbb{R}^n$;
  * T è una matrice $k \times k$ ad elementi reali *triangolare superiore*;

[1] 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 $Mx \cdot x \geqslant 0$. *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.

* $UT = A$.

La ricerca di una fattorizzazione QR può essere effettuata, se le colonne di A sono linearmente indipendenti, con la procedura GS definita nella Lezione 25. Esistono però procedure per la ricerca di una fattorizzazione *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 $n \geqslant k$ e:

  ```
  [U,T] = 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.

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

(1) *Il sistema delle equazioni normali* $A^{\mathsf{T}}Ax = A^{\mathsf{T}}b$ *è* equivalente *al sistema* $Tx = U^{\mathsf{T}}b$

Infatti, tenuto conto che $U^{\mathsf{T}}U = I$, si ricava:

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

dunque il sistema delle equazioni normali si riscrive:

$$T^{\mathsf{T}}Tx = T^{\mathsf{T}}U^{\mathsf{T}}b$$

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:

$$c_2(A^{\mathsf{T}}A) = \big(\, c_2(T)\,\big)^2$$

ovvero:

(2) *Le proprietà di condizionamento di* T *sono* (*quasi sempre*) migliori *di quelle di* $A^{\mathsf{T}}A$

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

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 $n = k$:
    * $[\mathtt{S}, \mathtt{D}, \mathtt{P}] = \mathtt{EGPP(A)}$;
    * $\mathtt{rcond}$ = una stima di $c_1(\mathtt{A})^{-1}$;
    * Se $\mathtt{rcond} > 20\,u$ allora: $\mathtt{A\backslash b} = \mathtt{SI}\big(\mathtt{D}, \mathtt{SA}(\mathtt{S}, \mathtt{Pb})\big)$;

  – Se $n \neq k$ oppure $\mathtt{rcond} \leqslant 20\,u$:
    * `A\b` = una colonna che approssima una soluzione di $\mathtt{A}x = \mathtt{b}$ nel senso dei minimi quadrati.

- *Esempio.*

  Siano:
  $$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
  La matrice A non è invertibile ma b è uguale alla prima colonna di A e il sistema ha infinite soluzioni:
  $$\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad , \quad \lambda \in \mathbb{R}$$
  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.
```

  Dopo aver avvisato l'utilizzatore che la stima di $c_1(A)^{-1}$ è inferiore a $20\,u \approx 2 \cdot 10^{-15}$ (e quindi $c_1(A)$ è maggiore di $(20\,u)^{-1} \approx 4 \cdot 10^{14}$), *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.
```

Calcolo delle funzioni che meglio approssimano i dati nel senso dei minimi quadrati

Siano I un intervallo non degenere, $\mathcal{F}$ un sottospazio vettoriale *di dimensione finita* dello spazio delle funzioni continue da I in $\mathbb{R}$, $x_0, \ldots, x_k$ numeri reali in I e $y_0, \ldots, y_k$ numeri reali. Ricordiamo che un elemento f^* di $\mathcal{F}$ è una *funzione che meglio approssima i dati* $(x_0, y_0), \ldots, (x_k, y_k)$ *nel senso dei minimi quadrati* se: *Per ogni* $f \in \mathcal{F}$ *si ha*:
$$\big(\, f^*(x_0) - y_0 \,\big)^2 + \cdots + \big(\, f^*(x_k) - y_k \,\big)^2 \leqslant \big(\, f(x_0) - y_0 \,\big)^2 + \cdots + \big(\, f(x_k) - y_k \,\big)^2$$
ovvero se f^* è un *minimo assoluto* della funzione $\mathrm{SQ} : \mathcal{F} \to \mathbb{R}$, *scarto quadratico*, definita da:
$$\mathrm{SQ}(f) = \big(\, f(x_0) - y_0 \,\big)^2 + \cdots + \big(\, f(x_k) - y_k \,\big)^2$$
Sia $f_1, \ldots, f_j$ una *base* di $\mathcal{F}$. Il problema si traduce allora nella ricerca di $a_1, \ldots, a_j \in \mathbb{R}$ tali che $a_1 f_1(x) + \cdots + a_j f_j(x)$ sia un minimo assoluto della funzione SQ. Poiché per ogni $f \in \mathcal{F}$ si ha:
$$\mathrm{SQ}(f) = \big(\, f(x_0) - y_0 \,\big)^2 + \cdots + \big(\, f(x_k) - y_k \,\big)^2 = \left\| \begin{bmatrix} f(x_0) - y_0 \\ \vdots \\ f(x_k) - y_k \end{bmatrix} \right\|_2^2$$

allora:

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

Posto:

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

si riscrive:

$$\mathrm{SQ}\big(a_1 f_1(x) + \cdots + a_j f_j(x)\big) = \| \, Aa - b \, \|_2^2$$

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 $\mathcal{F}$ 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.*

- *Esempio.*

  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:

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

  e l'elemento cercato è:

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

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.

- *Esempio.*

  Sia $\mathcal{F}$ un sottospazio vettoriale di dimensione due dello spazio delle funzioni continue da I in $\mathbb{R}$. Determinare gli elementi di $\mathcal{F}$ minimi assoluti della funzione $\mathrm{SQ} : \mathcal{F} \to \mathbb{R}$ definita da:

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

  Soluzione. Per ogni $f \in \mathcal{F}$ si ha:

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

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

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

si riscrive:

$$\mathrm{SQ}\big(a_1 f_1(x) + a_2 f_2(x)\big) = \| \, \Delta A a - \Delta b \, \|_2^2$$

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* $\Delta A x = \Delta b$. Queste ultime sono le soluzioni del sistema delle equazioni normali:

$$A^{\mathsf{T}} \Delta^2 A x = A^{\mathsf{T}} \Delta^2 b$$

Esercizi

---

1. Siano:

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

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.

2. Utilizzare la funzione `backslash` per 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 e per disegnare, su uno stesso piano cartesiano, i dati ed il grafico dell'elemento ottenuto.

3. Sia $F : \mathbb{R}^3 \to \mathbb{R}$ 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:

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

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

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

   * Dare un'interpretazione geometrica della funzione λ.
   * 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:

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

   Determinare il minimo assoluto di λ.

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

$$\mathrm{EP}(x) = \tfrac{1}{2}\, k_1 \, \| \, x - c_1 \, \|^2 + \cdots + \tfrac{1}{2}\, k_j \, \| \, x - c_j \, \|^2$$

* 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:

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

* Siano:

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

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