

Minimi quadrati

27 novembre 2008

In questa nota si indicherà sempre con V uno spazio vettoriale reale di dimensione finita n .

Definizione 1. Una *norma* su V è una funzione $\|\cdot\| : V \mapsto \mathbb{R}$ tale che $\forall \mathbf{x}, \mathbf{y} \in V$ e $\lambda \in \mathbb{R}$:

1. $\|\mathbf{x}\| \geq 0$;
2. $\|\mathbf{x}\| = 0$ implica $\mathbf{x} = \mathbf{0}$;
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (disuguaglianza triangolare).
4. $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$.

Esempio 1. Per ogni intero $p \geq 1$ la funzione

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$$

è una norma su \mathbb{R}^n , come pure

$$\|\mathbf{x}\|_\infty = \max_{i=1, \dots, n} \{|x_i|\}.$$

La norma $\|\cdot\|_2$ che si ottiene per $p = 2$ è detta *norma euclidea*.

Sistemi iperdeterminati

Si consideri il sistema lineare descritto dall'equazione matriciale

$$AX = B, \tag{1}$$

ove A è una matrice $n \times m$, X il vettore delle indeterminate e B un vettore colonna di \mathbb{R}^n . Tale sistema ammette soluzione se, e solamente se, il vettore colonna B appartiene all'immagine dell'applicazione lineare $f : \mathbb{R}^m \mapsto \mathbb{R}^n$ indotta dalla matrice A . È ben noto che per verificare l'esistenza di tale soluzione è possibile applicare il teorema di Rouché–Capelli, confrontando la dimensione dello spazio vettoriale generato dalle colonne di A con la dimensione dello spazio vettoriale generato da queste unitamente al vettore B ; in tale modo il calcolo da effettuare è un confronto fra ranghi.

Diremo che un sistema è *iperdeterminato* se esso non ammette soluzione. Vogliamo trovare una formulazione più generale rispetto (1), tale che:

1. Quando il sistema è risolubile fornisca la soluzione cercata.
2. Offra la risposta “migliore possibile” quando il sistema è iperdeterminato.

Si fissi una norma in \mathbb{R}^n e si consideri ora il problema di *trovare tutti i vettori* $X \in \mathbb{R}^m$ *tali che*

$$\|AX - B\| \tag{2}$$

sia minimo; ovvero trovare tutti gli X tali che per ogni $Z = AY$ appartenente all'immagine di A si abbia

$$\|AX - B\| \leq \|Z - B\|.$$

Osserviamo che $\|AX - B\| = \|\mathbf{0}\| = 0$ se, e solamente se X è soluzione di (1). Chiaramente 0 è in questo caso il minimo valore che la norma può assumere. Pertanto, per sistemi risolubili, i problemi (1) e (2) sono equivalenti ed è ben chiaro che cosa si possa intendere per soluzione “migliore possibile”.

In generale, l'eventuale soluzione di (2) quando (1) non è risolubile dipende strettamente dal tipo di norma considerato.

Esempio 2. Consideriamo il sistema

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} x = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

con $b_1 \geq b_2 \geq b_3 \geq 0$. Chiaramente, a meno che b_1 , b_2 e b_3 non siano tutti uguali, tale sistema non è risolubile. Consideriamo il problema (2) associato per 3 norme differenti:

- $\|Ax - b\|_1$: in questo caso cerchiamo di minimizzare la quantità

$$|x - b_1| + |x - b_2| + |x - b_3|.$$

Risolvendo i moduli si vede che la soluzione del problema è $x = b_2$.

- $\|Ax - b\|_2$: minimizzare la norma equivale a minimizzare la quantità

$$2\varphi(x) = (x-b_1)^2 + (x-b_2)^2 + (x-b_3)^2 = 3x^2 - 2(b_1 + b_2 + b_3)x + (b_1^2 + b_2^2 + b_3^2)$$

In particolare, calcolando la derivata si ottiene

$$\frac{d\varphi}{dx} = 3x - (b_1 + b_2 + b_3),$$

da cui $x = \frac{1}{3}(b_1 + b_2 + b_3)$, cioè x è la media aritmetica delle entrate del vettore dei termini noti.

- $\|Ax - b\|_\infty$: in questo caso dobbiamo considerare la quantità

$$\max\{|x - b_1|, |x - b_2|, |x - b_3|\}.$$

Poiché b_2 è compreso fra b_1 e b_3 , il massimo coincide con

$$\max\{|x - b_1|, |x - b_3|\}.$$

In particolare, il valore di x che minimizza quest'ultima espressione è proprio la media $x = \frac{1}{2}(b_1 + b_3)$.

Ci occuperemo ora in dettaglio delle tecniche risolutive per il problema (2) quando le stime siano effettuate rispetto la norma euclidea.

Questo è motivato da essenzialmente due fattori:

1. la funzione $\varphi(X) = \frac{1}{2}\|AX - B\|_2^2$ è in questo caso differenziabile (come pure per tutti i valori di p tali che $1 < p < \infty$);
2. esiste un prodotto scalare di \mathbb{R}^n che induce esattamente la norma considerata e tale prodotto consentirà di ben rappresentare la situazione geometrica in esame.

Distinguiamo un approccio di tipo analitico e un approccio di tipo geometrico al problema.

Approccio analitico

Consideriamo il funzionale $\mathbb{R}^n \mapsto \mathbb{R}$ dato da

$$\varphi(X) = \frac{1}{2}\|AX - B\|_2^2.$$

Chiaramente, se X che minimizza $\varphi(X)$, allora X è una soluzione del problema (2). A tal fine, studiamo i punti estremali del funzionale, ovvero cerchiamo di risolvere l'equazione

$$\nabla\varphi(X) = \mathbf{0}. \quad (3)$$

Un calcolo diretto mostra che

$$\begin{aligned} \frac{\partial\varphi}{\partial X_i} = & a_{1i}(a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m - b_1) + \\ & a_{2i}(a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m - b_2) + \\ & \vdots \\ & a_{ni}(a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nm}x_m - b_n) = \\ & (a_{1i} \ a_{2i} \ \dots \ a_{ni}) \left(A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} - B \right). \end{aligned}$$

Scrivendo ora in forma matriciale le equazioni

$$\frac{\partial\varphi}{\partial X_i} = 0,$$

al variare di $i = 1, \dots, m$ si ottiene che il gradiente di φ si annulla se, e solamente se, X è soluzione del sistema

$$(A^T A)X = A^T B. \quad (4)$$

Pertanto, per risolvere il problema (2) rispetto $\|\cdot\|_2$ si deve a risolvere il sistema lineare (4). Le m equazioni in m incognite di tale sistema sono dette *equazioni normali* del problema.

Osserviamo che, in generale le equazioni normali ammettono esattamente una soluzione se, e solamente se, la matrice quadrata $A^T A$ è non singolare. Questa condizione è esattamente quella per cui il problema (2) ammette una unica soluzione rispetto la norma $\|\cdot\|_2$.

Approccio algebrico

L'approccio analitico mostrato nel precedente paragrafo consente di ottenere le equazioni normali del problema in modo rapido; d'altro canto esso non rende evidente il motivo per cui il sistema dato dalle equazioni normali sia lineare.

Vediamo ora come sia possibile ottenere la medesima soluzione utilizzando strumenti prettamente algebrici.

Definizione 2. Un *proiettore* $\Pi : \mathbb{R}^n \mapsto \mathbb{R}^n$ è una applicazione lineare tale che

$$\Pi^2 = \Pi.$$

L'immagine $\Pi(\mathbf{x})$ di un vettore \mathbf{x} è detta *ombra* di \mathbf{x} secondo Π .

Osserviamo che la matrice P associata ad un proiettore, chiaramente, soddisfa la condizione

$$P^2 = P.$$

Una matrice di tale tipo è detta *idempotente*.

Teorema 1. Sia $\Pi : \mathbb{R}^n \mapsto \mathbb{R}^n$ un proiettore. Allora, $\mathfrak{I}\Pi = \ker(I - \Pi)$ e $\mathfrak{I}(I - \Pi) = \ker \Pi$. In particolare

$$\mathbb{R}^n = \mathfrak{I}\Pi \oplus \mathfrak{I}(I - \Pi).$$

Dimostrazione. Consideriamo un vettore $\mathbf{x} \in \mathfrak{I}\Pi$. Allora esiste $\mathbf{y} \in \mathbb{R}^n$ tale che $\mathbf{x} = \Pi\mathbf{y}$. Poiché $\Pi^2 = \Pi$, abbiamo

$$\Pi(\mathbf{x}) = \Pi^2(\mathbf{y}) = \Pi(\mathbf{y}) = \mathbf{x};$$

pertanto $\mathbf{x} \in \ker(I - \Pi)$. Viceversa, se $\mathbf{x} \in \ker(I - \Pi)$, si ha

$$\Pi(\mathbf{x}) = \mathbf{x};$$

dunque $\mathbf{x} \in \mathfrak{I}\Pi$ e la prima parte della tesi segue. La proprietà per $(I - \Pi)$ è conseguenza del fatto che

$$(I - \Pi)^2 = (I - \Pi)(I - \Pi) = (I - 2\Pi + \Pi^2) = (I - \Pi),$$

per cui anche $(I - \Pi)$ è un proiettore. □

L'idea alla base dell'approccio geometrico al problema (2) è quella di associare all'immagine \mathcal{U} dell'applicazione lineare indotta da A un proiettore $\Pi_{\mathcal{U}}$ tale che:

1. $\forall \mathbf{x} \in \mathcal{U} : \Pi_{\mathcal{U}}(\mathbf{x}) = \mathbf{x}$;
2. $\forall \mathbf{b} \in \mathbb{R}^n : \Pi_{\mathcal{U}}(\mathbf{b}) \in \mathcal{U}$ e, per ogni $\mathbf{x} \in \mathcal{U}$,

$$\|\Pi_{\mathcal{U}}(\mathbf{b}) - \mathbf{b}\| \leq \|\mathbf{x} - \mathbf{b}\|.$$

Osserviamo che noto il proiettore Π_U con le proprietà sopra elencate possiamo passare dal sistema lineare originario $AX = B$ ad un nuovo sistema, che sappiamo essere risolubile,

$$AX = \Pi(B) \quad (5)$$

e che, sotto le ipotesi, tale sistema è sicuramente lineare.

Supponiamo infatti che \tilde{X} sia una soluzione di (5) che U indichi l'immagine di A . Allora, per ogni $W \in \mathbb{R}^m$,

$$\begin{aligned} \|A\tilde{X} - B\| &= \|A\tilde{X} - \Pi(B) + \Pi(B) - B\| \leq \|A\tilde{X} - \Pi(B)\| + \|\Pi(B) - B\| = \\ &= \|\Pi(B) - B\| = \min_{Z \in U} \|Z - B\| \leq \|AW - B\|. \end{aligned}$$

Pertanto \tilde{X} minimizza la quantità $\|AX - B\|$ e, conseguentemente, esso è soluzione di (2).

Mostriamo ora come costruire il proiettore Π_U desiderato quando si considera la $\|\cdot\|_2$ e ottenere un sistema lineare equivalente alla (5).

Definizione 3. Sia V uno spazio vettoriale sul campo \mathbb{R} . Si dice *prodotto scalare* ogni applicazione $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{R}$ che è

1. bilineare, ovvero $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ e $\lambda \in \mathbb{R}$:

$$\langle \mathbf{x} + \lambda \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \lambda \langle \mathbf{y}, \mathbf{z} \rangle,$$

$$\langle \mathbf{x}, \lambda \mathbf{y} + \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle;$$

2. simmetrica, ovvero

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle;$$

3. definita positiva, ovvero $\forall \mathbf{x} \in V$,

$$\langle \mathbf{x}, \mathbf{x} \rangle \geq 0;$$

$$\langle \mathbf{x}, \mathbf{x} \rangle = 0 \text{ se, e solamente se } \mathbf{x} = \mathbf{0}.$$

Teorema 2. Assegnato un prodotto scalare $\langle \cdot, \cdot \rangle$, la funzione

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}}$$

è una norma su V .

Dimostrazione. I fatti che $\|\mathbf{x}\| \geq 0$ e $\|\mathbf{x}\| = 0$ se, e solamente se $\mathbf{x} = \mathbf{0}$ sono conseguenza immediata della definita positività del prodotto scalare. Inoltre, per la bilinearità, abbiamo

$$\|\lambda\mathbf{x}\|^2 = \langle \lambda\mathbf{x}, \lambda\mathbf{x} \rangle = \lambda^2 \langle \mathbf{x}, \mathbf{x} \rangle.$$

Resta da dimostrare la disuguaglianza triangolare. Calcoliamo dapprima

$$\|\mathbf{x} + \alpha\mathbf{y}\|^2 = \langle \mathbf{x} + \alpha\mathbf{y}, \mathbf{x} + \alpha\mathbf{y} \rangle = \|\mathbf{x}\|^2 + 2\alpha\langle \mathbf{x}, \mathbf{y} \rangle + \alpha^2\|\mathbf{y}\|^2.$$

Per la definita positività del prodotto scalare, la quantità scritta non può mai essere negativa; pertanto, il discriminante dell'equazione in α a sinistra deve essere negativo. Ne segue $\langle \mathbf{x}, \mathbf{y} \rangle^2 < \|\mathbf{x}\|^2\|\mathbf{y}\|^2$. Estrahendo le radici quadrate si ottiene

$$|\langle \mathbf{x}, \mathbf{y} \rangle| < \|\mathbf{x}\| \|\mathbf{y}\|. \quad (6)$$

La disuguaglianza (6) è detta *disuguaglianza di Schwarz*. Applicando tale disuguaglianza all'espansione di $\|\mathbf{x} + \mathbf{y}\|^2$ si ha

$$\|\mathbf{x} + \mathbf{y}\|^2 = \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle = \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2,$$

da cui, per la positività della norma,

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

La tesi segue. □

A meno di cambiamenti di riferimento, esiste essenzialmente un unico prodotto scalare in \mathbb{R}^n . Questi, in un opportuno riferimento, è dato da

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i.$$

Se \mathbf{x} e \mathbf{y} sono vettori colonna, tale prodotto si può anche scrivere come

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}.$$

La norma indotta è

$$\|\mathbf{x}\| = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}},$$

che coincide esattamente con $\|\mathbf{x}\|_2$.

Teorema 3. Sia V uno spazio vettoriale reale di dimensione finita e V^* il suo duale. L'applicazione $\Theta : V \mapsto V^*$ che associa ad ogni $\mathbf{x} \in V$

$$\Theta_{\mathbf{x}} : \begin{cases} V & \rightarrow \mathbb{R} \\ \mathbf{y} & \rightarrow \langle \mathbf{x}, \mathbf{y} \rangle \end{cases}$$

è un isomorfismo.

Dimostrazione. Poiché il prodotto scalare è bilineare, abbiamo che, per ogni \mathbf{x} fissato e per ogni $\mathbf{y}, \mathbf{z} \in V, \lambda \in \mathbb{R}$:

$$\Theta_{\mathbf{x}}(\mathbf{y} + \lambda \mathbf{z}) = \langle \mathbf{x}, \mathbf{y} + \lambda \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \lambda \langle \mathbf{x}, \mathbf{z} \rangle = \Theta_{\mathbf{x}}(\mathbf{y}) + \lambda \Theta_{\mathbf{x}}(\mathbf{z}).$$

Pertanto il codominio di Θ è proprio V^* . Inoltre, per ogni $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$,

$$\Theta_{\mathbf{x} + \lambda \mathbf{y}}(\mathbf{z}) = \langle \mathbf{x} + \lambda \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \lambda \langle \mathbf{y}, \mathbf{z} \rangle = \Theta_{\mathbf{x}}(\mathbf{z}) + \lambda \Theta_{\mathbf{y}}(\mathbf{z}) = (\Theta_{\mathbf{x}} + \lambda \Theta_{\mathbf{y}})(\mathbf{z});$$

quindi, Θ è una applicazione lineare. Siccome $\dim V = \dim V^*$, per dimostrare che Θ è un isomorfismo basta far vedere che $\ker \Theta = \{\mathbf{0}\}$. Infatti, se $\mathbf{x} \in \ker \Theta$, allora per ogni $\mathbf{y} \in V$,

$$\Theta_{\mathbf{x}}(\mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle = 0.$$

In particolare, $\Theta_{\mathbf{x}}(\mathbf{x}) = 0$, da cui, per la definita positività del prodotto scalare, $\mathbf{x} = \mathbf{0}$. □

Definizione 4. Sia U un sottospazio vettoriale di \mathbb{R}^n . Il *complemento ortogonale* di U , indicato con U^\perp è l'insieme di tutti i vettori $\mathbf{x} \in \mathbb{R}^n$ tali che $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ per ogni $\mathbf{y} \in U$.

Teorema 4. Il complemento ortogonale di U è un sottospazio vettoriale di \mathbb{R}^n . Inoltre

$$\mathbb{R}^n = U \oplus U^\perp.$$

Dimostrazione. I vettori di U^\perp sono esattamente l'annullatore dell'immagine di U in V^* mediante Θ . Pertanto, essi costituiscono un sottospazio vettoriale di dimensione $n - \dim U$. Inoltre, se $\mathbf{x} \in U \cap U^\perp$, allora il prodotto scalare di \mathbf{x} con ogni vettore di \mathbb{R}^n (incluso se stesso) è 0, per cui $\mathbf{x} = \mathbf{0}$. □

Usando la teoria degli annullatori, si vede rapidamente che $U^{\perp\perp} = U$.

Definizione 5. Per ogni $\mathbf{x} \in \mathbb{R}^n$ si dice *proiezione ortogonale su U di \mathbf{x}* ogni vettore $\mathbf{y} \in U$ tale che

$$\mathbf{x} - \mathbf{y} \in U^\perp.$$

Poiché $\mathbb{R}^n = U \oplus U^\perp$, ogni $\mathbf{x} \in \mathbb{R}^n$ si può scrivere in modo unico

$$\mathbf{x} = \mathbf{a} + \mathbf{b},$$

ove $\mathbf{a} \in U$ e $\mathbf{b} \in U^\perp$. Pertanto esiste per ogni vettore \mathbf{x} almeno una proiezione ortogonale su U .

Teorema 5. *La proiezione ortogonale \mathbf{y} di $\mathbf{x} \in \mathbb{R}^n$ su U è unica.*

Dimostrazione. Siano \mathbf{y}, \mathbf{y}' due proiezioni ortogonali di \mathbf{x} su U . Allora, $\mathbf{y} - \mathbf{y}' \in U^\perp$; d'altro canto $\mathbf{y}, \mathbf{y}' \in U$. Ne segue $\mathbf{y} - \mathbf{y}' \in U \cap U^\perp$ e dunque $\mathbf{y} = \mathbf{y}'$. \square

Conseguenza immediata del Teorema 5 è che per ogni $\mathbf{y} \in U$ si ha $\Pi_U(\mathbf{y}) = \mathbf{y}$, cioè che ogni vettore in U è fissato da Π_U . Per mostrare che Π_U è effettivamente un proiettore ai sensi della Definizione 2 dobbiamo ora far vedere che essa è una applicazione lineare.

Teorema 6. *La proiezione ortogonale su U è una applicazione lineare $\mathbb{R}^n \rightarrow U$.*

Dimostrazione. Siano $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$ e λ in \mathbb{R} . Per definizione, i vettori $\mathbf{x} - \Pi_U(\mathbf{x})$ e $\mathbf{z} - \Pi_U(\mathbf{z})$ appartengono ad U^\perp . In particolare, il vettore

$$\mathbf{x} - \Pi_U(\mathbf{x}) + \lambda(\mathbf{z} - \Pi_U(\mathbf{z})) = \mathbf{x} + \lambda\mathbf{z} - (\Pi_U(\mathbf{x}) + \lambda\Pi_U(\mathbf{z}))$$

appartiene anche esso a U^\perp . Conseguentemente $\Pi_U(\mathbf{x}) + \lambda\Pi_U(\mathbf{z}) \in U$ è una proiezione di $\mathbf{x} + \lambda\mathbf{z}$. Per il Teorema 5, la proiezione di $\mathbf{x} + \lambda\mathbf{z}$ è unica. Pertanto

$$\Pi_U(\mathbf{x} + \lambda\mathbf{z}) = \Pi_U(\mathbf{x}) + \lambda\Pi_U(\mathbf{z}),$$

cioè la proiezione ortogonale su U è lineare. \square

Usando ora il Teorema 1 e il fatto che $U^\perp = \ker \Pi_U$, si ha che ogni vettore $\mathbf{x} \in \mathbb{R}^n$ si può scrivere come somma

$$\mathbf{x} = \Pi_U(\mathbf{x}) + \Pi_{U^\perp}(\mathbf{x}).$$

Possiamo ora dimostrare che le proiezioni ortogonali godono della proprietà minimizzante che ci serve.

Teorema 7 (Teorema di approssimazione). *Sia $\mathbf{x} \in \mathbb{R}^n$ e U un sottospazio vettoriale. Allora, per ogni $\mathbf{y} \in U$ con $\mathbf{y} \neq \Pi_U(\mathbf{x})$ si ha*

$$\|\mathbf{x} - \Pi_U(\mathbf{x})\|_2 < \|\mathbf{x} - \mathbf{y}\|_2.$$

Dimostrazione. Poniamo $\mathbf{p} = \Pi_U(\mathbf{x})$. Per le proprietà delle proiezioni ortogonali, $\mathbf{x} - \mathbf{p} \in U^\perp$ e $\mathbf{p} - \mathbf{y} \in U$; pertanto,

$$\langle \mathbf{x} - \mathbf{p}, \mathbf{p} - \mathbf{y} \rangle = 0.$$

Allora,

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\|_2^2 &= \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle = \langle (\mathbf{x} - \mathbf{p}) + (\mathbf{p} - \mathbf{y}), (\mathbf{x} - \mathbf{p}) + (\mathbf{p} - \mathbf{y}) \rangle = \\ &= \|\mathbf{x} - \mathbf{p}\|_2^2 + 2\langle \mathbf{x} - \mathbf{p}, \mathbf{p} - \mathbf{y} \rangle + \|\mathbf{p} - \mathbf{y}\|_2^2 = \|\mathbf{x} - \mathbf{p}\|_2^2 + \|\mathbf{p} - \mathbf{y}\|_2^2 \end{aligned}$$

Ora, se $\mathbf{p} \neq \mathbf{y}$, si ha $\|\mathbf{p} - \mathbf{y}\|_2 > 0$ e, conseguentemente, $\|\mathbf{x} - \mathbf{y}\|_2 > \|\mathbf{x} - \mathbf{p}\|_2$. La tesi segue. \square

Il Teorema 7 mostra che la proiezione ortogonale gode di entrambe proprietà richieste per il proiettore Π da usarsi nella (5). Dunque, ogni vettore X tale che

$$AX = \Pi_U(B), \quad (7)$$

ove U è l'immagine della applicazione lineare associata ad A è soluzione del problema (2) rispetto $\|\cdot\|_2$.

Dobbiamo ora dimostrare che la (7) è equivalente alle equazioni normali del sistema. Premettiamo un teorema.

Teorema 8. *Sia A una matrice $n \times m$. Allora, per ogni $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$ vettori colonna si ha*

$$\langle \mathbf{x}, A\mathbf{y} \rangle = \langle A^T \mathbf{x}, \mathbf{y} \rangle.$$

In particolare, usando la notazione del Teorema 3 si ha

$$\Theta_{\mathbf{x}}(A\mathbf{y}) = \Theta_{A^T \mathbf{x}}(\mathbf{y}).$$

Si osservi che $\Theta_{\mathbf{x}}$ è un elemento del duale $(\mathbb{R}^n)^*$ di \mathbb{R}^n mentre $\Theta_{A^T \mathbf{x}}$ è un elemento del duale $(\mathbb{R}^m)^*$ di \mathbb{R}^m . In particolare, l'applicazione lineare $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$ induce una applicazione lineare $A^* : (\mathbb{R}^n)^* \rightarrow (\mathbb{R}^m)^*$ fra gli spazi vettoriali duali; tale applicazione lineare è detta *aggiunta* di A .

Dimostrazione. Per ogni $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$ si ha

$$\langle \mathbf{x}, A\mathbf{y} \rangle = \mathbf{x}^T A\mathbf{y} = (\mathbf{x}^T A\mathbf{y})^T = \mathbf{y}^T A^T \mathbf{x} = \langle \mathbf{y}, A^T \mathbf{x} \rangle = \langle A^T \mathbf{x}, \mathbf{y} \rangle.$$

\square

Se \tilde{X} soddisfa la (7), allora, per costruzione della proiezione ortogonale, $B - A\tilde{X} \in U^\perp$, cioè, usando il Teorema 8, per ogni $Z = AW \in U$

$$0 = \langle AW, B - A\tilde{X} \rangle = \langle W, A^T(B - A\tilde{X}) \rangle,$$

al variare di W in \mathbb{R}^m .

Ne segue $A^T(B - A\tilde{X}) = \mathbf{0}$, cioè che \tilde{X} deve essere soluzione delle equazioni normali

$$(A^T A)X = A^T B.$$

Viceversa, se \hat{X} è soluzione delle equazioni normali, allora per ogni $W \in \mathbb{R}^m$,

$$0 = \langle W, A^T(B - A\hat{X}) \rangle = \langle AW, B - A\hat{X} \rangle;$$

pertanto, per ogni $Y = AW \in \mathcal{I}A = U$ si ha $\langle Y, B - A\hat{X} \rangle = 0$, da cui $B - A\hat{X} \in U^\perp$.

Applicazioni ed esempi

Un esempio di applicazione dei sistemi iperdeterminati è dato dalla teoria dell'interpolazione.

Consideriamo dapprima un insieme di n punti (x_i, y_i) per $i = 1, \dots, n$ assegnati nel piano cartesiano ordinario. Vogliamo trovare la retta $y = mx + b$ che meglio approssima i punti assegnati, minimizzando la distanza euclidea rispetto l'insieme considerato.

A tal fine scriviamo il sistema lineare in nelle due indeterminate m, b dato dalle equazioni

$$mx_i + b = y_i$$

al variare di $i = 1, \dots, n$. La matrice incompleta di tale sistema è

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}.$$

Le equazioni normali sono date da

$$A^T A \begin{pmatrix} m \\ b \end{pmatrix} = A^T \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Svolgendo i calcoli si ottiene

$$\begin{pmatrix} x_1^2 + x_2^2 + \dots + x_n^2 & x_1 + x_2 + \dots + x_n \\ x_1 + x_2 + \dots + x_n & n \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} x_1 y_1 + x_2 y_2 + \dots + x_n y_n \\ y_1 + y_2 + \dots + y_n \end{pmatrix},$$

ovvero, in forma più compatta,

$$\begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix}.$$

Il caso più generale, è il seguente: consideriamo un insieme di funzioni reali di variabile reale $f_0(x), f_1(x), \dots, f_m(x)$ e assegnamo un insieme di n coppie ordinate (x_i, y_i) . Si vogliono trovare degli scalari a_0, a_1, \dots, a_m tali che la funzione

$$f(x) = a_0 f_0(x) + a_1 f_1(x) + \dots + a_m f_m(x)$$

sia la migliore approssimante possibile per le coppie assegnate, nel senso che la quantità

$$\left\| \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix} \right\|_2$$

sia la più piccola possibile. Per raggiungere il nostro obiettivo consideriamo il sistema lineare dato da

$$\begin{pmatrix} f_0(x_1) & f_1(x_1) & \dots & f_m(x_1) \\ f_0(x_2) & f_1(x_2) & \dots & f_m(x_2) \\ \vdots & \vdots & \dots & \vdots \\ f_0(x_n) & f_1(x_n) & \dots & f_m(x_n) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix}$$

e cerchiamo una soluzione approssimata. Scriviamo la matrice incompleta M di questo sistema come giustapposizione di vettori colonna. Pertanto, avremo $M = (M_0 \ M_1 \ \dots \ M_m)$ con

$$M_i = \begin{pmatrix} f_i(x_1) \\ f_i(x_2) \\ \vdots \\ f_i(x_n) \end{pmatrix}.$$

La posizione (i, j) della matrice $S = M^T M$ conterrà proprio il prodotto scalare $\langle M_i, M_j \rangle$. Similmente, la i -esima riga di $M^T Y$, ove Y è il vettore colonna degli y_i sarà data da $\langle M_i, Y \rangle$. Pertanto, tenuto conto della simmetria del prodotto scalare, il sistema da risolvere qualora si cerchi una approssimazione ai minimi quadrati è

$$\begin{pmatrix} \langle M_0, M_0 \rangle & \langle M_0, M_1 \rangle & \dots & \langle M_0, M_m \rangle \\ \langle M_0, M_1 \rangle & \langle M_1, M_1 \rangle & \dots & \langle M_1, M_m \rangle \\ \vdots & \vdots & \dots & \vdots \\ \langle M_0, M_m \rangle & \langle M_1, M_m \rangle & \dots & \langle M_m, M_m \rangle \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \langle M_0, Y \rangle \\ \langle M_1, Y \rangle \\ \vdots \\ \langle M_m, Y \rangle \end{pmatrix}$$

In particolare, il sistema ammette una unica soluzione se, e solamente se, il determinante della matrice simmetrica S è diverso da 0 .