RENATO LEONI

Note aggiuntive
di
Algebra lineare
per le applicazioni statistiche

Università di Firenze
Dipartimento di Statistica, Informatica, Applicazioni
"Giuseppe Parenti" (DiSIA)
Firenze, 2013

PREFAZIONE

In questo lavoro, ci proponiamo di fornire alcune integrazioni agli argomenti esposti nel volume *Algebra lineare per le applicazioni statistiche* (citato nel seguito con la dizione *Algebra lineare*).

Nel presentare tali integrazioni si daranno ovviamente per acquisiti i concetti espressi nel suddetto volume.

In qualche caso, le integrazioni servono a introdurre nuovi argomenti; in altri, consistono in un approfondimento o anche in una diversa presentazione di argomenti sostanzialmente noti.

Nell'una e nell'altra circostanza, si sono tenute presenti, implicitamente o esplicitamente, le applicazioni di carattere statistico di cui il volume citato dovrebbe costituire il supporto (*).

(*) Si avverte che uno stesso simbolo può denotare oggetti diversi a seconda del capitolo nel quale viene utilizzato.

INDICE

pag.

Cap. 1 SISTEMI DI EQUAZIONI LINEARI

1.1 Introduzione 7

1.2 Sottospazi fondamentali associati a una matrice 8

1.3 Sistemi lineari non omogenei compatibili 14

1.4 Sistemi lineari non omogenei incompatibili 19

APPENDICE 21

Cap. 2 FATTORIZZAZIONE MEDIANTE VALORI SINGOLARI DI UNA MATRICE E PSEUDOINVERSA

2.1 Introduzione 23

2.2 Fattorizzazione mediante valori singolari 23

2.3 Pseudoinversa 28

Cap. 3 FATTORIZZAZIONE QR DI UNA MATRICE E MINIMI QUADRATI

3.1 Introduzione 31

3.2 Fattorizzazione QR 31

3.3 Minimi quadrati 34

Cap. 4 SPAZIO DUALE DI R^p

4.1 Introduzione 37

4.2 Forme lineari e basi 38

pag.

4.3 Base duale 41

4.4 Isomorfismo canonico tra R^p e il suo duale 44

Cap. 1
SISTEMI DI EQUAZIONI LINEARI

1.1 INTRODUZIONE

Come è noto, un *sistema di* n *equazioni lineari in* p *incognite* è un oggetto del tipo

(1)
$$\begin{matrix} x_{11}a_1 + \dots + x_{1p}a_p = y_1 \\ \dots\dots\dots\dots\dots\dots \\ x_{n1}a_1 + \dots + x_{np}a_p = y_n \end{matrix}$$

in cui $a_1, \dots, a_p$ rappresentano le *incognite*, mentre $x_{11}, \dots, x_{np}$ e $y_1, \dots, y_n$ denotano numeri reali *assegnati* che ricevono la denominazione, rispettivamente, di *coefficienti* e *termini noti* del sistema.

Posto

$$\mathbf{X} = \begin{bmatrix} x_{11} \cdots x_{1p} \\ \cdots\cdots\cdots \\ x_{n1} \cdots x_{np} \end{bmatrix} \quad , \quad \mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} \quad , \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

il sistema di cui sopra si può scrivere, più compattamente, nella forma

(1')
$$\mathbf{X}\mathbf{a} = \mathbf{y} \ .$$

Qualora sia $\mathbf{y} \neq \mathbf{0}$ – nel caso, cioè, in cui i termini noti non siano tutti nulli – il sistema è detto *non omogeneo*; altrimenti, è detto *omogeneo*.

Si dice poi *vettore soluzione* (o, più semplicemente, *soluzione*) del sistema ogni vettore $\overline{\mathbf{a}} = \left[\overline{a}_j\right]_{(p,1)}$ tale che $\mathbf{X}\overline{\mathbf{a}} = \mathbf{y}$.

Il sistema è *compatibile* se esiste almeno un vettore soluzione; *incompatibile* nel caso contrario.

In questo capitolo, ci occuperemo di varie questioni riguardanti i sistemi lineari non omogenei siano essi compatibili o incompatibili.

Prima di ciò, conviene tuttavia esporre alcuni concetti i quali – oltre a rivestire un interesse di carattere generale – risultano rilevanti ai fini di una trattazione sufficientemente dettagliata delle specifiche questioni che intendiamo discutere.

1.2 SOTTOSPAZI FONDAMENTALI ASSOCIATI A UNA MATRICE

Consideriamo anzitutto la matrice **X** scritta nella forma

$$\mathbf{X} = \left[\, \mathbf{x}_1 \cdots \mathbf{x}_p \,\right]$$

dove $\mathbf{x}_1, \dots, \mathbf{x}_p$ denotano i p vettori colonna che compongono **X**.

Supposto per concretezza che sia $0 < r(\mathbf{X}) = r < p \leq n$, a tale matrice possiamo associare due sottospazi:

- il sottospazio $\mathcal{N}(\mathbf{X})$ di R^p, costituito dalle soluzioni del sistema lineare omogeneo $\mathbf{Xa} = \mathbf{0}$, di dimensione (nullità) $p-r$;
- il sottospazio $\mathcal{R}(\mathbf{X})$ di R^n, generato da $\mathbf{x}_1, \dots, \mathbf{x}_p$, di dimensione r.

Consideriamo poi la matrice **X** scritta nella forma

$$\mathbf{X} = \begin{bmatrix} \mathbf{z}_1' \\ \vdots \\ \mathbf{z}_n' \end{bmatrix}$$

dove $\mathbf{z}_1', \dots, \mathbf{z}_n'$ indicano gli n vettori riga che compongono **X**, e la matrice

$$\mathbf{X}' = \left[\, \mathbf{z}_1 \cdots \mathbf{z}_n \,\right]$$

ottenuta dalla precedente per trasposizione.

Poiché $0 < r(\mathbf{X}') = r < p \leq n$, a quest'ultima matrice possiamo associare due sottospazi:

- il sottospazio $\mathcal{N}(\mathbf{X}')$ di R^n, costituito dalle soluzioni del sistema lineare omogeneo $\mathbf{X}'\mathbf{c} = \mathbf{0}$ (1), di dimensione (nullità) $n-r$;
- il sottospazio $\mathcal{R}(\mathbf{X}')$ di R^p, generato da $\mathbf{z}_1, \dots, \mathbf{z}_n$, di dimensione r.

Ciò premesso, supposto che R^n e R^p siano dotati della metrica standard, vogliamo esaminare le relazioni che sussistono tra i sottospazi sopra definiti.

(a) Si considerino dapprima i sottospazi $\mathcal{N}(\mathbf{X})$ e $\mathcal{R}(\mathbf{X}')$ di R^p.

Poiché $\mathcal{N}(\mathbf{X})$ – sottospazio di R^p costituito dalle soluzioni del sistema lineare omogeneo $\mathbf{X}\mathbf{a} = \mathbf{0}$ – ha dimensione $p-r$, $\mathcal{N}(\mathbf{X})$ possiede una base formata da $p-r$ vettori $\overline{\mathbf{a}}_1, \dots, \overline{\mathbf{a}}_{p-r}$.

Pertanto, posto

$$\overline{\mathbf{A}} = \left[\, \overline{\mathbf{a}}_1 \cdots \overline{\mathbf{a}}_{p-r} \,\right],$$

risulta ($\mathbf{O}$: matrice nulla di ordine $(n, p-r)$)

$$\mathbf{X}\overline{\mathbf{A}} = \mathbf{O}\,. \tag{2}$$

A sua volta, poiché $\mathcal{R}(\mathbf{X}')$ – sottospazio di R^p generato dagli n vettori $\mathbf{z}_1, \dots, \mathbf{z}_n$ – ha dimensione r, $\mathcal{R}(\mathbf{X}')$ possiede una base formata da r vettori $\overline{\mathbf{z}}_{p-r+1}, \dots, \overline{\mathbf{z}}_p$.

Posto

$$\overline{\mathbf{Z}} = \left[\, \overline{\mathbf{z}}_{p-r+1} \cdots \overline{\mathbf{z}}_p \,\right],$$

possiamo dunque scrivere (Cfr. il punto 3 dei Complementi al Cap. IV di *Algebra lineare*)

$$\mathbf{X}' = \overline{\mathbf{Z}}\mathbf{K} \tag{3}$$

(1) Per semplicità, conveniamo di usare lo stesso simbolo per indicare sia il vettore nullo di ordine n sia il vettore nullo di ordine p.

dove $\mathbf{K}$ è una matrice di ordine (r,n) e rango r.

Ma, tenuto conto della (3) e del fatto che $\mathbf{KK}'$ è invertibile, si ha

$$(4)\qquad \{\mathbf{X}\overline{\mathbf{A}} = \mathbf{O}\} \Leftrightarrow \{\mathbf{K}'\,\overline{\mathbf{Z}}'\,\overline{\mathbf{A}} = \mathbf{O}\} \Leftrightarrow \{\mathbf{KK}'\,\overline{\mathbf{Z}}'\,\overline{\mathbf{A}} = \mathbf{O}\} \Leftrightarrow \{\overline{\mathbf{Z}}'\,\overline{\mathbf{A}} = \mathbf{O}\}$$

e ciò significa intanto che $\mathcal{N}(\mathbf{X})$ e $\mathcal{R}(\mathbf{X}')$ sono sottospazi ortogonali di R^p (Cfr. il Teorema 3 del Cap. VII di *Algebra lineare*).

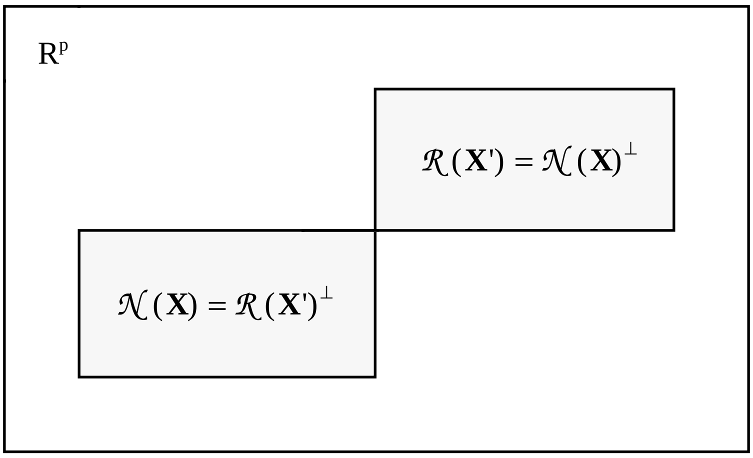
Di più, poiché la dimensione di $\mathcal{N}(\mathbf{X})$ è $p-r$ e quella di $\mathcal{R}(\mathbf{X}')$ è r, $\mathcal{N}(\mathbf{X})$ e $\mathcal{R}(\mathbf{X}')$ sono complementi ortogonali in R^p.

In definitiva, si ha che ($\mathcal{N}(\mathbf{X}) = \mathcal{R}(\mathbf{X}')^{\perp}$)

$$(5)\qquad R^p = \mathcal{R}(\mathbf{X}') \oplus \mathcal{N}(\mathbf{X}) = \mathcal{R}(\mathbf{X}') \oplus \mathcal{R}(\mathbf{X}')^{\perp} .$$

Quanto ora detto si può riassumere nel modo indicato in Fig. 1.

Fig.1



OSSERVAZIONE 1. In Matlab, la matrice $\overline{\mathbf{A}}$ i cui vettori colonna formano una base di $\mathcal{N}(\mathbf{X})$ si può determinare attraverso la funzione NB, mentre la matrice $\overline{\mathbf{Z}}$ i cui vettori colonna costituiscono una base di $\mathcal{R}(\mathbf{X}')$ si può calcolare mediante la funzione CB.

Tali funzioni sono riportate in Appendice al presente capitolo.

OSSERVAZIONE 2. Una volta scelta la matrice $\overline{\mathbf{Z}}$ che entra nella fattorizzazione $\mathbf{X}' = \overline{\mathbf{Z}}\mathbf{K}$, si può determinare $\mathbf{K}$ mediante l'espressione

$$\mathbf{K} = (\overline{\mathbf{Z}}'\,\overline{\mathbf{Z}})^{-1}\,\overline{\mathbf{Z}}'\,\mathbf{X}'\,.$$

ESEMPIO 1. Considerate le matrici

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad , \quad \mathbf{X}' = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 3 & 1 & 1 & 1 \end{bmatrix}$$

di rango pari a 2, una base di $\mathcal{N}(\mathbf{X})$ è fornita dal vettore

$$\overline{\mathbf{a}}_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

e una base di $\mathcal{R}(\mathbf{X}')$ è costituita dai vettori

$$\overline{\mathbf{z}}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad , \quad \overline{\mathbf{z}}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} .$$

∎

(b) Si considerino adesso i sottospazi $\mathcal{N}(\mathbf{X}')$ e $\mathcal{R}(\mathbf{X})$ di R^n.

Poiché $\mathcal{N}(\mathbf{X}')$ – sottospazio di R^n costituito dalle soluzioni del sistema lineare omogeneo $\mathbf{X}'\mathbf{c} = \mathbf{0}$ – ha dimensione $n-r$, $\mathcal{N}(\mathbf{X}')$ possiede una base formata da $n-r$ vettori $\overline{\mathbf{c}}_1, \dots, \overline{\mathbf{c}}_{n-r}$.

Pertanto, posto

$$\overline{\mathbf{C}} = \left[\, \overline{\mathbf{c}}_1 \cdots \overline{\mathbf{c}}_{n-r} \,\right] ,$$

risulta ($\overline{\mathbf{O}}$: matrice nulla di ordine $(p\,,n-r)$)

$$\mathbf{X}'\,\overline{\mathbf{C}} = \overline{\mathbf{O}}\,. \tag{6}$$

A sua volta, poiché $\mathcal{R}(\mathbf{X})$ – sottospazio di R^n generato dai p vettori $\mathbf{x}_1, \dots, \mathbf{x}_p$ – ha dimensione r, $\mathcal{R}(\mathbf{X})$ possiede una base formata da r vettori $\overline{\mathbf{x}}_{n-r+1}, \dots, \overline{\mathbf{x}}_n$.

Posto

$$\overline{\mathbf{X}} = [\,\overline{\mathbf{x}}_{n-r+1} \cdots \overline{\mathbf{x}}_n\,]$$

possiamo dunque scrivere (Cfr. il punto 3 dei Complementi al Cap. IV di *Algebra lineare*)

(7) $$\mathbf{X} = \overline{\mathbf{X}}\mathbf{H}$$

dove $\mathbf{H}$ è una matrice di ordine (r,p) e rango r.

Ma, tenuto conto della (7) e del fatto che $\mathbf{H}\mathbf{H}'$ è invertibile, si ha

(8) $$\{\mathbf{X}'\overline{\mathbf{C}} = \overline{\mathbf{O}}\} \Leftrightarrow \{\mathbf{H}'\overline{\mathbf{X}}'\overline{\mathbf{C}} = \overline{\mathbf{O}}\} \Leftrightarrow \{\mathbf{H}\mathbf{H}'\overline{\mathbf{X}}'\overline{\mathbf{C}} = \overline{\mathbf{O}}\} \Leftrightarrow \{\overline{\mathbf{X}}'\overline{\mathbf{C}} = \overline{\mathbf{O}}\}$$

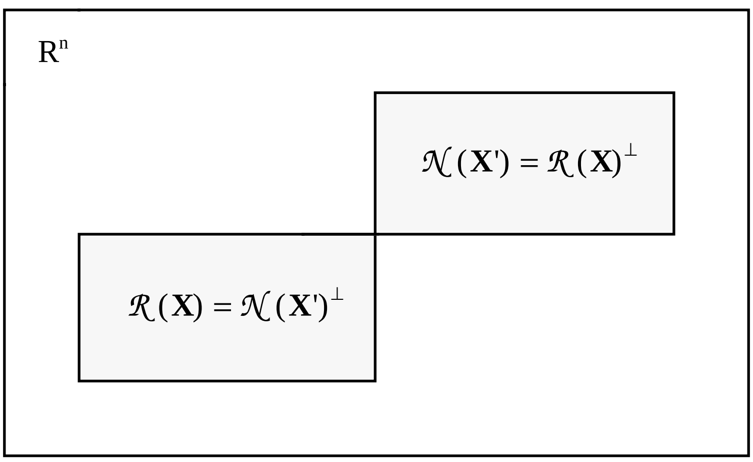
e ciò mostra che $\mathcal{N}(\mathbf{X}')$ e $\mathcal{R}(\mathbf{X})$ sono sottospazi ortogonali di R^n (Cfr. il Teorema 3 del Cap. VII di *Algebra lineare*).

Inoltre, poiché la dimensione di $\mathcal{N}(\mathbf{X}')$ è $n-r$ e quella di $\mathcal{R}(\mathbf{X})$ è r, $\mathcal{N}(\mathbf{X}')$ e $\mathcal{R}(\mathbf{X})$ sono complementi ortogonali in R^n e possiamo, quindi, concludere che ($\mathcal{N}(\mathbf{X}') = \mathcal{R}(\mathbf{X})^{\perp}$)

(9) $$R^n = \mathcal{R}(\mathbf{X}) \oplus \mathcal{N}(\mathbf{X}') = \mathcal{R}(\mathbf{X}) \oplus \mathcal{R}(\mathbf{X})^{\perp}\,.$$

Quanto ora detto si può riassumere nel modo indicato in Fig. 2.

Fig.2



OSSERVAZIONE 3. Una volta scelta la matrice $\bar{\mathbf{X}}$ che entra nella fattorizzazione $\mathbf{X} = \bar{\mathbf{X}}\mathbf{H}$, si può procedere al calcolo di $\mathbf{H}$ attraverso l'espressione

$$\mathbf{H} = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\mathbf{X}\,.$$

ESEMPIO 2. Considerate nuovamente le matrici di cui all'Esempio 1, una base di $\mathcal{N}(\mathbf{X}')$ è formata dai vettori

$$\bar{\mathbf{c}}_1 = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix} \quad , \quad \bar{\mathbf{c}}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

e una base di $\mathcal{R}(\mathbf{X})$ è costituita dai vettori

$$\bar{\mathbf{x}}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad , \quad \bar{\mathbf{x}}_4 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

∎

(c) A completamento di quanto precede, è opportuno osservare che – poiché ($\underline{\mathbf{O}}$: matrice nulla di ordine $(r, p-r)$; $\underline{\underline{\mathbf{O}}}$: matrice nulla di ordine $(r, n-r)$)

$$\mathbf{H}\bar{\mathbf{A}} = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\mathbf{X}\bar{\mathbf{A}} = \underline{\mathbf{O}} \quad , \quad \bar{\mathbf{Z}}'\bar{\mathbf{A}} = \underline{\mathbf{O}}$$

e

$$\mathbf{K}\bar{\mathbf{C}} = (\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{X}'\bar{\mathbf{C}} = \underline{\underline{\mathbf{O}}} \quad , \quad \bar{\mathbf{X}}'\bar{\mathbf{C}} = \underline{\underline{\mathbf{O}}}$$

– $\mathbf{H}'$ e $\mathbf{K}'$ sono matrici i cui vettori colonna formano, rispettivamente, una nuova base di $\mathcal{R}(\mathbf{X}')$, ovvero di $\mathcal{N}(\mathbf{X})^{\perp}$, e una nuova base di $\mathcal{R}(\mathbf{X})$, ovvero di $\mathcal{N}(\mathbf{X}')^{\perp}$.

Si noti anche che, indicando con $\mathbf{F}$ e $\mathbf{G}$ le matrici di passaggio dalle vecchie alle nuove basi ($\mathbf{H}' = \bar{\mathbf{Z}}\mathbf{F}$, $\mathbf{K}' = \bar{\mathbf{X}}\mathbf{G}$), risulta

$$\mathbf{F} = (\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{H}' \quad , \quad \mathbf{G} = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\mathbf{K}'\,.$$

Quanto detto circa i quattro sottospazi fondamentali presi in esame e le matrici i cui vettori colonna costituiscono le rispettive basi è parzialmente riassunto nella Tav. 1.

Tav. 1 - Sottospazi fondamentali e matrici i cui vettori colonna formano basi

Sottospazi	Matrici
$\mathcal{N}(\mathbf{X}) = \mathcal{R}(\mathbf{X}')^{\perp}$	$\overline{\mathbf{A}}$
$\mathcal{R}(\mathbf{X}') = \mathcal{N}(\mathbf{X})^{\perp}$	$\overline{\mathbf{Z}}$ oppure $\mathbf{H}'$
$\mathcal{N}(\mathbf{X}') = \mathcal{R}(\mathbf{X})^{\perp}$	$\overline{\mathbf{C}}$
$\mathcal{R}(\mathbf{X}) = \mathcal{N}(\mathbf{X}')^{\perp}$	$\overline{\mathbf{X}}$ oppure $\mathbf{K}'$

ESEMPIO 3. Considerati nuovamente gli Esempi 1 e 2, si ha

$$\overline{\mathbf{Z}} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \quad , \quad \overline{\mathbf{X}} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

da cui

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad , \quad \mathbf{K} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix}.$$

1.3 SISTEMI LINEARI NON OMOGENEI COMPATIBILI

Supponiamo che il sistema lineare non omogeneo $\mathbf{Xa} = \mathbf{y}$ sia compatibile e che, come in precedenza, sia $0 < \mathrm{r}(\mathbf{X}) = \mathrm{r} < \mathrm{p} \leq \mathrm{n}$ (2).

In tal caso, il sistema ammette *infinite* soluzioni tra le quali ci proponiamo di selezionarne *una* che soddisfi uno specifico criterio di scelta.

Poiché abbiamo supposto R^{p} munito della metrica standard, diviene naturale scegliere tra le infinite soluzioni del sistema quella di *minima norma*.

---

(2) Dato che $\mathcal{R}(\mathbf{X}) = \mathcal{N}(\mathbf{X}')^{\perp}$, dire che $\mathbf{y} \in \mathcal{R}(\mathbf{X})$, e cioè che il sistema $\mathbf{X\,a} = \mathbf{y}$ è compatibile, equivale ad affermare che $\mathbf{y} \in \mathcal{N}(\mathbf{X}')^{\perp}$ (*alternativa di Fredholm*).

Allo scopo di individuare tale soluzione, osserviamo intanto che ogni vettore $\bar{\mathbf{a}}$, soluzione del sistema in oggetto, può essere scomposto univocamente nella somma dei vettori $\tilde{\mathbf{a}} \in \mathcal{N}(\mathbf{X})$ e $\hat{\mathbf{a}} \in \mathcal{N}(\mathbf{X})^{\perp}$.

Ma, risultando

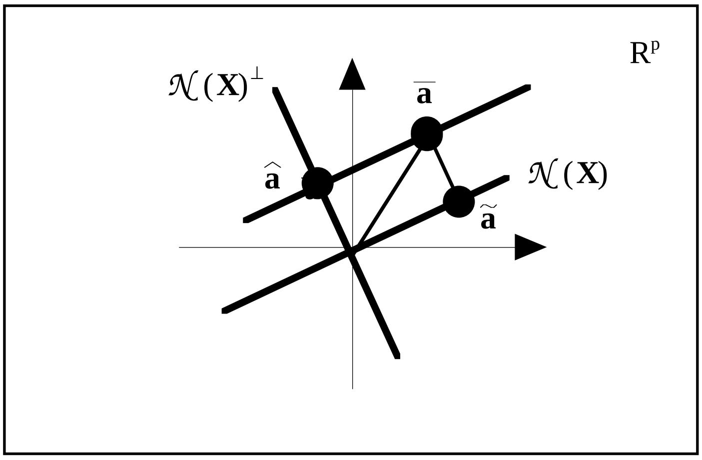
$$\mathbf{y} = \mathbf{X}\bar{\mathbf{a}} = \mathbf{X}(\tilde{\mathbf{a}} + \hat{\mathbf{a}}) = \mathbf{X}\tilde{\mathbf{a}} + \mathbf{X}\hat{\mathbf{a}} = \mathbf{X}\hat{\mathbf{a}}\ , \tag{10}$$

si riconosce subito che $\hat{\mathbf{a}}$ costituisce a sua volta una soluzione del sistema.

D'altro canto, $\tilde{\mathbf{a}}$ può essere visto come la proiezione ortogonale di $\bar{\mathbf{a}}$ su $\mathcal{N}(\mathbf{X})$ e, quindi, $\hat{\mathbf{a}} = \bar{\mathbf{a}} - \tilde{\mathbf{a}}$ è di minima norma (Cfr. il Teorema 10 del Cap. VII di *Algebra lineare*).

Rispetto alla base naturale di R^p, quanto ora detto ammette un'ovvia interpretazione geometrica, riassunta schematicamente in Fig. 3.

Fig.3



Ci proponiamo adesso di esporre un procedimento che consente di determinare $\hat{\mathbf{a}}$ in funzione delle matrici $\overline{\mathbf{X}}$ e $\mathbf{H}$ e del vettore $\mathbf{y}$.

Essenzialmente, il procedimento in questione si risolve nell'esprimere $\mathbf{y}$ in due modi diversi ed eguagliare i risultati.

Il primo modo consiste nell'osservare anzitutto che – poiché $\hat{\mathbf{a}} \in \mathcal{N}(\mathbf{X})^{\perp}$ e $\mathcal{N}(\mathbf{X})^{\perp} = \mathcal{R}(\mathbf{X}')$ – si ha che $\hat{\mathbf{a}} \in \mathcal{R}(\mathbf{X}')$.

Tenuto conto che i vettori colonna di $\mathbf{H}'$ formano una base di $\mathcal{R}(\mathbf{X}')$, ne consegue che $\hat{\mathbf{a}}$ si può scrivere nella forma

(11) $$\hat{\mathbf{a}} = \mathbf{H}'\mathbf{b}$$

per qualche **b** di ordine r, da cui

(12) $$\mathbf{y} = \mathbf{X}\,\hat{\mathbf{a}} = \overline{\mathbf{X}}\mathbf{H}\,\hat{\mathbf{a}} = \overline{\mathbf{X}}\mathbf{H}\,\mathbf{H}'\mathbf{b}\,.$$

Il secondo modo di esprimere **y** consiste semplicemente nel considerare l'identità

(13) $$\mathbf{y} = \overline{\mathbf{X}}\,(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\,\overline{\mathbf{X}}'\mathbf{y}\,.$$

Eguagliando la (12) e la (13) si ottiene

$$\overline{\mathbf{X}}\mathbf{H}\,\mathbf{H}'\mathbf{b} = \overline{\mathbf{X}}\,(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\,\overline{\mathbf{X}}'\mathbf{y}$$

da cui, dopo alcuni passaggi, si ha che

$$\mathbf{b} = (\mathbf{H}\,\mathbf{H}')^{-1}\,(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\,\overline{\mathbf{X}}'\mathbf{y}\,.$$

Inserita questa espressione di **b** nella (11), risulta dunque

(14) $$\hat{\mathbf{a}} = \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}\,(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\,\overline{\mathbf{X}}'\mathbf{y}$$

che, posto

(15) $$\mathbf{X}^{+} = \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}\,(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\,\overline{\mathbf{X}}'\,,$$

possiamo anche scrivere nel seguente modo

(14') $$\hat{\mathbf{a}} = \mathbf{X}^{+}\mathbf{y}\,.$$

La matrice $\mathbf{X}^{+}$ di ordine (p,n) di cui sopra è detta *pseudoinversa di Moore-Penrose* o, più semplicemente, *pseudoinversa* di **X** e ricopre lo stesso ufficio rivestito dalla matrice $\mathbf{X}^{-1}$ nel caso che **X** sia non singolare [(3)].

(3) La definizione di pseudoinversa ora data differisce da quella esposta al punto 5 dei Complementi al Cap. VII di *Algebra lineare*. Sull'argomento avremo tuttavia modo di tornare nel capitolo successivo.

ESEMPIO 4. Si consideri il sistema lineare non omogeneo

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \mathbf{a} = \begin{bmatrix} 6 \\ 2 \\ 2 \\ 2 \end{bmatrix}.$$

Come è subito visto, il suddetto sistema è compatibile.

Inoltre, tenuto conto di quanto risulta dall'Esempio 3, la matrice $\mathbf{X}$ dei coefficienti di tale sistema può essere fattorizzata nel prodotto delle matrici

$$\overline{\mathbf{X}} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad \mathbf{H} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Pertanto, in applicazione della (14'), si ha

$$\widehat{\mathbf{a}} = \begin{bmatrix} 0.083 & -0.250 & 0.583 & -0.250 \\ 0.083 & 0.250 & -0.416 & 0.250 \\ 0.166 & 0 & 0.166 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.666 \\ 0.666 \\ 1.333 \end{bmatrix}. \qquad \blacksquare$$

La pseudoinversa di $\mathbf{X}$ gode di alcune importanti proprietà che vogliamo adesso mettere in evidenza.

P 1. Gli elementi che intervengono nella definizione di pseudoinversa sono le matrici $\overline{\mathbf{X}}, \mathbf{H}$ che provengono dalla fattorizzazione $\mathbf{X} = \overline{\mathbf{X}}\mathbf{H}$.

Benché tale fattorizzazione non sia unica – nel senso che postmoltiplicando $\overline{\mathbf{X}}$ per una matrice invertibile $\mathbf{B}$ e premoltiplicando $\mathbf{H}$ per l'inversa di $\mathbf{B}$ si ottiene ancora una fattorizzazione di $\mathbf{X}$ (Cfr. il punto 3 dei Complementi al Cap. IV di *Algebra lineare*) – si può facilmente verificare che $\mathbf{X}^+$ è *unica*.

P 2. Allo stesso modo con cui, rispetto alle basi naturali di R^p e R^n, $\mathbf{X}$ si può interpretare come la matrice di una trasformazione lineare di R^p in R^n, $\mathbf{X}^+$ può esser vista come la matrice di una trasformazione lineare di R^n in R^p.

P 3. $\mathbf{X}\mathbf{X}^{+} = \overline{\mathbf{X}}\,(\overline{\mathbf{X}}'\,\overline{\mathbf{X}})^{-1}\,\overline{\mathbf{X}}'$ è la matrice rappresentativa, rispetto alla base naturale di R^n, del proiettore ortogonale su $\mathcal{R}(\mathbf{X}) = \mathcal{N}(\mathbf{X}')^{\perp}$.

P 4. $\mathbf{X}^{+}\mathbf{X} = \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}\,\mathbf{H}$ è la matrice rappresentativa, rispetto alla base naturale di R^p, del proiettore ortogonale su $\mathcal{R}(\mathbf{X}') = \mathcal{N}(\mathbf{X})^{\perp}$.

P 5. $\mathbf{I} - \mathbf{X}\mathbf{X}^{+} = \mathbf{I} - \overline{\mathbf{X}}\,(\overline{\mathbf{X}}'\,\overline{\mathbf{X}})^{-1}\,\overline{\mathbf{X}}'$ ($\mathbf{I}$: matrice unità di ordine (n,n)) è la matrice rappresentativa, rispetto alla base naturale di R^n, del proiettore ortogonale su $\mathcal{N}(\mathbf{X}') = \mathcal{R}(\mathbf{X})^{\perp}$.

P 6. $\overline{\mathbf{I}} - \mathbf{X}^{+}\mathbf{X} = \overline{\mathbf{I}} - \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}\,\mathbf{H}$ ($\overline{\mathbf{I}}$: matrice unità di ordine (p,p)) è la matrice rappresentativa, rispetto alla base naturale di R^p, del proiettore ortogonale su $\mathcal{N}(\mathbf{X}) = \mathcal{R}(\mathbf{X}')^{\perp}$.

OSSERVAZIONE 4. Tenuto conto della *P 6*, qualunque sia $\mathbf{h} \in R^p$, risulta che $(\overline{\mathbf{I}} - \mathbf{X}^{+}\mathbf{X})\,\mathbf{h} \in \mathcal{N}(\mathbf{X})$.

A sua volta, $\hat{\mathbf{a}} = \mathbf{X}^{+}\mathbf{y} \in \mathcal{N}(\mathbf{X})^{\perp}$ e, dunque, ciascuna soluzione del sistema lineare non omogeneo compatibile $\mathbf{X}\mathbf{a} = \mathbf{y}$ è della forma

$$\overline{\mathbf{a}} = \mathbf{X}^{+}\mathbf{y} + (\overline{\mathbf{I}} - \mathbf{X}^{+}\mathbf{X})\,\mathbf{h}\,.$$

OSSERVAZIONE 5. Nella esposizione precedente abbiamo supposto, per motivi di concretezza, che fosse $0 < r(\mathbf{X}) = r < p \leq n$.

I ragionamenti svolti, tuttavia, possono facilmente adattarsi a comprendere i casi seguenti.

- Se $0 < r(\mathbf{X}) = r = p \leq n$, cioè se $\mathbf{X}$ è una matrice di pieno rango per colonne, vale la fattorizzazione $\mathbf{X} = \mathbf{X}\overline{\mathbf{I}}$.

  Dunque, poiché $\mathbf{X}$ e $\overline{\mathbf{I}}$ sostituiscono rispettivamente $\overline{\mathbf{X}}$ e $\mathbf{H}$ nella (7), risulta $\mathbf{X}^{+} = (\mathbf{X}'\,\mathbf{X})^{-1}\,\mathbf{X}'$.

- Se $0 < r(\mathbf{X}) = r = n \leq p$, cioè se $\mathbf{X}$ è una matrice di pieno rango per righe, vale la fattorizzazione $\mathbf{X} = \mathbf{I}\mathbf{X}$.

  Quindi, poiché $\mathbf{I}$ e $\mathbf{X}$ sostituiscono rispettivamente $\overline{\mathbf{X}}$ e $\mathbf{H}$ nella (7), risulta $\mathbf{X}^{+} = \mathbf{X}'\,(\mathbf{X}\,\mathbf{X}')^{-1}$.

OSSERVAZIONE 6. Si noti esplicitmente che, in luogo delle matrici $\overline{\mathbf{X}}$ e $\mathbf{H}$ che compaiono nella (15), avremmo potuto usare le matrici $\mathbf{K}'$ e $\overline{\mathbf{Z}}'$, ottenendo l'espressione

$$\mathbf{X}^{+} = \overline{\mathbf{Z}}(\overline{\mathbf{Z}}'\,\overline{\mathbf{Z}})^{-1}\,(\mathbf{K}\mathbf{K}')^{-1}\,\mathbf{K}\,.$$

1.4 SISTEMI LINEARI NON OMOGENEI INCOMPATIBILI

Supponiamo adesso di trovarsi di fronte a un sistema lineare non omogeneo incompatibile, cioè tale che $\mathbf{y} \notin \mathcal{R}(\mathbf{X})$.

In questo caso, il procedimento a cui usualmente si ricorre consiste nel ricercare soluzioni approssimate, nel senso dei *minimi quadrati*, del vettore incognito $\mathbf{a}$.

Formalmente, si tratta di trovare

$$(16) \qquad \min_{\mathbf{a} \in \mathrm{R}^p} (\mathbf{X}\mathbf{a} - \mathbf{y})'(\mathbf{X}\mathbf{a} - \mathbf{y}) = \min_{\mathbf{a} \in \mathrm{R}^p} (\mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a} - 2\,\mathbf{y}'\mathbf{X}\mathbf{a} + \mathbf{y}'\mathbf{y})$$

e ciò, come è noto, conduce al sistema di *equazioni normali*

$$(17) \qquad \mathbf{X}'\mathbf{X}\mathbf{a} = \mathbf{X}'\mathbf{y}$$

in cui $\mathbf{X}'\mathbf{X}$ rappresenta la matrice dei coefficienti e $\mathbf{X}'\mathbf{y}$ è il vettore dei termini noti.

Vogliamo anzitutto dimostrare che il sistema (17) è equivalente al sistema

$$(18) \qquad \mathbf{X}\mathbf{a} = \hat{\mathbf{y}}$$

nel quale $\hat{\mathbf{y}}$ rappresenta la proiezione ortogonale di $\mathbf{y}$ su $\mathcal{R}(\mathbf{X})$.

A questo fine, si osservi intanto che, poiché $\hat{\mathbf{y}} \in \mathcal{R}(\mathbf{X})$, il vettore $\hat{\mathbf{y}}$ può essere espresso come combinazione lineare dei vettori colonna di $\mathbf{X}$ e ciò significa che esiste (almeno) un vettore $\overline{\mathbf{a}}$ tale che il sistema (18) risulti compatibile.

Ma, ogni soluzione del sistema (18) è anche una soluzione del sistema

(17) in quanto, posto $\mathbf{y} = \widehat{\mathbf{y}} + (\mathbf{y} - \widehat{\mathbf{y}})$ dove $(\mathbf{y} - \widehat{\mathbf{y}}) \in \mathcal{R}(\mathbf{X})^{\perp}$, si ha

$$\begin{aligned} &\{\mathbf{X}\overline{\mathbf{a}} = \widehat{\mathbf{y}}\} && \Rightarrow \{\mathbf{X}'\mathbf{X}\overline{\mathbf{a}} = \mathbf{X}'\widehat{\mathbf{y}}\} \\ \Leftrightarrow &\{\mathbf{X}'\mathbf{X}\overline{\mathbf{a}} = \mathbf{X}'\mathbf{y} - \mathbf{X}'(\mathbf{y} - \widehat{\mathbf{y}})\} && \Leftrightarrow \{\mathbf{X}'\mathbf{X}\overline{\mathbf{a}} = \mathbf{X}'\mathbf{y}\} \ . \end{aligned}$$

Ora, l'insieme S delle soluzioni del sistema (18) costituisce una varietà lineare di traslazione $\overline{\mathbf{a}}$ e direzione $\mathcal{N}(\mathbf{X})$ (Cfr. il punto 4 dei Complementi al Cap. IV di *Algebra lineare*), ovvero $\mathrm{S} = \overline{\mathbf{a}} + \mathcal{N}(\mathbf{X})$.

Analogamente, l'insieme $\overline{\mathrm{S}}$ delle soluzioni del sistema (17) forma una varietà lineare di traslazione $\overline{\mathbf{a}}$ e direzione $\mathcal{N}(\mathbf{X}'\mathbf{X})$, ovvero $\overline{\mathrm{S}} = \overline{\mathbf{a}} + \mathcal{N}(\mathbf{X}'\mathbf{X})$.

D'altra parte, risultando $\mathcal{N}(\mathbf{X}) = \mathcal{N}(\mathbf{X}'\mathbf{X})$ (Cfr. quanto detto nel corso della dimostrazione del Teorema 6 del Cap. IV di *Algebra lineare*), si conclude subito che $\mathrm{S} = \overline{\mathrm{S}}$ e cioè che i sistemi (17) e (18) sono equivalenti.

A questo punto, concentriamo l'attenzione sul sistema (18).

Supposto come in precedenza che sia $0 < \mathrm{r}(\mathbf{X}) = \mathrm{r} < \mathrm{p} \leq \mathrm{n}$, tale sistema – e altrettanto accade per il sistema (17) – ammette infinite soluzioni ciascuna delle quali dà luogo a un *minimo globale* per la funzione obiettivo che compare nella (16).

Tra queste infinite soluzioni possiamo scegliere quella di *minima norma*.

Per determinare tale soluzione basta applicare la (14') al sistema (18) ottenendo

$$\begin{aligned} \widehat{\mathbf{a}} &= \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\widehat{\mathbf{y}} \\ &= \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\overline{\mathbf{X}}\,(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\mathbf{y} \\ &= \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\mathbf{y} \\ &= \mathbf{X}^{+}\mathbf{y} \ . \end{aligned} \tag{19}$$

Si noti che la soluzione ora indicata è formalmente identica a quella individuata nel caso di un sistema lineare non omogeneo compatibile (Cfr. la (14')) in cui $\widehat{\mathbf{a}}$ denotava la soluzione di minima norma.

Nel caso che stiamo considerando, invece, $\widehat{\mathbf{a}}$ esprime la soluzione di minima norma nell'insieme delle soluzioni dei minimi quadrati.

Cap. 1
APPENDICE

```
% A è una matrice di ordine (m,q).
% NB(A) produce una matrice N le cui colonne formano
% una base del nucleo di A.
%
function N = nb(A)
[R,jb] = rref(A);
[m,q] = size(A);
r = length(jb);
qmr = q - r;
c = 1:q;
c(jb) = zeros(1,length(jb));
N = zeros(q,qmr);
c = 1:q;
c(jb) = zeros(1,length(jb));
kb = c(find(c));
for j = 1:qmr
    N(kb(j),j) = 1;
    N(jb,j) = -R(1:r,kb(j));
end
```

```
% A è una matrice di ordine (m,q).
% CB(A) produce una matrice C le cui colonne formano
% una base dello spazio generato dai vettori colonna
% di A.
%
function C = cb(A)
[R,jb] = rref(A);
C = A(:,jb);
```

Cap. 2

FATTORIZZAZIONE MEDIANTE VALORI SINGOLARI DI UNA MATRICE E PSEUDOIVERSA

2.1 INTRODUZIONE

Data una matrice **X** di ordine (n,p) e supposto per concretezza che sia $0 < r(\mathbf{X}) = r < p \leq n$, nel Cap. 1 ci siamo ampiamente avvalsi della possibilità di scomporre tale matrice nel prodotto di due matrici: $\overline{\mathbf{X}}$ (di ordine (n,r) e rango r) e **H** (di ordine (r,p) e rango r).

Una scomposizione di questo tipo è anche detta *fattorizzazione di pieno rango* di **X**.

Essa non è tuttavia l'unica disponibile e, in questo capitolo, ci proponiamo di esporre un altro metodo di scomposizione, mostrandone la rilevanza ai fini di una definizione di pseudoinversa alternativa a quella introdotta nel Cap. 1.

2.2 FATTORIZZAZIONE MEDIANTE VALORI SINGOLARI

Data la matrice **X** sopra indicata e qualora R^n e R^p siano dotati della metrica standard (1), il procedimento di scomposizione di **X** che ci apprestiamo a presentare è sostanzialmente noto (Cfr. il punto 4 dei Complementi al Cap. VII di *Algebra lineare*) ma, per ragioni di completezza, conviene qui riproporlo nelle sue linee essenziali.

(*a1*) Si consideri anzitutto la matrice (simmetrica) **X'X**.

---

(1) In realtà, quanto diremo vale qualunque siano la matrice **X** e la metrica (euclidea) adottata.

Chiaramente, rispetto alla base naturale di R^p, $\mathbf{X'X}$ si può interpretare come la matrice di una forma bilineare simmetrica tale che la forma quadratica a essa associata è semidefinita positiva (Cfr. il punto 3 (*i*) dei Complementi al Cap. VI di *Algebra lineare*).

Pertanto, definita la trasformazione autoaggiunta T_1 di R^p la cui matrice, rispetto alla suddetta base di R^p, è $\mathbf{X'X}$, T_1 ammette (Cfr. il paragrafo 8 del Cap. VII di *Algebra lineare*):

- r autovalori positivi $\lambda_1, \dots, \lambda_r$ ai quali possiamo associare r autovettori ortonormali $\mathbf{a}_1, \dots, \mathbf{a}_r$;
- un autovalore nullo di molteplicità $p-r$ a cui possiamo associare $p-r$ autovettori ortonormali $\mathbf{a}_{r+1}, \dots, \mathbf{a}_p$ tali che $\mathbf{a}_{r+h} \perp \mathbf{a}_k$ per tutti gli $h = 1, \dots, p$ e $k = 1, \dots, r$.

Posto

$$\mathbf{D}_1 = \text{diag}(\lambda_1, \dots, \lambda_r) \quad , \quad \mathbf{A}_1 = [\mathbf{a}_1 \cdots \mathbf{a}_r] \quad , \quad \mathbf{A}_2 = [\mathbf{a}_{r+1} \cdots \mathbf{a}_p]$$

risulta dunque

(1) $$\mathbf{X'XA}_1 = \mathbf{A}_1\mathbf{D}_1 \quad , \quad \mathbf{A}_1'\mathbf{A}_1 = \mathbf{I}_{(r,r)}$$

e anche

(2) $$\mathbf{X'XA}_2 = \mathbf{O}_{(p-r,p-r)} \quad , \quad \mathbf{A}_2'\mathbf{A}_2 = \mathbf{I}_{(p-r,p-r)} \, .$$

Inoltre, dalla prima delle eguaglianze in (2), premoltiplicando ambo i membri per $\mathbf{A}_2'$, si ottiene

$$\mathbf{A}_2'\mathbf{X'XA}_2 = \mathbf{O}_{(p-r,p-r)}$$

da cui

(3) $$\mathbf{XA}_2 = \mathbf{O}_{(n,p-r)} \, .$$

(*a2*) Si consideri adesso la matrice (simmetrica) $\mathbf{XX'}$.

Ovviamente, rispetto alla base naturale di R^n, $\mathbf{XX}'$ si può interpretare come la matrice di una forma bilineare simmetrica tale che la forma quadratica a essa associata è semidefinita positiva (Cfr. il punto 3 (*ii*) dei Complementi al Cap. VI di *Algebra lineare*).

Quindi, definita la trasformazione autoaggiunta T_2 di R^n la cui matrice, rispetto alla suddetta base di R^n, è $\mathbf{XX}'$, T_2 ammette (Cfr. il paragrafo 8 del Cap. VII di *Algebra lineare*):

- r autovalori positivi $\mu_1, \dots, \mu_r$ ai quali possiamo associare r autovettori ortonormali $\mathbf{b}_1, \dots, \mathbf{b}_r$;
- un autovalore nullo di molteplicità $n-r$ a cui possiamo associare $n-r$ autovettori ortonormali $\mathbf{b}_{r+1}, \dots, \mathbf{b}_n$ tali che $\mathbf{b}_{r+j} \perp \mathbf{b}_h$ per ogni $j = 1, \dots, n$ e $k = 1, \dots, r$.

Osservato che T_1 e T_2 hanno i medesimi autovalori positivi (Cfr. il punto 4 dei Complementi al Cap. V di *Algebra lineare*) e posto

$$\mathbf{B}_1 = [\mathbf{b}_1 \cdots \mathbf{b}_r] \quad , \quad \mathbf{B}_2 = [\mathbf{b}_{r+1} \cdots \mathbf{b}_n]$$

risulta dunque

$$\mathbf{XX}'\mathbf{B}_1 = \mathbf{B}_1\mathbf{D}_1 \quad , \quad \mathbf{B}_1'\mathbf{B}_1 = \mathbf{I}_{(r,r)} \tag{4}$$

e anche

$$\mathbf{XX}'\mathbf{B}_2 = \mathbf{O}_{(n-r,n-r)} \quad , \quad \mathbf{B}_2'\mathbf{B}_2 = \mathbf{I}_{(n-r,n-r)} \,. \tag{5}$$

Inoltre, dalla prima delle eguaglianze in (5), premoltiplicando ambo i membri per $\mathbf{B}_2'$, si ottiene

$$\mathbf{B}_2'\mathbf{XX}'\mathbf{B}_2 = \mathbf{O}_{(n-r,n-r)} \,.$$

da cui

$$\mathbf{X}'\mathbf{B}_2 = \mathbf{O}_{(p,n-r)} \,. \tag{6}$$

($a3$) A parziale completamento di quanto fin qui detto, vogliamo anzitutto dimostrare che tra le matrici $\mathbf{A}_1$ e $\mathbf{B}_1$ sopra definite sussistono le relazioni

$$(7) \qquad \mathbf{B}_1 = \mathbf{X}\mathbf{A}_1\mathbf{D}_1^{-1/2} \quad , \quad \mathbf{A}_1 = \mathbf{X}'\mathbf{B}_1\mathbf{D}_1^{-1/2} \; .$$

Infatti, poiché

$$\mathbf{X}'\mathbf{X}\mathbf{A}_1 = \mathbf{A}_1\mathbf{D}_1 \; ,$$

premoltiplicando ambo i membri per $\mathbf{X}$ e postmoltiplicando per $\mathbf{D}_1^{-1/2}$, si ha

$$\mathbf{X}\mathbf{X}'(\mathbf{X}\mathbf{A}_1\mathbf{D}_1^{-1/2}) = \mathbf{X}\mathbf{A}_1\mathbf{D}_1\mathbf{D}_1^{-1/2} = (\mathbf{X}\mathbf{A}_1\mathbf{D}_1^{-1/2})\,\mathbf{D}_1$$

e la prima delle relazioni in (7) risulta dimostrata.

Analogamente, poiché

$$\mathbf{X}\mathbf{X}'\mathbf{B}_1 = \mathbf{B}_1\mathbf{D}_1 \; ,$$

premoltiplicando ambo i membri per $\mathbf{X}'$ e postmoltiplicando per $\mathbf{D}_1^{-1/2}$, si ha

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{B}_1\mathbf{D}_1^{-1/2}) = \mathbf{X}'\mathbf{B}_1\mathbf{D}_1\mathbf{D}_1^{-1/2} = (\mathbf{X}'\mathbf{B}_1\mathbf{D}_1^{-1/2})\,\mathbf{D}_1$$

e anche la seconda delle relazioni in (7) è dimostrata.

Infine, considerata nuovamente la prima delle relazioni in (7) e la (1), risulta

$$(8) \qquad \mathbf{B}_1'\mathbf{X}\mathbf{A}_1 = \mathbf{D}_1^{-1/2}\mathbf{A}_1'\mathbf{X}'\mathbf{X}\mathbf{A}_1 = \mathbf{D}_1^{-1/2}\mathbf{A}_1'\mathbf{A}_1\mathbf{D}_1 = \mathbf{D}_1^{-1/2}\mathbf{D}_1 = \mathbf{D}_1^{1/2} \; .$$

($a4$) Tenuto conto di quanto precede, siamo adesso in grado di pervenire rapidamente alla desiderata scomposizione della matrice $\mathbf{X}$.

In effetti, posto

$$\mathbf{S} = \mathrm{diag}(s_1, \dots, s_r) = \mathrm{diag}\left(\lambda_1^{1/2}, \dots, \lambda_r^{1/2}\right) = \mathbf{D}_1^{1/2}$$

dove $s_1, \dots, s_r$ indicano i cosiddetti *valori singolari* di $\mathbf{X}$ e

$$\mathbf{A} = \left[\mathbf{A}_1 \ \mathbf{A}_2 \right] \quad , \quad \mathbf{B} = \left[\mathbf{B}_1 \ \mathbf{B}_2 \right]$$

si verifica immediatamente che

(9) $$\mathbf{B}'\mathbf{X}\mathbf{A} = \begin{bmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{bmatrix} \mathbf{X} \left[\mathbf{A}_1 \ \mathbf{A}_2 \right] = \begin{bmatrix} \mathbf{B}_1'\mathbf{X}\mathbf{A}_1 & \mathbf{B}_1'\mathbf{X}\mathbf{A}_2 \\ \mathbf{B}_2'\mathbf{X}\mathbf{A}_1 & \mathbf{B}_2'\mathbf{X}\mathbf{A}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{S} & \mathbf{O}_{(r,\,p-r)} \\ \mathbf{O}_{(n-r,\,r)} & \mathbf{O}_{(n-r,\,p-r)} \end{bmatrix}$$

da cui, essendo $\mathbf{B}' = \mathbf{B}^{-1}$ e $\mathbf{A}' = \mathbf{A}^{-1}$, si ottiene

(10) $$\mathbf{X} = \mathbf{B} \begin{bmatrix} \mathbf{S} & \mathbf{O}_{(r,\,p-r)} \\ \mathbf{O}_{(n-r,\,r)} & \mathbf{O}_{(n-r,\,p-r)} \end{bmatrix} \mathbf{A}'$$

e anche

(11) $$\mathbf{X} = \mathbf{B}_1 \mathbf{S} \mathbf{A}_1' \,.$$

La (10) e la (11) sono dette, rispettivamente, fattorizzazione *piena* e *ridotta* di **X** mediante valori singolari.

OSSERVAZIONE 1. Si noti che, mentre i valori singolari $s_1, \dots, s_r$ di **X** sono univocamente determinati, ciò non accade per i vettori colonna delle matrici **A** e **B** sopra definite.

In effetti, autovettori associati ad autovalori semplici sono univocamente determinati a meno del segno, mentre autovettori associati a un autovalore di molteplicità maggiore di uno sono determinati a meno di postmoltiplicazione per una matrice ortogonale (Cfr. l'Osservazione 6 del Cap. VII di *Algebra lineare*).

OSSERVAZIONE 2. In Matlab, la funzione SVD (acronimo di *S*ingular *V*alue *D*ecomposition), con le sue varie opzioni, consente di ottenere la fattorizzazione di una matrice mediante valori singolari (2).

(2) Si avverte tuttavia che la simbologia usata in Matlab differisce da quella che abbiamo introdotto.

ESEMPIO 1. Considerata la matrice

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

di rango pari a 2, risulta

$$\mathbf{S} = \begin{bmatrix} 4.3264 & 0 \\ 0 & 1.1323 \end{bmatrix}$$

$$\mathbf{A} = \left[\begin{array}{cc|c} 0.2563 & -0.7752 & 0.5774 \\ 0.5432 & 0.6096 & 0.5774 \\ 0.7995 & -0.1657 & -0.5774 \end{array}\right], \ \mathbf{B} = \left[\begin{array}{cc|cc} 0.8648 & -0.0469 & 0.4714 & -0.1667 \\ 0.3104 & 0.3920 & -0.7071 & -0.5000 \\ 0.2440 & -0.8309 & -0.4714 & 0.1667 \\ 0.3104 & 0.3920 & -0.2357 & 0.8333 \end{array}\right].$$

2.3 PSEUDOINVERSA

Le utilizzazioni del procedimento di fattorizzazione di una matrice $\mathbf{X}$ mediante valori singolari sono pressoché innumerevoli.

Una di queste riguarda la definizione di pseudoinversa di $\mathbf{X}$, alternativa a quella presentata nel Cap. 1, con una immediata ricaduta sulla questione della ricerca delle soluzioni di un sistema di equazioni lineari, sia esso compatibile o incompatibile.

Prima di affrontare l'argomento, conviene tuttavia mostrare, ricollegandosi a quanto detto nel Cap. 1 a proposito dei sottospazi fondamentali $\mathcal{R}(\mathbf{X}')$, $\mathcal{N}(\mathbf{X}), \mathcal{R}(\mathbf{X}), \mathcal{N}(\mathbf{X}')$ associati a $\mathbf{X}$, che i vettori colonna delle matrici $\mathbf{A}_1$, $\mathbf{A}_2, \mathbf{B}_1, \mathbf{B}_2$ rappresentano basi ortonormali di tali sottospazi.

In effetti, è anzitutto evidente, tenuto conto della (3) e della (6), che i vettori colonna di $\mathbf{A}_2$ e $\mathbf{B}_2$ costituiscono basi ortonormali, rispettivamente, di $\mathcal{N}(\mathbf{X})$ e $\mathcal{N}(\mathbf{X}')$.

Inoltre, poiché $\mathcal{R}(\mathbf{X}')$ è formato da tutti i vettori della forma $\mathbf{z} = \mathbf{X}'\mathbf{a}$ ottenuti facendo variare $\mathbf{a}$ in R^n, in considerazione della (11), si riconosce immediatamente che i vettori colonna di $\mathbf{A}_1$ costituiscono una base ortonormale di $\mathcal{R}(\mathbf{X}')$.

Infine, essendo $\mathcal{R}(\mathbf{X})$ è formato da tutti i vettori della forma $\mathbf{y} = \mathbf{X}\mathbf{b}$ ottenuti facendo variare $\mathbf{b}$ in R^p, sempre per la (11), è subito visto che i vettori colonna di $\mathbf{B}_1$ costituiscono una base ortonormale di $\mathcal{R}(\mathbf{X})$.

Quanto ora detto è riassunto nella Tav. 1 che segue, analoga alla corrispondente Tav. 1 del Cap. 1.

Tav. 1 - Sottospazi fondamentali e matrici i cui vettori colonna formano basi

Sottospazi	Matrici
$\mathcal{N}(\mathbf{X}) = \mathcal{R}(\mathbf{X}')^{\perp}$	$\mathbf{A}_2$
$\mathcal{R}(\mathbf{X}') = \mathcal{N}(\mathbf{X})^{\perp}$	$\mathbf{A}_1$
$\mathcal{N}(\mathbf{X}') = \mathcal{R}(\mathbf{X})^{\perp}$	$\mathbf{B}_2$
$\mathcal{R}(\mathbf{X}) = \mathcal{N}(\mathbf{X}')^{\perp}$	$\mathbf{B}_1$

Ciò premesso, riprendiamo la definizione di pseudoinversa di $\mathbf{X}$ data nella (15) del Cap. 1.

Posto $\overline{\mathbf{X}} = \mathbf{B}_1$ e $\mathbf{H} = \mathbf{S}\mathbf{A}_1'$, si ha

$$
\begin{aligned}
\mathbf{X}^{+} &= \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}' \\
&= \mathbf{A}_1\mathbf{S}(\mathbf{S}\mathbf{A}_1'\mathbf{A}_1\mathbf{S})^{-1}(\mathbf{B}_1'\mathbf{B}_1)^{-1}\mathbf{B}_1' \\
&= \mathbf{A}_1\mathbf{S}\mathbf{S}^{-2}\mathbf{B}_1' \\
&= \mathbf{A}_1\mathbf{S}^{-1}\mathbf{B}_1'
\end{aligned}
\tag{12}
$$

e questa costituisce una definizione alternativa di pseudoinversa di $\mathbf{X}$ ottenuta attraverso la fattorizzazione di $\mathbf{X}$ mediante valori singolari.

OSSERVAZIONE 3. Un'ovvia variante della (12) è data dall'espressione

$$
\mathbf{X}^{+}_{(p,n)} = \mathbf{A}\begin{bmatrix} \mathbf{S}^{-1} & \mathbf{O}_{(r,n-r)} \\ \mathbf{O}_{(p-r,r)} & \mathbf{O}_{(p-r,n-r)} \end{bmatrix}\mathbf{B}' .
$$

OSSERVAZIONE 4. In Matlab, la funzione PINV calcola la pseudoinversa di una matrice attraverso il procedimento ora descritto, ritenuto maggiormente affidabile, dal punto di vista numerico, di quello basata sulla fattorizzazione di pieno rango.

ESEMPIO 2. Considerato nuovamente l'Esempio 1 – poiché

$$\mathbf{S} = \begin{bmatrix} 4.3264 & 0 \\ 0 & 1.1323 \end{bmatrix}, \ \mathbf{A}_1 = \begin{bmatrix} 0.2563 & -0.7752 \\ 0.5432 & 0.6096 \\ 0.7995 & -0.1657 \end{bmatrix}, \mathbf{B}_1 = \begin{bmatrix} 0.8648 & -0.0469 \\ 0.3104 & 0.3920 \\ 0.2440 & -0.8309 \\ 0.3104 & 0.3920 \end{bmatrix}$$

– in applicazione della (12) si ottiene

$$\mathbf{X}^+ = \begin{bmatrix} 0.2563 & -0.7752 \\ 0.5432 & 0.6096 \\ 0.7995 & -0.1657 \end{bmatrix} \begin{bmatrix} 4.3264 & 0 \\ 0 & 1.1323 \end{bmatrix}^{-1} \begin{bmatrix} 0.8648 & -0.0469 \\ 0.3104 & 0.3920 \\ 0.2440 & -0.8309 \\ 0.3104 & 0.3920 \end{bmatrix}'$$

$$= \begin{bmatrix} 0.083 & -0.250 & 0.583 & -0.250 \\ 0.083 & 0.250 & -0.416 & 0.250 \\ 0.166 & 0 & 0.166 & 0 \end{bmatrix}.$$

Cap. 3
FATTORIZZAZIONE QR DI UNA MATRICE E MINIMI QUADRATI

3.1 INTRODUZIONE

Data una matrice $\mathbf{X}$ di ordine (n,p), supponiamo che questa abbia pieno rango per colonne, ovvero che sia $0 < r(\mathbf{X}) = p \leq n$.

In questo capitolo, ci proponiamo anzitutto di esporre un procedimento di scomposizione di tale matrice nel prodotto di due matrici, dotate di particolari caratteristiche.

Successivamente, vedremo una applicazione del procedimento in questione al problema della determinazione del vettore incognito dei coefficienti di un sistema di equazioni lineari non omogeneo incompatibile attraverso il metodo dei minimi quadrati.

3.2 FATTORIZZAZIONE QR

Data la matrice $\mathbf{X}$ di cui sopra e posto che R^n sia dotato della metrica standard [(1)], possiamo applicare il classico processo di ortonormalizzazione di Gram-Schmidt (Cfr. le sezioni 2.3 e 5.4 del Cap. VII di *Algebra lineare*) ai vettori colonna $\mathbf{x}_1, \ldots, \mathbf{x}_p$ che compogono $\mathbf{X}$ ottenendo i vettori ortonormali $\mathbf{q}_1, \ldots, \mathbf{q}_p$.

Nel dettaglio, il processo in questione conduce a definire i vettori ortogonali

(1) In realtà, quanto diremo vale qualunque sia la metrica (euclidea) adottata.

$$\mathbf{y}_1 = \mathbf{x}_1$$
$$\mathbf{y}_2 = \mathbf{x}_2 - \mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{x}_2$$
$$\dots \quad \dots \; \dots \; \dots \; \dots \; \dots \; \dots \; \dots \; \dots$$
$$\mathbf{y}_p = \mathbf{x}_p - [\mathbf{x}_1 \cdots \mathbf{x}_{p-1}]\left([\mathbf{x}_1 \cdots \mathbf{x}_{p-1}]'[\mathbf{x}_1 \cdots \mathbf{x}_{p-1}]\right)^{-1}[\mathbf{x}_1 \cdots \mathbf{x}_{p-1}]'\mathbf{x}_p$$

e da questi i vettori ortonormali

$$\mathbf{q}_1 = \frac{\mathbf{y}_1}{\sqrt{\mathbf{y}_1'\mathbf{y}_1}} \;,\; \mathbf{q}_2 = \frac{\mathbf{y}_2}{\sqrt{\mathbf{y}_2'\mathbf{y}_2}} \;,\; \dots ,\; \mathbf{q}_p = \frac{\mathbf{y}_p}{\sqrt{\mathbf{y}_p'\mathbf{y}_p}} \;.$$

Ciò premesso, indichiamo con $\widehat{\mathbf{Q}}$ la matrice di ordine (n,p) i cui vettori colonna sono $\mathbf{q}_1, \dots, \mathbf{q}_p$ e con $\widehat{\mathbf{R}}$ la matrice (invertibile) di ordine (p,p) del passaggio dalla base costituita da $\mathbf{x}_1, \dots, \mathbf{x}_p$ alla base ortonormale formata da $\mathbf{q}_1, \dots, \mathbf{q}_p$.

Ovviamente, risulta

(1) $$\mathbf{X} = \widehat{\mathbf{Q}}\widehat{\mathbf{R}} \,.$$

La (1) riceve la denominazione di fattorizzazione QR *ridotta* di $\mathbf{X}$.

Si noti che, essendo ($\widehat{\mathbf{Q}}'\widehat{\mathbf{Q}} = \mathbf{I}_{(p,p)}$)

$$\widehat{\mathbf{R}} = (\widehat{\mathbf{Q}}'\widehat{\mathbf{Q}})^{-1}\widehat{\mathbf{Q}}'\mathbf{X} = \widehat{\mathbf{Q}}'\mathbf{X} \,,$$

$\widehat{\mathbf{R}}$ ha una struttura particolare (*triangolare superiore*) in quanto, dato che per costruzione $\mathbf{q}_i \perp \mathbf{x}_j$ per ogni $i, j = 1, \dots, p$ e $i > j$, gli elementi al disotto della diagonale principale sono tutti nulli.

Si osservi anche che, a meno che sia $p = n$, $\widehat{\mathbf{Q}}$ *non* è una matrice ortogonale.

Supponiamo tuttavia di estendere, in un modo qualsiasi, la base di $\mathcal{R}(\mathbf{X})$ formata da $\mathbf{x}_1, \dots, \mathbf{x}_p$ a una base di R^n costituita da $\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{x}_{p+1}, \dots, \mathbf{x}_n$ (Cfr. il Teorema 8 del Cap. I di *Algebra lineare*).

Applicando il suddetto processo di ortonormalizzazione di Gram-Schmidt a $\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{x}_{p+1}, \dots, \mathbf{x}_n$, si ottengono i vettori ortonormali $\mathbf{q}_1, \dots, \mathbf{q}_p$,

$\mathbf{q}_{p+1}, \dots, \mathbf{q}_n$ che possiamo scrivere come vettori colonna di una matrice $\mathbf{Q}$ *ortogonale* di ordine (n,n).

Posto

$$\mathbf{Q} = \begin{bmatrix} \widehat{\mathbf{Q}} & \mathbf{Q}_1 \end{bmatrix} \quad , \quad \mathbf{R} = \begin{bmatrix} \widehat{\mathbf{R}} \\ \mathbf{O}_{(n-p,p)} \end{bmatrix}$$

è chiaro che

$$\mathbf{X} = \mathbf{QR}\,. \tag{2}$$

La (2) riceve la denominazione di fattorizzazione QR *piena* di **X**.

ESEMPIO 1. Si consideri la matrice

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

di rango pari a 2.

Applicando il processo di ortonormalizzazione di Gram-Schimidt ai due vettori colonna $\mathbf{x}_1, \mathbf{x}_2$ che compongono **X**, si verifica facilmente che

$$\mathbf{y}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad , \quad \mathbf{y}_2 = \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix} \quad , \quad \mathbf{q}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix} \quad , \quad \mathbf{q}_2 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$$

e, quindi, le matrici che entrano nella fattorizzazione QR ridotta di **X** sono

$$\widehat{\mathbf{Q}} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix} \quad , \quad \widehat{\mathbf{R}} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} \\ 0 & \sqrt{\frac{2}{3}} \end{bmatrix}.$$

Aggiungendo poi ai due vettori colonna $\mathbf{x}_1, \mathbf{x}_2$ della matrice $\mathbf{X}$ di cui sopra un terzo vettore colonna

$$\mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

in modo da ottenere una base di R^3, ripetendo il processo di ortonormalizzazione di Gram-Schimidt ai tre vettori colonna $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ che compongono la nuova matrice $\begin{bmatrix} \mathbf{X} & \mathbf{x}_3 \end{bmatrix}$ si perviene alle matrici

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{\sqrt{2}}{2} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{\sqrt{2}}{2} \end{bmatrix} \quad , \quad \mathbf{R} = \begin{bmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} \\ 0 & \sqrt{\frac{2}{3}} \\ 0 & 0 \end{bmatrix}$$

che entrano nella fattorizzazione QR piena di $\mathbf{X}$.

3.3 MINIMI QUADRATI

Supponiamo adesso di trovarsi di fronte a un sistema di equazioni lineari non omogeneo incompatibile.

In questo caso, il procedimento a cui usualmente si ricorre consiste nel ricercare soluzioni approssimate, nel senso dei *minimi quadrati*, del vettore incognito dei coefficienti.

Con le notazioni introdotte nel Cap. 1 e sempre nell'ipotesi che R^n sia dotato della metrica standard, si tratta di trovare

$$(3) \qquad \min_{\mathbf{a} \in R^p} (\mathbf{Xa} - \mathbf{y})'(\mathbf{Xa} - \mathbf{y}) = \min_{\mathbf{a} \in R^p} (\mathbf{a}'\mathbf{X}'\mathbf{Xa} - 2\mathbf{y}'\mathbf{Xa} + \mathbf{y}'\mathbf{y})$$

e ciò, come è noto, conduce al sistema di *equazioni normali*

$$(4) \qquad \mathbf{X}'\mathbf{Xa} = \mathbf{X}'\mathbf{y}$$

in cui $\mathbf{X}'\mathbf{X}$ rappresenta la matrice dei coefficienti e $\mathbf{X}'\mathbf{y}$ è il vettore dei termini noti.

Poiché abbiamo supposto che $\mathbf{X}$ sia di pieno rango per colonne, il sistema (4) ammette una soluzione (unica) data da

$$\widehat{\mathbf{a}} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y}\,. \tag{5}$$

Senonché, l'inversione della matrice $\mathbf{X}'\mathbf{X}$ può essere assai onerosa dal punto di vista computazionale [2].

Nell'intento di far fronte alla difficoltà ora accennata o quantomeno di alleviarne l'impatto, si può ricorrere al procedimento di fattorizzazione QR ridotta della matrice $\mathbf{X}$.

In tal caso, tenuta presente la (1), si riconosce subito che

$$\widehat{\mathbf{a}} = \widehat{\mathbf{R}}^{-1}\,\widehat{\mathbf{Q}}'\,\mathbf{y} \tag{5'}$$

dove la matrice $\widehat{\mathbf{R}}$, essendo triangolare superiore, risulta più facilmente invertibile.

ESEMPIO 2. Riprendendo l'Esempio 1 e posto

$$\widehat{\mathbf{y}} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix},$$

risulta

$$\widehat{\mathbf{a}} = \begin{bmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} \\ 0 & \sqrt{\frac{2}{3}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}' \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\sqrt{\frac{2}{3}} \\ 0 & \sqrt{\frac{3}{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3/2 \end{bmatrix}.$$

(2) Inoltre, se si lavora con un numero finito di decimali, sussiste il pericolo di introdurre errori che falsino il risultato.

Cap. 4
SPAZIO DUALE DI R^p

4.1 INTRODUZIONE

Riguardato R come uno spazio vettoriale su se stesso (di dimensione pari a 1), ha senso considerare lo spazio vettoriale costituito dalle trasformazioni lineari di R^p in R.

Questo spazio vettoriale − vale a dire, $\text{Hom}(R^p, R)$ − riceve la denominazione *spazio duale* di R^p, e risulta (Cfr. il punto 6 dei Complementi al Cap. V di *Algebra lineare*)

$$\dim(\text{Hom}(R^p, R)) = 1 \cdot p = p = \dim(R^p) \ .$$

Lo spazio duale di R^p è solitamente indicato con R^{p*} e i suoi elementi sono detti *forme* (*funzionali*) *lineari* su R^p.

Quindi, una forma lineare su R^p è una funzione f che associa a ciascun vettore $\mathbf{x} \in R^p$ un numero reale $f(\mathbf{x})$ − chiamato *immagine* di $\mathbf{x}$ − in modo tale che siano soddisfatte le proprietà seguenti ($\mathbf{x}_1, \mathbf{x}_2 \in R^p$; $a \in R$):

(i) $f(\mathbf{x}_1 + \mathbf{x}_2) = f(\mathbf{x}_1) + f(\mathbf{x}_2)$

(ii) $f(a\mathbf{x}_1) \quad = a f(\mathbf{x}_1)$.

ESEMPIO 1. La funzione f definita ponendo, per ogni $\mathbf{x} \in R^2$,

$$f(\mathbf{x}) = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

è, come subito si verifica, una forma lineare su R^2.

OSSERVAZIONE 1. Dati t vettori $\mathbf{x}_1, \dots, \mathbf{x}_t$ appartenenti a R^p e t numeri reali $a_1, \dots, a_t$, qualunque sia la forma lineare f su R^p, risulta

$$f(a_1\mathbf{x}_1 + \dots + a_t\mathbf{x}_t) = a_1 f(\mathbf{x}_1) + \dots + a_t f(\mathbf{x}_t) \ .$$

OSSERVAZIONE 2. Qualunque sia la forma lineare f su R^p, l'immagine del vettore nullo di R^p è il numero reale 0.

OSSERVAZIONE 3. La funzione che associa a ogni $\mathbf{x} \in R^p$ il numero reale 0 è una forma lineare su R^p (*forma lineare nulla*).

4.2 FORME LINEARI E BASI

4.2.1 Supponiamo che le basi prescelte di R^p e R siano formate, rispettivamente, da p vettori $\mathbf{x}_1, \dots, \mathbf{x}_p$ e dal numero 1 (base naturale di R).

Data una forma lineare f su R^p, sia

$$\left[f(\mathbf{x}_1) \cdots f(\mathbf{x}_p) \right] = \left[t_1 \cdots t_p \right] = \mathbf{t}'$$

il vettore riga costituito dalle immagini di $\mathbf{x}_1, \dots, \mathbf{x}_p$ mediante f.

Allora, per ogni

$$\mathbf{x} = a_1\mathbf{x}_1 + \dots + a_p\mathbf{x}_p = \left[\mathbf{x}_1 \cdots \mathbf{x}_p \right] \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} = \mathbf{X}\mathbf{a} \in R^p \ ,$$

tenuto conto dell'Osservazione 1, si ha che

$$(1) \qquad \begin{aligned} f(\mathbf{x}) = f(a_1\mathbf{x}_1 + \dots + a_p\mathbf{x}_p) &= a_1 f(\mathbf{x}_1) + \dots + a_p f(\mathbf{x}_p) \\ = \left[f(\mathbf{x}_1) \cdots f(\mathbf{x}_p) \right] \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} = \left[t_1 \cdots t_p \right] \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} &= \mathbf{t}'\mathbf{a} \ . \end{aligned}$$

Per mezzo della (1), l'immagine $f(\mathbf{x})$ di $\mathbf{x}$ risulta dunque espressa in funzione del vettore coordinato $\mathbf{a}$ di $\mathbf{x}$ e del vettore riga $\mathbf{t}'$.

Quest'ultimo vettore riga, univocamente associato alla forma lineare f su

R^p e alle suddette basi di R^p e R, riceve la denominazione di *vettore riga rappresentativo* di f rispetto a tali basi.

ESEMPIO 2. Rispetto alle basi naturali di R^2 e R, il vettore riga rappresentativo della forma lineare f su R^2 di cui all'Esempio 1 è

$$\mathbf{t}' = \begin{bmatrix} 1 & 2 \end{bmatrix} .$$

Invece, rispetto alla base di R^2 formata dai vettori

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad , \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

e alla base naturale di R – risultando

$$f(\mathbf{x}_1) = 3 \quad , \quad f(\mathbf{x}_2) = 4$$

– il vettore riga rappresentativo della stessa forma lineare f è

$$\mathbf{t}' = \begin{bmatrix} 3 & 4 \end{bmatrix} .$$

4.2.2 Supponiamo, come in precedenza, che le basi prescelte di R^p e R siano formate, rispettivamente, da p vettori $\mathbf{x}_1, \ldots, \mathbf{x}_p$ e dal numero 1.

Scelto *arbitrariamente* il vettore riga

$$\begin{bmatrix} t_1 \cdots t_p \end{bmatrix} = \mathbf{t}' ,$$

per ogni

$$\mathbf{x} = a_1 \mathbf{x}_1 + \ldots + a_p \mathbf{x}_p = \begin{bmatrix} \mathbf{x}_1 \cdots \mathbf{x}_p \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} = \mathbf{X}\mathbf{a} \in R^p ,$$

si definisca la funzione f ponendo

$$f(\mathbf{x}) = \begin{bmatrix} t_1 \cdots t_p \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} = \mathbf{t}'\mathbf{a} . \tag{2}$$

Si riconosce subito che, rispetto alle suddette basi di R^p e R, f è una forma lineare su R^p univocamente associata al vettore riga $\mathbf{t}'$ – tale che

$$\left[\, f(\mathbf{x}_1) \cdots f(\mathbf{x}_p) \right] = \left[t_1 \cdots t_p \right] = \mathbf{t}'$$

– e che il vettore riga rappresentativo di f rispetto a tali basi è $\mathbf{t}'$.

OSSERVAZIONE 4. Tenuto conto di quanto mostrato in questa e nella sezione precedente, possiamo affermare che esiste una corrispondenza biunivoca, dipendente dalle basi prescelte di R^p e R, tra l'insieme delle forme lineari su R^p e l'insieme dei vettori riga di ordine p.

4.2.3 Come si è visto – nell'ipotesi che le basi prescelte di R^p e R siano formate, rispettivamente, da p vettori $\mathbf{x}_1, \dots, \mathbf{x}_p$ e dal numero 1 – una forma lineare f su R^p risulta definita ponendo, per ogni $\mathbf{x} = \mathbf{X}\mathbf{a} \in R^p$, $f(\mathbf{x}) = \mathbf{t}'\mathbf{a}$.

Ferma restando la base di R [1], consideriamo una nuova base di R^p formata da p vettori $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_p$.

Posto $\tilde{\mathbf{X}} = \left[\tilde{\mathbf{x}}_1 \cdots \tilde{\mathbf{x}}_p \right] = \mathbf{X}\mathbf{A}$, domandiamoci come si trasformano $\mathbf{a}$ e $\mathbf{t}'$.

Ovviamente, $\mathbf{a}$ si trasforma in $\tilde{\mathbf{a}} = \mathbf{A}^{-1}\mathbf{a}$.

A sua volta, $\mathbf{t}'$, dovendo sussistere la relazione

$$f(\mathbf{x}) = \mathbf{t}'\mathbf{a} = \mathbf{t}'\mathbf{A}\mathbf{A}^{-1}\mathbf{a} \ ,$$

si trasforma in $\tilde{\mathbf{t}}' = \mathbf{t}'\mathbf{A}$.

Si usa esprimere questo diverso comportamento di $\mathbf{a}$ e $\mathbf{t}'$, in risposta a un cambiamento della base iniziale di R^p, dicendo che $\mathbf{a}$ si trasforma per *controvarianza* e $\mathbf{t}'$ per *covarianza*.

ESEMPIO 3. Rispetto alla base di R^2 formata dai vettori

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \ , \ \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

sia

(1) A tale ipotesi ci atterremo anche nel seguito.

$$\mathbf{t}' = \begin{bmatrix} 3 & 4 \end{bmatrix}$$

il vettore riga rappresentativo di una forma lineare f su R^2.

Rispetto alla base di R^2 formata dai vettori

$$\tilde{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad , \quad \tilde{\mathbf{x}}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

– risultando

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

– il vettore riga rappresentativo della stessa forma lineare f è

$$\tilde{\mathbf{t}}' = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 4 \end{bmatrix} .$$

4.3 BASE DUALE

4.3.1 Supponiamo, come in precedenza, che **X** sia la matrice i cui vettori colonna $\mathbf{x}_1, \dots, \mathbf{x}_p$ formano la base prescelta di R^p.

Posto

$$\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{t}'_1 \\ \vdots \\ \mathbf{t}'_p \end{bmatrix} ,$$

in vista di quanto detto nella sezione 4.2.2, a $\mathbf{t}'_1, \dots, \mathbf{t}'_p$ possiamo univocamente associare le p forme lineari $f_1, \dots, f_p$ su R^p tali che

$$(3) \qquad \begin{array}{l} \begin{bmatrix} f_1(\mathbf{x}_1) \cdots f_1(\mathbf{x}_p) \end{bmatrix} = \mathbf{t}'_1 \begin{bmatrix} \mathbf{x}_1 \cdots \mathbf{x}_p \end{bmatrix} = \begin{bmatrix} 1 \cdots 0 \end{bmatrix} \\ \qquad \cdots \qquad\qquad\qquad \cdots \\ \begin{bmatrix} f_p(\mathbf{x}_1) \cdots f_p(\mathbf{x}_p) \end{bmatrix} = \mathbf{t}'_p \begin{bmatrix} \mathbf{x}_1 \cdots \mathbf{x}_p \end{bmatrix} = \begin{bmatrix} 0 \cdots 1 \end{bmatrix} \end{array}$$

ovvero tali che, per ogni

$$\mathbf{x} = a_1\mathbf{x}_1 + \ldots + a_p\mathbf{x}_p = \begin{bmatrix}\mathbf{x}_1 \cdots \mathbf{x}_p\end{bmatrix}\begin{bmatrix}a_1\\ \vdots \\ a_p\end{bmatrix} = \mathbf{X}\mathbf{a} \in R^p ,$$

risulta

$$(4) \qquad \begin{array}{llll} f_1(\mathbf{x}) = \mathbf{t}'_1 \begin{bmatrix}\mathbf{x}_1 \cdots \mathbf{x}_p\end{bmatrix} \mathbf{a} = \begin{bmatrix}1 \cdots 0\end{bmatrix} \mathbf{a} = a_1 \\ \cdots \qquad \cdots \qquad \cdots \qquad \cdots \\ f_p(\mathbf{x}) = \mathbf{t}'_p \begin{bmatrix}\mathbf{x}_1 \cdots \mathbf{x}_p\end{bmatrix} \mathbf{a} = \begin{bmatrix}0 \cdots 1\end{bmatrix} \mathbf{a} = a_p . \end{array}$$

Ciò premesso, considerata una forma lineare f su R^p definita ponendo, per ogni $\mathbf{x} = \mathbf{X}\mathbf{a} \in R^p$, $f(\mathbf{x}) = \mathbf{t}'\mathbf{a}$, possiamo scrivere

$$(5) \qquad f(\mathbf{x}) = \mathbf{t}'\mathbf{a} = \mathbf{t}'\begin{bmatrix}a_1\\ \vdots \\ a_p\end{bmatrix} = \mathbf{t}'\begin{bmatrix}f_1(\mathbf{x})\\ \vdots \\ f_p(\mathbf{x})\end{bmatrix} .$$

In altri termini, f risulta esprimibile *in modo univoco* come combinazione lineare delle p forme lineari $f_1, \ldots, f_p$ su R^p con coefficienti dati dagli elementi del vettore riga $\mathbf{t}'$, ossia che

$$(6) \qquad f = \mathbf{t}'\begin{bmatrix}f_1\\ \vdots \\ f_p\end{bmatrix} .$$

Pertanto, $f_1, \ldots, f_p$ costituiscono una base di R^{p*} (Cfr. il punto 6 dei Complementi al Cap. V di *Algebra lineare*), detta *base duale* della base di R^p formata da $\mathbf{x}_1, \ldots, \mathbf{x}_p$.

ESEMPIO 4. Supponiamo che la base di R^2 sia formata dai vettori

$$\mathbf{x}_1 = \begin{bmatrix}1\\1\end{bmatrix} , \quad \mathbf{x}_2 = \begin{bmatrix}2\\1\end{bmatrix} .$$

Considerata la matrice $\mathbf{X}$ i cui vettori colonna costituiscono la suddetta base di R^2, sia

$$\mathbf{X}^{-1} = \begin{bmatrix} 1 & -2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{t}_1' \\ \mathbf{t}_2' \end{bmatrix} .$$

Allora, per ogni

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = [\mathbf{x}_1 \ \mathbf{x}_2]\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \mathbf{X}\mathbf{a} \in R^2 ,$$

la base duale di $\mathbf{x}_1$, $\mathbf{x}_2$ è costituita dalle forme lineari f_1 , f_2 su R^2 definite ponendo

$$f_1(\mathbf{x}) = \mathbf{t}_1'\mathbf{a} = [\ 1 \ \ -2\ ]\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \ a_1 - 2a_2$$

$$f_2(\mathbf{x}) = \mathbf{t}_2'\mathbf{a} = [-1 \ \ \ 1\ ]\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = -a_1 + \ a_2 .$$

4.3.2 Supponiamo che $\widetilde{\mathbf{X}}$ sia la matrice i cui vettori colonna $\widetilde{\mathbf{x}}_1, \dots, \widetilde{\mathbf{x}}_p$ formano una nuova base di R^p.

Posto

$$\widetilde{\mathbf{X}}^{-1} = \begin{bmatrix} \tilde{\mathbf{t}}_1' \\ \vdots \\ \tilde{\mathbf{t}}_p' \end{bmatrix} ,$$

con un ragionamento analogo a quello svolto nella sezione precedente, possiamo univocamente associare a $\tilde{\mathbf{t}}_1, \dots, \tilde{\mathbf{t}}_p$ le p forme lineari $\tilde{f}_1 , \dots , \tilde{f}_p$ le quali costituiscono la base duale della base di R^p formata da $\widetilde{\mathbf{x}}_1, \dots, \widetilde{\mathbf{x}}_p$.

Si noti adesso che, poiché la relazione che intercorre tra la matrice $\mathbf{X}$ i cui vettori colonna formano la base inizialmente prescelta di R^p e la matrice $\widetilde{\mathbf{X}}$ qui considerata è del tipo $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{A}$, risulta che

$$\widetilde{\mathbf{X}}^{-1} = \mathbf{A}^{-1}\mathbf{X}^{-1} ,$$

da cui si deduce subito che

(7) $$\begin{bmatrix} \tilde{f}_1 \\ \vdots \\ \tilde{f}_p \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} f_1 \\ \vdots \\ f_p \end{bmatrix} .$$

4.4 ISOMORFISMO CANONICO TRA $\mathbf{R}^p$ E IL SUO DUALE

Supponiamo che R^p sia munito della metrica (euclidea) rappresentata, rispetto alla base di R^p formata da p vettori $\mathbf{x}_1, \dots, \mathbf{x}_p$ (vettori colonna di $\mathbf{X}$), da una matrice $\mathbf{Q}$.

Fissato il vettore $\mathbf{y} = \mathbf{X}\mathbf{b} \in R^p$, possiamo univocamente associare a $\mathbf{b}$ il vettore riga $\mathbf{t}' = \mathbf{b}'\mathbf{Q}$, rappresentativo della forma lineare f su R^p definita ponendo, per ogni $\mathbf{x} = \mathbf{X}\mathbf{a} \in R^p$,

(8) $$f(\mathbf{x}) = \mathbf{t}'\mathbf{a} = \mathbf{b}'\mathbf{Q}\mathbf{a} \ .$$

Reciprocamente, dato il vettore riga $\mathbf{t}'$, rappresentativo della forma lineare f su R^p definita ponendo, per ogni $\mathbf{x} = \mathbf{X}\mathbf{a} \in R^p$, $f(\mathbf{x}) = \mathbf{t}'\mathbf{a}$, possiamo univocamente associare a $\mathbf{t}'$ il vettore $\mathbf{b} = \mathbf{Q}^{-1}\mathbf{t}$ tale che

(9) $$\mathbf{b}'\mathbf{Q}\mathbf{a} = \mathbf{t}'\mathbf{Q}^{-1}\mathbf{Q}\mathbf{a} = \mathbf{t}'\mathbf{a} = f(\mathbf{x}) \ .$$

Si può facilmente verificare che la corrispondenza biunivoca ora indicata tra vettori appartenenti a R^p e forme lineari su R^p, ottenuta mediante la introduzione di una metrica, non dipende dalla base prescelta di R^p, ovvero costituisce un isomorfismo canonico tra R^p e R^{p*}.