Introduzione all'algebra delle matrici

Lucio Barabesi
Lorenzo Fattorini

Prefazione

I piani di studio per lauree in discipline economiche o sociali contemplano corsi avanzati di statistica applicata che sono focalizzati su tecniche statistiche multivariate. Gli studenti che frequentano questo tipo di corsi non posseggono generalmente una conoscenza adeguata degli strumenti matematici necessari alla piena comprensione degli argomenti statistici proposti. Questa carenza è particolarmente evidente per quanto riguarda i concetti di algebra delle matrici. Per risolvere questa difficoltà sarebbe desiderabile introdurre la propedeuticità di un apposito corso di analisi matriciale, anche se questa soluzione non è usualmente praticabile a causa dei vincoli presenti nei piani di studio. Un ragionevole compromesso viene spesso ottenuto dai docenti dedicando circa un terzo del corso di statistica multivariata (supponendo sessanta ore di lezione frontale) all'introduzione dell'algebra delle matrici. Tuttavia, i testi di analisi matriciale disponibili sono generalmente di un livello troppo tecnico per gli studenti e coprono un insieme esteso di argomenti non tutti necessari per le finalità del corso. Di conseguenza, scopo del presente testo è quello di fornire i concetti di base dell'algebra delle matrici in modo estremamente succinto ed elementare.

Il testo non richiede nessun prerequisito da parte dello studente se non la conoscenza di elementi minimi di analisi matematica. Data la natura del testo, nessuna dimostrazione dei risultati presentati è stata considerata. Al contrario, al fine di migliorare la comprensione dello studente, i risultati sono estensivamente illustrati e verificati mediante numerosi esempi. In ogni caso, le notazioni e le terminologie sono state particolarmente curate cercando di adottare gli standard generalmente presenti in letteratura.

Gli studenti più preparati ed esigenti che desiderano approfondire gli argomenti proposti possono trovare una introduzione più completa all'algebra delle matrici nei testi di Bellman (1970), Bretscher (1997) e

Hadi (1996). Testi più avanzati sono quelli di Abadir e Magnus (2005), Horn e Johnson (1985) e Zhang (1999). I lettori particolarmente interessati alle applicazioni statistiche dell'algebra delle matrici dovrebbero considerare i testi di Graybill (1983), Harville (1997), Searle (1982) e Schott (2004). Le tematiche connesse alla generalizzazione dell'inversione di matrici sono analizzate in dettaglio nel testo di Rao e Mitra (1971). Infine, gli aspetti computazionali relativi all'analisi matriciale sono considerati da Gentle (2007).

Gli autori hanno beneficiato dei commenti e dei rilievi di alcuni colleghi sulle prime stesure del testo. I suggerimenti di questi colleghi hanno permesso di ottenere una versione finale notevolmente migliorata. In particolar modo, gli autori sono grati al Prof. Franco Fineschi, alla Dott.ssa Sara Franceschi, al Prof. Marco Lonzi, alla Prof.ssa Marzia Marcheselli e al Prof. Luca Pratelli. Ovviamente, errori ed omissioni sono da addebitare esclusivamente agli autori.

Lucio Barabesi
Lorenzo Fattorini

Siena
Marzo, 2010

Indice

Capitolo 1

Vettori

1.1. Definizioni e notazioni

Sia $\mathbb{R}$ l'insieme dei numeri reali e sia $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}$, ovvero $\mathbb{R}^n$ rappresenta il prodotto cartesiano di n copie di $\mathbb{R}$. Un elemento $\mathbf{x} \in \mathbb{R}^n$, ovvero

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

è detto *vettore* (o *vettore colonna*) di ordine n. I numeri reali $x_1, x_2, \ldots, x_n$ sono le *componenti* del vettore **x**. I vettori di ordine 1 sono gli usuali elementi di $\mathbb{R}$ e sono detti *scalari*.

Il *vettore opposto* a **x** è un vettore di ordine n definito come

$$-\mathbf{x} = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{pmatrix}.$$

Due vettori **x** e **y**, le cui i-esime componenti sono rispettivamente date da x_i e y_i, sono detti *uguali* se sono entrambi di ordine n e se $x_i = y_i$ per ogni $i = 1, 2, \ldots, n$. Per indicare che i due vettori sono uguali si adotta la notazione

$$\mathbf{x} = \mathbf{y}.$$

In caso contrario i vettori sono detti *diversi* e si scrive

$$\mathbf{x} \neq \mathbf{y}.$$

Esempio 1.1. Se $\mathbf{x} \in \mathbb{R}^3$ è un vettore tale che

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix},$$

allora il vettore opposto a $\mathbf{x}$ è dato da

$$-\mathbf{x} = \begin{pmatrix} -x_1 \\ -x_2 \\ -x_3 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ -4 \end{pmatrix}.$$

Se $\mathbf{y} \in \mathbb{R}^3$ è un vettore tale che

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix},$$

allora risulta $\mathbf{x} \neq \mathbf{y}$ in quanto si ha $x_3 \neq y_3$ pur essendo $x_1 = y_1$ e $x_2 = y_2$. $\square$

Il vettore con componenti nulle viene indicato con $\mathbf{0}$ o con $\mathbf{0}_n$ quando si vuole enfatizzare l'ordine del vettore. Analogamente, il vettore con componenti unitarie viene indicato con $\mathbf{1}$ o con $\mathbf{1}_n$.

I vettori di $\mathbb{R}^n$ con componenti nulle tranne una componente unitaria nell'i-esima posizione sono detti *vettori elementari*. Evidentemente esistono n distinti vettori elementari e l'i-esimo vettore elementare viene indicato con la notazione $\mathbf{e}_i$, ovvero

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Esempio 1.2. Il vettore di ordine 3 con componenti nulle è dato da

$$\mathbf{0} = \mathbf{0}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

mentre il vettore di ordine 3 con componenti unitarie risulta

$$\mathbf{1} = \mathbf{1}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} .$$

Infine, i tre vettori elementari di $\mathbb{R}^3$ sono dati da

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} , \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} , \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} . \qquad \square$$

Risulta interessante considerare i vettori anche da un punto di vista geometrico. In un sistema di riferimento cartesiano, un vettore può essere messo in corrispondenza biunivoca ad un *segmento orientato* (usualmente indicato con una freccia) che ha il punto di inizio coincidente con l'origine degli assi e il punto finale con coordinate date dalle componenti del vettore.

In questo ambito geometrico, la *lunghezza* (o *modulo*) del vettore è data dalla lunghezza del corrispondente segmento orientato. Inoltre, si dice che due vettori hanno la stessa *direzione* se i corrispondenti segmenti orientati appartengono ad una medesima retta passante per l'origine degli assi. Infine, si dice che due vettori hanno lo stesso *verso* se i corrispondenti segmenti orientati posseggono la medesima direzione ed uguale orientamento. Dunque il vettore opposto corrisponde ad un segmento orientato con la stessa lunghezza e direzione del vettore originale, sebbene con verso opposto.

Esempio 1.3. Si consideri il vettore $\mathbf{x} \in \mathbb{R}^2$ tale che

$$\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} .$$

Dunque si ha

$$-\mathbf{x} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} .$$

La Figura 1.1 riporta la rappresentazione geometrica dei due vettori $\mathbf{x}$ e $-\mathbf{x}$. $\square$

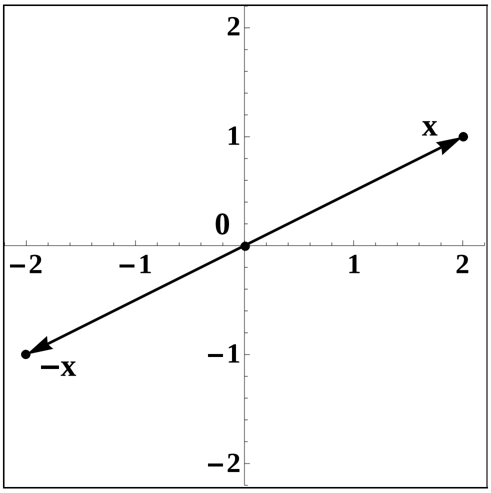

Figura 1.1.

1.2. Operazioni sui vettori

Se $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, il *vettore somma* è definito come

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} .$$

Risultato 1.1. Se $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, per la somma di vettori valgono le proprietà:
(i) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (proprietà commutativa);
(ii) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (proprietà associativa);
(iii) $\mathbf{x} + \mathbf{0} = \mathbf{x}$ (esistenza dell'elemento neutro);
(iv) $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ (esistenza dell'elemento inverso).

Per indicare il vettore somma $\mathbf{x} + (-\mathbf{y})$ si adotta semplicemente la notazione $\mathbf{x} - \mathbf{y}$. Il vettore $\mathbf{x} - \mathbf{y}$ è detto *vettore differenza*.

Esempio 1.4. Si considerino i vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ tali che

$$\mathbf{x} = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

Il vettore somma è quindi dato da

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ 1 \end{pmatrix},$$

mentre il vettore differenza risulta

$$\mathbf{x} - \mathbf{y} = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} - \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}. \qquad \square$$

Se $\mathbf{x} \in \mathbb{R}^n$, la *moltiplicazione* di $\mathbf{x}$ *per uno scalare* α è definita come il vettore

$$\alpha\mathbf{x} = \mathbf{x}\alpha = \alpha \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix}.$$

Risultato 1.2. Se α, β sono scalari e $\mathbf{x} \in \mathbb{R}^n$, per la moltiplicazione per uno scalare valgono le proprietà:
(i) $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ (proprietà associativa);
(ii) $1\mathbf{x} = \mathbf{x}$ (invarianza rispetto alla moltiplicazione per lo scalare 1).

Risultato 1.3. Se α, β sono scalari e $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, per la somma di vettori e la moltiplicazione per uno scalare valgono le proprietà:
(i) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ (proprietà distributiva rispetto alla somma di vettori);
(ii) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ (proprietà distributiva rispetto alla moltiplicazione per uno scalare).

Esempio 1.5. Si considerino lo scalare $\alpha = 2$ e il vettore $\mathbf{x} \in \mathbb{R}^3$ tale che

$$\mathbf{x} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} .$$

In questo caso, la moltiplicazione di $\mathbf{x}$ per uno scalare α risulta

$$\alpha\mathbf{x} = 2\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 6 \end{pmatrix} .$$ □

Dato uno scalare $\alpha \neq 0$, i vettori $\mathbf{x}$ e $\mathbf{y}$ di ordine n sono detti *collineari* se vale la relazione $\mathbf{y} = \alpha\mathbf{x}$. Evidentemente, i vettori $\mathbf{x}$ e $-\mathbf{x}$ sono collineari dal momento che si ha $-\mathbf{x} = \alpha\mathbf{x}$ con $\alpha = -1$.

Esempio 1.6. Si considerino i vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} , \mathbf{y} = \begin{pmatrix} 4 \\ 6 \end{pmatrix} .$$

Dal momento che

$$\mathbf{y} = \begin{pmatrix} 4 \\ 6 \end{pmatrix} = 2\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \alpha\mathbf{x}$$

con $\alpha = 2$, i vettori $\mathbf{x}$ e $\mathbf{y}$ sono collineari. □

Da un punto di vista geometrico il vettore $\mathbf{x} + \mathbf{y}$ può essere messo in corrispondenza ad un segmento orientato coincidente con la diagonale (avente punto di inizio nell'origine degli assi) del parallelogramma individuato dai vettori $\mathbf{x}$ e $\mathbf{y}$. Inoltre, il vettore $\alpha\mathbf{x}$ può essere messo in corrispondenza ad un segmento orientato con la stessa direzione di $\mathbf{x}$ e lunghezza pari ad $|\alpha|$ volte la lunghezza di $\mathbf{x}$ (e con lo stesso verso di $\mathbf{x}$ se $\alpha > 0$ e verso opposto ad $\mathbf{x}$ se $\alpha < 0$). Di conseguenza, se $\mathbf{x}$ e $\mathbf{y}$ sono collineari, i due vettori posseggono la stessa direzione, sebbene con lunghezza e verso eventualmente differenti.

Esempio 1.7. Si considerino i vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

per cui

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

e

$$\mathbf{x} - \mathbf{y} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Inoltre se $\alpha = 2$, si ha

$$\alpha\mathbf{y} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

Le Figure 1.2, 1.3 e 1.4 riportano rispettivamente le rappresentazioni geometriche del vettore somma $\mathbf{x} + \mathbf{y}$, del vettore differenza $\mathbf{x} - \mathbf{y}$ e del vettore $2\mathbf{y}$. □

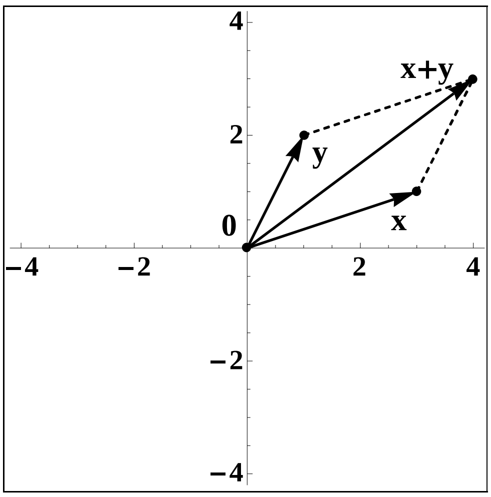

Figura 1.2.

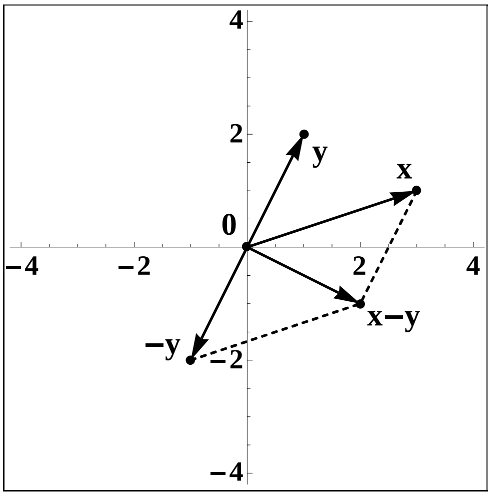

Figura 1.3.

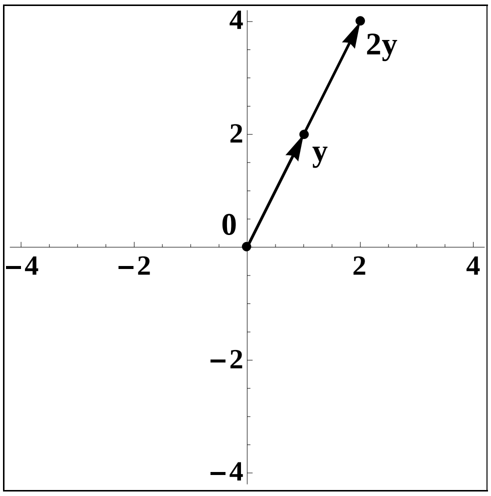

Figura 1.4.

In generale, un insieme non vuoto $\mathcal{V}$ per cui esistono una operazione di addizione sugli elementi di $\mathcal{V}$ e una operazione di moltiplicazione tra un qualsiasi numero reale e un qualsiasi elemento di $\mathcal{V}$ (che soddisfano alle proprietà elencate nei Risultati 1.1, 1.2 e 1.3) è detto *spazio vettoriale reale*. L'insieme dei vettori in $\mathbb{R}^n$ è dunque uno spazio vettoriale reale per ogni n.

Si consideri inoltre un sottoinsieme non vuoto $\mathcal{A}$ di uno spazio vettoriale $\mathcal{V}$. Se $\mathcal{A}$ è uno spazio vettoriale con le operazioni di $\mathcal{V}$ ristrette ad $\mathcal{A}$, allora $\mathcal{A}$ è detto *sottospazio vettoriale*. Ad esempio, $\mathbb{R}$ è un sottospazio vettoriale di $\mathbb{R}^2$, di $\mathbb{R}^3$ ed in generale di $\mathbb{R}^n$ $(n > 2)$, così come $\mathbb{R}^2$ è un sottospazio vettoriale di $\mathbb{R}^3$, di $\mathbb{R}^4$ ed in generale di $\mathbb{R}^n$ $(n > 3)$ e così via.

Esempio 1.8. Si consideri il sottoinsieme $\mathcal{A}$ di $\mathbb{R}^2$ dei vettori $\mathbf{x}$ che posseggono componenti identiche, ovvero

$$\mathbf{x} = \begin{pmatrix} x \\ x \end{pmatrix} .$$

Il sottoinsieme $\mathcal{A}$ di $\mathbb{R}^2$ costituisce un sottospazio vettoriale di $\mathbb{R}^2$ in quanto

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x \\ x \end{pmatrix} + \begin{pmatrix} y \\ y \end{pmatrix} = \begin{pmatrix} x + y \\ x + y \end{pmatrix} \in \mathcal{A} ,$$

mentre

$$\alpha\mathbf{x} = \alpha \begin{pmatrix} x \\ x \end{pmatrix} = \begin{pmatrix} \alpha x \\ \alpha x \end{pmatrix} \in \mathcal{A} . \qquad \square$$

1.3. Combinazioni lineari

Dati k vettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ di ordine n e k scalari $\alpha_1, \alpha_2, \dots, \alpha_k$, si dice *combinazione lineare* il vettore

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \ldots + \alpha_k\mathbf{x}_k = \sum_{i=1}^{k} \alpha_i\mathbf{x}_i .$$

I vettori $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ sono detti *linearmente dipendenti* se esistono k scalari $\alpha_1, \alpha_2, \ldots, \alpha_k$ non tutti nulli tali che

$$\sum_{i=1}^{k} \alpha_i \mathbf{x}_i = \mathbf{0} \, .$$

Se i vettori sono linearmente dipendenti, almeno uno dei k vettori può essere espresso come combinazione lineare dei rimanenti. Infatti, dal momento che esiste almeno un $\alpha_i \neq 0$ per qualche $i = 1, 2, \ldots, k$, si ha

$$\mathbf{x}_i = \sum_{j \neq i = 1}^{k} (-\alpha_i^{-1} \alpha_j) \mathbf{x}_j \, .$$

Due vettori collineari sono linearmente dipendenti in quanto la collinearità è un caso particolare della dipendenza lineare con $k = 2$. Se invece la relazione è verificata solo per $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$, i vettori sono detti *linearmente indipendenti* e nessuno di essi può essere espresso come combinazione lineare dei rimanenti.

Il massimo numero di vettori linearmente indipendenti in un sottospazio vettoriale $\mathcal{A}$ di $\mathbb{R}^n$ è detta *dimensione* di $\mathcal{A}$ e viene indicata con la notazione $\dim(\mathcal{A})$. Da un punto di vista geometrico, uno spazio vettoriale a dimensione 1 in $\mathbb{R}^n$ è una retta di $\mathbb{R}^n$ passante per l'origine, uno spazio vettoriale a dimensione 2 in $\mathbb{R}^n$ è un piano di $\mathbb{R}^n$ passante per l'origine, $\ldots$, uno spazio vettoriale a dimensione n in $\mathbb{R}^n$ coincide con $\mathbb{R}^n$. Infine $\{\mathbf{0}_n\}$ è l'unico spazio vettoriale di $\mathbb{R}^n$ a dimensione nulla.

Se A è un sottoinsieme non vuoto di vettori in $\mathbb{R}^n$, l'insieme $\mathcal{A}$ di combinazioni lineari di vettori in A è detto *spazio vettoriale generato* da A. Inoltre si dice che un insieme di k vettori $B = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ costituisce una *base* per $\mathcal{A}$ se $\mathcal{A}$ è generato da B ed i vettori $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ sono linearmente indipendenti. Facendo riferimento allo spazio $\mathbb{R}^n$, per verificare che un insieme di n vettori costituisce una base di $\mathbb{R}^n$ è sufficiente verificare una delle due condizioni, ovvero che $\mathbb{R}^n$ sia generato da B, oppure che gli n vettori di B siano linearmente indipendenti.

Esempio 1.9. Si considerino i vettori $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$ tali che

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

Un qualsiasi vettore in $\mathbb{R}^2$ può essere espresso come combinazione lineare di $\mathbf{x}_1$ e $\mathbf{x}_2$ dal momento che

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (2x_1 - x_2)\begin{pmatrix} 1 \\ 1 \end{pmatrix} + (x_2 - x_1)\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 ,$$

dove $\alpha_1 = 2x_1 - x_2$ e $\alpha_2 = x_2 - x_1$. Dunque, $A = \{\mathbf{x}_1, \mathbf{x}_2\}$ genera $\mathbb{R}^2$. Inoltre, A costituisce una base per $\mathbb{R}^2$. □

Esempio 1.10. Si considerino i vettori $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$ tali che

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

Essendo $\mathbf{x}_2 = 2\mathbf{x}_1$, le combinazioni lineari di $\mathbf{x}_1$ e $\mathbf{x}_2$ sono del tipo

$$\mathbf{x} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 = (\alpha_1 + 2\alpha_2)\mathbf{x}_1 .$$

Dunque, $A = \{\mathbf{x}_1, \mathbf{x}_2\}$ genera una retta $\mathcal{A}$ passante per l'origine che contiene tutti i vettori proporzionali ad $\mathbf{x}_1$. Quindi risulta dim$(\mathcal{A}) = 1$. In questo caso, A non è una base di $\mathbb{R}^2$. □

1.4. Prodotto interno e norma

Se $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, il *prodotto interno* (o *prodotto scalare*) di $\mathbf{x}$ e $\mathbf{y}$ è definito come

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

Risultato 1.4. Se α è uno scalare e $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, per il prodotto interno valgono le proprietà:
(i) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$;
(ii) $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$;
(iii) $\langle \alpha\mathbf{x}, \mathbf{y} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle$;
(iv) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, con $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ se e solo se $\mathbf{x} = \mathbf{0}$.

Vale inoltre la *disuguaglianza di Schwarz*, ovvero

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

dove l'uguaglianza si ha se e solo se **x** e **y** sono collineari.

Esempio 1.11. Si considerino i vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \mathbf{y} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} .$$

In questo caso si ha $\langle \mathbf{x}, \mathbf{y} \rangle = 5$, mentre $\langle \mathbf{x}, \mathbf{x} \rangle = 2$ e $\langle \mathbf{y}, \mathbf{y} \rangle = 13$. Risulta dunque verificata la disuguaglianza di Schwarz. □

Se $\mathbf{x} \in \mathbb{R}^n$, la *norma euclidea* (o semplicemente *norma*) di **x** è definita come

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

Da un punto di vista geometrico, $\|\mathbf{x}\|$ rappresenta la lunghezza del segmento orientato corrispondente al vettore **x**. Inoltre, $\|\mathbf{x}\|$ può essere anche interpretata come la distanza fra l'origine degli assi e il punto le cui coordinate sono le componenti del vettore **x**. Analogamente, $\|\mathbf{x} - \mathbf{y}\|$ può essere interpretata come la distanza fra due punti le cui coordinate sono rispettivamente date dalle componenti dei vettori **x** e **y**.

Risultato 1.5. Se α è uno scalare e $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, per la norma euclidea valgono le proprietà:
(i) $\|\alpha \mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$;
(ii) $\|\mathbf{x}\| \geq 0$, con $\|\mathbf{x}\| = 0$ se e solo se $\mathbf{x} = \mathbf{0}$;
(iii) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ con uguaglianza se e solo se **x** e **y** sono collineari (disuguaglianza triangolare).

Da un punto di vista geometrico, la disuguaglianza triangolare costituisce un'estensione a $\mathbb{R}^n$ del fatto che in un triangolo la somma delle lunghezze di due lati deve essere maggiore del terzo lato. In maniera equivalente, la disuguaglianza triangolare afferma che il percorso più breve fra due punti è dato dal segmento che congiunge i due punti.

Esempio 1.12. Si considerino i vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

In questo caso si ha $\|\mathbf{x}\| = 1$ e $\|\mathbf{y}\| = 5$, mentre $\|\mathbf{x} + \mathbf{y}\| = 4\sqrt{2}$. Risulta dunque verificata la disuguaglianza triangolare. □

Un vettore **x** per cui risulta $\|\mathbf{x}\| = 1$ è detto *normalizzato* (o *versore*). Un vettore **x** può essere normalizzato considerando la moltiplicazione di **x** per lo scalare $\|\mathbf{x}\|^{-1}$, ovvero

$$\mathbf{x}_\circ = \frac{1}{\|\mathbf{x}\|} \mathbf{x}.$$

In particolare, se **x** rappresenta un vettore di ordine 2, allora un vettore normalizzato può essere espresso come

$$\mathbf{x}_\circ = \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix},$$

dove $0 \leq \phi \leq 2\pi$.

Esempio 1.13. Si consideri il vettore $\mathbf{x} \in \mathbb{R}^2$ tale che

$$\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Dal momento che $\|\mathbf{x}\| = \sqrt{2}$, il vettore normalizzato è dato da

$$\mathbf{x}_\circ = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} = \begin{pmatrix} \cos(\pi/4) \\ \sin(\pi/4) \end{pmatrix}.$$

In questo caso risulta $\phi = \pi/4$. □

In generale uno spazio vettoriale reale dotato di una norma (che soddisfa all'insieme di proprietà elencate nel Risultato 1.5) è detto *spazio vettoriale reale normato*. L'insieme dei vettori in $\mathbb{R}^n$ è dunque uno spazio vettoriale reale normato con la norma euclidea.

1.5. Vettori ortogonali

Tenendo presente l'interpretazione geometrica dei vettori, sia θ l'arco relativo all'angolo convesso compreso fra i segmenti orientati corripondenti a due vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ tali che $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$. Si ha

$$\cos\theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \ ,$$

dove $0 \leq \theta \leq \pi$.

Il precedente risultato può essere facilmente verificato tenendo presente che dal Teorema di Carnot si ottiene

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos\theta \ .$$

Inoltre dalla definizione di norma e per le proprietà (i)-(iii) del Risultato 1.4 si ha

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

Combinando le precedenti relazioni si ottiene infine l'espressione data per $\cos\theta$.

Infine, per le proprietà di $\cos\theta$ si ha

$$-1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \leq 1 \ ,$$

che costituisce una interpretazione geometrica della disuguaglianza di Schwarz.

Esempio 1.14. Si considerino i vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} 2\sqrt{3} \\ 2 \end{pmatrix} , \mathbf{y} = \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix} .$$

Dunque risulta $\langle \mathbf{x}, \mathbf{y} \rangle = 4\sqrt{3}$, $\|\mathbf{x}\| = 4$ e $\|\mathbf{y}\| = 2$. Di conseguenza si ha che

$$\cos\theta = \frac{\sqrt{3}}{2}$$

e quindi risulta $\theta = \pi/6$. I vettori $\mathbf{x}$, $\mathbf{y}$ e l'arco θ sono rappresentati nella Figura 1.5. □

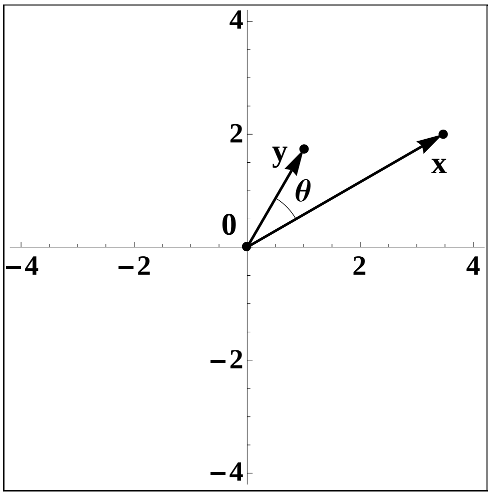

Figura 1.5.

Due vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ sono detti *ortogonali* se $\cos\theta = 0$ (e quindi $\theta = \pi/2$), ovvero se $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. Per indicare che due vettori $\mathbf{x}$ e $\mathbf{y}$ sono ortogonali si adotta la notazione $\mathbf{x} \perp \mathbf{y}$. Due vettori $\mathbf{x}$ e $\mathbf{y}$, per i quali risulta $\mathbf{x} \perp \mathbf{y}$ e $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$, sono detti *ortonormali*.

Infine, se $\mathbf{x} \perp \mathbf{y}$ si ha la relazione

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

che da un punto di vista geometrico può essere interpretata come una estensione del Teorema di Pitagora in $\mathbb{R}^n$.

Esempio 1.15. Si considerino i vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} 2\sqrt{3} \\ 2 \end{pmatrix} , \mathbf{y} = \begin{pmatrix} -3 \\ 3\sqrt{3} \end{pmatrix} .$$

Dal momento che si ha $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ risulta $\mathbf{x} \perp \mathbf{y}$. I vettori non sono ortonormali in quanto $\|\mathbf{x}\| = 4$ e $\|\mathbf{y}\| = 6$. Essendo $\|\mathbf{x} + \mathbf{y}\| = \sqrt{52}$, risulta verificato il Teorema di Pitagora. □

Esempio 1.16. Si considerino i vettori $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix}.$$

In questo caso **x** e **y** sono ortonormali in quanto si ha $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, mentre risulta $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. I vettori **x** e **y** sono rappresentati nella Figura 1.6. □

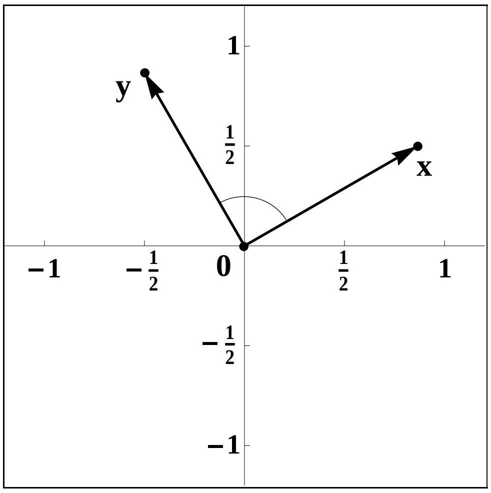

Figura 1.6.

In generale, si dice che k vettori $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ costituiscono un *insieme ortonormale* se $\mathbf{x}_i \perp \mathbf{x}_j$ per ogni $i \neq j = 1, 2, \ldots, k$ e $\|\mathbf{x}_i\| = 1$ per ogni $i = 1, 2, \ldots, k$.

Tenendo presente quanto detto nella Sezione 1.3, una base è detta *ortonormale* se è costituita da vettori ortonormali. In particolare, l'insieme

$$B_{0,n} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$$

è una base ortonormale di $\mathbb{R}^n$ in quanto gli elementi di $B_{0,n}$ costituiscono un insieme ortonormale e ogni vettore $\mathbf{x} \in \mathbb{R}^n$ può essere espresso come combinazione lineare degli $\mathbf{e}_i$, ovvero

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i \ ,$$

dove $\alpha_i = x_i$ per $i = 1, 2, \ldots, n$.

La base $B_{0,n}$ è detta *base canonica* di $\mathbb{R}^n$, anche se essa non costituisce l'unica base ortonormale di $\mathbb{R}^n$. In effetti, qualsiasi insieme di n vettori ortonormali è una base ortonormale di $\mathbb{R}^n$. In generale, se $B_n = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ è una qualsiasi base ortonormale, ogni vettore $\mathbf{x} \in \mathbb{R}^n$ può essere espresso come

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i \ ,$$

dove $\alpha_i = \langle \mathbf{x}, \mathbf{x}_i \rangle$ per $i = 1, 2, \ldots, n$.

Infine, si noti che ogni spazio vettoriale $\mathcal{A}$ a dimensione unitaria, ovvero tale che $\dim(\mathcal{A}) = 1$, ammette una sola base ortonormale costituita dall'unico vettore normalizato di $\mathcal{A}$ (a meno del segno).

Esempio 1.17. Mediante la base canonica $B_{0,2} = \{\mathbf{e}_1, \mathbf{e}_2\}$ un qualsiasi vettore $\mathbf{x} \in \mathbb{R}^2$ può essere espresso come

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 \ ,$$

dove $\alpha_1 = x_1$ e $\alpha_2 = x_2$. Inoltre, si considerino i vettori $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$

$$\mathbf{x}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} , \mathbf{x}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} .$$

La base $B = \{\mathbf{x}_1, \mathbf{x}_2\}$ costituisce un'ulteriore base ortonormale di $\mathbb{R}^2$ in quanto si ha $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = 0$, mentre $\|\mathbf{x}_1\| = \|\mathbf{x}_2\| = 1$. Dunque, un qualsiasi vettore $\mathbf{x} \in \mathbb{R}^2$ può essere espresso anche come

$$\mathbf{x} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 \ ,$$

dove

$$\alpha_1 = \langle \mathbf{x}, \mathbf{x}_1 \rangle = \frac{\sqrt{2}}{2}(x_1 + x_2) \; , \; \alpha_2 = \langle \mathbf{x}, \mathbf{x}_2 \rangle = \frac{\sqrt{2}}{2}(x_2 - x_1) \; . \quad \square$$

Esempio 1.18. Se $\mathcal{A}$ è lo spazio vettoriale in $\mathbb{R}^3$ di dimensione unitaria generato da

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \; ,$$

allora il vettore normalizzato

$$\mathbf{x}_\circ = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}$$

costituisce una base ortonormale di $\mathcal{A}$ ed esiste solo una altra base ortonormale di $\mathcal{A}$ data dal vettore $-\mathbf{x}_\circ$. $\square$

1.6. Proiezioni di vettori

Tenendo presente l'interpretazione geometrica dei vettori, dati due vettori $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$ tali che $\mathbf{x}, \mathbf{z} \neq \mathbf{0}$, si consideri la proiezione del segmento orientato corrispondente a $\mathbf{x}$ sulla retta che contiene il segmento orientato corrispondente a $\mathbf{z}$. Il vettore $\mathbf{y}$ che si ottiene dalla proiezione è dato da

$$\mathbf{y} = \frac{\langle \mathbf{x}, \mathbf{z} \rangle}{\|\mathbf{z}\|^2} \, \mathbf{z} \; .$$

Questo risultato può essere facilmente verificato tenendo presente innanzitutto che i vettori $\mathbf{y}$ e $\mathbf{z}$ deveno essere collineari. Inoltre, dal momento che

$$\|\mathbf{y}\| = \|\mathbf{x}\| \cdot |\cos \theta|$$

deve risultare

$$\mathbf{y} = \frac{\|\mathbf{x}\| \cos\theta}{\|\mathbf{z}\|}\, \mathbf{z} \, .$$

Considerando l'espressione di $\cos\theta$ si ottiene infine l'espressione del vettore **y**.

Esempio 1.19. Si considerino i vettori $\mathbf{x}, \mathbf{z} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} , \mathbf{z} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} .$$

In questo caso si ha $\langle \mathbf{x}, \mathbf{z} \rangle = 6$, mentre $\|\mathbf{x}\| = \|\mathbf{z}\| = \sqrt{10}$. Quindi l'arco relativo ad **x** e **z** risulta $\theta = \arccos(3/5)$. Il vettore **y** che si ottiene proiettando **x** su **z** è quindi dato da

$$\mathbf{y} = \frac{3}{5}\, \mathbf{z} = \begin{pmatrix} 9/5 \\ 3/5 \end{pmatrix} .$$

I vettori **x** e **z** e il relativo vettore **y** sono riportati nella Figura 1.7. $\square$

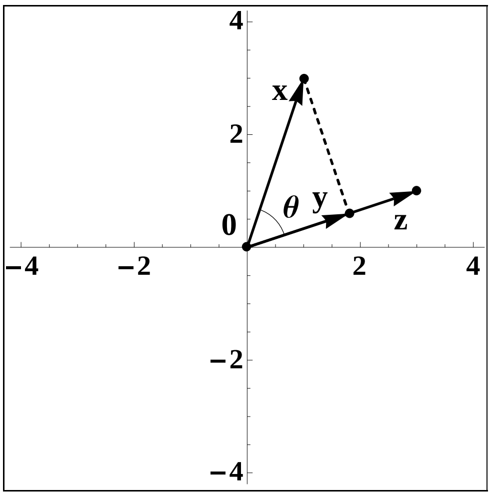

Figura 1.7.

Esempio 1.20. Si considerino i vettori $\mathbf{x}, \mathbf{z} \in \mathbb{R}^2$ tali che

$$\mathbf{x} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}, \mathbf{z} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

In questo caso si ha $\langle \mathbf{x}, \mathbf{z} \rangle = -4$, mentre $\|\mathbf{x}\| = 2\sqrt{2}$ e $\|\mathbf{z}\| = \sqrt{10}$. Quindi l'arco relativo ad $\mathbf{x}$ e $\mathbf{z}$ risulta $\theta = \arccos(-\sqrt{5}/5)$. Il vettore $\mathbf{y}$ che si ottiene proiettando $\mathbf{x}$ su $\mathbf{z}$ è quindi dato da

$$\mathbf{y} = -\frac{2}{5}\,\mathbf{z} = \begin{pmatrix} -6/5 \\ -2/5 \end{pmatrix}.$$

I vettori $\mathbf{x}$ e $\mathbf{z}$ e il relativo vettore $\mathbf{y}$ sono riportati nella Figura 1.8. $\square$

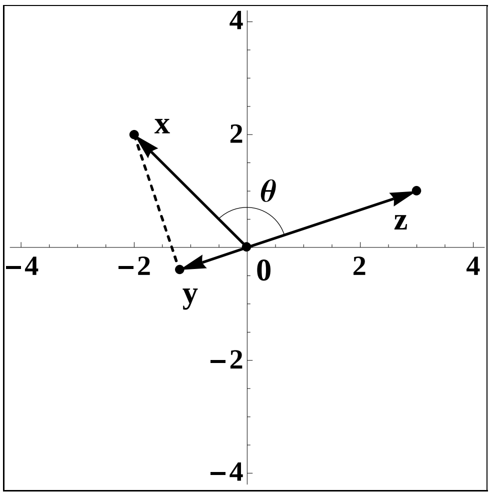

Figura 1.8.

1.6. Iperpiani

Dati un vettore $\boldsymbol{\alpha}$ di costanti e un vettore $\mathbf{x} \in \mathbb{R}^n$, lo spazio vettoriale

$$\mathcal{P} = \{\mathbf{x} : \langle \boldsymbol{\alpha}, \mathbf{x} \rangle = 0\}$$

è detto *iperpiano*. Dal momento che

$$\langle \boldsymbol{\alpha}, \mathbf{x} \rangle = \sum_{i=1}^{n} \alpha_i x_i = 0$$

dove α_i e x_i rappresentano rispettivamente le i-esime componenti di $\boldsymbol{\alpha}$ e **x**, supponendo $\alpha_i \neq 0$ per qualche $i = 1, 2, \ldots, n$, si ha

$$x_i = \sum_{j \neq i = 1}^{n} (- \alpha_i^{-1} \alpha_j) x_j \ .$$

Dunque, lo spazio vettoriale $\mathcal{P}$ è costituito da tutti i vettori di $\mathbb{R}^n$ la cui i-esima componente è data dalla combinazione lineare delle rimanenti $(n-1)$ componenti. Ogni iperpiano di $\mathbb{R}^n$ ha dimensione $(n-1)$, ovvero $\dim(\mathcal{P}) = n-1$. Da un punto di vista geometrico, ogni retta passante per l'origine è un iperpiano di $\mathbb{R}^2$, ogni piano passante per l'origine è un iperpiano di $\mathbb{R}^3$, e così via.

Dato un ulteriore scalare α_0, l'insieme

$$\mathcal{T} = \{\mathbf{x} : \langle \boldsymbol{\alpha}, \mathbf{x} \rangle = \alpha_0\}$$

è detto *iperpiano traslato* e non costituisce uno spazio vettoriale dal momento che non contiene il vettore **0**. Da un punto di vista geometrico, ogni retta è un iperpiano traslato di $\mathbb{R}^2$, ogni piano è un iperpiano traslato di $\mathbb{R}^3$, e così via.

Esempio 1.21. Dati $\alpha_1 = 4, \alpha_2 = -2$, si consideri l'iperpiano di $\mathbb{R}^2$

$$\mathcal{P} = \{\mathbf{x} : 4x_1 - 2x_2 = 0\} \ ,$$

che dunque è costituito dai vettori per cui $x_1 = x_2/2$, ovvero dai vettori del tipo

$$\mathbf{x} = \begin{pmatrix} \beta/2 \\ \beta \end{pmatrix} ,$$

dove $\beta \in \mathbb{R}$. Da un punto di vista geometrico $\mathcal{P}$ coincide con la retta di equazione $x_2 = 2x_1$. Analogamente, se $\alpha_0 = \ -1$, l'iperpiano traslato

$$\mathcal{T} = \{\mathbf{x} : 4x_1 - 2x_2 = \ -1\} \ ,$$

che è costituito dai vettori per cui $x_1 = x_2/2 - 1/4$, ovvero dai vettori del tipo

$$\mathbf{x} = \begin{pmatrix} \beta/2 - 1/4 \\ \beta \end{pmatrix},$$

dove $\beta \in \mathbb{R}$. Da un punto di vista geometrico $\mathcal{T}$ coincide con la retta di equazione $x_2 = 1/2 + 2x_1$. $\square$

Capitolo 2

Matrici

2.1. Definizioni e notazioni

Un insieme di numeri reali disposti in n righe e k colonne

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix}$$

è detto *matrice* di ordine $(n \times k)$. Una matrice di ordine $(n \times 1)$ è in effetti un vettore colonna di ordine n. Il numero reale a_{ij} rappresenta la *componente* di **A** che appartiene contemporaneamente alla i-esima riga e alla j-esima colonna. La matrice **A** di ordine $(n \times k)$ può essere indicata in forma più concisa tramite l'espressione

$$\mathbf{A} = (a_{ij}) \; .$$

La *matrice opposta* ad **A** è definita come

$$-\mathbf{A} = (-a_{ij}) = \begin{pmatrix} -a_{11} & -a_{12} & \dots & -a_{1k} \\ -a_{21} & -a_{22} & \dots & -a_{2k} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \dots & -a_{nk} \end{pmatrix} \; .$$

Due matrici $\mathbf{A} = (a_{ij})$ e $\mathbf{B} = (b_{ij})$ sono dette *uguali* se sono entrambe di ordine $(n \times k)$ e se $a_{ij} = b_{ij}$ per ogni $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, k$. In questo caso si adotta la notazione

$$\mathbf{A} = \mathbf{B} \; .$$

In caso contrario le matrici sono dette *diverse* e si scrive

$$\mathbf{A} \neq \mathbf{B}\,.$$

La matrice di ordine $(n \times k)$ con componenti tutte nulle viene indicata con $\mathbf{O}$ o con $\mathbf{O}_{n\times k}$ quando si vuole enfatizzare l'ordine della matrice.

Data una matrice $\mathbf{A}$ di ordine $(n \times k)$ si dice *trasposta* di $\mathbf{A}$ la matrice di ordine $(k \times n)$

$$\mathbf{A}^{\mathrm{T}} = (a_{ji}) = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1k} & a_{2k} & \dots & a_{nk} \end{pmatrix}$$

che si ottiene da $\mathbf{A}$ scambiando le righe con le colonne. In altri termini, la prima riga di $\mathbf{A}$ è la prima colonna di $\mathbf{A}^{\mathrm{T}}$, la seconda riga di $\mathbf{A}$ è la seconda colonna di $\mathbf{A}^{\mathrm{T}}$, $\dots$, la n-esima riga di $\mathbf{A}$ è la n-esima colonna di $\mathbf{A}^{\mathrm{T}}$. In generale, la componente di posto (i, j) di $\mathbf{A}$ è la componente di posto (j, i) di $\mathbf{A}^{\mathrm{T}}$. Inoltre si ha $(\mathbf{A}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{A}$.

Esempio 2.1. Si consideri la matrice $\mathbf{A}$ di ordine (2×3) data da

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 2 & 4 \end{pmatrix}.$$

La matrice opposta ad $\mathbf{A}$ di ordine (2×3) risulta

$$-\mathbf{A} = \begin{pmatrix} -3 & -1 & 0 \\ 0 & -2 & -4 \end{pmatrix},$$

mentre la trasposta di $\mathbf{A}$ è la matrice di ordine (3×2) data da

$$\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 3 & 0 \\ 1 & 2 \\ 0 & 4 \end{pmatrix}.$$

Infine, la matrice di ordine (2×3) con componenti nulle è data da

$$\mathbf{O} = \mathbf{O}_{2\times 3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \qquad \square$$

Se un vettore colonna di ordine n viene considerato come una matrice di ordine $(n \times 1)$, il corrispondente vettore trasposto è detto *vettore riga* di ordine n ed è in effetti una matrice di ordine $(1 \times n)$. Risulta adesso evidente la motivazione per la definizione di vettore colonna introdotta nella Sezione 1.1.

Una matrice di ordine $(n \times k)$ può essere espressa come un insieme di k vettori colonna di ordine n

$$\mathbf{A} = (\mathbf{a}_{\bullet 1} \ \mathbf{a}_{\bullet 2} \ \ldots \ \mathbf{a}_{\bullet k}) \ ,$$

dove

$$\mathbf{a}_{\bullet j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}$$

indica il j-esimo vettore colonna di **A**. Alternativamente, una matrice di ordine $(n \times k)$ può essere espressa come un insieme di n vettori riga di ordine k

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_{1\bullet}^{\mathrm{T}} \\ \mathbf{a}_{2\bullet}^{\mathrm{T}} \\ \vdots \\ \mathbf{a}_{n\bullet}^{\mathrm{T}} \end{pmatrix} \ ,$$

dove

$$\mathbf{a}_{i\bullet}^{\mathrm{T}} = (a_{i1} \ a_{i2} \ \ldots \ a_{ik})$$

indica l'i-esimo vettore riga di **A**. Di conseguenza, una matrice di ordine $(n \times k)$ può essere interpretata come un insieme di k vettori in $\mathbb{R}^n$, oppure come un insieme di n vettori in $\mathbb{R}^k$.

Esempio 2.2. Si consideri la matrice **A** di ordine (3×2) data da

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 0 \end{pmatrix} .$$

La matrice **A** può essere espressa come un insieme di 2 vettori colonna di ordine 3, ovvero

$$\mathbf{a}_{\bullet 1} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} , \mathbf{a}_{\bullet 2} = \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} ,$$

o alternativamente come un insieme di 3 vettori riga di ordine 2, ovvero

$$\mathbf{a}_{1\bullet}^{\mathrm{T}} = (2\ 1) , \mathbf{a}_{2\bullet}^{\mathrm{T}} = (1\ 4) , \mathbf{a}_{3\bullet}^{\mathrm{T}} = (1\ 0) . \qquad \square$$

2.2. Matrici a blocchi

Estendendo il concetto di rappresentazione per vettori colonna o vettori riga, ogni matrice può essere suddivisa in insiemi detti *sottomatrici* considerando gruppi di vettori colonna e vettori riga. Se i gruppi di vettori sono contigui, la matrice è detta *a blocchi*. In particolare, quando si considerano due gruppi contigui di vettori riga e due gruppi contigui di vettori colonna, ogni matrice di ordine $(n \times k)$ può essere espressa come matrice a blocchi del tipo

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} ,$$

dove $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$ e $\mathbf{A}_{22}$ sono sottomatrici di ordine rispettivamente pari a $(m \times h)$, $(m \times (k-h))$, $((n-m) \times h)$, $((n-m) \times (k-h))$, mentre m e h sono tali che $1 \leq m \leq n$ e $1 \leq h \leq k$.

Se $m = n$, allora la matrice è costituita da due blocchi del tipo

$$\mathbf{A} = (\mathbf{A}_{11}\ \mathbf{A}_{12}) ,$$

mentre se $h = k$, allora la matrice è costituita da due blocchi del tipo

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} \\ \mathbf{A}_{21} \end{pmatrix} .$$

Si noti che definendo l'ordine di $\mathbf{A}_{11}$ anche gli ordini delle restanti sottomatrici sono automaticamente definiti. Inoltre, per enfatizzare la separazione in blocchi della matrice $\mathbf{A}$ si utilizza eventualmente l'ulteriore notazione

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right) ,$$

o le notazioni

$$\mathbf{A} = (\mathbf{A}_{11} \,|\, \mathbf{A}_{12})$$

e

$$\mathbf{A} = \left(\begin{array}{c} \mathbf{A}_{11} \\ \hline \mathbf{A}_{21} \end{array} \right) ,$$

che possono risultare utili con matrici numeriche.

Infine, per quanto riguarda la trasposta di una matrice a blocchi si ha

$$\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} \mathbf{A}_{11}^{\mathrm{T}} & \mathbf{A}_{21}^{\mathrm{T}} \\ \mathbf{A}_{12}^{\mathrm{T}} & \mathbf{A}_{22}^{\mathrm{T}} \end{pmatrix} .$$

Esempio 2.3. Si consideri la matrice a blocchi di ordine (5×6) data

$$\mathbf{A} = \left(\begin{array}{cccc|cc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 2 & 1 & 1 \\ 1 & 3 & 2 & 2 & 0 & 1 \\ 2 & 1 & 4 & 0 & 0 & 1 \end{array} \right) .$$

In questo caso risulta $m = 2$ e $h = 4$. Inoltre si ha

$$\mathbf{A}_{11} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 0 \end{pmatrix} , \mathbf{A}_{12} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} ,$$

mentre

$$\mathbf{A}_{21} = \begin{pmatrix} 0 & 0 & 1 & 2 \\ 1 & 3 & 2 & 2 \\ 2 & 1 & 4 & 0 \end{pmatrix} , \mathbf{A}_{22} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} . \qquad \square$$

Esempio 2.4. Si consideri la matrice a blocchi di ordine (5×3) data da

$$\mathbf{A} = \left(\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 1 & 0 \\ \hline 0 & 0 & 1 \\ 1 & 3 & 1 \\ 2 & 1 & 1 \end{array}\right) .$$

In questo caso si ha $m = 2$ e $h = k = 3$, mentre

$$\mathbf{A}_{11} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \mathbf{A}_{21} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 3 & 1 \\ 2 & 1 & 1 \end{pmatrix} .$$

Inoltre la trasposta di **A** risulta

$$\mathbf{A}^{\mathrm{T}} = \left(\begin{array}{cc|ccc} 0 & 1 & 0 & 1 & 2 \\ 1 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{array}\right) ,$$

dove

$$\mathbf{A}_{11}^{\mathrm{T}} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{A}_{21}^{\mathrm{T}} = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix} . \qquad \square$$

2.3. Matrici quadrate

Una matrice per la quale il numero delle righe è uguale al numero delle colonne, ovvero quando $n = k$, è detta *matrice quadrata* di ordine k. Ovviamente, ogni matrice quadrata di ordine $k = 1$ è costituita da un solo numero reale, ovvero $\mathbf{A} = (a)$.

In una matrice quadrata **A** di ordine k le componenti $a_{11}, a_{22}, \ldots, a_{kk}$ costituiscono la *diagonale principale* della matrice. Inoltre, una matrice quadrata **A** è detta *simmetrica* quando $a_{ij} = a_{ji}$ per ogni $i \neq j = 1, 2, \ldots, k$, ovvero quando $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$. In particolare si ha $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$ se e solo se **A** è simmetrica.

Una matrice quadrata le cui componenti che non appartengono alla diagonale principale sono nulle, ovvero se $a_{ij} = 0$ per ogni

$i \neq j = 1, 2, \ldots, k$, è detta *matrice diagonale*. Ovviamente, ogni matrice diagonale è simmetrica.

Per ogni insieme di k scalari $a_1, a_2, \ldots, a_k$, si definisce con $\mathrm{dg}(a_1, a_2, \ldots, a_k)$ la matrice diagonale

$$\mathrm{dg}(a_1, a_2, \ldots, a_k) = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & a_k \end{pmatrix} .$$

Inoltre, se $\mathbf{A}$ è una matrice quadrata di ordine k, la matrice diagonale le cui componenti non nulle coincidono con le componenti della diagonale di $\mathbf{A} = (a_{ij})$ viene indicata con la notazione

$$\mathrm{dg}(\mathbf{A}) = \begin{pmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & a_{kk} \end{pmatrix} .$$

Una matrice diagonale nella quale le componenti della diagonale principale sono uguali allo scalare α, ovvero del tipo $\mathrm{dg}(\alpha, \alpha, \ldots, \alpha)$, è detta *matrice scalare*. Una matrice scalare di ordine k per la quale risulta $\alpha = 1$ è detta *matrice identità* e viene indicata con la notazione $\mathbf{I}_k$, ovvero

$$\mathbf{I}_k = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} .$$

Si noti che risulta $\mathbf{I}_k = (\mathbf{e}_1 \; \mathbf{e}_2 \; \ldots \; \mathbf{e}_k)$.

Infine, una matrice quadrata $\mathbf{A} = (a_{ij})$ è detta *triangolare inferiore* quando $a_{ij} = 0$ per ogni $i < j$. Una matrice quadrata $\mathbf{A} = (a_{ij})$ è detta *triangolare superiore* quando $a_{ij} = 0$ per ogni $i > j$.

Esempio 2.5. La matrice

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 4 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

è una matrice simmetrica di ordine 3 e si ha

$$\mathrm{dg}(\mathbf{A}) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

La matrice

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

è una matrice scalare di ordine 3 con $\alpha = 2$. La matrice

$$\mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

è una matrice identità di ordine 3. La matrice

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & 0 & 3 \end{pmatrix}$$

è una matrice triangolare inferiore di ordine 3, mentre la matrice

$$\mathbf{A} = \begin{pmatrix} 4 & 1 & 0 \\ 0 & 2 & 4 \\ 0 & 0 & 3 \end{pmatrix}$$

è una matrice triangolare superiore di ordine 3. □

2.4. Matrici ortogonali

Una matrice quadrata di ordine k le cui colonne sono costituite da k vettori ortonormali $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_k$ è detta *matrice ortogonale* e si indica generalmente con $\boldsymbol{\Gamma}$, ovvero

$$\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1 \; \boldsymbol{\gamma}_2 \; \ldots \; \boldsymbol{\gamma}_k) .$$

Si noti che una matrice ortogonale dovrebbe essere definita più correttamente come matrice ortonormale, anche se questa terminologia non è usualmente adottata.

Si può verificare che se $\mathbf{\Gamma}$ è ortogonale anche $\mathbf{\Gamma}^{\mathrm{T}}$ è ortogonale. Inoltre, nel caso particolare in cui $k = 2$, ogni matrice ortogonale può essere espressa nella forma

$$\mathbf{\Gamma} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix},$$

o nella forma

$$\mathbf{\Gamma} = \begin{pmatrix} \cos\phi & -\sin\phi \\ -\sin\phi & -\cos\phi \end{pmatrix},$$

dove $0 \le \phi \le 2\pi$.

Esempio 2.6. I vettori $\boldsymbol{\gamma}_1$ e $\boldsymbol{\gamma}_2$ di ordine 2

$$\boldsymbol{\gamma}_1 = \begin{pmatrix} \sqrt{3}/2 \\ 1/2 \end{pmatrix}, \boldsymbol{\gamma}_2 = \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix},$$

sono ortonormali in quanto $\|\boldsymbol{\gamma}_1\| = \|\boldsymbol{\gamma}_2\| = 1$, mentre $\langle \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \rangle = 0$. Quindi la matrice

$$\mathbf{\Gamma} = (\boldsymbol{\gamma}_1 \; \boldsymbol{\gamma}_2) = \begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix} = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix}$$

è ortogonale. È immediato verificare che

$$\mathbf{\Gamma}^{\mathrm{T}} = \begin{pmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{pmatrix}$$

è a sua volta una matrice ortogonale. □

2.5. Operazioni sulle matrici

Se **A** e **B** sono matrici di ordine $(n \times k)$, la *matrice somma* è definita come

$$\mathbf{A}+\mathbf{B}=(a_{ij})+(b_{ij})=(a_{ij}+b_{ij})\ .$$

Risultato 2.1. Se **A**, **B** e **C** sono matrici di ordine $(n\times k)$, per la somma di matrici valgono le proprietà:
(i) $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$ (proprietà commutativa);
(ii) $(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$ (proprietà associativa);
(iii) $\mathbf{A}+\mathbf{O}=\mathbf{A}$ (esistenza dell'elemento neutro);
(iv) $\mathbf{A}+(-\mathbf{A})=\mathbf{O}$ (esistenza dell'elemento inverso).

Per indicare la somma di matrici $\mathbf{A}+(-\mathbf{B})$ si adotta semplicemente la notazione $\mathbf{A}-\mathbf{B}$. La matrice $\mathbf{A}-\mathbf{B}$ è detta *matrice differenza*. Inoltre, per quanto riguarda la trasposta della matrice somma si ha

$$(\mathbf{A}+\mathbf{B})^{\mathrm{T}}=\mathbf{A}^{\mathrm{T}}+\mathbf{B}^{\mathrm{T}}\ .$$

Esempio 2.7. Si considerino le matrici **A** e **B** di ordine (3×2) tali che

$$\mathbf{A}=\begin{pmatrix}1&1\\2&1\\2&0\end{pmatrix},\mathbf{B}=\begin{pmatrix}0&1\\1&1\\3&1\end{pmatrix}.$$

In questo caso risulta

$$\mathbf{A}+\mathbf{B}=\begin{pmatrix}1&1\\2&1\\2&0\end{pmatrix}+\begin{pmatrix}0&1\\1&1\\3&1\end{pmatrix}=\begin{pmatrix}1&2\\3&2\\5&1\end{pmatrix},$$

mentre

$$\mathbf{A}-\mathbf{B}=\begin{pmatrix}1&1\\2&1\\2&0\end{pmatrix}-\begin{pmatrix}0&1\\1&1\\3&1\end{pmatrix}=\begin{pmatrix}1&0\\1&0\\-1&-1\end{pmatrix}. \qquad \square$$

Se **A** è una matrice di ordine $(n\times k)$, la *moltiplicazione* di **A** *per uno scalare* α è definita come

$$\alpha\mathbf{A}=\mathbf{A}\alpha=(\alpha a_{ij})\ .$$

Risultato 2.2. Se α, β sono scalari e **A** è una matrice di ordine $(n \times k)$, per la moltiplicazione per uno scalare valgono le proprietà:
(i) $\alpha(\beta\mathbf{A}) = (\alpha\beta)\mathbf{A}$ (proprietà associativa);
(ii) $1\mathbf{A} = \mathbf{A}$ (invarianza rispetto alla moltiplicazione per lo scalare 1).

Risultato 2.3. Se α, β sono scalari e **A** e **B** sono matrici di ordine $(n \times k)$, per la somma di matrici e la moltiplicazione per uno scalare valgono le proprietà:
(i) $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$ (proprietà distributiva rispetto alla somma di matrici);
(ii) $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$ (proprietà distributiva rispetto alla moltiplicazione per uno scalare).

Inoltre, per ogni matrice scalare di ordine k si ha

$$\mathrm{dg}(\alpha, \alpha, \ldots, \alpha) = \alpha\mathbf{I}_k \ .$$

Infine, per quanto detto nella Sezione 1.2, si noti che l'insieme delle matrici di ordine $(n \times k)$ costituisce uno *spazio vettoriale reale*.

Esempio 2.8. Si consideri la matrice **A** di ordine (2×3) tale che

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & 3 \end{pmatrix} .$$

In questo caso, se $\alpha = 3$ risulta

$$\alpha\mathbf{A} = 3\begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 3 & -3 \\ 6 & 0 & 9 \end{pmatrix} . \qquad \square$$

Date le matrici a blocchi **A** e **B** di ordine $(n \times k)$ definite come

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix},$$

con $\mathbf{A}_{11}$ e $\mathbf{B}_{11}$ sottomatrici dello stesso ordine, allora la matrice somma è data da

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{pmatrix} .$$

Inoltre la moltiplicazione di **A** per lo scalare α risulta

$$\alpha\mathbf{A} = \begin{pmatrix} \alpha\mathbf{A}_{11} & \alpha\mathbf{A}_{12} \\ \alpha\mathbf{A}_{21} & \alpha\mathbf{A}_{22} \end{pmatrix} .$$

Esempio 2.9. Si considerino le matrici quadrate **A** e **B** di ordine (3×4) tali che

$$\mathbf{A} = \left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \hline 0 & 2 & 1 & 1 \end{array}\right) , \mathbf{B} = \left(\begin{array}{cr|rr} 2 & 1 & -2 & 0 \\ 0 & 4 & 0 & 1 \\ \hline 0 & -1 & 0 & 1 \end{array}\right) .$$

In questo caso si ha

$$\mathbf{A} + \mathbf{B} = \left(\begin{array}{cc|rr} 2 & 2 & -2 & 0 \\ 1 & 4 & 1 & 1 \\ \hline 0 & 1 & 1 & 2 \end{array}\right) ,$$

mentre se $\alpha = 3$ risulta

$$\alpha\mathbf{A} = \left(\begin{array}{cc|cc} 0 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ \hline 0 & 6 & 3 & 3 \end{array}\right) . \qquad \square$$

2.6. Traccia

La somma delle componenti della diagonale principale di una matrice quadrata $\mathbf{A} = (a_{ij})$ di ordine k è detta *traccia* e viene indicata con la notazione

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^{k} a_{ii} .$$

La traccia è un operatore lineare, nel senso che se **A** e **B** sono matrici quadrate dello stesso ordine e α è uno scalare, allora si ha:
(i) $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$;
(ii) $\text{tr}(\alpha\mathbf{A}) = \alpha\text{tr}(\mathbf{A})$.

Inoltre se a è uno scalare si pone

$$\mathrm{tr}(a) = a\ .$$

Infine, data la matrice quadrata a blocchi **A** di ordine k con $\mathbf{A}_{11}$ sottomatrice quadrata di ordine h, la traccia di **A** risulta

$$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}_{11}) + \mathrm{tr}(\mathbf{A}_{22})\ .$$

Esempio 2.10. Si considerino le matrici quadrate **A** e **B** di ordine 3 tali che

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 3 \\ 5 & 1 & -3 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 3 & 0 & 1 \\ 1 & 2 & -2 \\ 2 & 0 & 3 \end{pmatrix}\ .$$

In questo caso si ha $\mathrm{tr}(\mathbf{A}) = -2$ e $\mathrm{tr}(\mathbf{B}) = 8$. Dal momento che

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 4 & 1 & 2 \\ 3 & 2 & 1 \\ 7 & 1 & 0 \end{pmatrix}\ ,$$

allora risulta

$$\mathrm{tr}(\mathbf{A} + \mathbf{B}) = 6 = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})\ .$$

Inoltre, se $\alpha = 2$ si ha

$$\alpha\mathbf{A} = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 0 & 6 \\ 10 & 2 & -6 \end{pmatrix}\ ,$$

per cui

$$\mathrm{tr}(\alpha\mathbf{A}) = -4 = \alpha\mathrm{tr}(\mathbf{A})\ . \qquad \square$$

Capitolo 3

Prodotto di matrici

3.1. Prodotto di Cayley

Data una matrice $\mathbf{A} = (a_{il})$ di ordine $(n \times p)$ e una matrice $\mathbf{B} = (b_{lj})$ di ordine $(p \times k)$, la *matrice prodotto di Cayley* (o *matrice prodotto righe per colonne*, o semplicemente *matrice prodotto*) è definita come

$$\mathbf{AB} = (\langle \mathbf{a}_{i\bullet}, \mathbf{b}_{\bullet j} \rangle) = \left(\sum_{l=1}^{p} a_{il} b_{lj} \right) ,$$

dove $\mathbf{a}_{i\bullet}^{\mathrm{T}}$ e $\mathbf{b}_{\bullet j}$ rappresentano rispettivamente l'i-esimo vettore riga di **A** e il j-esimo vettore colonna di **B**. Dunque la componente di posto (i, j) della matrice **AB** è ottenuta effettuando il prodotto interno dell'i-esimo vettore riga di **A** con il j-esimo vettore colonna di **B**. Evidentemente la matrice **AB** è di ordine $(n \times k)$.

Affinchè il prodotto tra due matrici sia possibile, occorre che il numero delle colonne di **A** sia uguale al numero delle righe di **B**. In questo caso le due matrici sono dette *conformabili* rispetto al prodotto.

Nel prodotto di Cayley è fondamentale considerare l'ordine in cui compaiono le matrici coinvolte. In effetti la matrice **BA** può non essere definita o essere diversa dalla matrice **AB**. In particolare i prodotti **AB** e **BA** esistono entrambi solamente quando le matrici **A** e **B** sono rispettivamente di ordine $(n \times k)$ e $(k \times n)$. In questo caso i prodotti danno luogo rispettivamente a matrici quadrate di ordine k o di ordine n. Dunque in generale non vale per il prodotto di Cayley la proprietà commutativa, ovvero

$$\mathbf{AB} \neq \mathbf{BA} .$$

Per questo motivo con **AB** si specifica che **A** premoltiplica **B** o che **B** postmoltiplica **A**.

Esempio 3.1. Si considerino le matrici **A** e **B** rispettivamente di ordine (2×3) e (3×2) tali che

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 0 \end{pmatrix}.$$

In questo caso risulta

$$\mathbf{AB} = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix},$$

mentre

$$\mathbf{BA} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ 5 & 0 & 3 \\ 2 & 0 & 2 \end{pmatrix}.$$

Anche se i prodotti **AB** e **BA** esistono, tuttavia danno luogo a matrici quadrate diverse, rispettivamente di ordine 2 e 3. $\square$

Esempio 3.2. Si considerino le matrici triangolari **A** e **B** di ordine 2 tali che

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

In questo caso risulta

$$\mathbf{AB} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$$

mentre

$$\mathbf{BA} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Anche se i prodotti **AB** e **BA** esistono e danno luogo a matrici quadrate di ordine 2, tuttavia risulta $\mathbf{AB} \neq \mathbf{BA}$. $\square$

Risultato 3.1. Se **A**, **B** e **C** sono matrici opportunamente conformabili, per il prodotto di Cayley valgono le proprietà:
(i) $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ (proprietà associativa);
(ii) $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$ (proprietà distributiva a destra);
(iii) $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA}+\mathbf{CA}$ (proprietà distributiva a sinistra).

Dalla definizione di prodotto di Cayley si può verificare che, premoltiplicando o postmoltiplicando una matrice **A** di ordine $(n \times k)$ per una matrice identità conformabile rispetto al prodotto con **A**, si ottiene nuovamente la matrice **A**, ovvero

$$\mathbf{I}_n\mathbf{A} = \mathbf{A}\mathbf{I}_k = \mathbf{A} \, ,$$

mentre premoltiplicando o postmoltiplicando per una matrice con componenti nulle conformabile rispetto al prodotto con **A** si ha

$$\mathbf{OA} = \mathbf{AO} = \mathbf{O} \, .$$

Se **A** rappresenta una matrice quadrata di ordine k, si definisce $\mathbf{A}^2 = \mathbf{AA}$. Inoltre **A** è detta *idempotente* se risulta $\mathbf{A}^2 = \mathbf{A}$.

Esempio 3.3. Si consideri la matrice **A** di ordine (3×2) tale che

$$\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 2 \end{pmatrix} .$$

Si ha

$$\mathbf{I}_3\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 2 \end{pmatrix} = \mathbf{A}$$

e

$$\mathbf{A}\mathbf{I}_2 = \begin{pmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{A} \, .$$

Inoltre risulta

$$\mathbf{O}_{2\times 3}\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 2 \end{pmatrix} = \mathbf{O}_{2\times 2}$$

e

$$\mathbf{A}\mathbf{O}_{2\times 4} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \mathbf{O}_{3\times 4} \, . \qquad \square$$

Esempio 3.4. Si consideri la matrice quadrata **A** di ordine 3 tale che

$$\mathbf{A} = \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix} .$$

Si ha

$$\mathbf{A}^2 = \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix} = \mathbf{A}$$

e dunque la matrice **A** è idempotente. $\square$

Per quanto riguarda la trasposta della matrice prodotto risulta

$$(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} \, ,$$

mentre per quanto riguarda la traccia della matrice prodotto, se **A** e **B** sono matrici di ordine $(n \times k)$ si ha

$$\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{B}) = \mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{B}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{B}\mathbf{A}^{\mathrm{T}}) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_{ij}b_{ij} \, .$$

Date infine le matrici a blocchi **A** e **B** rispettivamente di ordine $(n \times p)$ e $(p \times k)$ con $\mathbf{A}_{11}$ e $\mathbf{B}_{11}$ sottomatrici rispettivamente di ordine $(m \times q)$ e $(q \times h)$, allora la matrice prodotto è data da

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix} .$$

Esempio 3.5. Si considerino le matrici **A** e **B** di ordine (2×3) tali che

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 5 \end{pmatrix}.$$

In questo caso risulta

$$\mathbf{A}\mathbf{B}^{\mathrm{T}} = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ 3 & 5 \end{pmatrix},$$

mentre

$$\mathbf{B}\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 5 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 7 & 5 \end{pmatrix},$$

da cui

$$(\mathbf{A}\mathbf{B}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{B}\mathbf{A}^{\mathrm{T}}.$$

Inoltre si ha

$$\mathbf{A}^{\mathrm{T}}\mathbf{B} = \begin{pmatrix} 3 & 2 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 5 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 13 \\ 1 & 0 & 1 \\ 1 & 2 & 6 \end{pmatrix},$$

mentre

$$\mathbf{B}^{\mathrm{T}}\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 3 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 1 \\ 4 & 0 & 2 \\ 13 & 1 & 6 \end{pmatrix},$$

da cui si ha

$$(\mathbf{A}^{\mathrm{T}}\mathbf{B})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}.$$

Infine si noti che

$$\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{B}) = \mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{B}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{B}\mathbf{A}^{\mathrm{T}}) = 9\,. \qquad \square$$

3.2. Prodotto interno e norma di matrici

Se **A** e **B** sono matrici di ordine $(n \times k)$, il *prodotto interno* (o *prodotto scalare*) di **A** e **B** è definito come

$$\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{B}) \ ,$$

mentre la *norma di Frobenius* di **A** è definita come

$$\|\mathbf{A}\|_{\mathrm{F}} = \langle \mathbf{A}, \mathbf{A} \rangle^{1/2} \ .$$

Queste definizioni estendono alle matrici i concetti di prodotto interno e norma. In effetti, per $k = 1$ si ottengono di nuovo le definizioni introdotte nella Sezione 1.4 per i vettori.

Risultato 3.2. Se α è uno scalare, mentre **A**, **B** e **C** sono matrici di ordine $(n \times k)$, per il prodotto interno di matrici valgono le proprietà:
(i) $\langle \mathbf{A}, \mathbf{B} \rangle = \langle \mathbf{B}, \mathbf{A} \rangle$;
(ii) $\langle \mathbf{A}, \mathbf{B} + \mathbf{C} \rangle = \langle \mathbf{A}, \mathbf{B} \rangle + \langle \mathbf{A}, \mathbf{C} \rangle$;
(iii) $\langle \alpha\mathbf{A}, \mathbf{B} \rangle = \alpha\langle \mathbf{A}, \mathbf{B} \rangle$;
(iv) $\langle \mathbf{A}, \mathbf{A} \rangle \geq 0$, con $\langle \mathbf{A}, \mathbf{A} \rangle = 0$ se e solo se $\mathbf{A} = \mathbf{O}$.

Risultato 3.3. Se α è uno scalare, mentre **A** e **B** sono matrici di ordine $(n \times k)$, per la norma di Frobenius valgono le proprietà:
(i) $\|\alpha\mathbf{A}\|_{\mathrm{F}} = |\alpha| \cdot \|\mathbf{A}\|_{\mathrm{F}}$;
(ii) $\|\mathbf{A}\|_{\mathrm{F}} \geq 0$, con $\|\mathbf{A}\|_{\mathrm{F}} = 0$ se e solo se $\mathbf{A} = \mathbf{O}$;
(iii) $\|\mathbf{A} + \mathbf{B}\|_{\mathrm{F}} \leq \|\mathbf{A}\|_{\mathrm{F}} + \|\mathbf{B}\|_{\mathrm{F}}$ con uguaglianza se e solo se $\mathbf{B} = \alpha\mathbf{A}$ (disuguaglianza triangolare).

L'insieme delle matrici di ordine $(n \times k)$ costituisce dunque uno spazio *vettoriale reale normato* con la norma di Frobenius.

Esempio 3.6. Si considerino le matrici **A** e **B** di ordine (3×2) tali che

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix} , \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 1 \end{pmatrix} .$$

Dal momento che

$$\mathbf{A}^{\mathrm{T}}\mathbf{B} = \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} ,$$

mentre

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} ,$$

e

$$\mathbf{B}^{\mathrm{T}}\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} ,$$

allora si ha $\langle \mathbf{A}, \mathbf{B} \rangle = 4$, $\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{7}$ e $\|\mathbf{B}\|_{\mathrm{F}} = \sqrt{3}$. □

3.3. Alcuni prodotti particolari

Se $\mathbf{a} = (a_1\ a_2\ \dots\ a_n)^{\mathrm{T}}$ e $\mathbf{b} = (b_1\ b_2\ \dots\ b_n)^{\mathrm{T}}$ sono due vettori di ordine n, il prodotto interno dei due vettori è un caso particolare di prodotto di Cayley dal momento che

$$\mathbf{a}^{\mathrm{T}}\mathbf{b} = \sum_{i=1}^{n} a_i b_i = \langle \mathbf{a}, \mathbf{b} \rangle .$$

Ovviamente risulta

$$\mathbf{a}^{\mathrm{T}}\mathbf{b} = \mathbf{b}^{\mathrm{T}}\mathbf{a} .$$

Se $\mathbf{b} = \mathbf{1}$, il prodotto interno

$$\mathbf{a}^{\mathrm{T}}\mathbf{1} = \sum_{i=1}^{n} a_i$$

coincide con la somma delle componenti di $\mathbf{a}$. Si ha inoltre $\mathbf{1}^{\mathrm{T}}\mathbf{1} = n$.

Anche il quadrato della norma di un vettore è un caso particolare di prodotto di Cayley, dal momento che

$$\mathbf{a}^{\mathrm{T}}\mathbf{a} = \sum_{i=1}^{n} a_i^2 = \|\mathbf{a}\|^2 .$$

Analogamente, il prodotto interno

$$(\mathbf{a}-\mathbf{b})^{\mathrm{T}}(\mathbf{a}-\mathbf{b}) = \sum_{i=1}^{n}(a_i-b_i)^2 = \|\mathbf{a}-\mathbf{b}\|^2$$

rappresenta il quadrato della norma $\|\mathbf{a}-\mathbf{b}\|$.

Considerati i vettori $\mathbf{a} = (a_1\ a_2\ \dots\ a_n)^{\mathrm{T}}$ e $\mathbf{b} = (b_1\ b_2\ \dots\ b_k)^{\mathrm{T}}$, il prodotto $\mathbf{ab}^{\mathrm{T}}$ (detto *prodotto esterno*) è la matrice di ordine $(n \times k)$ data da

$$\mathbf{ab}^{\mathrm{T}} = \begin{pmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_k \\ a_2b_1 & a_2b_2 & \dots & a_2b_k \\ \vdots & \vdots & & \vdots \\ a_nb_1 & a_nb_2 & \dots & a_nb_k \end{pmatrix} = (a_ib_j)\ .$$

Si noti che $\mathbf{ab}^{\mathrm{T}} \neq \mathbf{ba}^{\mathrm{T}}$ anche se risulta $\mathbf{ab}^{\mathrm{T}} = (\mathbf{ba}^{\mathrm{T}})^{\mathrm{T}}$. Se $\mathbf{b} = \mathbf{1}$, il prodotto esterno è dato da

$$\mathbf{a1}^{\mathrm{T}} = \begin{pmatrix} a_1 & a_1 & \dots & a_1 \\ a_2 & a_2 & \dots & a_2 \\ \vdots & \vdots & & \vdots \\ a_n & a_n & \dots & a_n \end{pmatrix} = (\mathbf{a}\ \mathbf{a}\ \dots\ \mathbf{a})\ ,$$

ovvero una matrice in cui tutti i vettori colonna sono uguali ad **a**. Analogamente, il prodotto esterno $\mathbf{1a}^{\mathrm{T}}$ risulta

$$\mathbf{1a}^{\mathrm{T}} = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ a_1 & a_2 & \dots & a_n \\ \vdots & \vdots & & \vdots \\ a_1 & a_2 & \dots & a_n \end{pmatrix} = \begin{pmatrix} \mathbf{a}^{\mathrm{T}} \\ \mathbf{a}^{\mathrm{T}} \\ \vdots \\ \mathbf{a}^{\mathrm{T}} \end{pmatrix}\ ,$$

ovvero una matrice in cui tutti i vettori riga sono uguali ad $\mathbf{a}^{\mathrm{T}}$. Infine, il prodotto esterno $\mathbf{aa}^{\mathrm{T}}$ dà luogo a una matrice simmetrica di ordine n

$$\mathbf{aa}^{\mathrm{T}} = \begin{pmatrix} a_1^2 & a_1a_2 & \dots & a_1a_n \\ a_2a_1 & a_2^2 & \dots & a_2a_n \\ \vdots & \vdots & & \vdots \\ a_na_1 & a_na_2 & \dots & a_n^2 \end{pmatrix} = (a_ia_j)\ .$$

Esempio 3.7. Si considerino i vettori di ordine 3 dati da

$$\mathbf{a} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} .$$

In questo caso si ha

$$\mathbf{ab}^{\mathrm{T}} = \begin{pmatrix} 0 & 4 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \end{pmatrix} .$$

Inoltre risulta

$$\mathbf{a1}_3^{\mathrm{T}} = \begin{pmatrix} 2 & 2 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

e

$$\mathbf{1}_3\mathbf{a}^{\mathrm{T}} = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 1 & 0 \\ 2 & 1 & 0 \end{pmatrix} ,$$

mentre

$$\mathbf{aa}^{\mathrm{T}} = \begin{pmatrix} 4 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \qquad \square$$

Data una matrice **A** di ordine $(n \times k)$ si considerino i prodotti $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ e $\mathbf{A}\mathbf{A}^{\mathrm{T}}$. Ovviamente $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ e $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ sono matrici simmetriche di ordine k e n. Si noti inoltre che

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = (\mathbf{a}_{\bullet i}^{\mathrm{T}}\mathbf{a}_{\bullet j}) = \sum_{i=1}^{n} \mathbf{a}_{i\bullet}\mathbf{a}_{i\bullet}^{\mathrm{T}} ,$$

dove $\mathbf{a}_{i\bullet}^{\mathrm{T}}$ e $\mathbf{a}_{j\bullet}^{\mathrm{T}}$ rappresentano rispettivamente l'i-esimo e il j-esimo vettore riga di **A**, mentre $\mathbf{a}_{\bullet i}$ e $\mathbf{a}_{\bullet j}$ rappresentano rispettivamente l'i-esimo e il j-esimo vettore colonna di **A**. In modo del tutto analogo si ha che

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = (\mathbf{a}_{i\bullet}^{\mathrm{T}}\mathbf{a}_{j\bullet}) = \sum_{j=1}^{k} \mathbf{a}_{\bullet j}\mathbf{a}_{\bullet j}^{\mathrm{T}} \,.$$

Infine, se $\mathbf{A}$ è una matrice simmetrica, si ha $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{A}^2$.

Esempio 3.8. Si consideri la matrice $\mathbf{A}$ di ordine (3×2) tale che

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 2 & 1 \\ 0 & 1 \end{pmatrix} .$$

Si ha

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix} ,$$

mentre

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

e quindi $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ e $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ sono matrici simmetriche di ordine 2 e 3. □

Esempio 3.9. Si consideri la matrice quadrata $\mathbf{A}$ di ordine 2 tale che

$$\mathbf{A} = (\mathbf{a}_{\bullet 1}\ \mathbf{a}_{\bullet 2}) = \begin{pmatrix} \mathbf{a}_{1\bullet}^{\mathrm{T}} \\ \mathbf{a}_{2\bullet}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} .$$

In questo caso risulta

$$\mathbf{a}_{1\bullet}\mathbf{a}_{1\bullet}^{\mathrm{T}} + \mathbf{a}_{2\bullet}\mathbf{a}_{2\bullet}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$$

e

$$\mathbf{a}_{\bullet 1}\mathbf{a}_{\bullet 1}^{\mathrm{T}} + \mathbf{a}_{\bullet 2}\mathbf{a}_{\bullet 2}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \mathbf{A}\mathbf{A}^{\mathrm{T}} \,. \quad \square$$

3.4. Prodotto di Kronecker

Date le matrici **A** di ordine $(n \times k)$ e **B** di ordine $(p \times q)$, si dice *prodotto di Kronecker* la matrice a blocchi di ordine $(np \times kq)$

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1k}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2k}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \dots & a_{nk}\mathbf{B} \end{pmatrix} .$$

Il prodotto di Kronecker $\mathbf{A} \otimes \mathbf{B}$ è sempre fattibile. In generale, per il prodotto di Kronecker non vale la proprietà commutativa, ovvero

$$\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A} .$$

Esempio 3.10. Si considerino le matrici **A** e **B** rispettivamente di ordine (2×2) e (2×3) tali che

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} , \mathbf{B} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 4 \end{pmatrix} .$$

In questo caso risulta

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 1\mathbf{B} & 1\mathbf{B} \\ 0\mathbf{B} & 2\mathbf{B} \end{pmatrix} = \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 2 & 4 & 0 & 2 & 4 \\ \hline 0 & 0 & 0 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 4 & 8 \end{array}\right) . \qquad \square$$

Risultato 3.4. Date le matrici **A**, **B**, **C** e **D**, se **B** e **C** sono dello stesso ordine e se i prodotti di Cayley **AC** e **BD** esistono, per il prodotto di Kronecker valgono le proprietà:
(i) $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$ (proprietà associativa);
(ii) $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$ (proprietà distributiva a destra);
(iii) $(\mathbf{B} + \mathbf{C}) \otimes \mathbf{A} = \mathbf{B} \otimes \mathbf{A} + \mathbf{C} \otimes \mathbf{A}$ (proprietà distributiva a sinistra);
(iv) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$ (proprietà distributiva rispetto al prodotto di Cayley).

Il prodotto di Kronecker di **A** con la matrice con componenti nulle risulta

$$\mathbf{O} \otimes \mathbf{A} = \mathbf{A} \otimes \mathbf{O} = \mathbf{O} \, .$$

Per quanto riguarda la trasposta della matrice prodotto di Kronecker si ha

$$(\mathbf{A} \otimes \mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}} \otimes \mathbf{B}^{\mathrm{T}} \, .$$

Se $\mathbf{A}$ e $\mathbf{B}$ sono matrici quadrate, per la traccia della matrice prodotto di Kronecker risulta

$$\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B}) \, .$$

Data la matrice a blocchi $\mathbf{A}$, la matrice prodotto di Kronecker è data da

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} \mathbf{A}_{11} \otimes \mathbf{B} & \mathbf{A}_{12} \otimes \mathbf{B} \\ \mathbf{A}_{21} \otimes \mathbf{B} & \mathbf{A}_{22} \otimes \mathbf{B} \end{pmatrix} .$$

Infine, se $\mathbf{a}$ e $\mathbf{b}$ sono vettori rispettivamente di ordine n e k, si noti che il prodotto esterno è un caso particolare di prodotto di Kronecker, ovvero $\mathbf{a}\mathbf{b}^{\mathrm{T}} = \mathbf{a} \otimes \mathbf{b}^{\mathrm{T}}$.

Esempio 3.11. Si considerino le matrici $\mathbf{A}$ e $\mathbf{B}$ rispettivamente di ordine (2×2) e (3×3) tali che

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} , \mathbf{B} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 4 \\ 0 & 0 & 1 \end{pmatrix} .$$

In questo caso risulta

$$\mathbf{A} \otimes \mathbf{B} = \left(\begin{array}{ccc|ccc} 2 & 0 & 2 & 1 & 0 & 1 \\ 0 & 4 & 8 & 0 & 2 & 4 \\ 0 & 0 & 2 & 0 & 0 & 1 \\ \hline 1 & 0 & 1 & 2 & 0 & 2 \\ 0 & 2 & 4 & 0 & 4 & 8 \\ 0 & 0 & 1 & 0 & 0 & 2 \end{array} \right) .$$

Dunque si ha $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = 16 = \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})$. □

Capitolo 4

Rango e determinante

4.1. Operazioni elementari

Data una matrice $\mathbf{A}$ di ordine $(n \times k)$, sono dette *operazioni elementari* sui vettori riga (risp. colonna) di $\mathbf{A}$:
(i) lo scambio di due vettori riga (risp. colonna);
(ii) la moltiplicazione di un vettore riga (risp. colonna) per uno scalare $\alpha \neq 0$;
(iii) l'addizione ad un vettore riga (risp. colonna) di un altro vettore riga (risp. colonna) moltiplicato per uno scalare $\alpha \neq 0$.

Le operazioni elementari sui vettori riga (risp. colonna) producono una nuova matrice i cui vettori riga (risp. colonna) appartengono allo spazio vettoriale generato dai vettori riga (risp. colonna) della matrice originale.

È conveniente considerare inizialmente le singole operazioni elementari sulla matrice identità. Dunque, sia $\mathbf{E}_{ij}$ la matrice ottenuta scambiando l'i-esimo e il j-esimo vettore riga di $\mathbf{I}_n$, sia $\mathbf{E}_i(\alpha)$ la matrice ottenuta moltiplicando per α l'i-esimo vettore riga di $\mathbf{I}_n$ e sia $\mathbf{E}_i(\alpha|j)$ la matrice ottenuta sommando all'i-esimo vettore riga di $\mathbf{I}_n$ il j-esimo vettore riga di $\mathbf{I}_n$ moltiplicato per α. Le matrici $\mathbf{E}_{ij}$, $\mathbf{E}_i(\alpha)$ e $\mathbf{E}_i(\alpha|j)$ sono dette *matrici elementari di riga* e sono esprimibili attraverso la matrice identità e gli elementi della base canonica, ovvero

$$\mathbf{E}_{ij} = \mathbf{I}_n - (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^{\mathrm{T}} ,$$

$$\mathbf{E}_i(\alpha) = \mathbf{I}_n + (\alpha - 1)\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}} ,$$

e

$$\mathbf{E}_i(\alpha|j) = \mathbf{I}_n + \alpha\mathbf{e}_i\mathbf{e}_j^{\mathrm{T}} .$$

Esempio 4.1. Se $n = 3$ e $\alpha = 5$, si ha

$$\mathbf{E}_{12} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} (1 \quad -1 \quad 0) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

e

$$\mathbf{E}_1(5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 4 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} (1 \quad 0 \quad 0) = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} ,$$

mentre

$$\mathbf{E}_2(5|3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 5 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} (0 \quad 0 \quad 1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{pmatrix} . \ \square$$

Inoltre, in modo analogo, sia $\mathbf{F}_{ij}$ la matrice ottenuta scambiando l'i-esimo e il j-esimo vettore colonna di $\mathbf{I}_k$, sia $\mathbf{F}_i(\alpha)$ la matrice ottenuta moltiplicando per α l'i-esimo vettore colonna di $\mathbf{I}_k$ e sia $\mathbf{F}_i(\alpha|j)$ la matrice ottenuta sommando all'i-esimo vettore colonna di $\mathbf{I}_k$ il j-esimo vettore colonna di $\mathbf{I}_k$ moltiplicato per α. Le matrici $\mathbf{F}_{ij}$, $\mathbf{F}_i(\alpha)$ e $\mathbf{F}_i(\alpha|j)$ sono dette *matrici elementari di colonna* e risulta

$$\mathbf{F}_{ij} = \mathbf{I}_k - (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^{\mathrm{T}} ,$$

$$\mathbf{F}_i(\alpha) = \mathbf{I}_k + (\alpha - 1)\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}$$

e

$$\mathbf{F}_i(\alpha|j) = \mathbf{I}_k + \alpha\mathbf{e}_j\mathbf{e}_i^{\mathrm{T}} .$$

Se $n = k$, si ha inoltre

$$\mathbf{F}_{ij} = \mathbf{E}_{ij} ,$$

$$\mathbf{F}_i(\alpha) = \mathbf{E}_i(\alpha)$$

e

$$\mathbf{F}_i(\alpha|j) = \mathbf{E}_j(\alpha|i) = \mathbf{E}_i(\alpha|j)^{\mathrm{T}}.$$

Esempio 4.2. Se $n = k = 3$, è immediato verificare che

$$\mathbf{F}_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{F}_1(5) = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{F}_2(5|3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{pmatrix}$$

e considerando l'Esercizio 4.1 si ha $\mathbf{F}_{12} = \mathbf{E}_{12}$, $\mathbf{F}_1(5) = \mathbf{E}_1(5)$, mentre

$$\mathbf{F}_2(5|3) = \mathbf{E}_2(5|3)^{\mathrm{T}} = \mathbf{E}_3(5|2)\ . \qquad \square$$

Se $\mathbf{B}$ è la matrice ottenuta mediante una operazione elementare sui vettori riga di $\mathbf{A}$ e se $\mathbf{E}$ è una matrice del tipo $\mathbf{E}_{ij}$, $\mathbf{E}_i(\alpha)$ o $\mathbf{E}_i(\alpha|j)$, allora risulta

$$\mathbf{B} = \mathbf{EA}\ ,$$

mentre se $\mathbf{B}$ è la matrice ottenuta mediante una operazione elementare sui vettori colonna di $\mathbf{A}$ e se $\mathbf{F}$ è una matrice del tipo $\mathbf{F}_{ij}$, $\mathbf{F}_i(\alpha)$ o $\mathbf{F}_i(\alpha|j)$, allora risulta

$$\mathbf{B} = \mathbf{AF}\ .$$

Esempio 4.3. Data la matrice

$$\mathbf{A} = \begin{pmatrix} 1 & 4 & 1 \\ 0 & 1 & 3 \\ 4 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix},$$

lo scambio del secondo vettore riga con il quarto vettore riga di $\mathbf{A}$ può essere espresso come

$$\mathbf{E}_{24}\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 4 & 1 \\ 0 & 1 & 3 \\ 4 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 1 \\ 2 & 0 & 2 \\ 4 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix},$$

mentre la somma del primo vettore colonna moltiplicato per 2 con il terzo vettore colonna di **A** può essere espressa come

$$\mathbf{AF}_3(2|1) = \begin{pmatrix} 1 & 4 & 1 \\ 0 & 1 & 3 \\ 4 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 3 \\ 0 & 1 & 3 \\ 4 & 1 & 8 \\ 2 & 0 & 6 \end{pmatrix} . \qquad \square$$

4.2. Matrici a scala

Una matrice $\mathbf{R} = (r_{ij})$ di ordine $(n \times k)$ è detta *matrice a scala* se:
(i) $r_{ij} = 0$ per ogni $i > j$;
(ii) $r_{(i+1)l} = 0$ per ogni $l \leq j$, quando $r_{ij} \neq 0$ e $r_{il} = 0$ per ogni $l < j$.

Dunque una matrice è a scala se gli eventuali vettori riga uguali a $\mathbf{0}^{\mathrm{T}}$ sono gli ultimi e la prima componente non nulla di un qualsiasi vettore riga successivo al primo è in uno dei vettori colonna successivi, rispetto alla prima componente non nulla del precedente vettore riga. Evidentemente, le matrici triangolari superiori sono casi particolari di matrici a scala. Inoltre le prime componenti non nulle di ogni vettore riga diverso da $\mathbf{0}^{\mathrm{T}}$ di **R** sono detti *pivot*.

Esempio 4.4. La matrice triangolare superiore **R** di ordine 3 data da

$$\mathbf{R} = \begin{pmatrix} 6 & 4 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix} ,$$

è una matrice a scala. I pivot sono rispettivamente dati da $r_{11} = 6$, $r_{22} = 2$ e $r_{33} = 2$. Anche la matrice **R** di ordine (5×6) data da

$$\mathbf{R} = \begin{pmatrix} 0 & 3 & 1 & 5 & 0 & 1 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} ,$$

è una matrice a scala. I pivot sono rispettivamente dati da $r_{12} = 3$, $r_{24} = 2$ e $r_{35} = 4$. $\square$

Risultato 4.1. Per ogni matrice $\mathbf{A}$ di ordine $(n \times k)$ valgono le proprietà:
(i) la matrice $\mathbf{A}$ può essere ridotta ad una matrice a scala $\mathbf{R}$ mediante una successione opportuna di operazioni elementari sui vettori riga, ovvero premoltiplicando $\mathbf{A}$ per una successione opportuna di matrici elementari di riga;
(ii) la rappresentazione di $\mathbf{A}$ mediante una matrice a scala $\mathbf{R}$ non è unica;
(iii) tutte le rappresentazioni di $\mathbf{A}$ mediante matrici a scala hanno lo stesso numero di vettori riga diversi da $\mathbf{0}^{\mathrm{T}}$;
(iv) i vettori riga della matrice $\mathbf{A}$ sono linearmente dipendenti se e solo se la rappresentazione di $\mathbf{A}$ mediante una matrice a scala $\mathbf{R}$ contiene almeno un vettore riga pari a $\mathbf{0}^{\mathrm{T}}$;
(v) il numero di vettori di $\mathbf{A}$ (riga o colonna) linearmente indipendenti sono pari al numero di vettori riga diversi da $\mathbf{0}^{\mathrm{T}}$ nella rappresentazione di $\mathbf{A}$ mediante una matrice a scala $\mathbf{R}$.

In base alla proprietà (i) del Risultato 4.1, per la rappresentazione di $\mathbf{A}$ mediante una matrice a scala $\mathbf{R}$ vale l'espressione

$$\mathbf{R} = \mathbf{PA} \, ,$$

dove

$$\mathbf{P} = \mathbf{E}_s \mathbf{E}_{s-1} \ldots \mathbf{E}_1 \, ,$$

mentre $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_s$ sono s opportune matrici del tipo $\mathbf{E}_{ij}$, $\mathbf{E}_i(\alpha)$ o $\mathbf{E}_i(\alpha|j)$. Inoltre, se r rappresenta il numero di vettori di $\mathbf{A}$ (riga o colonna) linearmente indipendenti si dimostra in modo analogo che

$$\mathbf{RQ} = \mathbf{H} \, ,$$

dove $\mathbf{H}$ è una matrice diagonale a blocchi di ordine $(n \times k)$ data da

$$\mathbf{H} = \begin{pmatrix} \mathbf{I}_r & \mathbf{O}_{r \times (k-r)} \\ \mathbf{O}_{(n-r) \times r} & \mathbf{O}_{(n-r) \times (k-r)} \end{pmatrix} ,$$

mentre

$$\mathbf{Q} = \mathbf{F}_1 \mathbf{F}_2 \ldots \mathbf{F}_t \, ,$$

e $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_t$ sono t opportune matrici del tipo $\mathbf{F}_{ij}$, $\mathbf{F}_i(\alpha)$ o $\mathbf{F}_i(\alpha|j)$. Equivalentemente si ha dunque la rappresentazione (non unica) di $\mathbf{A}$

$$\mathbf{PAQ} = \mathbf{H}\,.$$

Per quanto riguarda il calcolo pratico di una matrice a scala **R** a partire dalla matrice **A**, si supponga che il primo vettore colonna di **A** diverso da **0** sia il j-esimo. In questo caso, mediante un eventuale scambio di due vettori riga di **A**, si può ricavare una nuova matrice la cui la prima componente del j-esimo vettore colonna sia non nulla. Inoltre, mediante operazioni elementari di tipo (iii), si può ottenere una nuova matrice nella quale tutte le componenti della j-esima colonna (eccetto la prima) siano nulle. Si può applicare il medesimo procedimento alla matrice costituita dai restanti $(n-1)$ vettori riga e l'algoritmo ha termine quando i vettori riga rimanenti sono tutti pari a $\mathbf{0}^{\mathrm{T}}$ oppure non vi sono ulteriori vettori riga. La matrice infine ottenuta è a scala.

Per quanto riguarda invece il calcolo pratico della matrice **P** è sufficiente considerare la matrice a blocchi $(\mathbf{A}\,|\,\mathbf{I}_n)$ ed effettuare su questa matrice le stesse operazioni elementari necessarie ad ottenere una matrice a scala **R** a partire dalla matrice **A**, dal momento che premoltiplicando $(\mathbf{A}\,|\,\mathbf{I}_n)$ per **P** si ha

$$\mathbf{P}(\mathbf{A}\,|\,\mathbf{I}_n) = (\mathbf{PA}\,|\,\mathbf{P}) = (\mathbf{R}\,|\,\mathbf{P})$$

e quindi la matrice a blocchi risultante dal prodotto contiene nel secondo blocco la matrice **P**.

Analogamente, per il calcolo pratico delle matrice **Q** è sufficiente considerare la matrice a blocchi $(\mathbf{R}^{\mathrm{T}}\,|\,\mathbf{I}_k)^{\mathrm{T}}$ ed effettuare su questa matrice le stesse operazioni elementari necessarie ad ottenere la matrice **H**, dal momento che

$$\begin{pmatrix} \mathbf{R} \\ \mathbf{I}_k \end{pmatrix}\mathbf{Q} = \begin{pmatrix} \mathbf{RQ} \\ \mathbf{Q} \end{pmatrix} = \begin{pmatrix} \mathbf{H} \\ \mathbf{Q} \end{pmatrix},$$

e quindi la matrice a blocchi risultante dal prodotto contiene nel blocco inferiore la matrice **Q**.

Esempio 4.5. Data la matrice **A** di ordine (3×4)

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 2 \\ 2 & 4 & -2 & 0 \\ 2 & 4 & -1 & 2 \end{pmatrix},$$

si può considerare la seguente matrice a blocchi

$$\mathbf{A}_1 = (\mathbf{A}\,|\,\mathbf{I}_3) = \left(\begin{array}{cccc|ccc} 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 2 & 4 & -2 & 0 & 0 & 1 & 0 \\ 2 & 4 & -1 & 2 & 0 & 0 & 1 \end{array}\right).$$

Scambiando il primo e il secondo vettore riga di $\mathbf{A}_1$ si ottiene

$$\mathbf{A}_2 = \mathbf{E}_{12}\mathbf{A}_1 = \left(\begin{array}{cccc|ccc} 2 & 4 & -2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 2 & 4 & -1 & 2 & 0 & 0 & 1 \end{array}\right),$$

mentre sottraendo il primo vettore riga dal terzo vettore riga di $\mathbf{A}_2$ si ha

$$\mathbf{A}_3 = \mathbf{E}_3(-1|1)\mathbf{A}_2 = \left(\begin{array}{cccc|ccc} 2 & 4 & -2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & -1 & 1 \end{array}\right).$$

Inoltre, sottraendo il secondo vettore riga dal terzo vettore riga di $\mathbf{A}_3$ si ottiene

$$\mathbf{A}_4 = \mathbf{E}_3(-1|2)\mathbf{A}_3 = \left(\begin{array}{cccc|ccc} 2 & 4 & -2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right),$$

per cui una rappresentazione di $\mathbf{A}$ mediante una matrice a scala risulta

$$\mathbf{R} = \mathbf{P}\mathbf{A} = \begin{pmatrix} 2 & 4 & -2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

dove

$$\mathbf{P} = \mathbf{E}_3(-1|2)\mathbf{E}_3(-1|1)\mathbf{E}_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & -1 & 1 \end{pmatrix}.$$

Inoltre, si ha

$$\mathbf{R}_1 = \begin{pmatrix} \mathbf{R} \\ \mathbf{I}_4 \end{pmatrix} = \left(\begin{array}{cccc} 2 & 4 & -2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right),$$

per cui scambiando il secondo e il terzo vettore colonna di $\mathbf{R}_1$ si ottiene

$$\mathbf{R}_2 = \mathbf{R}_1 \mathbf{F}_{23} = \left(\begin{array}{cccc} 2 & -2 & 4 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

Sommando il primo vettore colonna al secondo vettore colonna di $\mathbf{R}_2$ si ha

$$\mathbf{R}_3 = \mathbf{R}_2 \mathbf{F}_2(1|1) = \left(\begin{array}{cccc} 2 & 0 & 4 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right),$$

mentre sottraendo il primo vettore colonna moltiplicato per 2 al terzo vettore colonna di $\mathbf{R}_3$ si ottiene

$$\mathbf{R}_4 = \mathbf{R}_3 \mathbf{F}_3(-2|1) = \left(\begin{array}{cccc} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

Sottraendo il secondo vettore colonna moltiplicato per 2 al quarto vettore colonna di $\mathbf{R}_4$ si ottiene

$$\mathbf{R}_5 = \mathbf{R}_4\mathbf{F}_4(-2|2) = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 1 & -2 & -2 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

mentre dividendo per 2 il primo vettore colonna di $\mathbf{R}_5$ si ha

$$\mathbf{R}_6 = \mathbf{R}_5\mathbf{F}_1(1/2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 1/2 & 1 & -2 & -2 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{H} \\ \mathbf{Q} \end{pmatrix} .$$

Dunque si ha

$$\mathbf{Q} = \mathbf{F}_{23}\mathbf{F}_2(1|1)\mathbf{F}_3(-2|1)\mathbf{F}_4(-2|2)\mathbf{F}_1(1/2) \, ,$$

ovvero

$$\mathbf{Q} = \begin{pmatrix} 1/2 & 1 & -2 & -2 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \qquad \square$$

4.3. Rango

Dal momento che una matrice $\mathbf{A}$ di ordine $(n \times k)$ è costituita da n vettori riga di ordine k o da k vettori colonna di ordine n, si definisce *rango* (o *caratteristica*) di $\mathbf{A}$ e si indica con la notazione $\mathrm{r}(\mathbf{A})$, il massimo numero di vettori riga o di vettori colonna linearmente indipendenti in $\mathbf{A}$.

Dalle proprietà (iii)-(iv) del Risultato 4.1 si ha che il rango di una matrice **A** è unico, per cui il massimo numero di vettori riga linearmente indipendenti coincide con il massimo numero di vettori colonna linearmente indipendenti. Inoltre, $\mathrm{r}(\mathbf{A})$ è pari al numero di vettori riga diversi dal vettore $\mathbf{0}^{\mathrm{T}}$ nella rappresentazione di **A** mediante una matrice a scala.

Dalla definizione segue che il rango di una matrice **A** non può superare il numero delle sue righe o delle sue colonne, per cui

$$\mathrm{r}(\mathbf{A}) \leq \min(n, k) \ .$$

Se $\mathrm{r}(\mathbf{A}) = \min(n, k)$ allora **A** assume rango massimo e **A** è detta a *rango pieno*. Ogni matrice quadrata **A** di ordine k è a rango pieno quando $\mathrm{r}(\mathbf{A}) = k$.

Esempio 4.6. Si consideri di nuovo la matrice **A** dell'Esempio 4.5. Dal momento che **A** è di ordine (3×4) deve essere $\mathrm{r}(\mathbf{A}) \leq \min(3, 4) = 3$. Inoltre, tenendo presente la rappresentazione di **A** mediante una matrice a scala, la matrice **R** possiede due vettori riga diversi dal vettore $\mathbf{0}^{\mathrm{T}}$ e quindi $\mathrm{r}(\mathbf{A}) = 2$. Inoltre, dal momento che $\mathrm{r}(\mathbf{A}) < 3$, allora **A** non è a rango pieno. □

Risultato 4.2. Per il rango valgono le proprietà:
(i) $\mathrm{r}(\mathbf{A}^{\mathrm{T}}) = \mathrm{r}(\mathbf{A})$;
(ii) se **A** e **B** sono matrici dello stesso ordine, allora si ha $\mathrm{r}(\mathbf{A} + \mathbf{B}) \leq \mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B})$;
(iii) se **A** e **B** sono matrici conformabili si ha $\mathrm{r}(\mathbf{AB}) \leq \min(\mathrm{r}(\mathbf{A}), \mathrm{r}(\mathbf{B}))$;
(iv) $\mathrm{r}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) = \mathrm{r}(\mathbf{A}\mathbf{A}^{\mathrm{T}}) = \mathrm{r}(\mathbf{A})$;
(v) $\mathrm{r}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{r}(\mathbf{A})\mathrm{r}(\mathbf{B})$.

Dalla proprietà (iv) del Risultato 4.2 si deduce che le matrici $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ e $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ sono a rango pieno se e solo se si verifica rispettivamente $\mathrm{r}(\mathbf{A}) = k$ o $\mathrm{r}(\mathbf{A}) = n$.

Esempio 4.7. Si consideri la matrice **A** di ordine (3×2) data da

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 1 & 2 \end{pmatrix} .$$

Dal momento che i due vettori colonna di $\mathbf{A}$ sono linearmente dipendenti, allora risulta $\mathrm{r}(\mathbf{A}) = 1$. Inoltre si ha

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix} ,$$

mentre

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 5 & 0 & 5 \\ 0 & 0 & 0 \\ 5 & 0 & 5 \end{pmatrix} .$$

Le matrici $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ e $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ hanno entrambe un solo vettore colonna linearmente indipendente e quindi $\mathrm{r}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) = \mathrm{r}(\mathbf{A}\mathbf{A}^{\mathrm{T}}) = 1 = \mathrm{r}(\mathbf{A})$. □

4.4. Determinante

Il *determinante* è un numero reale, indicato con la notazione $\det(\mathbf{A})$, associato a ogni matrice quadrata $\mathbf{A} = (\mathbf{a}_{\bullet 1}\ \mathbf{a}_{\bullet 2}\ \ldots\ \mathbf{a}_{\bullet k})$ di ordine k in modo che siano soddisfatte le proprietà:
(i) $\det(\mathbf{A})$ è una funzione lineare di ogni vettore colonna (risp. riga), nel senso che se il j-esimo vettore colonna (risp. riga) di $\mathbf{A}$ è somma di due vettori colonna (risp. riga), ovvero se $\mathbf{a}_{\bullet j} = \mathbf{b} + \mathbf{c}$, si ha

$$\det(\mathbf{a}_{\bullet 1}\ \ldots\ \mathbf{a}_{\bullet j}\ \ldots\ \mathbf{a}_{\bullet k}) = \det(\mathbf{a}_{\bullet 1}\ \ldots\ \mathbf{b}\ \ldots\ \mathbf{a}_{\bullet k}) + \det(\mathbf{a}_{\bullet 1}\ \ldots\ \mathbf{c}\ \ldots\ \mathbf{a}_{\bullet k})$$

così come se il j-esimo vettore colonna (risp. riga) è moltiplicato per uno scalare α, allora

$$\det(\mathbf{a}_{\bullet 1}\ \ldots\ \alpha\mathbf{a}_{\bullet j}\ \ldots\ \mathbf{a}_{\bullet k}) = \alpha \det(\mathbf{a}_{\bullet 1}\ \ldots\ \mathbf{a}_{\bullet j}\ \ldots\ \mathbf{a}_{\bullet k}) ;$$

(ii) se si scambiano tra loro due vettori colonna (risp. riga) qualunque $\det(\mathbf{A})$ cambia di segno, ovvero

$$\det(\mathbf{a}_{\bullet 1}\ \ldots\ \mathbf{a}_{\bullet j}\ \ldots\ \mathbf{a}_{\bullet l}\ \ldots\ \mathbf{a}_{\bullet k}) = -\det(\mathbf{a}_{\bullet 1}\ \ldots\ \mathbf{a}_{\bullet l}\ \ldots\ \mathbf{a}_{\bullet j}\ \ldots\ \mathbf{a}_{\bullet k}) ;$$

(iii) se $\mathbf{I}$ è una matrice identità di qualsiasi ordine allora

$$\det(\mathbf{I}) = 1 .$$

Da queste tre proprietà il determinante risulta univocamente definito come

$$\det(\mathbf{A}) = \sum (-1)^{\varphi(j_1,j_2,\ldots,j_k)}\, a_{1j_1}a_{2j_2}\ldots a_{kj_k} \;,$$

dove la sommatoria è estesa a tutte le possibili permutazioni $j_1, j_2, \ldots, j_k$ degli indici $1, 2, \ldots, k$ e $\varphi(j_1, j_2, \ldots, j_k)$ è il numero di trasposizioni che si devono eseguire sulla permutazione $j_1, j_2, \ldots, j_k$ per riportarla alla permutazione di base $1, 2, \ldots, k$. Da questa definizione si ottiene che se la matrice è di ordine 1 il determinante coincide con l'unica componente della matrice stessa, mentre se la matrice è di ordine 2 il determinante risulta

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21} \;.$$

I procedimenti per calcolare il determinante di una generica matrice quadrata verranno introdotti nella Sezione 4.5.

Da un punto di vista geometrico, il valore assoluto del determinante può essere interpretato come il volume dell'*iper-parallelogramma* in $\mathbb{R}^k$ individuato dai vettori colonna $\mathbf{a}_{\bullet 1}, \mathbf{a}_{\bullet 2}, \ldots, \mathbf{a}_{\bullet k}$ di $\mathbf{A}$. Questo volume è identico a quello dell'iper-parallelogramma individuato dai vettori riga $\mathbf{a}_{1\bullet}^{\mathrm{T}}, \mathbf{a}_{2\bullet}^{\mathrm{T}}, \ldots, \mathbf{a}_{k\bullet}^{\mathrm{T}}$ di $\mathbf{A}$. Si consideri infatti come caso particolare la matrice quadrata $\mathbf{A}$ di ordine 2 data da

$$\mathbf{A} = (\mathbf{a}_{\bullet 1}\ \mathbf{a}_{\bullet 2}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} .$$

Le lunghezze della base e dell'altezza del parallelogramma individuato da $\mathbf{a}_{\bullet 1}$ e $\mathbf{a}_{\bullet 2}$ risultano rispettivamente $\|\mathbf{a}_{\bullet 1}\|$ e $\|\mathbf{a}_{\bullet 2}\|\sin\theta$, dove θ rappresenta l'arco relativo ad $\mathbf{a}_{\bullet 1}$ e $\mathbf{a}_{\bullet 2}$ $(0 \le \theta \le \pi)$. Tenendo presente l'espressione di $\cos\theta$, l'area del parallelogramma è data da

$$\|\mathbf{a}_{\bullet 1}\| \cdot \|\mathbf{a}_{\bullet 2}\| \sin\theta = \|\mathbf{a}_{\bullet 1}\| \cdot \|\mathbf{a}_{\bullet 2}\|(1 - \cos^2\theta)^{1/2} = |a_{11}a_{22} - a_{12}a_{21}|$$

e quindi

$$|\det(\mathbf{A})| = \|\mathbf{a}_{\bullet 1}\| \cdot \|\mathbf{a}_{\bullet 2}\| \sin\theta \;,$$

ovvero l'area del parallelogramma corrisponde al valore assoluto del determinante.

Esempio 4.8. Si consideri la matrice quadrata $\mathbf{A}$ di ordine 2 data da

$$\mathbf{A} = (\mathbf{a}_{\bullet 1}\ \mathbf{a}_{\bullet 2}) = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$$

per cui $\det(\mathbf{A}) = 5$. Si ha $\|\mathbf{a}_{\bullet 1}\| = \sqrt{10}$ e $\|\mathbf{a}_{\bullet 2}\| = \sqrt{5}$ mentre $\langle \mathbf{a}_{\bullet 1}, \mathbf{a}_{\bullet 2} \rangle = 5$. Inoltre, se θ rappresenta l'arco relativo ad $\mathbf{a}_{\bullet 1}$ e $\mathbf{a}_{\bullet 2}$, allora si ha $\cos\theta = \sqrt{2}/2$ e quindi $\sin\theta = \sqrt{2}/2$. Dunque, l'area del parallelogramma individuato da $\mathbf{a}_{\bullet 1}$ e $\mathbf{a}_{\bullet 2}$ (vedi Figura 4.1) è pari a $\|\mathbf{a}_{\bullet 1}\| \cdot \|\mathbf{a}_{\bullet 2}\| \sin\theta = 5$ e coincide con $|\det(\mathbf{A})|$. □

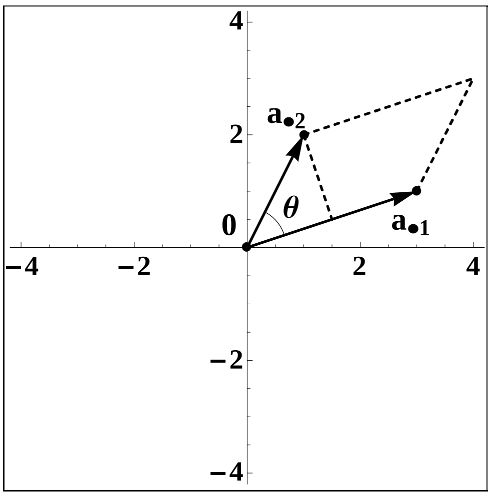

Figura 4.1

Risultato 4.3. Se $\mathbf{A}$ è una matrice quadrata di ordine k, per il determinante valgono le proprietà:
(i) se la matrice ha due vettori riga (risp. colonna) uguali il determinante è nullo;
(ii) se un vettore riga (risp. colonna) è pari al vettore $\mathbf{0}^{\mathrm{T}}$ (risp. $\mathbf{0}$), il determinante è nullo;
(iii) $\det(\mathbf{A}^{\mathrm{T}}) = \det(\mathbf{A})$;
(iv) $\det(\alpha\mathbf{A}) = \alpha^k \det(\mathbf{A})$;
(v) se $\mathbf{A} = \mathrm{dg}(a_1, a_2, \ldots, a_k)$ allora si ha

$$\det(\mathbf{A}) = \prod_{i=1}^{k} a_i \ ;$$

(vi) se $\mathbf{A} = (a_{ij})$ è una matrice triangolare allora si ha

$$\det(\mathbf{A}) = \det(\mathrm{dg}(\mathbf{A})) = \prod_{i=1}^{k} a_{ii} \ ;$$

(vii) se $\mathbf{B}$ è una matrice quadrata di ordine k, allora

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}) \ ;$$

(viii) se $\mathbf{B}$ è una matrice quadrata di ordine q, allora

$$\det(\mathbf{A} \otimes \mathbf{B}) = \det(\mathbf{A})^k \det(\mathbf{B})^q \ .$$

Una matrice quadrata $\mathbf{A}$ è detta *non singolare* se $\det(\mathbf{A}) \neq 0$. Ogni matrice quadrata $\mathbf{A}$ non singolare è a rango pieno. Inoltre, il prodotto con matrici non singolari non altera il rango, ovvero se $\mathbf{A}$ e $\mathbf{B}$ sono matrici opportunamente conformabili rispetto al prodotto, allora

$$\mathrm{r}(\mathbf{AB}) = \mathrm{r}(\mathbf{B})$$

e

$$\mathrm{r}(\mathbf{BA}) = \mathrm{r}(\mathbf{B}) \ ,$$

se $\det(\mathbf{A}) \neq 0$.

Esempio 4.9. Si considerino le matrici quadrate $\mathbf{A}$ e $\mathbf{B}$ di ordine 2 definite come

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} , \mathbf{B} = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix} ,$$

per cui $\det(\mathbf{A}) = \det(\mathrm{dg}(\mathbf{A})) = 2$ e $\det(\mathbf{B}) = 4$. Inoltre si ha

$$\mathbf{AB} = \begin{pmatrix} 5 & 3 \\ 4 & 4 \end{pmatrix} ,$$

da cui $\det(\mathbf{AB}) = 8 = \det(\mathbf{A})\det(\mathbf{B})$. Dal momento che $\det(\mathbf{A}) \neq 0$, risulta $\mathrm{r}(\mathbf{AB}) = \mathrm{r}(\mathbf{B}) = 2$. □

Esempio 4.10. Si considerino le matrici quadrate **A** e **B** di ordine 2 definite come

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} , \mathbf{B} = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix} ,$$

per cui $\det(\mathbf{A}) = 0$ e quindi la matrice **A** risulta singolare. Inoltre si ha

$$\mathbf{AB} = \begin{pmatrix} 5 & 3 \\ 0 & 0 \end{pmatrix} ,$$

per cui $\mathrm{r}(\mathbf{AB}) = 1 \neq \mathrm{r}(\mathbf{B}) = 2.$ □

4.5. Calcolo del determinante

In generale il determinante di una matrice quadrata può essere calcolato per mezzo della *regola di Laplace*. Data una matrice quadrata $\mathbf{A} = (a_{ij})$ di ordine k, si dice *minore complementare* della componente a_{ij} la matrice quadrata $\mathbf{A}_{(ij)}$ di ordine $(k-1)$ ottenuta eliminando l'i-esimo vettore riga e il j-esimo vettore colonna di **A**. Il *complemento algebrico* (o *cofattore*) di una componente a_{ij} è lo scalare

$$A_{ij} = (-1)^{i+j} \det(\mathbf{A}_{(ij)}) ,$$

ovvero il determinante del minore complementare cambiato di segno se $(i+j)$ è dispari.

La regola di Laplace afferma che il valore di un determinante è uguale alla somma dei prodotti delle componenti di un qualunque vettore riga o colonna per i rispettivi complementi algebrici, ovvero

$$\det(\mathbf{A}) = \sum_{i=1}^{k} a_{ij} A_{ij} = \sum_{j=1}^{k} a_{ij} A_{ij} .$$

Esempio 4.11. Data la matrice quadrata **A** di ordine 3

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 4 & -2 \\ 2 & 4 & -1 \end{pmatrix} ,$$

si vuole sviluppare il determinante mediante la regola di Laplace applicata al primo vettore riga. In questo caso si ha

$$A_{11} = (-1)^2 \det(\mathbf{A}_{(11)}) = \det\begin{pmatrix} 4 & -2 \\ 4 & -1 \end{pmatrix} = 4\,,$$

$$A_{12} = (-1)^3 \det(\mathbf{A}_{(12)}) = -\det\begin{pmatrix} 2 & -2 \\ 2 & -1 \end{pmatrix} = -2\,,$$

$$A_{13} = (-1)^4 \det(\mathbf{A}_{(13)}) = \det\begin{pmatrix} 2 & 4 \\ 2 & 4 \end{pmatrix} = 0\,,$$

da cui

$$\det(\mathbf{A}) = 0 \times 4 + 1 \times (-2) + 1 \times 0 = -2\,.$$

Se invece si vuole sviluppare il determinante mediante la regola di Laplace applicata al secondo vettore colonna, si ha inoltre

$$A_{22} = (-1)^4 \det(\mathbf{A}_{(22)}) = \det\begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} = -2\,,$$

$$A_{32} = (-1)^5 \det(\mathbf{A}_{(32)}) = -\det\begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} = 2\,,$$

da cui di nuovo risulta

$$\det(\mathbf{A}) = 1 \times (-2) + 4 \times (-2) + 4 \times 2 = -2\,. \qquad \square$$

Per quanto riguarda il calcolo pratico del determinante di $\mathbf{A}$, è conveniente adoperare la rappresentazione di $\mathbf{A}$ mediante una matrice a scala. In effetti, si tenga presente che vale l'espressione

$$\mathbf{R} = \mathbf{E}_s\mathbf{E}_{s-1}\ldots\mathbf{E}_1\mathbf{A}\,,$$

dove $\mathbf{E}_s\mathbf{E}_{s-1}\ldots\mathbf{E}_1$ sono s opportune matrici del tipo $\mathbf{E}_{ij}$, $\mathbf{E}_i(\alpha)$ o $\mathbf{E}_i(\alpha|j)$ (vedi Sezione 4.2). Per la proprietà (vii) del Risultato 4.3 si ha

$$\det(\mathbf{R}) = \det(\mathbf{E}_s)\det(\mathbf{E}_{s-1})\ldots\det(\mathbf{E}_1)\det(\mathbf{A})\,,$$

ovvero

$$\det(\mathbf{A}) = \det(\mathbf{E}_s)^{-1}\det(\mathbf{E}_{s-1})^{-1}\ldots\det(\mathbf{E}_1)^{-1}\det(\mathbf{R}) \ ,$$

Da questa relazione è immediato ricavare $\det(\mathbf{A})$ tenendo presente che $\mathbf{R} = (r_{ij})$ è una matrice triangolare superiore quando $\mathbf{A}$ è quadrata. Quindi dalla proprietà (vi) del Risultato 4.3 si ha

$$\det(\mathbf{R}) = \prod_{i=1}^{k} r_{ii} \ ,$$

mentre risulta $\det(\mathbf{E}_{ij}) = -1$, $\det(\mathbf{E}_i(\alpha)) = \alpha$ e $\det(\mathbf{E}_i(\alpha|j)) = 1$ dalle proprietà (i)-(iii) che definiscono il determinante.

Esempio 4.12. Data la matrice $\mathbf{A}$ dell'Esempio 4.11, scambiando il primo e il secondo vettore riga di $\mathbf{A}$ si ha

$$\mathbf{A}_1 = \mathbf{E}_{12}\mathbf{A} = \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 2 & 4 & -1 \end{pmatrix} ,$$

mentre sottraendo il primo vettore riga dal terzo vettore riga di $\mathbf{A}_1$ si ottiene

$$\mathbf{A}_2 = \mathbf{E}_3(-1|1)\mathbf{A}_1 = \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} ,$$

per cui

$$\mathbf{R} = \mathbf{A}_2 = \mathbf{E}_3(-1|1)\mathbf{E}_{12}\mathbf{A} \ .$$

Dal momento che $\det(\mathbf{R}) = 2$, $\det(\mathbf{E}_{12}) = -1$ e $\det(\mathbf{E}_3(-1|1)) = 1$, allora $\det(\mathbf{A}) = -2$, che conferma il risultato ottenuto nell'Esempio 4.11. □

Capitolo 5

Matrici inverse

5.1. Definizioni e notazioni

Data una matrice quadrata $\mathbf{A}$ di ordine k, è detta *inversa* di $\mathbf{A}$ e si indica con $\mathbf{A}^{-1}$ una matrice dello stesso ordine tale che

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_k \ .$$

Se A_{ij} rappresenta il complemento algebrico della componente a_{ij} di $\mathbf{A}$, si può dimostrare che

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\,\mathbf{A}' \ ,$$

dove $\mathbf{A}'$ rappresenta la *matrice aggiunta*, ovvero la trasposta della matrice dei complementi algebrici data da

$$\mathbf{A}' = \begin{pmatrix} A_{11} & A_{21} & \dots & A_{k1} \\ A_{12} & A_{22} & \dots & A_{k2} \\ \vdots & \vdots & & \vdots \\ A_{1k} & A_{2k} & \dots & A_{kk} \end{pmatrix} .$$

Dunque, $\mathbf{A}^{-1}$ esiste se e solo se $\det(\mathbf{A}) \neq 0$, ovvero quando $\mathbf{A}$ è non singolare. Inoltre, se $\mathbf{A}^{-1}$ esiste allora è unica. Infine, se $\mathbf{A}$ è una matrice quadrata di ordine 2 non singolare, allora risulta

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} .$$

Esempio 5.1. Se si considera la matrice quadrata di ordine 2

$$\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} ,$$

allora dal momento che det$(\mathbf{A}) = 1$, risulta

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} .$$

In effetti si ha

$$\mathbf{A}\mathbf{A}^{-1} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2$$

e

$$\mathbf{A}^{-1}\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2 . \qquad \square$$

Risultato 5.1. Se **A** è una matrice quadrata non singolare di ordine k e α è uno scalare, per la matrice inversa valgono le proprietà:
(i) $\mathbf{I}_k^{-1} = \mathbf{I}_k$;
(ii) det$(\mathbf{A}^{-1}) =$ det$(\mathbf{A})^{-1}$;
(iii) $(\mathbf{A}^{\mathrm{T}})^{-1} = (\mathbf{A}^{-1})^{\mathrm{T}}$;
(iv) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$;
(v) $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$;
(vi) dg$(a_1, a_2, \ldots, a_k)^{-1} =$ dg$(a_1^{-1}, a_2^{-1}, \ldots, a_k^{-1})$.

Esempio 5.2. Se si considera la matrice quadrata di ordine 2 data da

$$\mathbf{A} = \begin{pmatrix} 6 & 4 \\ 2 & 2 \end{pmatrix} ,$$

allora si ha

$$\mathbf{A}^{-1} = \begin{pmatrix} 1/2 & -1 \\ -1/2 & 3/2 \end{pmatrix} .$$

Di conseguenza risulta

$$\text{det}(\mathbf{A}^{-1}) = 1/4 = \text{det}(\mathbf{A})^{-1} .$$

Inoltre si ha

$$(\mathbf{A}^{\mathrm{T}})^{-1} = \begin{pmatrix} 1/2 & -1/2 \\ -1 & 3/2 \end{pmatrix} = (\mathbf{A}^{-1})^{\mathrm{T}} . \qquad \square$$

Esempio 5.3. Se si considera la matrice diagonale di ordine 2 data da

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} ,$$

allora si ha

$$\mathbf{A}^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} . \qquad \square$$

Risultato 5.2. Per la matrice inversa valgono le ulteriori proprietà:
(i) se **A** e **B** sono due matrici quadrate non singolari dello stesso ordine, allora

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} ;$$

(ii) se **A** e **B** sono due matrici quadrate non singolari rispettivamente di ordine k e q, allora

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} ;$$

(iii) se **A** è una matrice quadrata di ordine k e **B**, **C** e **D** sono matrici di ordine rispettivamente pari a $(k \times n)$, $(n \times n)$ e $(n \times k)$

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1} ;$$

(iv) se **A** è una matrice quadrata di ordine k e **B** e **C** sono due matrici di ordine rispettivamente pari a $(k \times n)$ ed $(n \times k)$

$$\det(\mathbf{A} + \mathbf{BC}) = \frac{\det(\mathbf{I}_k + \mathbf{A}^{-1}\mathbf{BC})}{\det(\mathbf{A})} = \frac{\det(\mathbf{I}_n + \mathbf{CA}^{-1}\mathbf{B})}{\det(\mathbf{A})} .$$

Esempio 5.4. Se si considerano le matrici quadrate di ordine 2

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

allora si ha

$$\mathbf{AB} = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix},$$

da cui

$$(\mathbf{AB})^{-1} = \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix}.$$

Tuttavia risulta anche che

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \mathbf{B}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

da cui

$$\mathbf{B}^{-1}\mathbf{A}^{-1} = \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix} = (\mathbf{AB})^{-1}. \qquad \square$$

Esempio 5.5. Se si considerano le matrici quadrate di ordine 2

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \mathbf{C} = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix},$$

allora risulta

$$\begin{aligned} \mathbf{I}_2 + \mathbf{A}^{-1}\mathbf{BC} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -1 \\ 6 & 4 \end{pmatrix}, \end{aligned}$$

per cui

$$\det(\mathbf{A} + \mathbf{BC}) = \frac{\det(\mathbf{I}_2 + \mathbf{A}^{-1}\mathbf{BC})}{\det(\mathbf{A})} = 6. \qquad \square$$

5.2. Inverse di alcune matrici particolari

Se $\mathbf{A}$ è una matrice quadrata a blocchi di ordine k

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

tale che $\mathbf{A}_{11}$ e $\mathbf{A}_{22}$ sono rispettivamente matrici quadrate di ordine h e $(k-h)$ non singolari, allora risulta

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}'_{11} & \mathbf{A}'_{12} \\ \mathbf{A}'_{21} & \mathbf{A}'_{22} \end{pmatrix} ,$$

dove

$$\mathbf{A}'_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} ,$$

$$\mathbf{A}'_{12} = -\mathbf{A}'_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}'_{22} ,$$

$$\mathbf{A}'_{21} = -\mathbf{A}'_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} = -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}'_{11} ,$$

$$\mathbf{A}'_{22} = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} .$$

Inoltre se $\mathbf{A}_{11}$ e $\mathbf{A}_{22}$ sono rispettivamente matrici quadrate di ordine h e $(k-h)$ non singolari, allora si ha

$$\det\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \det(\mathbf{A}_{11})\det(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}) ,$$

e

$$\det\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \det(\mathbf{A}_{22})\det(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) ,$$

e quindi se $\mathbf{A}_{12} = \mathbf{O}$ oppure se $\mathbf{A}_{21} = \mathbf{O}$

$$\det\begin{pmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \det\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{O} & \mathbf{A}_{22} \end{pmatrix} = \det(\mathbf{A}_{11})\det(\mathbf{A}_{22}) .$$

Esempio 5.6. Se si considera la matrice quadrata a blocchi di ordine 3

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \left(\begin{array}{cc|c} 2 & 1 & 0 \\ 1 & 1 & 1 \\ \hline 1 & 0 & 1 \end{array}\right) ,$$

allora si ha

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}'_{11} & \mathbf{A}'_{12} \\ \mathbf{A}'_{21} & \mathbf{A}'_{22} \end{pmatrix} = \left(\begin{array}{cc|c} 1/2 & -1/2 & 1/2 \\ 0 & 1 & -1 \\ \hline -1/2 & 1/2 & 1/2 \end{array}\right) .$$

Inoltre, dal momento che $\det(\mathbf{A}_{11}) = 1$ e $\det(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}) = 2$, allora risulta $\det(\mathbf{A}) = 2$. □

Data una matrice ortogonale $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1\ \boldsymbol{\gamma}_2\ \ldots\ \boldsymbol{\gamma}_k)$ di ordine k si ha

$$\mathbf{\Gamma}^{\mathrm{T}}\mathbf{\Gamma} = (\boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\gamma}_j) = \mathbf{I}_k ,$$

in quanto $\boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\gamma}_j$ vale 1 se $i = j$ e vale 0 se $i \neq j$. In modo analogo si può dimostrare che $\mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{I}_k$ e quindi

$$\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}^{\mathrm{T}} .$$

Adoperando questa relazione è immediato ricavare inoltre che

$$\det(\mathbf{\Gamma}) = \pm 1 .$$

Esempio 5.7. Se si considera la matrice ortogonale di ordine 2

$$\mathbf{\Gamma} = \begin{pmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{pmatrix} ,$$

allora si ha

$$\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}^{\mathrm{T}} = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix} ,$$

mentre risulta $\det(\mathbf{\Gamma}) = 1$. Analogamente, se si considera la matrice ortogonale di ordine 2

$$\mathbf{\Gamma} = \begin{pmatrix} 3/5 & 4/5 \\ 4/5 & -3/5 \end{pmatrix} ,$$

allora si ha

$$\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}^{\mathrm{T}} = \begin{pmatrix} 3/5 & 4/5 \\ 4/5 & -3/5 \end{pmatrix},$$

mentre det($\mathbf{\Gamma}$) $= -1$. □

5.3. Calcolo della matrice inversa

Il procedimento basato sull'espressione della matrice inversa dato nella Sezione 5.1 può risultare piuttosto laborioso e per il calcolo pratico di $\mathbf{A}^{-1}$ è conveniente adottare la rappresentazione di $\mathbf{A}$ introdotta nella Sezione 4.2, ovvero

$$\mathbf{PAQ} = \mathbf{I}_k ,$$

dove si è tenuto presente che quando $r(\mathbf{A}) = k$ risulta $\mathbf{H} = \mathbf{I}_k$.

Dal momento che per $n = k$ si ha $\mathbf{F}_{ij} = \mathbf{E}_{ij}$, $\mathbf{F}_i(\alpha) = \mathbf{E}_i(\alpha)$ e $\mathbf{F}_i(\alpha|j) = \mathbf{E}_i(\alpha|j)^{\mathrm{T}}$ (vedi Sezione 4.1), allora

$$\mathbf{Q} = \mathbf{F}_1\mathbf{F}_2\ldots\mathbf{F}_t = \mathbf{E}_1\mathbf{E}_2\ldots\mathbf{E}_t$$

è un prodotto di t opportune matrici del tipo $\mathbf{E}_{ij}$, $\mathbf{E}_i(\alpha)$ o $\mathbf{E}_i(\alpha|j)$ (vedi Sezione 4.2). Dunque per la proprietà (i) del Risultato 5.2 si ha

$$\mathbf{Q}^{-1} = \mathbf{E}_t^{-1}\mathbf{E}_{t-1}^{-1}\ldots\mathbf{E}_1^{-1} .$$

Inoltre, adoperando la proprietà (iii) del Risultato 5.2 si ottiene che

$$\mathbf{E}_{ij}^{-1} = \mathbf{E}_{ij} ,$$

$$\mathbf{E}_i(\alpha)^{-1} = \mathbf{E}_i(\alpha^{-1}) ,$$

$$\mathbf{E}_i(\alpha|j)^{-1} = \mathbf{E}_i(-\alpha|j) ,$$

e allora $\mathbf{Q}^{-1}$ esiste. Quindi, premoltiplicando per $\mathbf{Q}$ e postmoltiplicando per $\mathbf{Q}^{-1}$ la rappresentazione di $\mathbf{A}$ si ha

$$\mathbf{QPAQQ}^{-1} = \mathbf{QQ}^{-1} ,$$

ovvero

$$\mathbf{QPA} = \mathbf{I}_k \ ,$$

e per definizione risulta $\mathbf{A}^{-1} = \mathbf{QP}$.

Per quanto riguarda il calcolo della matrice $\mathbf{A}^{-1}$ è dunque sufficiente considerare la matrice a blocchi $(\mathbf{A} \,|\, \mathbf{I}_k)$ ed effettuare su questa matrice le stesse operazioni elementari sui vettori riga necessarie ad ottenere la matrice identità da $\mathbf{A}$, dal momento che premoltiplicando $(\mathbf{A} \,|\, \mathbf{I}_k)$ per $\mathbf{QP}$ si ha

$$\mathbf{QP}(\mathbf{A} \,|\, \mathbf{I}_k) = (\mathbf{I}_k \,|\, \mathbf{QP}) = (\mathbf{I}_k \,|\, \mathbf{A}^{-1})$$

e quindi la matrice a blocchi risultante dal prodotto contiene nel secondo blocco la matrice $\mathbf{A}^{-1}$.

Esempio 5.8. Si consideri la matrice quadrata di ordine 3

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 4 \end{pmatrix} .$$

In questo caso, formando la matrice a blocchi si ha

$$\mathbf{A}_1 = (\mathbf{A} \,|\, \mathbf{I}_3) = \left(\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 2 & 1 & 4 & 0 & 0 & 1 \end{array}\right) ,$$

da cui scambiando il primo e il secondo vettore riga di $\mathbf{A}_1$ risulta

$$\mathbf{A}_2 = \mathbf{E}_{12}\mathbf{A}_1 = \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 2 & 1 & 4 & 0 & 0 & 1 \end{array}\right) ,$$

mentre sottraendo il primo vettore riga moltiplicato per 2 al terzo vettore riga di $\mathbf{A}_2$ si ottiene

$$\mathbf{A}_3 = \mathbf{E}_3(-2|1)\mathbf{A}_2 = \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & -2 & 1 \end{array}\right) .$$

Sommando il secondo vettore riga al terzo vettore riga di $\mathbf{A}_3$ si ha

$$\mathbf{A}_4 = \mathbf{E}_3(1|2)\mathbf{A}_3 = \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right) ,$$

mentre sottraendo il terzo vettore riga al secondo vettore riga di $\mathbf{A}_4$ si ottiene

$$\mathbf{A}_5 = \mathbf{E}_2(-1|3)\mathbf{A}_4 = \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right) .$$

Sottraendo il terzo vettore riga moltiplicato per 2 al primo vettore riga di $\mathbf{A}_5$ risulta

$$\mathbf{A}_6 = \mathbf{E}_1(-2|3)\mathbf{A}_5 = \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & -2 & 5 & -2 \\ 0 & 1 & 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right) ,$$

mentre sottraendo il secondo vettore riga al primo vettore riga di $\mathbf{A}_6$ si ha

$$\mathbf{A}_7 = \mathbf{E}_1(-1|2)\mathbf{A}_6 = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 3 & -1 \\ 0 & 1 & 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right) = (\mathbf{I}_3 \,|\, \mathbf{A}^{-1}) .$$

Dunque la matrice inversa di $\mathbf{A}$ è data da

$$\mathbf{A}^{-1} = \begin{pmatrix} -2 & 3 & -1 \\ 0 & 2 & -1 \\ 1 & -2 & 1 \end{pmatrix} . \qquad \square$$

5.4. Matrici inverse generalizzate

È utile estendere il concetto di matrice inversa anche a matrici non quadrate. In questo caso, data la matrice $\mathbf{A}$ di ordine $(n \times k)$, la matrice $\mathbf{A}^{-}$ di ordine $(k \times n)$ è detta *inversa generalizzata* se

$$\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A} .$$

La matrice $\mathbf{A}^{-}$ esiste sempre e non è unica.

Risultato 5.3. Se $\mathbf{A}$ è una matrice di ordine $(n \times k)$ tale che $r = r(\mathbf{A})$, per la matrice inversa generalizzata $\mathbf{A}^-$ valgono le proprietà:
(i) $r \leq r(\mathbf{A}^-) \leq \min(n, k)$;
(ii) $\mathbf{A}^-\mathbf{A}$ e $\mathbf{A}\mathbf{A}^-$ sono matrici idempotenti di rango r;
(iii) $(\mathbf{I}_k - \mathbf{A}^-\mathbf{A})$ e $(\mathbf{I}_n - \mathbf{A}\mathbf{A}^-)$ sono idempotenti rispettivamente di rango $(k - r)$ e $(n - r)$;
(iv) se $\mathbf{A}$ è una matrice idempotente allora si ha $\mathbf{A}^- = \mathbf{A}$.

Fra tutte le possibili inverse generalizzate è conveniente considerare quella che si ottiene partendo dalla rappresentazione

$$\mathbf{PAQ} = \mathbf{H}$$

introdotta nella Sezione 4.2. Innanzitutto si noti che la matrice $\mathbf{H}^{\mathrm{T}}$ è una inversa generalizzata di $\mathbf{H}$, essendo

$$\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{H} = \mathbf{H} \, .$$

Inoltre, $\mathbf{P}^{-1}$ e $\mathbf{Q}^{-1}$ sono definite in quanto inverse di prodotti di matrici elementari (vedi Sezione 5.3). Dal momento che premoltiplicando per $\mathbf{P}^{-1}$ e postmoltiplicando per $\mathbf{Q}^{-1}$ la rappresentazione introdotta nella Sezione 4.2 si ottiene

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{H}\mathbf{Q}^{-1} \, ,$$

una matrice inversa generalizzata di $\mathbf{A}$ è data da

$$\mathbf{A}^- = \mathbf{Q}\mathbf{H}^{\mathrm{T}}\mathbf{P} \, .$$

In effetti risulta

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = (\mathbf{P}^{-1}\mathbf{H}\mathbf{Q}^{-1})(\mathbf{Q}\mathbf{H}^{\mathrm{T}}\mathbf{P})(\mathbf{P}^{-1}\mathbf{H}\mathbf{Q}^{-1}) = \mathbf{P}^{-1}\mathbf{H}\mathbf{Q}^{-1} = \mathbf{A} \, .$$

Per il calcolo pratico della matrice $\mathbf{A}^- = \mathbf{Q}\mathbf{H}^{\mathrm{T}}\mathbf{P}$, è sufficiente quindi ottenere le matrici $\mathbf{P}$ e $\mathbf{Q}$ come esemplificato nella Sezione 4.2.

Esempio 5.9. Si consideri la matrice $\mathbf{A}$ di ordine (3×2) data da

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

In questo caso, formando la matrice a blocchi si ha

$$\mathbf{A}_1 = (\mathbf{A}\,|\,\mathbf{I}_3) = \left(\begin{array}{cc|ccc} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{array}\right) ,$$

da cui sottraendo il primo vettore riga dal secondo vettore riga di $\mathbf{A}_1$ si ottiene

$$\mathbf{A}_2 = \mathbf{E}_2(-1|1)\mathbf{A}_1 = \left(\begin{array}{cc|rcc} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{array}\right) ,$$

mentre sottraendo il secondo vettore riga dal terzo vettore riga di $\mathbf{A}_2$ risulta

$$\mathbf{A}_3 = \mathbf{E}_3(-1|2)\mathbf{A}_2 = \left(\begin{array}{cc|rrc} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 \end{array}\right) = (\mathbf{H}\,|\,\mathbf{P}) .$$

Dalla precedente rappresentazione è inoltre evidente che $\mathbf{Q} = \mathbf{I}_2$ essendo $\mathbf{R} = \mathbf{H}$. In definitiva si ha

$$\begin{aligned} \mathbf{A}^- = \mathbf{Q}\mathbf{H}^\mathrm{T}\mathbf{P} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \end{pmatrix} . \end{aligned}$$

In effetti $\mathbf{A}^-$ è una inversa generalizzata di $\mathbf{A}$ in quanto è immediato verificare che $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$. $\square$

Assumendo un ulteriore insieme di proprietà è possibile ottenere una matrice inversa generalizzata definita in modo univoco. Data la matrice $\mathbf{A}$ di ordine $(n \times k)$ tale che $r = r(\mathbf{A})$, la matrice $\mathbf{A}^+$ di ordine $(k \times n)$ è detta *inversa di Moore-Penrose* se soddisfa alle condizioni:
(i) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$;
(ii) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$;
(iii) $(\mathbf{A}\mathbf{A}^+)^\mathrm{T} = \mathbf{A}\mathbf{A}^+$;
(iv) $(\mathbf{A}^+\mathbf{A})^\mathrm{T} = \mathbf{A}^+\mathbf{A}$.

Risultato 5.4. Se $\mathbf{A}$ è una matrice di ordine $(n \times k)$, per la matrice inversa di Moore-Penrose $\mathbf{A}^{+}$ valgono le proprietà:
(i) $\mathbf{A}^{+}$ esiste sempre ed è unica;
(ii) se $r(\mathbf{A}) = k$ allora $\mathbf{A}^{+} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}$ e $\mathbf{A}^{+}\mathbf{A} = \mathbf{I}_k$;
(iii) se $r(\mathbf{A}) = n$ allora $\mathbf{A}^{+} = \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}$ e $\mathbf{A}\mathbf{A}^{+} = \mathbf{I}_n$.

Le proprietà (ii) e (iii) del Risultato 5.4 permettono il calcolo pratico della matrice inversa di Moore-Penrose quando $r(\mathbf{A}) = \min(n, k)$. In caso contrario si può applicare il risultato che verrà presentato nella Sezione 8.3.

Esempio 5.10. Si consideri di nuovo la matrice $\mathbf{A}$ dell'Esempio 5.9. Dal momento che $r(\mathbf{A}) = 2$, allora

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

da cui si ha

$$(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1} = \begin{pmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{pmatrix},$$

per cui

$$\mathbf{A}^{+} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 2/3 & 1/3 & -1/3 \\ -1/3 & 1/3 & 2/3 \end{pmatrix}.$$

Si noti che $\mathbf{A}^{+}\mathbf{A} = \mathbf{I}_2$. □

Capitolo 6

Applicazioni lineari

6.1. Definizioni e notazioni

Si dice *applicazione* da $\mathbb{R}^k$ a $\mathbb{R}^n$ ogni corrispondenza che associa ad un vettore $\mathbf{x} \in \mathbb{R}^k$ uno ed un solo vettore $\mathbf{y} \in \mathbb{R}^n$. Una applicazione $\mathbf{f}$ da $\mathbb{R}^k$ a $\mathbb{R}^n$ viene indicata con la notazione $\mathbf{f} : \mathbb{R}^k \to \mathbb{R}^n$. Il corrispondente $\mathbf{y} \in \mathbb{R}^n$ di un vettore $\mathbf{x} \in \mathbb{R}^k$ si dice *immagine* di $\mathbf{x}$ tramite $\mathbf{f}$ e si scrive

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) \ .$$

L'insieme $D \subseteq \mathbb{R}^k$ dei vettori sui quali è definita $\mathbf{f}$ è detto *dominio* di $\mathbf{f}$, mentre l'insieme $C \subseteq \mathbb{R}^n$ delle corrispondenti immagini è detto *codominio* di $\mathbf{f}$.

Una applicazione da $\mathbb{R}^k$ a $\mathbb{R}^n$ è un vettore di ordine n di *funzioni di variabile vettoriale* di ordine k, ovvero l'i-esima componente del vettore $\mathbf{y} = (y_1 \ y_2 \ \dots \ y_n)^{\mathrm{T}}$ è una funzione del vettore $\mathbf{x} = (x_1 \ x_2 \ \dots \ x_k)^{\mathrm{T}}$ del tipo

$$y_i = f_i(x_1, x_2, \dots, x_k) = f_i(\mathbf{x}) \ ,$$

dove $f_i : \mathbb{R}^k \to \mathbb{R}$. Quindi si ha

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix} \ .$$

Se ogni vettore $\mathbf{y} \in C$ è l'immagine di un solo vettore $\mathbf{x} \in D$, l'applicazione $\mathbf{f}$ è detta *biunivoca* ed esiste l'*immagine inversa* di $\mathbf{y}$ tramite $\mathbf{f}$ che viene denotata come

$$\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y}) \ .$$

Una applicazione $\mathbf{f} : \mathbb{R}^k \to \mathbb{R}^n$ è detta *lineare* se:
(i) $\mathbf{f}(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{f}(\mathbf{x}_1) + \mathbf{f}(\mathbf{x}_2)$ per ogni $\mathbf{x}_1, \mathbf{x}_2 \in D$;
(ii) $\mathbf{f}(\alpha\mathbf{x}) = \alpha\mathbf{f}(\mathbf{x})$ per ogni $\mathbf{x} \in D$ e $\alpha \in \mathbb{R}$.

Si può dimostrare che per ogni applicazione lineare esiste una e una sola matrice $\mathbf{A} = (a_{ij})$ di ordine $(n \times k)$ tale che l'applicazione stessa può essere rappresentata nella forma

$$\mathbf{y} = \mathbf{A}\mathbf{x} \, .$$

Esplicitando la precedente relazione, l'i-esima componente di $\mathbf{y}$ risulta

$$y_i = \mathbf{a}_{i\bullet}^{\mathrm{T}}\mathbf{x} = \sum_{j=1}^{k} a_{ij}x_j \, ,$$

ovvero ogni applicazione lineare da $\mathbb{R}^k$ a $\mathbb{R}^n$ è costituita da n combinazioni lineari delle componenti di $\mathbf{x}$, i cui coefficienti sono dati dalle componenti dei vettori riga $\mathbf{a}_{1\bullet}^{\mathrm{T}}, \mathbf{a}_{2\bullet}^{\mathrm{T}}, \ldots, \mathbf{a}_{n\bullet}^{\mathrm{T}}$ della matrice $\mathbf{A}$.

Nell'ambito delle applicazioni lineari sono di particolare importanza le applicazioni da $\mathbb{R}^k$ a $\mathbb{R}^k$. In questo caso la matrice $\mathbf{A}$ associata ad ogni applicazione lineare è evidentemente una matrice quadrata di ordine k. Una applicazione lineare da $\mathbb{R}^k$ a $\mathbb{R}^k$ è detta *non singolare* se la matrice $\mathbf{A}$ è non singolare. Se $\mathbf{A}$ è non singolare l'applicazione è *biunivoca*. Infatti, esistendo $\mathbf{A}^{-1}$, esiste l'immagine inversa $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$, che risulta ancora lineare.

6.2. Alcune applicazioni lineari particolari

Esistono alcune applicazioni lineari da $\mathbb{R}^k$ a $\mathbb{R}^k$ di particolare interesse, anche dette *trasformazioni*, le cui proprietà sono ovviamente determinate dalla conformazione della matrice $\mathbf{A}$.

La *trasformazione di scala* si ha quando $\mathbf{A} = \mathrm{dg}(a_1, a_2, \ldots, a_k)$ con $a_i > 0$ per ogni $i = 1, 2, \, \ldots, k$. Dunque la trasformazione di scala

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \begin{pmatrix} a_1x_1 \\ a_2x_2 \\ \vdots \\ a_kx_k \end{pmatrix}$$

determina in pratica una variazione nelle unità di misura di ogni singola componente di $\mathbf{x}$.

La *trasformazione di riflessione* si ha quando $\mathbf{A} = \mathrm{dg}(a_1, a_2, \dots, a_k)$ con $a_j = 1$ per ogni $j \neq i = 1, 2, \dots, k$ e $a_i = -1$, ovvero

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ -x_i \\ \vdots \\ x_k \end{pmatrix} .$$

Questa trasformazione produce dunque un cambiamento di segno dell'i-esima componente di $\mathbf{x}$, che da un punto di vista geometrico corrisponde a una riflessione dell'i-esima componente del vettore $\mathbf{x}$.

La *trasformazione ortogonale* si ha quando $\mathbf{A} = \mathbf{\Gamma}$ è una matrice ortogonale. La trasformazione ortogonale

$$\mathbf{y} = \mathbf{\Gamma}\mathbf{x}$$

produce una rotazione rigida dei vettori lasciando invariata la loro norma dal momento che

$$\|\mathbf{y}\| = (\mathbf{y}^{\mathrm{T}}\mathbf{y})^{1/2} = (\mathbf{x}^{\mathrm{T}}\mathbf{\Gamma}^{\mathrm{T}}\mathbf{\Gamma}\mathbf{x})^{1/2} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{1/2} = \|\mathbf{x}\| .$$

Ad esempio, data la matrice ortogonale di ordine 2

$$\mathbf{\Gamma} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

dove $0 \leq \phi \leq 2\pi$, la trasformazione corrispondente produce una rotazione antioraria pari a un arco ϕ. Si noti inoltre che $\mathbf{y} = \mathbf{\Gamma}^{\mathrm{T}}\mathbf{x}$ è ancora una trasformazione ortogonale.

Infine, si consideri il caso in cui la matrice $\mathbf{A}$ è singolare e quindi esiste almeno un vettore riga di $\mathbf{A}$ che può essere espresso come combinazione lineare dei vettori riga rimanenti. Si supponga dunque che per qualche $i = 1, 2, \dots, k$ risulti

$$\mathbf{a}_{i\bullet}^{\mathrm{T}} = \sum_{j \neq i = 1}^{k} \alpha_j \mathbf{a}_{j\bullet}^{\mathrm{T}} ,$$

per cui l'i-esima componente del vettore $\mathbf{y}$ è data da

$$y_i = \mathbf{a}_{i\bullet}^{\mathrm{T}}\mathbf{x} = \sum_{j \neq i=1}^{k} \alpha_j (\mathbf{a}_{j\bullet}^{\mathrm{T}}\mathbf{x}) = \sum_{j \neq i=1}^{k} \alpha_j y_j \ .$$

Dunque l'i-esima componente del vettore **y** è una combinazione lineare delle rimanenti componenti. In altri termini, se **A** è singolare il codominio dell'applicazione si trova su un iperpiano di $\mathbb{R}^k$.

Esempio 6.1. Al fine di valutare l'effetto delle applicazioni lineari da $\mathbb{R}^2$ a $\mathbb{R}^2$ per diverse scelte di **A**, si consideri una figura contenuta in un quadrato di lato unitario e centrato nell'origine. Il quadrato è stato riportato nella Figura 6.1. Se si assume la trasformazione di scala con $\mathbf{A} = \mathbf{I}_2$, il quadrato non viene ovviamente modificato dalla trasformazione. Inoltre, se si considera la trasformazione di scala con

$$\mathbf{A} = \begin{pmatrix} 3/2 & 0 \\ 0 & 3/2 \end{pmatrix} = \frac{3}{2}\,\mathbf{I}_2 \ ,$$

le dimensioni del quadrato vengono aumentate di una volta e mezzo (vedi Figura 6.2). Nel caso in cui la trasformazione di scala sia basata sulla matrice diagonale

$$\mathbf{A} = \begin{pmatrix} 3/2 & 0 \\ 0 & 1 \end{pmatrix} = \mathrm{dg}(3/2, 1) \ ,$$

la scala di ascissa viene aumentata di una volta e mezzo, mentre la scala di ordinata rimane inalterata (vedi Figura 6.3). Per la traformazione di riflessione con

$$\mathbf{A} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \mathrm{dg}(-1, 1) \ ,$$

si ottiene una riflessione del quadrato rispetto all'asse delle ordinate (vedi Figura 6.4). Considerando la trasformazione ortogonale con

$$\mathbf{A} = \mathbf{\Gamma} = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} = \begin{pmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{pmatrix} ,$$

si verifica una rotazione antioraria del quadrato pari all'arco $\phi = \pi/4$ (vedi Figura 6.5). Quando **A** è una matrice generica del tipo

Figura 6.1.

Figura 6.2.

Figura 6.3.

Figura 6.4.

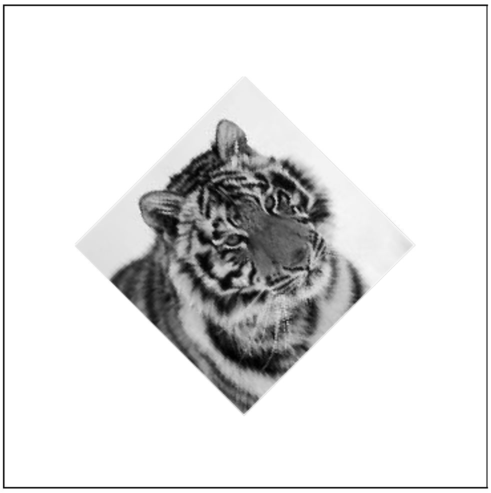

Figura 6.5.

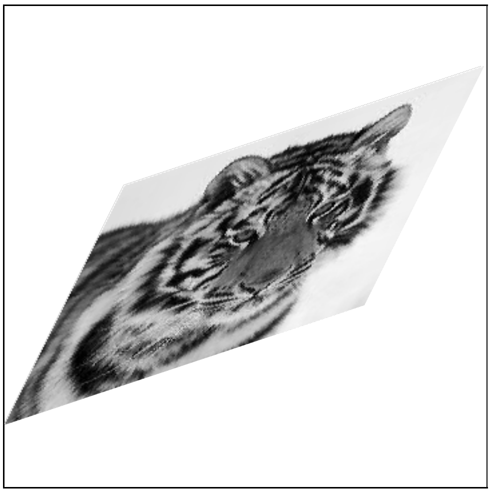

Figura 6.6.

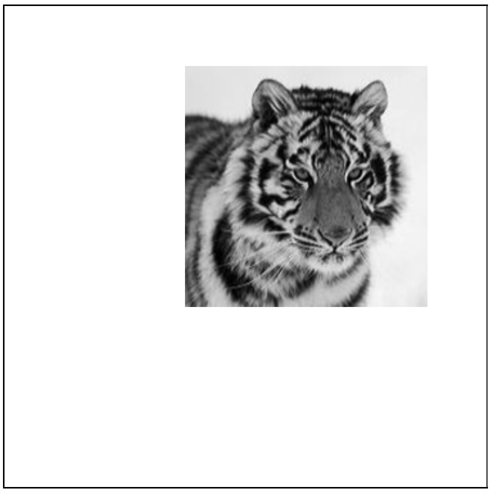

Figura 6.7.

Figura 6.8.

$$\mathbf{A} = \begin{pmatrix} 3/2 & 1/2 \\ 1/2 & 1 \end{pmatrix} ,$$

il quadrato viene contemporaneamente deformato, ruotato e variato d'ampiezza (vedi Figura 6.6). Infine, quando **A** è singolare il quadrato viene compresso in una retta. Ad esempio se si considera la matrice singolare **A** data da

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/8 \end{pmatrix}$$

la trasformazione comprime l'intero quadrato nella retta $y_2 = y_1/2$. □

6.3. Traslazioni

Dato un vettore **b** di ordine n si dice *traslazione* l'applicazione $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ tale che

$$\mathbf{y} = \mathbf{x} + \mathbf{b} .$$

La traslazione è una applicazione invertibile in quanto

$$\mathbf{x} = \mathbf{y} - \mathbf{b} ,$$

ma non è una applicazione lineare.

Si consideri una applicazione lineare $\mathbf{f} : \mathbb{R}^k \to \mathbb{R}^n$ associata a una traslazione, ovvero

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} ,$$

dove $\mathbf{b} \in \mathbb{R}^n$. Nella terminologia corrente, la precedente applicazione continua a essere detta, anche se impropriamente, applicazione lineare. Se **A** è matrice quadrata di ordine k non singolare, l'applicazione inversa della precedente applicazione risulta

$$\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) .$$

Esempio 6.2. Se si considera di nuovo l'immagine contenuta nel quadrato dell'Esempio 6.1, la traslazione con

$$\mathbf{b} = \begin{pmatrix} 1/4 \\ 1/4 \end{pmatrix} ,$$

produce uno spostamento del quadrato di un quarto di unità nelle ascisse e di un quarto di unità nelle ordinate (vedi Figura 6.7). L'applicazione lineare associata alla traslazione con

$$\mathbf{A} = \begin{pmatrix} 1 & -1/4 \\ 1/4 & 1/2 \end{pmatrix} , \mathbf{b} = \begin{pmatrix} 1/4 \\ -1/4 \end{pmatrix} ,$$

produce una deformazione, rotazione e variazione d'ampiezza unite ad uno spostamento del quadrato (vedi Figura 6.8). □

6.4. Proiezioni ortogonali

Data una matrice $\mathbf{Z}$ di ordine $(n \times k)$ tale che $\mathrm{r}(\mathbf{Z}) = k$, una applicazione lineare $\mathbf{y} = \mathbf{A}\mathbf{x}$ con

$$\mathbf{A} = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$$

matrice simmetrica di ordine n, è detta *proiezione ortogonale*. In effetti, questa applicazione proietta il vettore $\mathbf{x}$ sullo spazio vettoriale generato dalle colonne di $\mathbf{Z}$. Data la natura della matrice $\mathbf{A}$, non vi è perdita di generalità nell'assumere $\mathrm{r}(\mathbf{Z}) = k$, dal momento che in caso contrario è sufficiente ottenere l'insieme delle colonne linearmente indipendenti di $\mathbf{Z}$ e considerare questa nuova matrice al posto di $\mathbf{Z}$.

Quando $k = 1$, si ottiene per $\mathbf{y}$ l'espressione introdotta nella Sezione 1.6 nel caso in cui $\mathbf{x}$ venga proiettato su un vettore $\mathbf{z}$. Infatti in questo caso si ha

$$\mathbf{A} = \mathbf{z}(\mathbf{z}^{\mathrm{T}}\mathbf{z})^{-1}\mathbf{z}^{\mathrm{T}} = \frac{1}{\|\mathbf{z}\|^2}\,\mathbf{z}\mathbf{z}^{\mathrm{T}} ,$$

da cui

$$\mathbf{y} = \frac{1}{\|\mathbf{z}\|^2}\,\mathbf{z}\mathbf{z}^{\mathrm{T}}\mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{z} \rangle}{\|\mathbf{z}\|^2}\,\mathbf{z} .$$

Risultato 6.1. Se **Z** rappresenta un matrice di ordine $(n \times k)$ tale che $\mathrm{r}(\mathbf{Z}) = k$, per la matrice $\mathbf{A} = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$ valgono le proprietà:
(i) $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$;
(ii) **A** è idempotente;
(iii) $\mathrm{tr}(\mathbf{A}) = \mathrm{r}(\mathbf{A}) = k$;
(iv) $(\mathbf{I}_n - \mathbf{A})$ è idempotente;
(v) $\mathrm{tr}(\mathbf{I}_n - \mathbf{A}) = \mathrm{r}(\mathbf{I}_n - \mathbf{A}) = n - k$;
(vi) per ogni vettore **x** di ordine n si ha $\mathbf{Ax} \perp (\mathbf{I}_n - \mathbf{A})\mathbf{x}$.

Esempio 6.3. Se si considera la matrice

$$\mathbf{Z} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$$

è evidente che le colonne di **Z** generano $\mathbb{R}^2$, inteso in questo caso come sottospazio vettoriale di $\mathbb{R}^3$. Inoltre risulta

$$\mathbf{A} = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Di conseguenza qualsiasi vettore $\mathbf{x} = (x_1\ x_2\ x_3)^{\mathrm{T}}$ di $\mathbb{R}^3$ viene proiettato su $\mathbb{R}^2$ mediante l'applicazione lineare $\mathbf{y} = \mathbf{Ax}$. In effetti si ha

$$\mathbf{Ax} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}.$$

Inoltre si noti che

$$(\mathbf{I}_3 - \mathbf{A})\mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}$$

e quindi $(\mathbf{I}_3 - \mathbf{A})$ proietta qualsiasi vettore **x** su uno spazio i cui elementi sono ortogonali agli elementi dello spazio generato dalle proiezioni attraverso **A**, confermando in questo modo la proprietà (vi) del Risultato 6.1. □

Esempio 6.4. Se si considera la matrice

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$$

allora risulta

$$\mathbf{A} = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} .$$

Si noti che **A** è idempotente e risulta

$$\mathrm{tr}(\mathbf{A}) = 2 = \mathrm{r}(\mathbf{Z}) .$$

Inoltre

$$\mathbf{I}_3 - \mathbf{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{pmatrix}$$

è a sua volta una matrice idempotente con

$$\mathrm{tr}(\mathbf{I}_3 - \mathbf{A}) = 1 = 3 - \mathrm{r}(\mathbf{Z}) . \qquad \square$$

Capitolo 7

Sistemi lineari

7.1. Definizioni e notazioni

Un *sistema lineare* di n equazioni lineari in k incognite

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1k}x_k = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2k}x_k = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nk}x_k = b_n \end{array}$$

può essere scritto più semplicemente come

$$\mathbf{A}\mathbf{x} = \mathbf{b} ,$$

dove

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nk} \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} .$$

In questo caso, $\mathbf{A}$ è detta *matrice dei coefficienti*, $\mathbf{x}$ è detto *vettore delle incognite*, mentre $\mathbf{b}$ è detto *vettore dei termini noti*.

Obiettivo dello studio dei sistemi lineari è quello di determinare l'esistenza ed eventualmente l'unicità di un vettore $\mathbf{x}$ che soddisfi alla precedente relazione e che viene detto *soluzione del sistema*.

Sulla base della rappresentazione di $\mathbf{A}$ in termini di vettori riga, da un punto di vista geometrico un sistema lineare può essere considerato come un insieme di n iperpiani traslati in $\mathbb{R}^k$, il cui i-esimo iperpiano traslato è dato da

$$\mathcal{T}_i = \{\mathbf{x} : \mathbf{a}_{i\bullet}^{\mathrm{T}}\mathbf{x} = b_i\} ,$$

dove $\mathbf{a}_{i\bullet}^{\mathrm{T}}$ rappresenta l'i-esimo vettore riga di $\mathbf{A}$. Evidentemente la soluzione del sistema lineare deve appartenere all'intersezione degli n iperpiani traslati, ovvero si deve avere $\mathbf{x} \in \mathcal{T}_1 \cap \mathcal{T}_2 \cap \ldots \cap \mathcal{T}_n$.

Esempio 7.1. Si consideri il sistema lineare di 3 equazioni in 2 incognite

$$\begin{aligned} 2x_1 - 2x_2 &= -2 \\ x_1 - x_2 &= -1 \\ x_1 + x_2 &= 3 \,. \end{aligned}$$

La matrice dei coefficienti e il vettore dei termini noti sono quindi dati da

$$\mathbf{A} = \begin{pmatrix} 2 & -2 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -2 \\ -1 \\ 3 \end{pmatrix} .$$

Da un punto di vista geometrico il sistema lineare può essere descritto da tre rette in $\mathbb{R}^2$. Si noti che le prime due rette del sistema sono coincidenti e quindi l'intersezione delle tre rette è data dall'intersezione della prima e della terza retta, ovvero dal punto di coordinate

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} .$$

In questo caso la soluzione del sistema è unica. □

Esempio 7.2. Si consideri il sistema lineare di 3 equazioni in 2 incognite

$$\begin{aligned} 2x_1 - 2x_2 &= -2 \\ x_1 - x_2 &= -1 \\ 4x_1 - 4x_2 &= -4 \,. \end{aligned}$$

La matrice dei coefficienti e il vettore dei termini noti sono quindi dati da

$$\mathbf{A} = \begin{pmatrix} 2 & -2 \\ 1 & -1 \\ 4 & -4 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -2 \\ -1 \\ -4 \end{pmatrix} .$$

Di nuovo, da un punto di vista geometrico il sistema lineare può essere descritto da tre rette in $\mathbb{R}^2$. Le tre rette sono coincidenti e la loro intersezione è data dalla stessa retta. Di conseguenza esistono infinite soluzioni, ovvero tutti i punti che giacciono su questa retta, e queste soluzioni sono del tipo

$$\mathbf{x} = \begin{pmatrix} \beta - 1 \\ \beta \end{pmatrix},$$

con $\beta \in \mathbb{R}$. □

Esempio 7.3. Si consideri il sistema lineare di 3 equazioni in 2 incognite

$$\begin{aligned} &2x_1 - 2x_2 = -2 \\ &x_1 - x_2 = 1 \\ &3x_1 + x_2 = 3 \, . \end{aligned}$$

Dunque la matrice dei coefficienti e il vettore dei termini noti sono dati da

$$\mathbf{A} = \begin{pmatrix} 2 & -2 \\ 1 & -1 \\ 3 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix}.$$

Da un punto di vista geometrico il sistema lineare può essere descritto da tre rette in $\mathbb{R}^2$. Le tre rette non hanno nessun punto congiuntamente in comune e quindi il sistema non ammette soluzioni. □

7.2. Sistemi lineari omogenei

Se per il vettore dei termini noti risulta $\mathbf{b} = \mathbf{0}$, il sistema lineare è detto *omogeneo* e ovviamente ammette almeno la soluzione $\mathbf{x} = \mathbf{0}$. Considerando la rappresentazione di $\mathbf{A}$ in termini di vettori colonna, se il sistema lineare omogeneo ammette soluzioni diverse dal vettore $\mathbf{0}$, per $\mathbf{x} \neq \mathbf{0}$ si ha

$$\sum_{j=1}^{k} x_j \mathbf{a}_{\bullet j} = \mathbf{0} \, ,$$

dove $\mathbf{a}_{\bullet j}$ rappresenta il j-esimo vettore colonna di $\mathbf{A}$. Dunque, i vettori colonna di $\mathbf{A}$ sono tra loro linearmente dipendenti, ovvero risulta $\mathrm{r}(\mathbf{A}) < k$.

In ultima analisi, un sistema lineare omogeneo ammette soluzioni non nulle se e solo se $\mathrm{r}(\mathbf{A}) < k$, ovvero se il rango della matrice dei coefficienti è inferiore al numero delle incognite. In particolare, se il numero delle equazioni è uguale al numero delle incognite, ovvero se **A** è quadrata, il sistema lineare omogeneo ammette soluzioni non nulle se e solo se $\det(\mathbf{A}) = 0$.

Quando $\mathrm{r}(\mathbf{A}) < k$, l'insieme delle soluzioni di un sistema lineare omogeneo

$$\mathcal{A} = \{\mathbf{x} : \mathbf{Ax} = \mathbf{0}\}$$

costituisce uno spazio vettoriale con $\dim(\mathcal{A}) = k - \mathrm{r}(\mathbf{A})$. In questo caso, l'insieme delle soluzioni può essere ottenuto attraverso l'uso delle matrici inverse generalizzate. Dal momento che risulta $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$, se $\mathbf{x}$ è una soluzione del sistema lineare omogeneo anche $\mathbf{A}^{-}\mathbf{Ax}$ è una soluzione essendo

$$\mathbf{AA}^{-}\mathbf{Ax} = \mathbf{Ax} = \mathbf{0} \, .$$

Dunque, se $\boldsymbol{\beta} = (\beta_1 \, \beta_2 \, \ldots \, \beta_k)^{\mathrm{T}}$ rappresenta un vettore arbitrario di ordine k, una *soluzione generale* del sistema lineare omogeneo è data da

$$\mathbf{x} = (\mathbf{I}_k - \mathbf{A}^{-}\mathbf{A})\boldsymbol{\beta} \, .$$

In effetti risulta

$$\mathbf{A}(\mathbf{I}_k - \mathbf{A}^{-}\mathbf{A})\boldsymbol{\beta} = (\mathbf{A} - \mathbf{AA}^{-}\mathbf{A})\boldsymbol{\beta} = \mathbf{0} \, .$$

Esempio 7.4. Si consideri il sistema lineare omogeneo di 3 equazioni in 3 incognite

$$\begin{aligned} &x_1 - 2x_2 - 2x_3 = 0 \\ &x_1 - x_2 + x_3 = 0 \\ &2x_1 - 2x_2 + 2x_3 = 0 \, . \end{aligned}$$

La matrice dei coefficienti

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & -2 \\ 1 & -1 & 1 \\ 2 & -2 & 2 \end{pmatrix}$$

è tale che $\mathrm{r}(\mathbf{A}) = 2 < 3$ e quindi il sistema lineare omogeneo ammette anche soluzioni diverse dal vettore $\mathbf{0}$. Di conseguenza lo spazio vettoriale delle soluzioni ha dimensione $\dim(\mathcal{A}) = 3 - 2 = 1$. Inoltre una inversa generalizzata $\mathbf{A}^-$ di $\mathbf{A}$, ottenuta come evidenziato nella Sezione 5.4, è data da

$$\mathbf{A}^- = \begin{pmatrix} -1 & 2 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} .$$

Quindi, una soluzione generale del sistema lineare omogeneo risulta

$$\mathbf{x} = (\mathbf{I}_3 - \mathbf{A}^-\mathbf{A})\boldsymbol{\beta} = \begin{pmatrix} -4\beta_3 \\ -3\beta_3 \\ \beta_3 \end{pmatrix} ,$$

dove $\boldsymbol{\beta} = (\beta_1\ \beta_2\ \beta_3)^{\mathrm{T}}$ rappresenta un vettore arbitrario di ordine 3. □

7.3. Sistemi lineari non omogenei

Nel caso in cui il vettore dei termini noti risulti $\mathbf{b} \neq \mathbf{0}$, il sistema lineare è detto *non omogeneo*. Per quanto riguarda la soluzione di sistemi lineari non omogenei, è opportuno considerare la matrice a blocchi $(\mathbf{A}\,|\,\mathbf{b})$. In effetti, se il sistema lineare non omogeneo ammette una soluzione $\mathbf{x}$, allora si ha

$$\sum_{j=1}^{k} x_j \mathbf{a}_{\bullet j} = \mathbf{b} ,$$

ovvero $\mathbf{b}$ è una combinazione lineare dei vettori colonna di $\mathbf{A}$ e dunque risulta $\mathrm{r}(\mathbf{A}\,|\,\mathbf{b}) = \mathrm{r}(\mathbf{A})$. In questo caso vi sono tre possibili situazioni, ovvero si ha

$$\mathrm{r}(\mathbf{A}\,|\,\mathbf{b}) = \mathrm{r}(\mathbf{A}) = k$$

e quindi il sistema lineare non omogeneo ammette una sola soluzione, oppure si ha

$$\mathrm{r}(\mathbf{A}\,|\,\mathbf{b}) = \mathrm{r}(\mathbf{A}) < k$$

e quindi il sistema lineare non omogeneo ammette infinite soluzioni. Al contrario, se

$$\mathrm{r}(\mathbf{A}\,|\,\mathbf{b}) = \mathrm{r}(\mathbf{A}) + 1 \,,$$

allora il sistema lineare non omogeneo non ammette soluzioni.

Nel caso particolare in cui il numero delle equazioni coincida con il numero delle incognite, ovvero se $n = k$, il sistema lineare ammette un'unica soluzione quando **A** è non singolare. Premoltiplicando per $\mathbf{A}^{-1}$ la relazione $\mathbf{Ax} = \mathbf{b}$ si ottiene

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \,.$$

In generale, quando $\mathrm{r}(\mathbf{A}\,|\,\mathbf{b}) = \mathrm{r}(\mathbf{A})$ l'insieme delle soluzioni può essere ottenuto considerando una inversa generalizzata $\mathbf{A}^{-}$. Tenendo presente che dalla proprietà (ii) del Risultato 5.3 si ha $\mathrm{r}(\mathbf{AA}^{-}) = \mathrm{r}(\mathbf{A})$, premoltiplicando per $\mathbf{AA}^{-}$ la relazione $\mathbf{Ax} = \mathbf{b}$ si ottiene

$$\mathbf{AA}^{-}\mathbf{Ax} = \mathbf{AA}^{-}\mathbf{b} \,.$$

Dal momento che $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$, allora dalla precedente relazione deve risultare $\mathbf{Ax} = \mathbf{AA}^{-}\mathbf{b}$, ovvero

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{b} \,.$$

Si noti che deve sussistere anche la relazione $\mathbf{b} = \mathbf{AA}^{-}\mathbf{b}$. Una *soluzione generale* del sistema è quindi data da

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{b} + (\mathbf{I}_k - \mathbf{A}^{-}\mathbf{A})\boldsymbol{\beta} \,,$$

dove $\boldsymbol{\beta} = (\beta_1\ \beta_2\ \dots\ \beta_k)^{\mathrm{T}}$ rappresenta un vettore arbitrario di ordine k. In effetti si ha

$$\mathbf{A}(\mathbf{A}^{-}\mathbf{b} + (\mathbf{I}_k - \mathbf{A}^{-}\mathbf{A})\boldsymbol{\beta}) = \mathbf{AA}^{-}\mathbf{b} + (\mathbf{A} - \mathbf{AA}^{-}\mathbf{A})\boldsymbol{\beta} = \mathbf{b} \,.$$

Se $\mathrm{r}(\mathbf{A}\,|\,\mathbf{b}) = \mathrm{r}(\mathbf{A}) = k$, allora l'inversa generalizzata può essere scelta come l'inversa di Moore-Penrose, dal momento che in questo caso $\mathbf{A}^{+} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}$ e quindi $\mathbf{A}^{+}\mathbf{A} = \mathbf{I}_k$ (vedi proprietà (ii) del Risultato 5.4). Di conseguenza si ha un'unica soluzione che può essere ottenuta premoltiplicando per $\mathbf{A}^{+}$ la relazione $\mathbf{Ax} = \mathbf{b}$, ovvero

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b} \,.$$

Esempio 7.5. Si consideri il sistema lineare non omogeneo di 2 equazioni in 2 incognite dato da

$$x_1 + 3x_2 = 5$$
$$x_1 - x_2 = 1\ .$$

La matrice dei coefficienti e il vettore dei termini noti sono quindi dati da

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 1 & -1 \end{pmatrix}\ ,\ \mathbf{b} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}\ .$$

Il numero delle incognite è dunque pari al numero delle equazioni e **A** è non singolare essendo $\det(\mathbf{A}) = -4$. Dal momento che

$$\mathbf{A}^{-1} = \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & -1/4 \end{pmatrix}\ ,$$

l'unica soluzione del sistema lineare risulta

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}\ .$$ □

Esempio 7.6. Si consideri di nuovo il sistema lineare dell'Esempio 7.2. Dal momento che risulta $\mathrm{r}(\mathbf{A}\,|\,\mathbf{b}) = \mathrm{r}(\mathbf{A}) = 1$, allora essendo

$$\mathbf{A}^{-} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\ ,$$

una soluzione del sistema lineare è data da

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{b} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}\ .$$

Se $\boldsymbol{\beta} = (\beta_1\ \beta_2)^{\mathrm{T}}$ rappresenta un vettore arbitrario di ordine 2, tutte le restanti soluzioni sono date da

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{b} + (\mathbf{I}_2 - \mathbf{A}^{-}\mathbf{A})\boldsymbol{\beta} = \begin{pmatrix} \beta_2 - 1 \\ \beta_2 \end{pmatrix}\ ,$$

confermando il risultato dell'Esempio 7.2. □

Esempio 7.7. Si consideri di nuovo il sistema lineare dell'Esempio 7.1. Dal momento che risulta $\mathrm{r}(\mathbf{A}\,|\,\mathbf{b}) = \mathrm{r}(\mathbf{A}) = 2$, allora si ha

$$\mathbf{A}^{+} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 2/10 & 1/10 & 5/10 \\ -2/10 & -1/10 & 5/10 \end{pmatrix}$$

e quindi l'unica soluzione del sistema lineare è data da

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} ,$$

confermando il risultato dell'Esempio 7.1. □

7.4. Calcolo della soluzione di sistemi lineari

Al fine di risolvere un sistema lineare non omogeneo quando $n = k$ e il sistema ammette un'unica soluzione, ovvero $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, da un punto di vista pratico è conveniente adottare un procedimento basato sulla matrice a blocchi $(\mathbf{A} \,|\, \mathbf{b})$. Considerando il metodo di calcolo della matrice inversa introdotto nella Sezione 5.3, è dunque sufficiente effettuare sulla matrice a blocchi $(\mathbf{A} \,|\, \mathbf{b})$ le stesse operazioni elementari di riga necessarie ad ottenere la matrice identità da $\mathbf{A}$, dal momento che premoltiplicando $(\mathbf{A} \,|\, \mathbf{b})$ per $\mathbf{QP} = \mathbf{A}^{-1}$ si ha

$$\mathbf{QP}(\mathbf{A} \,|\, \mathbf{b}) = (\mathbf{I}_k \,|\, \mathbf{QPb}) = (\mathbf{I}_k \,|\, \mathbf{A}^{-1}\mathbf{b}) = (\mathbf{I}_k \,|\, \mathbf{x})$$

e quindi la matrice a blocchi risultante dal prodotto contiene nel secondo blocco esattamente l'unica soluzione del sistema lineare.

Esempio 7.8. Si consideri il sistema lineare non omogeneo di 3 equazioni in 3 incognite

$$\begin{aligned} x_1 + x_2 - x_3 &= 0 \\ 2x_1 + x_2 - x_3 &= 1 \\ x_1 + 2x_2 - x_3 &= 2 \,. \end{aligned}$$

La matrice dei coefficienti e il vettore dei termini noti sono quindi dati da

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & -1 \\ 1 & 2 & -1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}.$$

Dal momento che $\det(\mathbf{A}) = -1 \neq 0$ allora si ha un'unica soluzione. In questo caso, formando la matrice a blocchi $(\mathbf{A} \,|\, \mathbf{b})$ si ottiene

$$\mathbf{A}_1 = (\mathbf{A} \,|\, \mathbf{b}) = \left(\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 2 & 1 & -1 & 1 \\ 1 & 2 & -1 & 2 \end{array}\right),$$

da cui sottraendo il primo vettore riga moltiplicato per 2 dal secondo vettore riga di $\mathbf{A}_1$ si ottiene

$$\mathbf{A}_2 = \mathbf{E}_2(-2|1)\mathbf{A}_1 = \left(\begin{array}{rrr|c} 1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 \\ 1 & 2 & -1 & 2 \end{array}\right),$$

mentre sottraendo il primo vettore riga dal terzo vettore riga di $\mathbf{A}_2$ risulta

$$\mathbf{A}_3 = \mathbf{E}_3(-1|1)\mathbf{A}_2 = \left(\begin{array}{rrr|c} 1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 \\ 0 & 1 & 0 & 2 \end{array}\right).$$

Inoltre, sommando il secondo vettore riga al terzo vettore riga di $\mathbf{A}_3$ si ha

$$\mathbf{A}_4 = \mathbf{E}_3(1|2)\mathbf{A}_3 = \left(\begin{array}{rrr|c} 1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right),$$

mentre sommando il secondo vettore riga al primo vettore riga di $\mathbf{A}_4$ risulta

$$\mathbf{A}_5 = \mathbf{E}_1(1|2)\mathbf{A}_4 = \left(\begin{array}{rrr|c} 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right).$$

Inoltre, cambiando di segno al secondo vettore riga di $\mathbf{A}_5$ si ottiene

$$\mathbf{A}_6 = \mathbf{E}_2(-1)\mathbf{A}_5 = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 3 \end{array}\right) .$$

Infine, sommando il terzo vettore riga al secondo vettore riga di $\mathbf{A}_6$ risulta

$$\mathbf{A}_7 = \mathbf{E}_2(1|3)\mathbf{A}_6 = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right) = (\mathbf{I}_3 \,|\, \mathbf{x}) ,$$

per cui l'unica soluzione del sistema lineare è data da

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} . \qquad \square$$

Capitolo 8

Autovalori e autovettori

8.1. Definizioni e notazioni

Data una matrice simmetrica **A** di ordine k, si considerino lo scalare λ e il vettore $\mathbf{x} \neq \mathbf{0}$ di ordine k per cui vale la relazione

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \ .$$

In questo caso λ e **x** si dicono rispettivamente *autovalore* e *autovettore* di **A**. Identiche definizioni possono essere date anche per una matrice quadrata non simmetrica. Tuttavia, al fine di evitare alcune complessità che sorgono in un approccio generale, nel seguito vengono considerate per semplicità solo matrici simmetriche. In ogni caso, la teoria relativa agli autovalori ed autovettori di matrici simmetriche riveste importanza fondamentale, ad esempio nelle applicazioni di tipo statistico ed econometrico.

La precedente relazione può essere riscritta come

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \ ,$$

che rappresenta un sistema lineare omogeneo nel vettore delle incognite **x**. Questo sistema ammette soluzioni diverse dal vettore $\mathbf{x} = \mathbf{0}$ se e solo se la matrice $(\mathbf{A} - \lambda\mathbf{I})$ è singolare, ovvero se si ha

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

(vedi Sezione 7.2). Poichè lo sviluppo di questo determinante produce un polinomio di grado k nella variabile λ, detto *polinomio caratteristico*, la precedente relazione rappresenta un'equazione di grado k nell'incognita λ, ovvero

$$c_k\lambda^k + c_{k-1}\lambda^{k-1} + \ldots + c_1\lambda + c_0 = 0 \ ,$$

che viene detta *equazione caratteristica*. Per il Teorema Fondamentale dell'Algebra, l'equazione caratteristica ammette k soluzioni $\lambda_1, \lambda_2, \dots, \lambda_k$ distinte o coincidenti che sono dette autovalori di **A**. In particolare, per $k = 2$ risulta

$$\lambda^2 - \mathrm{tr}(\mathbf{A})\lambda + \det(\mathbf{A}) = 0\ .$$

Se λ_i è un autovalore di **A** si ha dunque

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x} = \mathbf{0}$$

e quindi ogni vettore **x** che soddisfa questa relazione è detto autovettore di **A** associato all'autovalore λ_i. Per quanto detto nella Sezione 7.2 riguardo alle soluzioni di sistemi lineari omogenei, l'insieme degli autovettori associati all'autovalore λ_i costituisce uno spazio vettoriale $\mathcal{A}_i$ che in questo caso viene detto *autospazio* di **A** associato a λ_i.

Esempio 8.1. Si consideri la matrice simmetrica di ordine 2 data da

$$\mathbf{A} = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\ .$$

L'equazione caratteristica

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\begin{pmatrix} 5-\lambda & 2 \\ 2 & 2-\lambda \end{pmatrix} = \lambda^2 - 7\lambda + 6 = 0$$

ha come soluzioni $\lambda_1 = 6$ e $\lambda_2 = 1$. Gli autovettori associati all'autovalore $\lambda_1 = 6$ si ottengono risolvendo il sistema lineare omogeneo

$$(\mathbf{A} - 6\mathbf{I})\mathbf{x} = \mathbf{0}$$

con $\mathbf{x} = (x_1\ x_2)^{\mathrm{T}}$, ovvero

$$\begin{aligned} -x_1 + 2x_2 &= 0 \\ 2x_1 - 4x_2 &= 0\ , \end{aligned}$$

per cui l'autospazio $\mathcal{A}_1$ è costituito da vettori del tipo $\mathbf{x} = (2\beta\ \beta)^{\mathrm{T}}$ con $\beta \in \mathbb{R}$ e geometricamente coincide con la retta $x_2 = x_1/2$. Analogamente gli autovettori associati all'autovalore $\lambda_2 = 1$ si ottengono risolvendo il sistema lineare omogeneo

$$(\mathbf{A}-\mathbf{I})\mathbf{x}=\mathbf{0}\ ,$$

ovvero

$$\begin{aligned}4x_1+2x_2&=0\\2x_1+x_2&=0\ ,\end{aligned}$$

da cui l'autospazio $\mathcal{A}_2$ è costituito da vettori del tipo $\mathbf{x}=(-\beta/2\ \beta)^{\mathrm{T}}$ con $\beta\in\mathbb{R}$ e geometricamente coincide con la retta $x_2=\ -2x_1$. □

8.2. Proprietà degli autovalori e autovettori

Gli autovalori e gli autovettori risultano fondamentali nello studio di una matrice simmetrica per la caratterizzazione che queste quantità possono dare della matrice stessa, come si può evincere dal seguente risultato.

Risultato 8.1. Se **A** è una matrice simmetrica di ordine k, per gli autovalori e gli autovettori valgono le proprietà:
(i) gli autovalori di **A** sono reali;
(ii) se gli autovalori di **A** sono $\lambda_1,\lambda_2,\dots,\lambda_k$, allora risulta

$$\det(\mathbf{A})=\prod_{i=1}^{k}\lambda_i\ ,\ \mathrm{tr}(\mathbf{A})=\sum_{i=1}^{k}\lambda_i\ ;$$

(iii) dato $\alpha\in\mathbb{R}$, la matrice $\mathbf{A}+\alpha\mathbf{I}$ possiede autovalori pari a $\lambda_i+\alpha$ e gli stessi autovettori di **A**;
(iv) il numero degli autovalori non nulli di **A** è uguale al rango di **A**;
(v) se $\lambda_i\neq\lambda_j$ sono due autovalori distinti di **A** gli autospazi corrispondenti $\mathcal{A}_i$ e $\mathcal{A}_j$ sono tra loro ortogonali, ovvero $\mathbf{x}^{\mathrm{T}}\mathbf{y}=0$ per ogni $\mathbf{x}\in\mathcal{A}_i$ e $\mathbf{y}\in\mathcal{A}_j$;
(vi) se λ_i è un autovalore di molteplicità r_i della matrice **A**, allora l'autospazio corrispondente $\mathcal{A}_i$ ha dimensione $\dim(\mathcal{A}_i)=r_i$.

Dal momento che per la proprietà (i) del Risultato 8.1 gli autovalori $\lambda_1,\lambda_2,\dots,\lambda_k$ di una matrice simmetrica **A** sono reali, per quanto riguarda la loro numerazione è convenzione scegliere l'ordinamento decrescente, ovvero si assume che $\lambda_1\geq\lambda_2\geq\dots\geq\lambda_k$.

Esempio 8.2. Si consideri la matrice simmetrica di ordine 3 data da

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} .$$

L'equazione caratteristica risulta

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det \begin{pmatrix} 1-\lambda & 0 & 0 \\ 0 & 2-\lambda & 1 \\ 0 & 1 & 2-\lambda \end{pmatrix} = 0 \, ,$$

ovvero

$$(1-\lambda)(\lambda^2 - 4\lambda + 3) = 0$$

ed ha come soluzioni $\lambda_1 = 3$ e $\lambda_2 = \lambda_3 = 1$. Risulta quindi $\det(\mathbf{A}) = 3 = \lambda_1\lambda_2\lambda_3$ e $\mathrm{tr}(\mathbf{A}) = 5 = \lambda_1 + \lambda_2 + \lambda_3$. □

Esempio 8.3. Si consideri la matrice simmetrica di ordine 3 data da

$$\mathbf{A} = \begin{pmatrix} 6 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 3 \end{pmatrix} .$$

L'equazione caratteristica risulta

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det \begin{pmatrix} 6-\lambda & 0 & 2 \\ 0 & -\lambda & 0 \\ 2 & 0 & 3-\lambda \end{pmatrix} = 0$$

ovvero

$$-\lambda(\lambda^2 - 9\lambda + 14) = 0$$

ed ha come soluzioni $\lambda_1 = 7$ e $\lambda_2 = 2$ e $\lambda_3 = 0$. Si ha inoltre $\mathrm{r}(\mathbf{A}) = 2$ ed in effetti vi sono due autovalori non nulli. □

Si indichi con $\mathbf{\Lambda}$ la matrice diagonale contenente sulla diagonale principale gli autovalori $\lambda_1, \lambda_2, \ldots, \lambda_k$, eventualmente non tutti distinti, di una matrice simmetrica $\mathbf{A}$, ovvero $\mathbf{\Lambda} = \mathrm{dg}(\lambda_1, \lambda_2, \ldots, \lambda_k)$. Inoltre, siano $\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_h}$ (dove $i_1, i_2, \ldots, i_h$ rappresenta una scelta ordinata di h dei primi k numeri interi) gli h autovalori distinti di $\mathbf{A}$ con molteplicità $r_1, r_2, \ldots, r_h$ e tali che $\sum_{j=1}^{h} r_j = k$. Sulla base delle proprietà (v) e (vi) del Risultato 8.1, scegliendo h basi ortonormali degli autospazi $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_h$ associati agli h autovalori distinti, si ottiene un insieme di k autovettori ortonormali $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_k$ dei quali, i primi r_1 sono relativi all'autovalore λ_{i_1}, i successivi r_2 sono relativi all'autovalore λ_{i_2}, $\ldots$, gli ultimi r_h sono relativi all'autovalore λ_{i_h}.

Se si considera la matrice ortogonale $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1\ \boldsymbol{\gamma}_2\ \ldots\ \boldsymbol{\gamma}_k)$, si ha

$$\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{A}\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_j^{\mathrm{T}}\mathbf{A}\boldsymbol{\gamma}_i) = (\boldsymbol{\gamma}_j^{\mathrm{T}}\lambda_i\boldsymbol{\gamma}_i) = (\lambda_i\boldsymbol{\gamma}_j^{\mathrm{T}}\boldsymbol{\gamma}_i) = \boldsymbol{\Lambda} ,$$

in quanto, essendo $\boldsymbol{\gamma}_i$ un autovettore di $\mathbf{A}$ associato a λ_i, esso soddisfa alla relazione

$$\mathbf{A}\boldsymbol{\gamma}_i = \lambda_i\boldsymbol{\gamma}_i ,$$

e $\boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\gamma}_j$ vale 1 se $i = j$ e 0 altrimenti. Premoltiplicando e postmoltiplicando $\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{A}\boldsymbol{\Gamma}$ rispettivamente per $\boldsymbol{\Gamma}$ e $\boldsymbol{\Gamma}^{\mathrm{T}}$ si ottiene infine

$$\mathbf{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{\mathrm{T}} = \sum_{i=1}^{k} \lambda_i\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} ,$$

che è detta *rappresentazione diagonale* di una matrice simmetrica $\mathbf{A}$.

La rappresentazione diagonale è valida qualunque sia l'ordine in cui gli autovalori compaiono nella matrice $\boldsymbol{\Lambda}$, purchè gli autovettori corrispondenti seguano lo stesso ordine nella matrice $\boldsymbol{\Gamma}$. Tuttavia, come specificato in precedenza è usuale convenzione scegliere l'ordinamento decrescente degli autovalori.

Si noti inoltre che se la matrice $\mathbf{A}$ ha autovalori $\lambda_1, \lambda_2, \ldots, \lambda_k$ tutti distinti, ogni autospazio $\mathcal{A}_i$ ha dimensione unitaria e quindi ammette una sola base ortonormale $\boldsymbol{\gamma}_i$ (a meno del segno). Quindi la matrice $\boldsymbol{\Gamma}$ risulta determinata univocamente da $\mathbf{A}$ (a meno del segno) e a loro volta le matrici $\boldsymbol{\Lambda}$ e $\boldsymbol{\Gamma}$ determinano univocamente la rappresentazione diagonale della matrice $\mathbf{A}$. Al contrario, se la matrice $\mathbf{A}$ ha autovalori coincidenti la matrice $\boldsymbol{\Gamma}$ non è unica e di conseguenza la rappresentazione diagonale della matrice $\mathbf{A}$ non è unica.

Esempio 8.4. Si consideri la matrice simmetrica di ordine 2 data da

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} .$$

L'equazione caratteristica

$$\det(\mathbf{A}-\lambda\mathbf{I}) = \det\begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} = \lambda^2 - 4\lambda + 3 = 0$$

ha come soluzioni $\lambda_1 = 3$ e $\lambda_2 = 1$. Poichè entrambi gli autovalori hanno molteplicità unitaria, gli autospazi associati $\mathcal{A}_1$ e $\mathcal{A}_2$ hanno dimensione unitaria, ovvero $\dim(\mathcal{A}_1) = \dim(\mathcal{A}_2) = 1$. Gli autovettori associati all'autovalore $\lambda_1 = 3$ si ottengono risolvendo il sistema lineare omogeneo

$$(\mathbf{A} - 3\mathbf{I})\mathbf{x} = \mathbf{0}$$

con $\mathbf{x} = (x_1\ x_2)^{\mathrm{T}}$, ovvero

$$\begin{aligned} -\,x_1 + x_2 &= 0 \\ x_1 - x_2 &= 0\ , \end{aligned}$$

da cui l'autospazio $\mathcal{A}_1$ è costituito da vettori del tipo $\mathbf{x} = (\beta\ \beta)^{\mathrm{T}}$ con $\beta \in \mathbb{R}$ e geometricamente coincide con la retta $x_2 = x_1$. In questo caso l'unica base ortonormale di $\mathcal{A}_1$ (a meno del segno) è data dal vettore normalizzato che si ottiene ponendo

$$\|\mathbf{x}\|^2 = 2\beta^2 = 1\ ,$$

da cui $\beta = \pm\sqrt{2}/2$ e quindi $\boldsymbol{\gamma}_1 = \pm(\sqrt{2}/2\ \sqrt{2}/2)^{\mathrm{T}}$. Analogamente gli autovettori associati all'autovalore $\lambda_2 = 1$ si ottengono risolvendo il sistema lineare

$$(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}\ ,$$

ovvero

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_1 + x_2 &= 0\ , \end{aligned}$$

da cui l'autospazio $\mathcal{A}_2$ è costituito da vettori del tipo $\mathbf{x} = (-\beta\ \beta)^{\mathrm{T}}$ con $\beta \in \mathbb{R}$ e geometricamente coincide con la retta $x_2 = -x_1$. Anche in questo caso l'unica base ortonormale di $\mathcal{A}_2$ (a meno del segno) è data dal vettore normalizzato che si ottiene ponendo $2\beta^2 = 1$, per cui di nuovo $\beta = \pm\sqrt{2}/2$ e quindi $\boldsymbol{\gamma}_2 = \pm(-\sqrt{2}/2\ \ \sqrt{2}/2)^{\mathrm{T}}$. In pratica la matrice ortogonale

$$\mathbf{\Gamma} = \pm \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$$

risulta univocamente determinata a meno del segno. □

Esempio 8.5. Si consideri la matrice simmetrica di ordine 2 data da

$$\mathbf{A} = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} .$$

L'equazione caratteristica

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\begin{pmatrix} 5-\lambda & 0 \\ 0 & 5-\lambda \end{pmatrix} = (5-\lambda)^2 = 0$$

ha come soluzioni $\lambda_1 = \lambda_2 = 5$. Poichè l'autovalore $\lambda_1 = 5$ ha molteplicità $r_1 = 2$, per l'autospazio associato $\mathcal{A}_1$ risulta $\dim(\mathcal{A}_1) = 2$. Gli autovettori associati a questo autovalore si ottengono risolvendo il sistema lineare

$$(\mathbf{A} - 5\mathbf{I})\mathbf{x} = \mathbf{0}$$

con $\mathbf{x} = (x_1 \; x_2)^{\mathrm{T}}$, ovvero

$$\begin{aligned} 0x_1 + 0x_2 &= 0 \\ 0x_1 + 0x_2 &= 0 \, , \end{aligned}$$

da cui l'autospazio corrispondente coincide con tutto $\mathbb{R}^2$ in quanto ogni vettore $\mathbf{x}$ di $\mathbb{R}^2$ soddisfa il sistema. In questo caso esistono infinite coppie di vettori ortonormali $\boldsymbol{\gamma}_1$ e $\boldsymbol{\gamma}_2$ che costituiscono altrettante basi di $\mathbb{R}^2$. Dunque, $\mathbf{\Gamma}$ può essere una qualsiasi matrice ortogonale di ordine 2 (vedi Sezione 2.4). □

Risultato 8.2. Per gli autovalori e gli autovettori di particolari matrici simmetriche valgono le proprietà:
(i) gli autovalori di una matrice diagonale sono le componenti stesse della diagonale principale; inoltre se le componenti diagonali sono tutte distinte allora risulta $\mathbf{\Gamma} = \mathbf{I}_k$, ovvero $\boldsymbol{\gamma}_j = \mathbf{e}_j$ per $j = 1, 2, \ldots, k$; in caso contrario $\mathbf{\Gamma} = \mathbf{I}_k$ non è l'unica scelta possibile per $\mathbf{\Gamma}$;
(ii) gli autovalori di una matrice idempotente sono uguali a 1 o 0;
(iii) gli autovalori di una matrice ortogonale simmetrica sono 1 o -1.

Esempio 8.6. Si consideri la matrice diagonale di ordine 3 data da

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

L'equazione caratteristica

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\begin{pmatrix} 2-\lambda & 0 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix} = (2-\lambda)^2(1-\lambda) = 0$$

ha come soluzioni $\lambda_1 = \lambda_2 = 2$ e $\lambda_3 = 1$. Gli autovettori associati all'autovalore $\lambda_1 = 2$, con molteplicità $r_1 = 2$, si ottengono risolvendo il sistema lineare

$$(\mathbf{A} - 2\mathbf{I})\mathbf{x} = \mathbf{0}$$

con $\mathbf{x} = (x_1\ x_2\ x_3)^{\mathrm{T}}$, ovvero

$$\begin{aligned} 0x_1 + 0x_2 + 0x_3 &= 0 \\ 0x_1 + 0x_2 + 0x_3 &= 0 \\ 0x_1 + 0x_2 - x_3 &= 0\ , \end{aligned}$$

da cui l'autospazio corrispondente $\mathcal{A}_1$, tale che $\dim(\mathcal{A}_1) = 2$, è costituito da vettori del tipo $\mathbf{x} = (\beta_1\ \beta_2\ 0)^{\mathrm{T}}$ con $\beta_1, \beta_2 \in \mathbb{R}$. In questo caso esistono infinite coppie di vettori ortonormali $\boldsymbol{\gamma}_1$ e $\boldsymbol{\gamma}_2$ che costituiscono altrettante basi dell'autospazio e tenendo presente il vincolo di normalizzazione

$$\|\mathbf{x}\|^2 = \beta_1^2 + \beta_2^2 = 1$$

possono essere scelte come

$$\boldsymbol{\gamma}_1 = \pm \begin{pmatrix} \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} , \boldsymbol{\gamma}_2 = \pm \begin{pmatrix} -\sin\phi \\ \cos\phi \\ 0 \end{pmatrix} ,$$

dove $0 \leq \phi \leq 2\pi$. Inoltre, gli autovettori associati all'autovalore $\lambda_3 = 1$, con molteplicità $r_2 = 1$, si ottengono risolvendo il sistema lineare

$$(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}\ ,$$

ovvero

$$\begin{aligned} 1x_1 + 0x_2 + 0x_3 &= 0 \\ 0x_1 + 1x_2 + 0x_3 &= 0 \\ 0x_1 + 0x_2 + 0x_3 &= 0\ , \end{aligned}$$

da cui l'autospazio corrispondente $\mathcal{A}_2$, tale che $\dim(\mathcal{A}_2) = 1$, è costituito da vettori del tipo $\mathbf{x} = (0\ 0\ \beta_3)^{\mathrm{T}}$ con $\beta_3 \in \mathbb{R}$. In questo caso l'unica base ortonormale di $\mathcal{A}_2$ (a meno del segno) è data dal vettore normalizzato che si ottiene ponendo

$$\|\mathbf{x}\|^2 = \beta_3^2 = 1\ ,$$

da cui $\beta_3 = \pm 1$ e quindi $\boldsymbol{\gamma}_3 = \pm\, \mathbf{e}_3$. Si ha dunque

$$\boldsymbol{\Gamma} = \pm \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

Evidentemente, se $\phi = 0$ risulta $\boldsymbol{\Gamma} = \mathbf{I}_3$. □

Esempio 8.7. Si consideri la matrice idempotente simmetrica di ordine 3 data da

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} .$$

L'equazione caratteristica

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\begin{pmatrix} 1-\lambda & 0 & 0 \\ 0 & 1/2-\lambda & 1/2 \\ 0 & 1/2 & 1/2-\lambda \end{pmatrix} = \lambda(1-\lambda)^2 = 0$$

ha come soluzioni $\lambda_1 = \lambda_2 = 1$ e $\lambda_3 = 0$. □

Esempio 8.8. Si consideri la matrice ortogonale simmetrica di ordine 2 data da

$$\mathbf{A} = \begin{pmatrix} 4/5 & 3/5 \\ 3/5 & -4/5 \end{pmatrix} .$$

L'equazione caratteristica

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\begin{pmatrix} 4/5 - \lambda & 3/5 \\ 3/5 & -4/5 - \lambda \end{pmatrix} = \lambda^2 - 1 = 0$$

ha come soluzioni $\lambda_1 = 1$ e $\lambda_2 = -1$. □

8.3. Rappresentazione dei valori singolari

Se si considera una generica matrice **A** di ordine $(n \times k)$ tale che $\mathrm{r}(\mathbf{A}) = r$, si può ottenere una rappresentazione analoga a quella diagonale per matrici simmetriche discussa nella Sezione 8.2.

Si indichino con $\tau_1, \tau_2, \ldots, \tau_r$ gli autovalori non nulli della matrice simmetrica $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ e con $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_r$ i relativi autovettori ortonormali. Inoltre si considerino la matrice diagonale $\boldsymbol{\Delta} = \mathrm{dg}(\tau_1^{1/2}, \tau_2^{1/2}, \ldots, \tau_r^{1/2})$ di ordine r e le matrici $\mathbf{T} = (\boldsymbol{\delta}_1\ \boldsymbol{\delta}_2 \ldots \boldsymbol{\delta}_r)$ e $\mathbf{S} = \mathbf{AT}\boldsymbol{\Delta}^{-1}$ rispettivamente di ordine $(k \times r)$ e $(n \times r)$. Si può dimostrare che **S** è una matrice le cui colonne sono gli autovettori ortonormali corrispondenti agli autovalori non nulli della matrice $\mathbf{AA}^{\mathrm{T}}$. La *rappresentazione dei valori singolari* è data da

$$\mathbf{A} = \mathbf{S}\boldsymbol{\Delta}\mathbf{T}^{\mathrm{T}}\ .$$

Questa rappresentazione contiene come caso particolare la rappresentazione diagonale $\mathbf{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{\mathrm{T}}$ di una matrice simmetrica **A** con autovalori non nulli. In effetti, essendo in questo caso

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{AA}^{\mathrm{T}} = \mathbf{A}^2$$

e

$$\mathbf{A}^2 = (\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{\mathrm{T}})(\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{\mathrm{T}}) = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{\Gamma}^{\mathrm{T}}\ ,$$

allora $\mathbf{A}^2$ possiede gli stessi autovettori di **A** ed autovalori che sono gli autovalori di **A** al quadrato. Di conseguenza risulta $\mathbf{S} = \mathbf{T} = \boldsymbol{\Gamma}$ e $\boldsymbol{\Delta} = \boldsymbol{\Lambda}$.

La rappresentazione dei valori singolari permette anche di ottenere una espressione per il calcolo della matrice inversa di Moore-Penrose di **A**, ovvero

$$\mathbf{A}^+ = \mathbf{T}\boldsymbol{\Delta}^{-1}\mathbf{S}^{\mathrm{T}}\ .$$

In effetti le condizioni sull'inversa di Moore-Penrose risultano soddisfatte per questa espressione di $\mathbf{A}^+$, dal momento che essendo

$$\mathbf{T}^{\mathrm{T}}\mathbf{T} = \mathbf{S}^{\mathrm{T}}\mathbf{S} = \mathbf{I}_r$$

risulta

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = (\mathbf{S}\boldsymbol{\Delta}\mathbf{T}^{\mathrm{T}})(\mathbf{T}\boldsymbol{\Delta}^{-1}\mathbf{S}^{\mathrm{T}})(\mathbf{S}\boldsymbol{\Delta}\mathbf{T}^{\mathrm{T}}) = \mathbf{S}\boldsymbol{\Delta}\mathbf{T}^{\mathrm{T}} = \mathbf{A} \, ,$$

$$\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = (\mathbf{T}\boldsymbol{\Delta}^{-1}\mathbf{S}^{\mathrm{T}})(\mathbf{S}\boldsymbol{\Delta}\mathbf{T}^{\mathrm{T}})(\mathbf{T}\boldsymbol{\Delta}^{-1}\mathbf{S}^{\mathrm{T}}) = \mathbf{T}\boldsymbol{\Delta}^{-1}\mathbf{S}^{\mathrm{T}} = \mathbf{A}^+ \, ,$$

$$(\mathbf{A}\mathbf{A}^+)^{\mathrm{T}} = (\mathbf{S}\boldsymbol{\Delta}^{-1}\mathbf{T}^{\mathrm{T}})(\mathbf{T}\boldsymbol{\Delta}\mathbf{S}^{\mathrm{T}}) = \mathbf{S}\mathbf{S}^{\mathrm{T}} = (\mathbf{S}\boldsymbol{\Delta}\mathbf{T}^{\mathrm{T}})(\mathbf{T}\boldsymbol{\Delta}^{-1}\mathbf{S}^{\mathrm{T}}) = \mathbf{A}\mathbf{A}^+ \, ,$$

$$(\mathbf{A}^+\mathbf{A})^{\mathrm{T}} = (\mathbf{T}\boldsymbol{\Delta}\mathbf{S}^{\mathrm{T}})(\mathbf{S}\boldsymbol{\Delta}^{-1}\mathbf{T}^{\mathrm{T}}) = \mathbf{T}\mathbf{T}^{\mathrm{T}} = (\mathbf{T}\boldsymbol{\Delta}^{-1}\mathbf{S}^{\mathrm{T}})(\mathbf{S}\boldsymbol{\Delta}\mathbf{T}^{\mathrm{T}}) = \mathbf{A}^+\mathbf{A} \, .$$

Esempio 8.9. Si consideri di nuovo la matrice $\mathbf{A}$ di ordine (3×2) dell'Esempio 5.9. Evidentemente si ha $\mathrm{r}(\mathbf{A}) = 2$, mentre

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} .$$

Gli autovalori della matrice $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ risultano $\tau_1 = 3$ e $\tau_2 = 1$ e i relativi autovettori (a meno del segno) sono dati da $\boldsymbol{\delta}_1 = (\sqrt{2}/2 \ \sqrt{2}/2)^{\mathrm{T}}$ e $\boldsymbol{\delta}_2 = (-\sqrt{2}/2 \ \ \sqrt{2}/2)^{\mathrm{T}}$. Dunque si ha

$$\boldsymbol{\Delta} = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} , \mathbf{T} = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} ,$$

mentre

$$\mathbf{S} = \mathbf{A}\mathbf{T}\boldsymbol{\Delta}^{-1} = \begin{pmatrix} \sqrt{6}/6 & -\sqrt{2}/2 \\ \sqrt{6}/3 & 0 \\ \sqrt{6}/6 & \sqrt{2}/2 \end{pmatrix} .$$

La rappresentazione del valore singolare è quindi data da

$$\mathbf{A} = \mathbf{S\Delta T}^{\mathrm{T}} = \begin{pmatrix} \sqrt{6}/6 & -\sqrt{2}/2 \\ \sqrt{6}/3 & 0 \\ \sqrt{6}/6 & \sqrt{2}/2 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} .$$

Inoltre la matrice inversa di Moore-Penrose di **A** risulta

$$\mathbf{A}^{+} = \mathbf{T\Delta}^{-1}\mathbf{S}^{\mathrm{T}} = \begin{pmatrix} 2/3 & 1/3 & -1/3 \\ -1/3 & 1/3 & 2/3 \end{pmatrix} ,$$

coerentemente con i risultati dell'Esempio 5.10. □

8.4. Applicazioni di matrici

Il problema di definire una applicazione $\mathbf{F} : \mathbb{R}^{nk} \to \mathbb{R}^{nk}$, che associa ad una matrice $\mathbf{X}$ di ordine $(n \times k)$ una matrice $\mathbf{Y} = \mathbf{F}(\mathbf{X})$ dello stesso ordine, risulta piuttosto complesso. Per semplicità conviene ridurre di nuovo l'analisi alla definizione di applicazione di matrice simmetrica.

In primo luogo, data una matrice diagonale $\mathbf{A} = \mathrm{dg}(a_1, a_2, \ldots, a_k)$ di ordine k e data una funzione $f : \mathbb{R} \to \mathbb{R}$, si assume come $f(\mathbf{A})$ la matrice diagonale

$$f(\mathbf{A}) = \mathrm{dg}(f(a_1), f(a_2), \ldots, f(a_k)) .$$

Questa notazione (non troppo coerente) viene comunemente adottata in letteratura, anche se $f(\mathbf{A})$ rappresenta in effetti una matrice quadrata di ordine k. Si noti che la matrice $f(\mathbf{A})$ è definita per tutte le matrici diagonali per le quali risultano definiti $f(a_1), f(a_2), \ldots, f(a_k)$.

Esempio 8.10. Si consideri la matrice simmetrica di ordine 2

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \mathrm{dg}(1, 2) ,$$

per cui risulta

$$\exp(\mathbf{A}) = \begin{pmatrix} \exp(1) & 0 \\ 0 & \exp(2) \end{pmatrix} = \mathrm{dg}(\exp(1), \exp(2)) .$$ □

Estendendo la precedente definizione, data una matrice simmetrica **A** di ordine k con rappresentazione diagonale $\mathbf{A} = \mathbf{\Gamma\Lambda\Gamma}^{\mathrm{T}}$ e data una funzione $f : \mathbb{R} \to \mathbb{R}$, si assume come $f(\mathbf{A})$ la matrice

$$f(\mathbf{A}) = \mathbf{\Gamma} f(\mathbf{\Lambda})\mathbf{\Gamma}^{\mathrm{T}} \; .$$

La matrice $f(\mathbf{A})$ risulta definita per tutte le matrici con autovalori per i quali risultano definiti $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_k)$. Poichè la precedente espressione costituisce ancora una rappresentazione diagonale, $f(\mathbf{A})$ è una matrice simmetrica che ha lo stesso insieme di k autovettori ortonormali della matrice **A** e autovalori $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_k)$.

Se si considera la funzione $f(x) = x^{-1}$, applicando la precedente definizione si ottiene che

$$\mathbf{A}^{-1} = \mathbf{\Gamma}\mathrm{dg}(\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_k^{-1})\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{\Gamma\Lambda}^{-1}\mathbf{\Gamma}^{\mathrm{T}} \; .$$

Poichè

$$\mathbf{A}^{-1}\mathbf{A} = (\mathbf{\Gamma\Lambda}^{-1}\mathbf{\Gamma}^{\mathrm{T}})(\mathbf{\Gamma\Lambda\Gamma}^{\mathrm{T}}) = \mathbf{\Gamma\Gamma}^{\mathrm{T}} = \mathbf{I}_k \; ,$$

si deduce che l'inversa di una matrice simmetrica è una matrice simmetrica con lo stesso insieme di autovettori ortonormali e autovalori che sono i reciproci degli autovalori della matrice originale. Ovviamente $\mathbf{A}^{-1}$ esiste se e solo se **A** è a rango pieno dal momento che solo in questo caso si ha $\lambda_i \neq 0$ per ogni $i = 1, 2, \ldots, k$. Quindi questi risultati sono coerenti con la definizione di matrice inversa data nella Sezione 5.1.

Esempio 8.11. Si consideri la matrice simmetrica di ordine 2

$$\mathbf{A} = \begin{pmatrix} 10 & 4 \\ 4 & 4 \end{pmatrix} ,$$

per cui risulta

$$\mathbf{\Lambda} = \begin{pmatrix} 12 & 0 \\ 0 & 2 \end{pmatrix} , \mathbf{\Gamma} = \pm \begin{pmatrix} 2\sqrt{5}/5 & -\sqrt{5}/5 \\ \sqrt{5}/5 & 2\sqrt{5}/5 \end{pmatrix} .$$

In questo caso si ha

$$\mathbf{A}^{-1} = \mathbf{\Gamma\Lambda}^{-1}\mathbf{\Gamma}^{\mathrm{T}} = \begin{pmatrix} 1/6 & -1/6 \\ -1/6 & 5/12 \end{pmatrix} ,$$

che ovviamente coincide con l'usuale matrice inversa. □

Se si considera la funzione $f(x) = x^{1/2}$, applicando la definizione si ottiene che

$$\mathbf{A}^{1/2} = \mathbf{\Gamma}\mathrm{dg}(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_k^{1/2})\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{\Gamma}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}^{\mathrm{T}} \; .$$

Poichè

$$\mathbf{A}^{1/2}\mathbf{A}^{1/2} = (\mathbf{\Gamma}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}^{\mathrm{T}})(\mathbf{\Gamma}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}^{\mathrm{T}}) = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{A} \; ,$$

la matrice $\mathbf{A}^{1/2}$ è detta radice quadrata di $\mathbf{A}$. Evidentemente $\mathbf{A}^{1/2}$ esiste se e solo se $\lambda_i \geq 0$ per ogni $i = 1, 2, \ldots, k$.

Esempio 8.12. Si consideri la matrice simmetrica di ordine 2

$$\mathbf{A} = \begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{pmatrix} ,$$

per cui si ha

$$\mathbf{\Lambda} = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} , \mathbf{\Gamma} = \pm \begin{pmatrix} \sqrt{6}/3 & -\sqrt{3}/3 \\ \sqrt{3}/3 & \sqrt{6}/3 \end{pmatrix} .$$

In questo caso risulta

$$\mathbf{A}^{1/2} = \mathbf{\Gamma}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}^{\mathrm{T}} = \begin{pmatrix} 5/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 4/3 \end{pmatrix} .$$

In effetti si ha

$$\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \begin{pmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{pmatrix} = \mathbf{A} \; .$$ □

Se si considera la funzione $f(x) = x^{-1/2}$, applicando la definizione si ottiene che

$$\mathbf{A}^{-1/2} = \mathbf{\Gamma}\mathrm{dg}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_k^{-1/2})\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{\Gamma}\mathbf{\Lambda}^{-1/2}\mathbf{\Gamma}^{\mathrm{T}} \; .$$

Poichè

$$\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = (\mathbf{\Gamma}\mathbf{\Lambda}^{-1/2}\mathbf{\Gamma}^{\mathrm{T}})(\mathbf{\Gamma}\mathbf{\Lambda}^{-1/2}\mathbf{\Gamma}^{\mathrm{T}}) = \mathbf{\Gamma}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{A}^{-1}$$

$\mathbf{A}^{-1/2}$ è detta radice quadrata di $\mathbf{A}^{-1}$. Evidentemente $\mathbf{A}^{-1/2}$ esiste se e solo se $\lambda_i > 0$ per ogni $i = 1, 2, \ldots, k$.

Esempio 8.13. Si consideri di nuovo la matrice dell'Esempio 8.12. In questo caso risulta

$$\mathbf{A}^{-1/2} = \mathbf{\Gamma}\mathbf{\Lambda}^{-1/2}\mathbf{\Gamma}^{\mathrm{T}} = \begin{pmatrix} 2/3 & -\sqrt{2}/6 \\ -\sqrt{2}/6 & 5/6 \end{pmatrix} .$$

In effetti si ha

$$\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \begin{pmatrix} 1/2 & -\sqrt{2}/4 \\ -\sqrt{2}/4 & 3/4 \end{pmatrix} = \mathbf{A}^{-1} . \qquad \square$$

Capitolo 9

Forme quadratiche

9.1. Definizioni e notazioni

Dato un vettore $\mathbf{x} = (x_1 \; x_2 \; \ldots \; x_k)^{\mathrm{T}}$ di ordine k e una matrice simmetrica $\mathbf{A} = (a_{ij})$ dello stesso ordine, il prodotto

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}x_i x_j$$

è detto *forma quadratica*.

Nel trattamento delle forme quadratiche non è riduttivo assumere $\mathbf{A}$ matrice simmetrica. In effetti, se $\mathbf{A}$ non è simmetrica, essendo

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{x}$$

si può scrivere

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \frac{1}{2}\,\mathbf{x}^{\mathrm{T}}(\mathbf{A} + \mathbf{A}^{\mathrm{T}})\mathbf{x}$$

e tenere presente che $(\mathbf{A} + \mathbf{A}^{\mathrm{T}})$ è una matrice simmetrica.

Se $\mathbf{A} = \mathrm{dg}(a_1, a_2, \ldots, a_k)$ la forma quadratica corrispondente è una somma ponderata di quadrati in quanto

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \sum_{i=1}^{k} a_i x_i^2 \; ,$$

mentre se $\mathbf{A} = \mathbf{I}$ la forma quadratica coincide con il prodotto interno $\mathbf{x}^{\mathrm{T}}\mathbf{x}$. Inoltre, per le proprietà del prodotto di Cayley (vedi Sezione 3.1), valgono le identità

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathrm{tr}(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \mathrm{tr}(\mathbf{A}\mathbf{x}\mathbf{x}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{A}) \; .$$

Una forma quadratica e la matrice corrispondente sono dette *definite positive* se per ogni $\mathbf{x} \neq \mathbf{0}$ si ha

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} > 0 \; ,$$

mentre sono dette *semidefinite positive* se per ogni $\mathbf{x}$ si ha

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \geq 0 \; .$$

Se una matrice $\mathbf{A}$ è definita positiva si scrive $\mathbf{A} > \mathbf{O}$, mentre se una matrice $\mathbf{A}$ è semidefinita positiva si scrive $\mathbf{A} \geq \mathbf{O}$. Queste definizioni estendono il concetto di positività nell'ambito delle matrici.

In modo analogo, una forma quadratica e la matrice corrispondente sono dette *definite negative* se per ogni $\mathbf{x} \neq \mathbf{0}$ si ha

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} < 0 \; ,$$

mentre sono dette *semidefinite negative* se per ogni $\mathbf{x}$ si ha

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \leq 0 \; .$$

Se una matrice $\mathbf{A}$ è definita negativa si scrive $\mathbf{A} < \mathbf{O}$, mentre se una matrice $\mathbf{A}$ è semidefinita negativa si scrive $\mathbf{A} \leq \mathbf{O}$. Una forma quadratica e la matrice corrispondente sono dette indefinite se non possono essere classificate nè come (semi)definite positive nè come (semi)definite negative.

Infine, con la notazione $\mathbf{A} > \mathbf{B}$ si indica che $\mathbf{A} - \mathbf{B} > \mathbf{O}$. Simili definizioni valgono per le notazioni $\mathbf{A} \geq \mathbf{B}$, $\mathbf{A} < \mathbf{B}$ e $\mathbf{A} \leq \mathbf{B}$.

Esempio 9.1. La forma quadratica associata alla matrice simmetrica di ordine 2

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

è definita positiva in quanto per ogni $\mathbf{x} \in \mathbb{R}^2$ tale che $\mathbf{x} \neq \mathbf{0}$ si ha

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = x_1^2 + x_2^2 > 0 \; .$$

La forma quadratica associata alla matrice

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

è semidefinita positiva in quanto per ogni $\mathbf{x} \in \mathbb{R}^2$ si ha

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = x_1^2 + x_2^2 - 2x_1x_2 = (x_1 - x_2)^2 \geq 0 \; ,$$

essendo $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = 0$ per i vettori $\mathbf{x}$ per cui risulta $x_1 = x_2$. La forma quadratica associata alla matrice

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} ,$$

ovvero

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = x_1^2 - x_2^2 \; ,$$

è indefinita in quanto si ha $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \geq 0$ se $|x_1| \geq |x_2|$ e $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} < 0$ altrimenti. □

Esempio 9.2. Date le matrici simmetriche di ordine 2

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} , \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} ,$$

per ogni $\mathbf{x} \in \mathbb{R}^2$ si ha

$$\mathbf{x}^{\mathrm{T}}(\mathbf{A} - \mathbf{B})\mathbf{x} = x_1^2 + x_2^2 + 2x_1x_2 = (x_1 + x_2)^2 \geq 0 \; ,$$

essendo $\mathbf{x}^{\mathrm{T}}(\mathbf{A} - \mathbf{B})\mathbf{x} = 0$ per i vettori $\mathbf{x}$ per cui risulta $x_1 = -x_2$. Si ha quindi $\mathbf{A} \geq \mathbf{B}$. □

Risultato 9.1. Per le matrici definite (semi)positive e (semi)negative si hanno le proprietà:
(i) se $\mathbf{A} \geq \mathbf{O}$ si ha

$$0 \leq \det(\mathbf{A}) \leq \det(\mathrm{dg}(\mathbf{A})) = \prod_{i=1}^{k} a_{ii} \; ;$$

(ii) gli autovalori di $\mathbf{A}$ sono positivi se e solo se $\mathbf{A} > \mathbf{O}$;
(iii) gli autovalori di $\mathbf{A}$ sono non negativi se e solo se $\mathbf{A} \geq \mathbf{O}$;
(iv) gli autovalori di $\mathbf{A}$ sono negativi se e solo se $\mathbf{A} < \mathbf{O}$;
(v) gli autovalori di $\mathbf{A}$ sono non positivi se e solo se $\mathbf{A} \leq \mathbf{O}$.

Dalle proprietà (ii) e (iii) del Risultato 9.1 risulta che $\mathbf{A}^{1/2}$ esiste se e solo se $\mathbf{A} \geq \mathbf{O}$, mentre $\mathbf{A}^{-1/2}$ esiste se e solo se $\mathbf{A} > \mathbf{O}$.

Esempio 9.3. Si consideri la matrice simmetrica

$$\mathbf{A} = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix} .$$

Dal momento che per ogni vettore **x** di ordine 2 tale che $\mathbf{x} \neq \mathbf{O}$ si ha

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = 5x_1^2 + 2x_2^2 + 4x_1x_2 = x_1^2 + x_2^2 + (2x_1 + x_2)^2 > 0 ,$$

la forma quadratica associata alla matrice **A** è definita positiva. Inoltre, la corrispondente equazione caratteristica

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\begin{pmatrix} 5-\lambda & 2 \\ 2 & 2-\lambda \end{pmatrix} = \lambda^2 - 7\lambda + 6 = 0$$

ha come soluzioni gli autovalori $\lambda_1 = 6$ e $\lambda_2 = 1$ e quindi in effetti risulta $\lambda_1, \lambda_2 > 0$. □

Infine, data la matrice **A** di ordine $(n \times k)$, per ogni **x** la forma quadratica corrispondente ad $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ può essere scritta come

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^{\mathrm{T}}(\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2 \geq 0 .$$

Inoltre risulta $\|\mathbf{A}\mathbf{x}\|^2 = 0$ se e solo se $\mathbf{A}\mathbf{x} = \mathbf{0}$, ma questo si verifica se e solo se $\mathrm{r}(\mathbf{A}) = \mathrm{r}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) < k$ (vedi Risultato 4.2 e Sezione 7.2). Dunque, si ha che $\mathbf{A}^{\mathrm{T}}\mathbf{A} > \mathbf{O}$ se $\mathrm{r}(\mathbf{A}) = k$, mentre $\mathbf{A}^{\mathrm{T}}\mathbf{A} \geq \mathbf{O}$ se $\mathrm{r}(\mathbf{A}) < k$. Analogamente, si può verificare che $\mathbf{A}\mathbf{A}^{\mathrm{T}} > \mathbf{O}$ se $\mathrm{r}(\mathbf{A}) = n$, mentre $\mathbf{A}\mathbf{A}^{\mathrm{T}} \geq \mathbf{O}$ se $\mathrm{r}(\mathbf{A}) < n$.

9.2. Ellissoidi

Data una matrice simmetrica $\mathbf{A} > \mathbf{O}$ di ordine k e un vettore **b** dello stesso ordine, si consideri l'equazione nel vettore **x** di ordine k tale che

$$(\mathbf{x} - \mathbf{b})^{\mathrm{T}}\mathbf{A}^{-1}(\mathbf{x} - \mathbf{b}) = c$$

con $c > 0$. Il luogo geometrico che si ottiene al variare del vettore $\mathbf{x}$ rappresenta un *ellissoide* in $\mathbb{R}^k$. Se $k = 2$ il luogo geometrico corrisponde all'usuale *ellisse* in $\mathbb{R}^2$.

L'ellissoide ha centro in $\mathbf{b}$ ed assi le cui direzioni sono determinate dalle direzioni dei k autovettori ortonormali $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_k$ associati alla matrice $\mathbf{A}$. Inoltre, le lunghezze degli assi sono pari a $2\sqrt{c\lambda_i}$ dove $\lambda_1, \lambda_2, \ldots, \lambda_k$ sono gli autovalori associati alla matrice $\mathbf{A}$. Si noti inoltre che al variare di c si ha una famiglia di ellissoidi concentriche la cui "ampiezza" è funzione crescente di c.

Dal momento che $\mathbf{A}^{-1}$ è l'inversa di una matrice simmetrica, tenendo presente che $\mathbf{A}^{-1} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}$ (vedi Sezione 8.4), l'equazione dell'ellissoide può essere riscritta nel seguente modo

$$(\mathbf{x} - \mathbf{b})^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}(\mathbf{x} - \mathbf{b}) = c \; .$$

Operando dunque la trasformazione ortogonale associata ad una traslazione, ovvero

$$\mathbf{y} = \boldsymbol{\Gamma}^{\mathrm{T}}(\mathbf{x} - \mathbf{b}) \; ,$$

l'equazione dell'ellissoide si riduce a

$$\mathbf{y}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\mathbf{y} = c \; .$$

La precedente espressione rappresenta ancora un ellissoide con centro $\mathbf{b} = \mathbf{0}$, con assi le cui direzioni sono determinate dall'insieme dei k autovettori ortonormali associati alla matrice diagonale $\boldsymbol{\Lambda}$, ovvero $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$, e le cui lunghezze sono determinate dagli autovalori di $\boldsymbol{\Lambda}$, ovvero $\lambda_1, \lambda_2, \ldots, \lambda_k$.

In pratica quindi questa trasformazione modifica l'ellissoide originale in un ellissoide centrato nell'origine, della stessa forma e ampiezza del precedente ma ruotato in modo che le direzioni dei suoi assi coincidano con quelle degli assi cartesiani. Per questo motivo la trasformazione è detta di *diagonalizzazione* delle forme quadratiche.

Dal momento che $\mathbf{A} > \mathbf{O}$, esiste la matrice $\mathbf{A}^{-1/2}$ tale che $\mathbf{A}^{-1} = \mathbf{A}^{-1/2}\mathbf{A}^{-1/2}$ (vedi Sezione 8.4). Quindi l'espressione dell'ellissoide può essere riscritta come

$$(\mathbf{x} - \mathbf{b})^{\mathrm{T}}\mathbf{A}^{-1/2}\mathbf{A}^{-1/2}(\mathbf{x} - \mathbf{b}) = c \; .$$

Operando la seguente applicazione lineare associata a traslazione

$$\mathbf{u} = \mathbf{A}^{-1/2}(\mathbf{x} - \mathbf{b})$$

si ottiene inoltre

$$\mathbf{u}^{\mathrm{T}}\mathbf{u} = c\,,$$

che geometricamente coincide con una *ipersfera* di raggio $\sqrt{c}$ centrata in $\mathbf{0}$. In pratica quindi questa trasformazione modifica l'ellissoide originale in un ellissoide centrato nell'origine, ruotato in modo che le direzioni dei suoi assi coincidano con quelle degli assi cartesiani e deformato in modo che le lunghezze di tutti gli assi risultino uguali. Per questo motivo la trasformazione è detta di *sfericizzazione* delle forme quadratiche.

Esempio 9.4. Si consideri l'ellisse caratterizzata da

$$\mathbf{A} = \begin{pmatrix} 13/4 & 3\sqrt{3}/4 \\ 3\sqrt{3}/4 & 7/4 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

con $c = 1$. Gli autovalori di $\mathbf{A}$ sono dati da $\lambda_1 = 4$ e $\lambda_2 = 1$, mentre i corrispondenti autovettori sono dati da $\boldsymbol{\gamma}_1 = (\sqrt{3}/2\ \ 1/2)^{\mathrm{T}}$ e $\boldsymbol{\gamma}_2 = (-1/2\ \ \sqrt{3}/2)^{\mathrm{T}}$. Di conseguenza l'ellisse è centrata in $\mathbf{b}$ ed ha il primo asse di lunghezza $2\sqrt{c\lambda_1} = 4$ con un inclinazione pari all'arco $\pi/6$ rispetto alle ascisse, mentre ha il secondo asse di lunghezza $2\sqrt{c\lambda_2} = 2$ con una inclinazione pari all'arco $2\pi/3$ rispetto alle ascisse (vedi Figura 9.1). Nel caso in cui $c = 9/4$ le dimensioni degli assi dell'ellisse diventano rispettivamente $2\sqrt{c\lambda_1} = 6$ e $2\sqrt{c\lambda_2} = 3$, ovvero le dimensioni dell'ellisse aumentano di $\sqrt{c} = 3/2$ (vedi Figura 9.2). Operando la trasformazione di diagonalizzazione delle forme quadratiche, l'ellisse originale viene trasformata in una ellisse centrata nell'origine, della stessa forma e ampiezza della precedente ma ruotata in modo che le direzioni dei suoi assi coincidano con quelle degli assi cartesiani (vedi Figura 9.3). Operando la trasformazione di sfericizzazione delle forme quadratiche, l'ellisse originale viene trasformata in una ellisse centrata nell'origine, ruotata in modo che le direzioni dei suoi assi coincidano con quelle degli assi cartesiani e deformata in modo che le lunghezze di tutti gli assi risultino uguali e pari a $\sqrt{c} = 1$ (vedi Figura 9.4). □

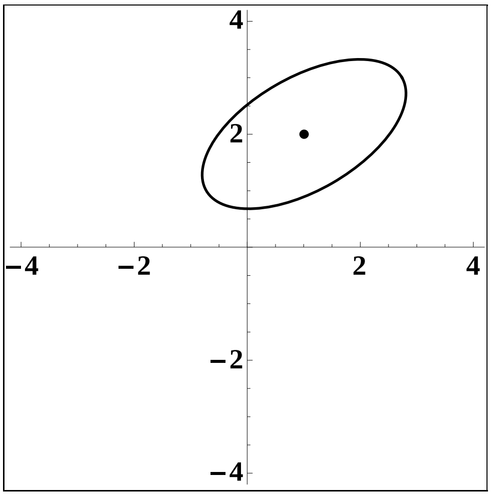

Figura 9.1.

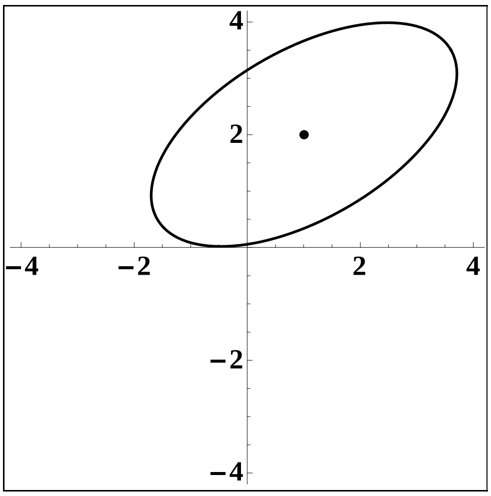

Figura 9.2.

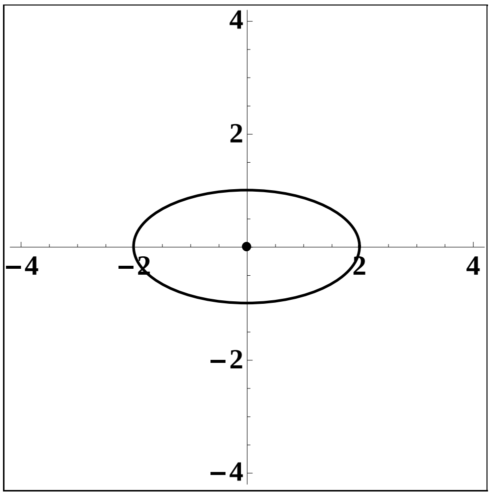

Figura 9.3.

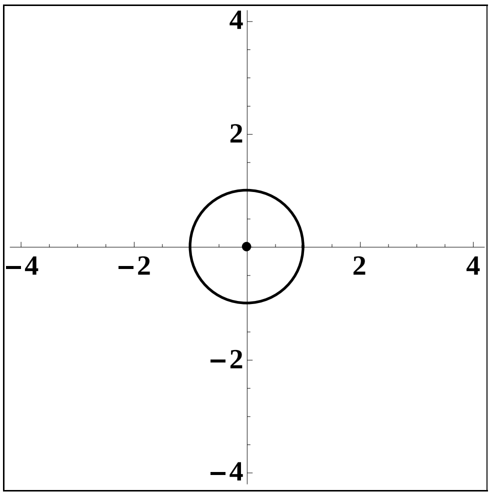

Figura 9.4.

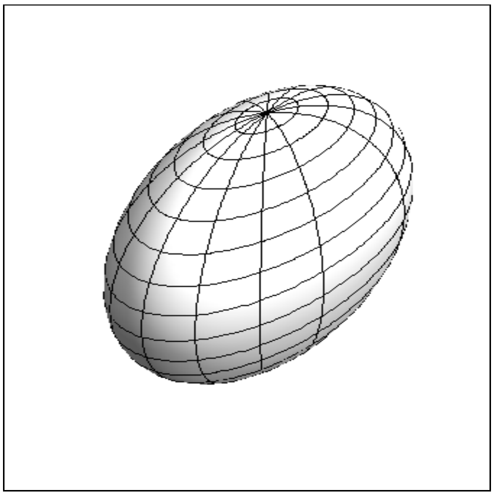

Figura 9.5.

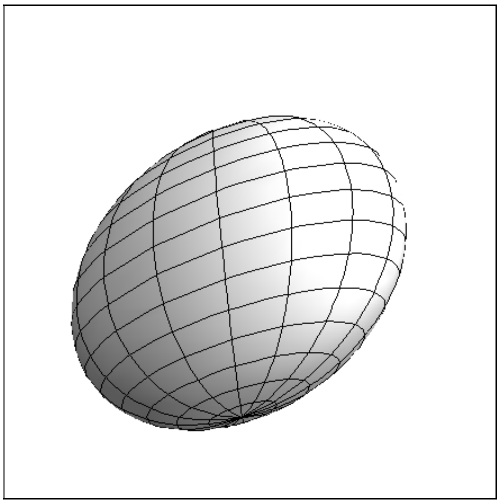

Figura 9.6.

Esempio 9.5. Si consideri l'ellissoide in $\mathbb{R}^3$ caratterizzato da

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9/4 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

con $c = 1$. Gli autovalori di **A** sono dati da $\lambda_1 = 1$, $\lambda_2 = 4$ e $\lambda_3 = 9/4$, mentre i corrispondenti autovettori risultano $\mathbf{e}_1$, $\mathbf{e}_2$ e $\mathbf{e}_3$. In questo caso gli autovalori non sono stati ordinati per semplicità di notazione degli autovettori. Di conseguenza, l'ellissoide è centrato in **b** ed ha gli assi di lunghezza $2\sqrt{c\lambda_1} = 1$, $2\sqrt{c\lambda_2} = 2$ e $2\sqrt{c\lambda_3} = 3/2$, mentre le direzioni dei suoi assi coincidono con quelle degli assi cartesiani. Due grafici dell'ellissoide da differenti punti di vista sono riportati nelle Figure 9.5 e 9.6. □

Capitolo 10

Calcolo differenziale per matrici

10.1. Derivazione di funzioni di variabile vettoriale

Si considerino un vettore $\mathbf{x} = (x_1\ x_2\ \dots\ x_k)^{\mathrm{T}}$ di ordine k ed una funzione $f(\mathbf{x})$ tale che $f : \mathbb{R}^k \to \mathbb{R}$, ovvero f rappresenta una *funzione di variabile vettoriale*. Si definisce *gradiente*

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_k} \end{pmatrix} ,$$

ovvero $\partial f(\mathbf{x})/\partial \mathbf{x}$ rappresenta un vettore colonna di ordine k la cui i-esima componente è data dalla derivata di f fatta rispetto a x_i, mentre si definisce

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{T}}} = \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right)^{\mathrm{T}} .$$

Esempio 10.1. Se $f(\mathbf{x}) = \mathbf{a}^{\mathrm{T}}\mathbf{x}$ dove $\mathbf{a} = (a_1\ a_2\ \dots\ a_k)^{\mathrm{T}}$ rappresenta un vettore di costanti di ordine k, tenendo presente che il prodotto interno è dato da

$$\mathbf{a}^{\mathrm{T}}\mathbf{x} = \sum_{j=1}^{k} a_j x_j ,$$

allora si ha

$$\frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{x})}{\partial x_i} = \frac{\partial}{\partial x_i}\sum_{j=1}^{k} a_j x_j = a_i\ ,$$

da cui segue che

$$\frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}\ ,$$

mentre

$$\frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{x})}{\partial \mathbf{x}^{\mathrm{T}}} = \mathbf{a}^{\mathrm{T}}\ .$$

È immediato verificare inoltre che

$$\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{a})}{\partial \mathbf{x}} = \mathbf{a}\ ,$$

mentre

$$\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{a})}{\partial \mathbf{x}^{\mathrm{T}}} = \mathbf{a}^{\mathrm{T}}\ . \qquad \square$$

Esempio 10.2. Se $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x}$, tenendo presente la definizione di prodotto interno di un vettore per se stesso

$$\mathbf{x}^{\mathrm{T}}\mathbf{x} = \sum_{j=1}^{k} x_j^2\ ,$$

si ha

$$\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{x})}{\partial x_i} = \frac{\partial}{\partial x_i}\sum_{j=1}^{k} x_j^2 = 2x_i\ ,$$

da cui segue che

$$\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{x}\ ,$$

mentre

$$\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{x})}{\partial\mathbf{x}^{\mathrm{T}}} = 2\mathbf{x}^{\mathrm{T}} . \qquad \square$$

Esempio 10.3. Se $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ con $\mathbf{A} = (a_{ij})$ matrice simmetrica di ordine k di costanti, esplicitando la forma quadratica si ottiene

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \sum_{j=1}^{k} a_{jj}x_j^2 + 2\sum_{j=1}^{k-1}\sum_{l=j+1}^{k} a_{jl}x_j x_l ,$$

per cui

$$\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial x_i} = 2\sum_{l=1}^{k} a_{il}x_l = 2\mathbf{a}_{i\bullet}^{\mathrm{T}}\mathbf{x} ,$$

dove $\mathbf{a}_{i\bullet}^{\mathrm{T}}$ rappresenta l'i-esimo vettore riga di $\mathbf{A}$. Quindi si ha

$$\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial\mathbf{x}} = 2\begin{pmatrix} \mathbf{a}_{1\bullet}^{\mathrm{T}}\mathbf{x} \\ \mathbf{a}_{2\bullet}^{\mathrm{T}}\mathbf{x} \\ \vdots \\ \mathbf{a}_{k\bullet}^{\mathrm{T}}\mathbf{x} \end{pmatrix} = 2\mathbf{A}\mathbf{x} ,$$

mentre

$$\frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial\mathbf{x}^{\mathrm{T}}} = 2\mathbf{x}^{\mathrm{T}}\mathbf{A} . \qquad \square$$

Esempio 10.4. Se **a** e **b** sono vettori di costanti di ordine k, si ha

$$\frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{b})}{\partial\mathbf{x}} = \frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{a})}{\partial\mathbf{x}} = \frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{a})}{\partial\mathbf{x}} = \frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{b})}{\partial\mathbf{x}} = \mathbf{0} ,$$

mentre

$$\frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{b})}{\partial\mathbf{x}^{\mathrm{T}}} = \frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{a})}{\partial\mathbf{x}^{\mathrm{T}}} = \frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{a})}{\partial\mathbf{x}^{\mathrm{T}}} = \frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{b})}{\partial\mathbf{x}^{\mathrm{T}}} = \mathbf{0}^{\mathrm{T}} . \qquad \square$$

10.2. Derivazione di funzioni di variabile matriciale

Si considerino una matrice $\mathbf{X} = (x_{ij})$ di ordine $(n \times k)$ ed una funzione $f(\mathbf{X})$ tale che $f : \mathbb{R}^{nk} \to \mathbb{R}$, ovvero f rappresenta una *funzione di variabile matriciale*. Ad esempio, una funzione di questo tipo è il determinante, il quale associa a ogni matrice quadrata $\mathbf{X}$ di ordine k un numero reale $\det(\mathbf{X})$, ovvero risulta $\det(\,\cdot\,) : \mathbb{R}^{k^2} \to \mathbb{R}$. In questo caso, si definisce

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \begin{pmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \frac{\partial f(\mathbf{X})}{\partial x_{12}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1k}} \\ \frac{\partial f(\mathbf{X})}{\partial x_{21}} & \frac{\partial f(\mathbf{X})}{\partial x_{22}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{2k}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{n1}} & \frac{\partial f(\mathbf{X})}{\partial x_{n2}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{nk}} \end{pmatrix} ,$$

ovvero $\partial f(\mathbf{X})/\partial \mathbf{X}$ rappresenta una matrice di ordine $(n \times k)$ la cui componente di posto (i, j) è data dalla derivata di f fatta rispetto alla variabile x_{ij}, mentre si definisce

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}} = \left(\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \right)^{\mathrm{T}} .$$

Risultato 10.1. Per quanto riguarda la derivazione di alcune particolari funzioni di variabile matriciale, valgono i seguenti risultati:
(i) se $\mathbf{X}$ è una matrice simmetrica di ordine k

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}' - \mathrm{dg}(\mathbf{X}') ,$$

dove $\mathbf{X}'$ indica la matrice aggiunta di $\mathbf{X}$;
(ii) se $\mathbf{X}$ è una matrice simmetrica di ordine k tale che $\mathbf{X} > \mathbf{O}$

$$\frac{\partial \ln \det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - \mathrm{dg}(\mathbf{X}^{-1}) ;$$

(iii) se $\mathbf{X}$ ed $\mathbf{A}$ sono due matrici simmetriche di ordine k

$$\frac{\partial\, \mathrm{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \frac{\partial\, \mathrm{tr}(\mathbf{XA})}{\partial \mathbf{X}} = 2\mathbf{A} - \mathrm{dg}(\mathbf{A})$$

e quindi se $\mathbf{A} = \mathbf{I}_k$ si ottiene

$$\frac{\partial\, \mathrm{tr}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I}_k \ ;$$

(iv) se $\mathbf{X}$ è una matrice simmetrica di ordine k ed $\mathbf{a}$ è un vettore dello stesso ordine

$$\frac{\partial(\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{a})}{\partial \mathbf{X}} = 2\mathbf{a}\mathbf{a}^{\mathrm{T}} - \mathrm{dg}(\mathbf{a}\mathbf{a}^{\mathrm{T}}) \ ;$$

(v) se $\mathbf{X}$ è una matrice diagonale di ordine k, $\mathbf{A}$ è una matrice di ordine $(n \times k)$ e $\mathbf{B}$ è una matrice quadrata di ordine k

$$\frac{\partial \ln \det(\mathbf{X} + \mathbf{A}^{\mathrm{T}}\mathbf{A})}{\partial \mathbf{X}} = \mathrm{dg}((\mathbf{X} + \mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1})$$

e

$$\frac{\partial\, \mathrm{tr}((\mathbf{X} + \mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{B})}{\partial \mathbf{X}} = \mathrm{dg}((\mathbf{X} + \mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{B}(\mathbf{X} + \mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}) \ ;$$

(vi) se $\mathbf{A}$ è una matrice diagonale di ordine k, $\mathbf{X}$ è una matrice di ordine $(n \times k)$ e $\mathbf{B}$ è una matrice quadrata di ordine k

$$\frac{\partial \ln \det(\mathbf{A} + \mathbf{X}^{\mathrm{T}}\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}(\mathbf{A} + \mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \ ,$$

e

$$\frac{\partial\, \mathrm{tr}((\mathbf{A} + \mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{B})}{\partial \mathbf{X}} = 2\mathbf{X}(\mathbf{X} + \mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{B}(\mathbf{X} + \mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1} \ .$$

Esempio 10.5. Si consideri la matrice simmetrica di ordine 2

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} ,$$

dove $x_{21} = x_{12}$, per cui risulta $\det(\mathbf{X}) = x_{11}x_{22} - x_{12}^2$. Quindi si ha

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \begin{pmatrix} \frac{\partial \det(\mathbf{X})}{\partial x_{11}} & \frac{\partial \det(\mathbf{X})}{\partial x_{12}} \\ \frac{\partial \det(\mathbf{X})}{\partial x_{21}} & \frac{\partial \det(\mathbf{X})}{\partial x_{22}} \end{pmatrix} = \begin{pmatrix} x_{22} & -2x_{21} \\ -2x_{12} & x_{11} \end{pmatrix} ,$$

ovvero

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = 2\begin{pmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{pmatrix} - \text{dg}\begin{pmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{pmatrix} = 2\mathbf{X}' - \text{dg}(\mathbf{X}')$$

dove

$$\mathbf{X}' = \begin{pmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{pmatrix}$$

rappresenta la matrice aggiunta di **X**. Evidentemente il risultato è un caso particolare di quello generale per $k = 2$. □

Esempio 10.6. Si consideri di nuovo la matrice **X** dell'Esempio 10.5 e la matrice simmetrica **A** di ordine 2

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} .$$

In questo caso, essendo $a_{21} = a_{12}$, si ha

$$\text{tr}(\mathbf{AX}) = a_{11}x_{11} + 2a_{12}x_{12} + a_{22}x_{22}$$

e quindi

$$\frac{\partial \,\text{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \begin{pmatrix} \frac{\partial \,\text{tr}(\mathbf{AX})}{\partial x_{11}} & \frac{\partial \,\text{tr}(\mathbf{AX})}{\partial x_{12}} \\ \frac{\partial \,\text{tr}(\mathbf{AX})}{\partial x_{21}} & \frac{\partial \,\text{tr}(\mathbf{AX})}{\partial x_{22}} \end{pmatrix} = \begin{pmatrix} a_{11} & 2a_{12} \\ 2a_{12} & a_{22} \end{pmatrix} = 2\mathbf{A} - \text{dg}(\mathbf{A}) .$$

Evidentemente il risultato è un caso particolare di quello generale per $k = 2$. □

10.3. Derivazione di applicazioni vettoriali

Si consideri una applicazione vettoriale $\mathbf{f} : \mathbb{R}^k \to \mathbb{R}^k$. Dal momento che **f** è un vettore composto da k funzioni f_i della variabile vettoriale $\mathbf{x} = (x_1\ x_2\ \dots\ x_k)^{\mathrm{T}}$, allora è possibile definire k^2 derivate. In questo caso si assume che

$$\mathbf{J}(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_1(\mathbf{x})}{\partial x_k} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_2(\mathbf{x})}{\partial x_k} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_k(\mathbf{x})}{\partial x_1} & \frac{\partial f_k(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial f_k(\mathbf{x})}{\partial x_k} \end{pmatrix} ,$$

ovvero $\mathbf{J}(\mathbf{x})$ rappresenta una matrice quadrata di ordine k la cui componente di posto (i, j) è la derivata di $f_i(\mathbf{x})$ fatta rispetto a x_j. Il determinante di questa matrice è detto *jacobiano* di f e si indica con la notazione $J(\mathbf{x})$.

Esempio 10.7. Data la matrice quadrata $\mathbf{A} = (a_{ij})$ di ordine k e un vettore $\mathbf{b} = (b_1\ b_2\ \dots\ b_k)^{\mathrm{T}}$ dello stesso ordine, si consideri l'applicazione lineare associata ad una traslazione $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$. Dal momento che

$$y_i = \mathbf{a}_{i\bullet}^{\mathrm{T}}\mathbf{x} + b_i = \sum_{l=1}^{k} a_{il}x_l + b_i ,$$

la componente di posto (i, j) di $\mathbf{J}(\mathbf{x})$ è data da

$$\frac{\partial(\mathbf{a}_{i\bullet}^{\mathrm{T}}\mathbf{x} + b_i)}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\sum_{l=1}^{k} a_{il}x_l + b_i\right) = a_{ij} ,$$

per cui si può concludere che

$$\mathbf{J}(\mathbf{x}) = \mathbf{A} ,$$

e quindi che $J(\mathbf{x}) = \det(\mathbf{A})$. □

10.4. Ottimizzazione di funzioni vettoriali e matriciali

Si considerino un vettore $\mathbf{x} = (x_1\ x_2\ \dots\ x_k)^{\mathrm{T}}$ di ordine k e la funzione di variabile vettoriale $f(\mathbf{x})$ tale che $f : \mathbb{R}^k \to \mathbb{R}$. Innanzitutto, si noti che esistono k^2 derivate seconde di f, ovvero tante quante sono le possibili coppie di variabili (x_i, x_j). In questo caso si definisce *matrice hessiana*

$$\mathbf{H}(\mathbf{x}) = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}} = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_k} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_k} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_k \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_k \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_k^2} \end{pmatrix},$$

ovvero $\mathbf{H}(\mathbf{x})$ è una matrice quadrata di ordine k, la cui componente di posto (i, j) è data dalla derivata parziale seconda di $f(\mathbf{x})$ fatta rispetto a x_i e x_j. La matrice hessiana risulta simmetrica dal momento che nella derivazione è indifferente l'ordine delle variabili.

Esempio 10.8. Si consideri la funzione di variabile vettoriale $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ con $\mathbf{A} = (a_{ij})$ matrice simmetrica di ordine k. Tenendo presente l'Esempio 10.3, la componente di posto (i, j) di $\mathbf{H}(\mathbf{x})$ risulta

$$\frac{\partial^2(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j} \frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial x_i} = \frac{\partial}{\partial x_j}\left(2\sum_{l=1}^{k} a_{il}x_l\right) = 2a_{ij}\,,$$

da cui

$$\mathbf{H}(\mathbf{x}) = 2\mathbf{A}\,. \qquad \square$$

Si supponga di dover determinare il vettore che massimizza o minimizza la funzione $f(\mathbf{x})$. A questo fine, in modo analogo a quanto viene fatto nel caso di funzioni di una sola variabile, si considera l'equazione vettoriale

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0}\,.$$

Si può dimostrare che ogni soluzione $\mathbf{x}_m$ della precedente equazione può essere un punto di massimo o di minimo locale per $f(\mathbf{x})$. Se $\mathbf{H}(\mathbf{x}_m) < \mathbf{O}$, allora $\mathbf{x}_m$ è un punto di massimo per $f(\mathbf{x})$. Ovviamente ogni minimizzazione può essere ricondotta a una massimizzazione dal momento che ogni vettore che minimizza $f(\mathbf{x})$ massimizza $-f(\mathbf{x})$. In ogni caso, se $\mathbf{H}(\mathbf{x}_m) > \mathbf{O}$, allora $\mathbf{x}_m$ è un punto di minimo per $f(\mathbf{x})$.

Esempio 10.9. Sia $\mathbf{A} = (a_{ij})$ una matrice di ordine $(n \times k)$ con $\mathrm{r}(\mathbf{A}) = k$ e sia $\mathbf{b} = (b_1\ b_2\ \ldots\ b_n)^{\mathrm{T}}$ un vettore di ordine n. Si vuole determinare il vettore $\mathbf{x}$ di ordine k che minimizza la distanza $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|$ tra $\mathbf{b}$ e $\mathbf{A}\mathbf{x}$. Equivalentemente, si desidera ottenere il vettore $\mathbf{x}$ che minimizza la funzione di variabile vettoriale

$$f(\mathbf{x}) = (\mathbf{b} - \mathbf{A}\mathbf{x})^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\mathbf{x}) \ .$$

Sviluppando il prodotto interno, tenendo presente che $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{b} = \mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, si ottiene

$$f(\mathbf{x}) = \mathbf{b}^{\mathrm{T}}\mathbf{b} - 2\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x} \ ,$$

dove $\mathbf{C} = (c_{ij}) = \mathbf{A}^{\mathrm{T}}\mathbf{A}$ è una matrice simmetrica. Si noti che $\mathbf{C} > \mathbf{O}$ essendo $\mathrm{r}(\mathbf{A}) = k$ (vedi Sezione 9.1). In questo caso si ha

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{b})}{\partial \mathbf{x}} - 2\,\frac{\partial(\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial(\mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x})}{\partial \mathbf{x}} = -\,2\mathbf{A}^{\mathrm{T}}\mathbf{b} + 2\mathbf{C}\mathbf{x} \ ,$$

da cui uguagliando la precedente espressione al vettore $\mathbf{0}$ si ottiene il sistema lineare

$$\mathbf{C}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{b} \ ,$$

che ha come unica soluzione

$$\mathbf{x}_m = \mathbf{C}^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} \ .$$

Dal momento che $\mathbf{x}_m$ è unico esso rappresenta senz'altro un punto di massimo o di minimo assoluto. Per determinare se $\mathbf{x}_m$ è un punto di minimo si consideri la matrice hessiana. In questo caso, dal momento che l'i-esima componente del vettore $\nabla f(\mathbf{x})$ è data da

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = -\,2\mathbf{a}_{\bullet i}^{\mathrm{T}}\mathbf{b} + 2\mathbf{c}_{i\bullet}^{\mathrm{T}}\mathbf{x} = -\,2\sum_{l=1}^{n} a_{li}b_l + 2\sum_{l=1}^{k} c_{il}x_l \ ,$$

mentre $\mathbf{a}_{\bullet i}$ e $\mathbf{c}_{i\bullet}^{\mathrm{T}}$ rappresentano rispettivamente l'i-esimo vettore colonna di $\mathbf{A}$ e l'i-esimo vettore riga di $\mathbf{C}$, allora

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_j}\left(-\,2\sum_{l=1}^{n} a_{li}b_l + 2\sum_{l=1}^{k} c_{il}x_l\right) = 2c_{ij} \ .$$

Si ha infine

$$\mathbf{H}(\mathbf{x}) = 2\mathbf{C} \ .$$

Poichè $\mathbf{H}(\mathbf{x}) > \mathbf{O}$ per ogni $\mathbf{x}$, allora $\mathbf{H}(\mathbf{x})$ risulta definita positiva anche per il vettore $\mathbf{x}_m$, che quindi rappresenta un punto di minimo. Si noti infine che il vettore $\mathbf{x}_m$ è la proiezione ortogonale del vettore $\mathbf{b}$ sullo spazio vettoriale generato dalle colonne di $\mathbf{A}$ (vedi Sezione 6.4). □

La massimizzazione o la minimizzazione di una funzione di variabile vettoriale può essere ottenuta senza l'uso di metodi di differenziazione. In effetti, quando possibile, può essere più efficace a questo fine applicare appropriate disuguaglianze, che possono permettere anche la massimizzazione o la minimizzazione di funzioni di variabile matriciale.

Esempio 10.10. Data la matrice simmetrica $\mathbf{A} > \mathbf{O}$ di ordine k di costanti, si consideri il cosiddetto rapporto di Rayleigh, ovvero la funzione di variabile vettoriale

$$f(\mathbf{x}) = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}} \ .$$

Poichè per ogni scalare $\alpha \neq 0$ risulta $f(\mathbf{x}) = f(\alpha\mathbf{x})$, la funzione f non dipende dalla lunghezza del vettore $\mathbf{x}$ ma solo dalla sua direzione e quindi si può imporre che la lunghezza di $\mathbf{x}$ sia tale che $\|\mathbf{x}\| = 1$, ovvero che $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$. Dunque massimizzare o minimizzare $f(\mathbf{x})$ equivale a massimizzare o minimizzare la funzione $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ soggetta al vincolo $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$. Poichè $\mathbf{A}$ è una matrice simmetrica si può considerare la matrice diagonale dei suoi autovalori $\mathbf{\Lambda} = \mathrm{dg}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ e la matrice ortogonale dei suoi autovettori $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1\ \boldsymbol{\gamma}_2\ \ldots\ \boldsymbol{\gamma}_k)$. Sfruttando quindi la rappresentazione diagonale $\mathbf{A} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^{\mathrm{T}}$ (vedi Sezione 8.2), si deve massimizzare o minimizzare la funzione $\mathbf{x}^{\mathrm{T}}\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^{\mathrm{T}}\mathbf{x}$ soggetta a $\mathbf{x}^{\mathrm{T}}\mathbf{x} = 1$. In questo caso, posto $\mathbf{y} = (y_1\ y_2\ \ldots\ y_k)^{\mathrm{T}}$ e considerando la trasformazione ortogonale $\mathbf{y} = \mathbf{\Gamma}^{\mathrm{T}}\mathbf{x}$ da cui $\mathbf{x} = \mathbf{\Gamma}\mathbf{y}$ (vedi Sezione 6.2), si ottiene la funzione

$$\mathbf{y}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{y} = \sum_{i=1}^{k} \lambda_i y_i^2 \ ,$$

soggetta a $\mathbf{y}^{\mathrm{T}}\mathbf{\Gamma}^{\mathrm{T}}\mathbf{\Gamma}\mathbf{y} = \mathbf{y}^{\mathrm{T}}\mathbf{y} = 1$. Essendo gli autovalori ordinati in senso non decrescente risulta che

$$\sum_{i=1}^{k} \lambda_i y_i^2 \leq \sum_{i=1}^{k} \lambda_1 y_i^2 = \lambda_1 \sum_{i=1}^{k} y_i^2 = \lambda_1 \ ,$$

dal momento che

$$\mathbf{y}^{\mathrm{T}}\mathbf{y} = \sum_{i=1}^{k} y_i^2 = 1 \ .$$

Inoltre

$$\sum_{i=1}^{k} \lambda_i y_i^2 \geq \sum_{i=1}^{k} \lambda_k y_i^2 = \lambda_k \sum_{i=1}^{k} y_i^2 = \lambda_k \ .$$

Dunque, il massimo valore del rapporto, ovvero λ_1, si ottiene per $\mathbf{y}_m = \mathbf{e}_1$ per cui il punto di massimo di $f(\mathbf{x})$ è dato da $\mathbf{x}_m = \mathbf{\Gamma}\mathbf{e}_1 = \boldsymbol{\gamma}_1$. In modo del tutto analogo, il minimo valore del rapporto, ovvero λ_k, si ottiene per $\mathbf{y}_m = \mathbf{e}_k$ e quindi il punto di minimo di $f(\mathbf{x})$ è dato da $\mathbf{x}_m = \mathbf{\Gamma}\mathbf{e}_k = \boldsymbol{\gamma}_k$. □

Esempio 10.11. Data la matrice simmetrica $\mathbf{X} > \mathbf{O}$ di ordine k e la matrice simmetrica $\mathbf{A} > \mathbf{O}$ di ordine k con componenti costanti, si consideri la minimizzazione della funzione di variabile matriciale

$$f(\mathbf{X}) = \ln\det(\mathbf{X}) + \mathrm{tr}(\mathbf{X}^{-1}\mathbf{A}) \ .$$

Aggiungendo e sottraendo la quantità $\ln\det(\mathbf{A})$ si ha

$$f(\mathbf{X}) = \ln\det(\mathbf{A}) - \ln\det(\mathbf{X}^{-1}\mathbf{A}) + \mathrm{tr}(\mathbf{X}^{-1}\mathbf{A}) \ .$$

La matrice $\mathbf{X}^{-1}\mathbf{A}$ è simmetrica in quanto prodotto di matrici simmetriche. Inoltre, se si considera gli autovalori della matrice $\mathbf{X}^{-1}\mathbf{A}$ si ha

$$\mathbf{X}^{-1}\mathbf{A}\boldsymbol{\gamma} = \lambda\boldsymbol{\gamma} \ ,$$

per cui premoltiplicando quest'espressione per $\mathbf{A}^{1/2}$ si ottiene inoltre

$$\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\delta} = \lambda\boldsymbol{\delta} \ ,$$

dove $\mathbf{C} = \mathbf{X}^{-1/2}\mathbf{A}^{1/2}$, mentre $\boldsymbol{\delta} = \mathbf{A}^{1/2}\boldsymbol{\gamma}$. Dal momento che $\mathbf{C}^{\mathrm{T}}\mathbf{C} > \mathbf{O}$, allora dalla precedente relazione risulta $\lambda > 0$ (vedi proprietà (iii) del

Risultato 9.1) e quindi si ha anche $\mathbf{X}^{-1}\mathbf{A} > \mathbf{O}$. Se inoltre $\lambda_1, \lambda_2, \ldots, \lambda_k$ rappresentano gli autovalori della matrice $\mathbf{X}^{-1}\mathbf{A}$, allora dalla proprietà (i) del Risultato 8.1 si ottiene

$$f(\mathbf{X}) = \ln\det(\mathbf{A}) - \ln\prod_{i=1}^{k}\lambda_i + \sum_{i=1}^{k}\lambda_i = \ln\det(\mathbf{A}) + \sum_{i=1}^{k}(-\ln\lambda_i + \lambda_i)$$

Inoltre per un qualsiasi scalare α si ha $-\ln\alpha + \alpha \geq 1$ con uguaglianza se $\alpha = 1$ e dunque

$$f(\mathbf{X}) = \ln\det(\mathbf{A}) + \sum_{i=1}^{k}(-\ln\lambda_i + \lambda_i) \geq \ln\det(\mathbf{A}) + k\ .$$

Nella precedente relazione si ha l'uguaglianza se $\lambda_i = 1$ per ogni $i = 1, 2, \ldots, k$ e quindi, se $\mathbf{\Gamma}$ rappresenta la matrice degli autovettori di $\mathbf{X}^{-1}\mathbf{A}$, il minimo di $f(\mathbf{X})$ viene raggiunto quando

$$\mathbf{X}^{-1}\mathbf{A} = \mathbf{\Gamma}\mathbf{I}_k\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}} = \mathbf{I}_k\ ,$$

ovvero quando $\mathbf{X}_m = \mathbf{A}$. □

Bibliografia

Abadir, K.M. e Magnus, J.R. (2005) *Matrix Algebra*, Cambridge University Press, New York.

Bellman, R. (1970) *Introduction to Matrix Analysis*, seconda edizione, McGraw-Hill, New York.

Bretscher, O. (1997) *Linear Algebra with Applications*, Prentice-Hall, Upper Saddle River.

Gentle, J.E. (2007) *Matrix Algebra: Theory, Computations, and Applications in Statistics*, Springer, New York.

Graybill, F.A. (1983) *Introduction to Matrices with Applications in Statistics*, Wadsworth Publishing Company, Belmont.

Hadi, A.S. (1996) *Matrix Algebra as a Tool*, Duxbury Press, Belmont.

Hadley, G. (1961) *Linear Algebra*, Addison-Wesley, Reading.

Harville, D.A. (1997) *Matrix Algebra from a Statistician's Point of View*, Springer, New York.

Horn, R.A. e Johnson, C.R. (1985) *Matrix Analysis*, Cambridge University Press, New York.

Rao, C.R. e Mitra, S.K. (1971) *Generalized Inverse of Matrices and its Applications*, Wiley, New York.

Searle, S.R. (1982) *Matrix Algebra Useful for Statistics*, Wiley, New York.

Schott, J.R. (2004) *Matrix Analysis for Statistics*, Wiley, New York.

Zhang, F. (1999) *Matrix Theory: Basic Results and Techniques*, Springer, New York.

Indice analitico