progetto
didattica in rete

Appunti di algebra lineare

Massimo Ferrarotti

Dipartimento di Matematica
Politecnico di Torino, dicembre 2001

otto editore

APPUNTI DI ALGEBRA LINEARE

MASSIMO FERRAROTTI

POLITECNICO DI TORINO - DIPARTIMENTO DI MATEMATICA
WWW.POLITO.IT

INDICE

Note preliminari 5

1. Sistemi lineari **7**

2. Matrici **11**

2.1. Definizione e notazioni 11

2.2. Somma e prodotto per scalare 16

2.3. Prodotto tra matrici 18

2.4. Prodotto tra matrici quadrate e invertibilità 23

2.5. Matrici e sistemi 26

3. Matrici equivalenti e algoritmo di Gauss **29**

3.1. Operazioni elementari 29

3.2. Algoritmo di Gauss 33

3.3. Studio e risoluzione di sistemi 38

3.4. Rango di una matrice 42

3.5. Matrici elementari e matrici invertibili 44

4. Determinanti **49**

4.1. Definizione e proprietà fondamentali 49

4.2. Determinante e operazioni elementari 54

4.3. Formula dell'inversa e regola di Cramer 56

4.4. Determinante e rango 57

5. Struttura lineare di K^n **59**

5.1. Vettori 59

5.2. Sottospazi lineari 59

5.3. Insiemi liberi e basi 66

5.4. Indipendenza lineare e rango 73

5.5. Coordinate in K^n e cambiamenti di base 76

6. Similitudine e diagonalizzazione **79**

6.1. Matrici e applicazioni 79

6.2. Similitudine 81

6.3. Autovettori e autovalori 84

7. Struttura metrica di R^n **95**

7.1. Prodotto scalare e distanza 95

7.2. Ortogonalità e basi ortonormali 97

7.3. Matrici ortogonali e matrici simmetriche 102

7.4. Matrici simmetriche e Teorema Spettrale 104

Note preliminari

I presenti appunti sono divisi in sezioni, a loro volta suddivise in sottosezioni.

Il simbolo K indica indifferentemente il campo dei numeri reali R o quello dei numeri complessi C. Gli elementi di K sono chiamati genericamente «numeri» o, talvolta, «scalari»: il termine «scalare» viene comunemente usato in algebra lineare (teoria degli spazi vettoriali) e in fisica come sinonimo di numero.

1. SISTEMI LINEARI

Se $n \in N$ è un intero positivo, un'*equazione lineare* su K in n incognite è un'espressione formale

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

dove i *coefficienti* $a_1, \ldots, a_n$ e il *termine noto* b sono elementi di K.

Osserviamo che «lineare» è sinonimo di «di primo grado», nel senso che in una equazione lineare le incognite compaiono solo con grado 1.

Quando il numero delle incognite è minore o uguale a 3 si usano sovente le lettere $x,\ y,\ z$ al posto di $x_1,\ x_2, x_3$.

Esempi

Le equazioni

$$3x_1 - x_2 + \sqrt{2}x_3 = \pi \qquad \text{e} \qquad ix_1 - x_2 = 1 + \sqrt{3}i$$

sono lineari su R e su C rispettivamente (naturalmente la prima può anche essere considerata su C).

Invece l'equazione $x_1^2 + x_1x_2 - x_3^3 + 2x_1 = 0$ non è lineare.

Osservazione

> Normalmente, quando un coefficiente di un'equazione lineare è nullo, si omette di scrivere l'addendo dell'espressione ad esso relativo e quindi il numero delle incognite va specificato a parte. Per esempio $x_1 + x_2 = 0$ può essere considerata in 3, 4,... incognite ponendo $a_3 = 0, a_4 = 0, \ldots$; se si usano indici per le incognite come nel nostro caso e se non altrimenti specificato, il numero delle incognite sarà il massimo indice.

Se $m \in N$, un *sistema lineare* S su K di m equazioni in n incognite è un insieme di m equazioni lineari su K in n incognite. Si rappresenta S nel seguente modo:

$$S : \begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2 \\ \dots \\ \dots \\ \dots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \end{cases}$$

I numeri $a_{i,j}$ e i b_i per $i = 1, \dots, m$, $j = 1, \dots, n$ sono rispettivamente i *coefficienti* e i termini noti di S.
Se $b_i = 0$ per ogni i, S si dice *omogeneo*. Se $m = n$, S si dice quadrato.
Si osservi che per $m = 1$, S si riduce a un'equazione lineare, quindi lo studio delle equazioni lineari è un caso particolare di quello dei sistemi.
Ricordiamo che K^n indica il prodotto cartesiano $K \times K \times \dots \times K$ di K per se stesso n volte. Dunque

$$K^n = \{(x_1, x_2, \dots, x_n) \mid x_i \in K,\ i = 1, \dots n\}$$

è l'insieme delle n-uple formate con elementi di K.
Un elemento $\overline{X} = (\overline{x}_1, \dots, \overline{x}_n)$ di K^n è *soluzione* di un'equazione lineare

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

se sostituendo per ogni $i = 1, \dots, n$ il numero $\overline{x}_i$ all'incognita x_i si ottiene una uguaglianza.
Diciamo che $\overline{X}$ è soluzione di S se è soluzione di ogni equazione componente S . Indichiamo il sottoinsieme di K^n formato dalle soluzioni di S con il simbolo $sol(S)$. Risolvere un sistema S significa determinare $sol(S)$. Vedremo che sono possibili solo tre casi:

1. $sol(S) = \emptyset$: S è *impossibile*.
2. $sol(S)$ è formato da un solo elemento: S è *determinato*.
3. $sol(S)$ è formato da infiniti elementi: S è *indeterminato*.

Nei casi (2) o (3) diremo che S è *risolubile*.

Osservazione

Se S è omogeneo, $O = (0, \ldots, 0) \in sol(S)$, quindi S è risolubile.

Esempi

Consideriamo il sistema

$$S : \begin{cases} x + y + z = 1 \\ x - y = 0 \end{cases}$$

S è indeterminato, in quanto $sol(S) = \{(t, t, 1 - 2t)\}$ al variare di t in R. Aggiungendo equazioni a S possiamo ottenere sistemi di tipo diverso:

$$S' : \begin{cases} x + y + z = 1 \\ x - y = 0 \\ x + y = -2 \end{cases} \qquad S'' : \begin{cases} x + y + z = 1 \\ x - y = 0 \\ 2x + z = 1 \end{cases} \qquad S''' : \begin{cases} x + y + z = 1 \\ x - y = 0 \\ 2x + z = -1 \end{cases}$$

Allora S' è determinato con $sol(S') = \{(-1, -1, 3)\}$, S'' è indeterminato con le stesse soluzioni di S mentre S''' è impossibile.
Osserviamo che se un sistema S' è ottenuto dal sistema S aggiungendo equazioni, allora $sol(S') \subseteq sol(S)$.

Definizione 1.1. *Se S e S' sono sistemi lineari a n incognite, diciamo che S e S' sono equivalenti se $sol(S) = sol(S')$.*

Osservazione

Nella precedente definizione S e S' possono avere un numero diverso di equazioni, ma $sol(S)$ e $sol(S')$ sono comunque in K^n.

Esempio

I sistemi

$$S = \begin{cases} x + y + z = 1 \\ y + z = 0 \\ x - y - z = 1 \end{cases} \qquad S' = \begin{cases} x = 1 \\ y + z = 0 \end{cases}$$

sono equivalenti poiché $sol(S) = sol(S') = \{(1, t, -t) \mid t \in R\}$. D'altra parte è evidente che S' è più semplice da risolvere rispetto a S.

L'esempio precedente suggerisce la seguente strategia: cercare un metodo, possibilmente algoritmico, per passare da un sistema dato a uno a esso equivalente ma di soluzione più immediata.
Uno strumento indispensabile per lo studio di questo e di altri aspetti dell'algebra lineare è dato dalle matrici, che introduciamo nella prossima sezione.

2. MATRICI

2.1. Definizione e notazioni

Dati m, n interi positivi, una *matrice* A su K di tipo (m,n) (o matrice $m \times n$) è definita assegnando ad ogni coppia (i,j) di numeri interi positivi con $1 \leq i \leq m$, $1 \leq j \leq n$ un elemento di K, detto *elemento di* A *in posizione* $i,\ j$ oppure *elemento di* A *con indici* $i,\ j$.

Quindi A si può rappresentare come una tabella di numeri

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}$$

Data una matrice A, indichiamo con $[A]_{i,j}$ l'elemento in posizione $i,\ j$. Per semplificare la notazione, l'elemento in posizione $i,\ j$ di A viene designato talvolta usando la corrispondente lettera minuscola, cioè $a_{i,j}$.

Alternativamente, dato un sottoinsieme in K di mn elementi $a_{i,j}$ indicizzati da $1 \leq i \leq m$ e $1 \leq j \leq n$, si indica con $\{a_{i,j}\}$ la matrice determinata da tali numeri.

Due matrici A e B $m \times n$ sono uguali se $[A]_{i,j} = [B]_{i,j}$ per ogni coppia di indici i e j. Se A è una matrice $m \times n$, fissato i tra 1 e m, gli elementi $[A]_{i,1},\ [A]_{i,2}, \dots [A]_{i,n}$ formano la i*-esima riga* di A, indicata con $[A]_i$. Analogamente, fissato j tra 1 e n, gli elementi $[A]_{1,j},\ [A]_{2,j}, \dots [A]_{m,j}$ formano la j*-esima colonna* di A, indicata con $[A]^j$. Per questo motivo, il primo e il secondo indice si dicono rispettivamente *indice di riga* e *indice di colonna* e la matrice A si dice *matrice a* m *righe e* n *colonne*. Osserviamo

che le righe di A sono esempi di matrici $1 \times n$ mentre le colonne sono matrici $m \times 1$. Possiamo anche pensare l'elemento $[A]_{i,j}$ come intersezione di $[A]_i$ e $[A]^j$.
Una matrice $n \times n$ si dice *quadrata di ordine* n. Se A è quadrata di ordine n, gli elementi $[A]_{i,i}$ formano la *diagonale principale* di A.

Esempi

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 1 \\ 0 & 5 & -4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 5 & -4 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & 0 \\ 3 & -1 \\ 0 & 5 \end{pmatrix}, \qquad D = \begin{pmatrix} 3 & -1 \\ 0 & 5 \end{pmatrix}$$

sono matrici 3×3, 2×3, 3×2 e 2×2 rispettivamente.
Con le notazioni definite, abbiamo per esempio $[A]_{1,3} = 2$, $[B]_{2,2} = 5$, $[C]_2 = \begin{pmatrix} 3 & -1 \end{pmatrix}$ e $[D]^1 = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$.
A e D sono quadrate di ordine 3 e 2, e la diagonale principale di A è data da $1,\ -1,\ -4$.

Denotiamo l'insieme di tutte le matrici $m \times n$ su K con il simbolo $M_{m,n}$. Nel caso di matrici quadrate di ordine n poniamo $M_{n,n} = M_n$. Dovendo considerare solo matrici su R (rispettivamente su C), scriviamo $M_{m,n}(R)$ (rispettivamente $M_{m,n}(C)$).
Gli elementi di $M_{1,n}$ si dicono n*-vettori riga*, mentre quelli di $M_{m,1}$ si dicono m*-vettori colonna*. Più genericamente, si parla di «vettori riga » e «vettori colonna ».
Osserviamo che l'unica differenza tra $M_{1,n}$ e K^n deriva da convenzioni notazionali: $(1, 2, -1)$ è un elemento di K^3 mentre $\begin{pmatrix} 1 & 2 & -1 \end{pmatrix}$ appartiene a $M_{1,3}$. Identifichiamo quindi $M_{1,n}$ con K^n adottando la notazione di K^n (in particolare $M_{1,1} = K$).
Nel caso di vettori riga (o colonna), possiamo identificare i singoli elementi con le colonne (o le righe) della matrice e utilizzare un solo indice per individuare un elemento: $A = ([A]^1, \ldots, [A]^n)$ oppure $A = \begin{pmatrix} [A]_1 \\ \vdots \\ [A]_m \end{pmatrix}$.

Esempi

$X = (1, 0, 2)$ è un 3-vettore riga, mentre $Y = \begin{pmatrix} 0 \\ -4 \end{pmatrix}$ è un 2-vettore colonna.

Introduciamo ora, utilizzando le notazioni introdotte, due tipi di matrici di particolare importanza.

1. Dati m e n, la *matrice nulla* $m \times n$ è la matrice $O_{m,n} \in M_{m,n}$ definita da $[O_{m,n}]_{i,j} = 0$ per ogni coppia $i,\ j$.

   $O_{m,n}$ è in pratica una tabella $m \times n$ formata da zeri. Se $m = n$, si pone $O_{n,n} = O_n$. Inoltre, quando non vi è possibilità di equivoco, è comodo scrivere O al posto di $O_{m,n}$ (per esempio, se $A \in M_{m,n}$, l'uguaglianza $A = O$ ha senso solo se $O = O_{m,n}$).

2. Dato n, la matrice identica di ordine n è la matrice quadrata $I_n \in M_n$ definita da

$$[I_n]_{i,j} = \begin{cases} 1, & \text{se } i = j; \\ 0, & \text{se } i \neq j \end{cases}.$$

   I_n è una matrice quadrata con 1 sulla diagonale principale e zero altrove. Come nel caso precedente, quando non vi è possibilità di equivoco, si può scrivere I al posto di I_n.

Esempi

$$O_{2,3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Se A è una matrice $m \times n$, si considerino p righe e q colonne di A, con $1 \leq p \leq m$, $1 \leq q \leq n$. Allora gli elementi di A che sono intersezione di una riga e una colonna tra quelle considerate formano una matrice $p \times q$ detta *sottomatrice* $p \times q$ di A.
Un sottomatrice formata da righe e colonne con indici successivi si dice *blocco*. In particolare le righe e le colonne di A sono blocchi.

Esempi

Con riferimento al primo esempio di questa sottosezione, se

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

le matrici C e D sono blocchi di A, mentre B e la matrice $\begin{pmatrix} 1 & 2 \\ 0 & -4 \end{pmatrix}$ ottenuta intersecando prima e terza riga con prima e terza colonna sono sottomatrici ma non blocchi.

Una *rappresentazione a blocchi* di una matrice $A \in M$ è data da una famiglia $\{A_{i,j}\}$ di blocchi di A individuati da due indici $1 \leq i \leq h$, $1 \leq j \leq k$, tali che

1. I blocchi $A_{i,1}, \ldots, A_{i,k}$ sono ottenuti intersecando la stesse righe con insiemi disgiunti di colonne.
2. I blocchi $A_{1,j}, \ldots, A_{h,j}$ sono ottenuti intresecando la stesse colonne con insiemi disgiunti di righe.
3. Ogni elemento di A appartiene a uno e un solo blocco.

In pratica, possiamo pensare una matrice rappresentata a blocchi come una tabella i cui elementi non sono numeri ma matrici.
Per esempio, se $h = k = 2$, una rappresentazione a 4 blocchi di A si scrive

$$A = \begin{pmatrix} A_{1,1} & | & A_{1,2} \\ - & - & - \\ A_{2,1} & | & A_{2,2} \end{pmatrix}$$

dove $A_{1,1}$ è $p \times q$, $A_{1,2}$ è $p \times q'$, $A_{2,1}$ è $p' \times q$, $A_{2,2}$ è $p' \times q'$ e $p + p' = m$, $q + q' = n$.

Esempi

Ecco due diverse rappresentazioni a 4 blocchi della stessa matrice:

$$\begin{pmatrix} 0 & 1 & | & 0 & 2 \\ -2 & 3 & | & -1 & 1 \\ - & - & - & - & - \\ 1 & 0 & | & 5 & -4 \\ 0 & 0 & | & 0 & 1 \end{pmatrix}$$

e

$$\begin{pmatrix} 0 & | & 1 & 0 & 2 \\ - & - & - & - & - \\ -2 & | & 3 & -1 & 1 \\ 1 & | & 0 & 5 & -4 \\ 0 & | & 0 & 0 & 1 \end{pmatrix}$$

Utili rappresentazioni a blocchi di un matrice $A \in M_{m,n}$ sono la *rappresentazione per righe*:

$$A = \begin{pmatrix} [A]_1 \\ \vdots \\ [A]_m \end{pmatrix}$$

e la *rappresentazione per colonne*:

$$A = \begin{pmatrix} [A]^1 & \dots & [A]^m \end{pmatrix}$$

Se $A \in M_{m,n}$, la matrice ${}^tA \in M_{n,m}$ definita da $[\,{}^tA]_{i,j} = [A]_{j,i}$ si dice *trasposta* di A. In pratica A si ottiene da A scambiando le righe con le colonne.

Esempio

$${}^t\begin{pmatrix} 1 & 0 & 2 \\ 0 & 5 & -4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 5 \\ 2 & -4 \end{pmatrix}.$$

L'applicazione di trasposizione $T : M_{m,n} \to M_{n,m}$ definita da $T(A) = {}^tA$ (si associa a ogni matrice la sua trasposta) è biunivoca. Infatti si vede subito che ${}^t({}^tA) = A$, quindi l'inversa della trasposizione su $M_{m,n}$ è la trasposizione su $M_{n,m}$.
Una matrice $A \in M_n$ si dice *simmetrica* se $A = {}^tA$, cioè se $[A]_{i,j} = [A]_{j,i}$ per $1 \leq i,\ j \leq n$.
Il termine «simmetrica» deriva dal fatto che una tale matrice rappresentata esplicitamente come tabella risulta simmetrica rispetto alla diagonale principale.

Nota: una matrice 1×1, cioè uno scalare, è sempre simmetrica!

Esempi

1. O_n e I_n sono matrici simmetriche.

2. Le matrici

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \qquad \text{e} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

sono simmetriche. La seconda ci suggerisce che la tabellina pitagorica è una matrice simmmetrica di ordine 10.

Le considerazioni precedenti ci permettono, quando ciò non crea confusione, di identificare $K^n = M_{1,n}$ con $M_{n,1}$ chiamando semplicemente n-*vettori* (o «vettori») gli elementi di K^n senza specificare se stiamo considerando vettori riga o colonna e scrivendoli per esteso come righe o colonne a seconda dell'opportunità (per esempio qui su questa riga è più comodo scrivere $(1, 0, 1)$ che $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$).

Per convenzione, salvo diverso avviso, un vettore denotato da un simbolo (per esempio X) è considerato un vettore colonna (e dunque ${}^tX$ è un vettore riga). Naturalmente, se A è una matrice, le righe $[A]_i$ sono considerate vettori riga!

2.2. Somma e prodotto per scalare

Se $A,\ B \in M_{m,n}$, si definisce *somma* di A e B la matrice $A + B \in M_{m,n}$ data da $[A + B]_{i,j} = [A]_{i,j} + [B]_{i,j}$.

Di fatto la somma di due matrici è la matrice che ha come elemento di indici $i,\ j$ la somma degli elementi di indici $i,\ j$ delle matrici considerate.

Esempi

Se

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 3 & -1 & 1 \\ 0 & 5 & -4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}, \qquad C = \begin{pmatrix} -1 & 0 & -2 \\ -3 & 1 & -1 \\ 0 & -5 & 4 \end{pmatrix}$$

allora

$$A + B = \begin{pmatrix} 2 & 2 & 5 \\ 5 & 3 & 7 \\ 3 & 11 & 5 \end{pmatrix} \qquad \text{e} \qquad A + C = O.$$

La somma è un'operazione interna dell'insieme $M_{m,n}$ che gode delle seguenti proprietà: date $A,\ B,\ C \in M_{m,n}$

S1. Associativa: $A + (B + C) = (A + B) + C$;

S2. Commutativa: $A + B = B + A$;

S3. Esistenza elemento neutro: $A + O_{m,n} = O_{m,n} + A = A$;

S4. Esistenza opposto: per ogni $A \in M_{m,n}$ esiste $-A \in M_{m,n}$ tale che $A + (-A) = -A + A = O_{m,n}$.

Le proprietà elencate derivano direttamente dalla definizione. Per (4) basta definire $[-A]_{i,j} = -[A]_{i,j}$, cioè cambiare il segno a tutti gli elementi di A.
Possiamo osservare che $O_{m,n}$ e $-A$ sono le uniche matrici che soddisfano a (3) e (4) rispettivamente.
Se ora $a \in K$ e $A \in M_{m,n}$ definiamo il *prodotto per lo scalare* a di A come la matrice $aA \in M_{m,n}$ data da $[aA]_{i,j} = a[A]_{i,j}$.
Di fatto aA si ottiene moltiplicando ogni elemento di A per a.

Esempio

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

$$-2A = \begin{pmatrix} -2 & 0 & -4 \\ -6 & 2 & -2 \end{pmatrix}, \qquad \frac{1}{2}A = \begin{pmatrix} \frac{1}{2} & 0 & 1 \\ \frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Il prodotto per scalare non è evidentemente un'operazione interna e gode delle seguenti proprietà che, come quelle della somma, sono di verifica immediata: se $A,\ B \in M_{m,n}$ e $a,\ b \in K$.

P1. $(a + b)A = aA + bA$;

P2. $a(A + B) = aA + aB$;

P3. $a(bA) = (ab)A$;

P4. $1A = A$.

Inoltre abbiamo evidentemente $(-1)A = -A$, $0A = O_{m,n}$ e $aO_{m,n} = O_{m,n}$.
Le operazioni introdotte sono compatibili rispetto alla trasposizione, nel senso della seguente proposizione.

Proposizione 2.1. *Se* $A,\ B \in M_{m,n}$ *e* $a \in K$ *allora* ${}^t(A+B) = {}^tA + {}^tB$ *e* ${}^t(aA) = a\,{}^tA$.

PROVA

Applicando direttamente le definizioni abbiamo

$$[\,{}^t(A+B)]_{i,j} = [A+B]_{j,i} = [A]_{j,i} + [B]_{j,i} = [\,{}^tA]_{i,j} + [\,{}^tB]_{i,j}.$$

Analogo per il prodotto per scalare.

2.3. Prodotto tra matrici

Date due matrici $A \in M_{m,n}$ e $B \in M_{n,p}$, è possibile definire la *matrice prodotto* AB (o semplicemente il *prodotto*) di A e B come elemento di $M_{m,p}$; cioè il prodotto di due matrici è definito solo quando la matrice A ha tante colonne quante sono le righe della matrice B. Una coppia (A, B) di matrici con tale proprietà viene detta *moltiplicabile*. Evidentemente tale operazione non è in generale interna né ha senso chiedersi se è commutativa.
Definiamo AB dapprima in un caso particolare per poi dare la definizione generale utilizzando tale caso particolare.

Caso particolare

Sia $A \in M_{1,n}$ e $B \in M_{n,1}$ (cioè A è un n-vettore riga e B è un n-vettore colonna). Allora la matrice prodotto AB deve appartenere a K e si ottiene mediante la formula

$$AB = [A]^1[B]_1 + \cdots + [A]^n[B]_n = \sum_{i=1}^{n} [A]^i[B]_i.$$

Esempio

Se $A = (1, 0, 2)$ e $B = \begin{pmatrix} 3 \\ 5 \\ -4 \end{pmatrix}$, allora $AB = 1 \cdot 3 + 0 \cdot 5 + 2 \cdot (-4) = -5$

Caso generale

Se $A \in M_{m,n}$ e $B \in M_{n,p}$, le righe di A formano con le colonne di B coppie moltiplicabili. Definiamo dunque AB come la matrice i cui elementi sono dati da

$$[AB]_{i,j} = [A]_i[B]^j = [A]_{i,1}[B]_{1,j} + [A]_{i,2}[B]_{2,j} + \cdots + [A]_{i,n}[B]_{n,j} = \sum_{k=1}^{n} [A]_{i,k}[B]_{k,j}.$$

Quindi l'elemento di indici $i,\ j$ di AB è il prodotto della riga i di A con la colonna j di B effettuato con la regola data per il caso particolare. In virtù di questa definizione il prodotto tra matrici viene spesso indicato come *prodotto righe per colonne*.

Esempio

Se $A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 5 & -4 \end{pmatrix}$ e $B = \begin{pmatrix} 1 & 2 \\ 3 & -1 \\ 0 & -1 \end{pmatrix}$, allora

$$[AB]_{1,1} = (1, 1, 2) \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} = 4$$

$$[AB]_{1,2} = (1, 1, 2) \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} = -1$$

$$[AB]_{2,1} = (0, 5, -4) \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} = 15$$

$$[AB]_{2,2} = (0, 5, -4) \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} = -1$$

dunque $AB = \begin{pmatrix} 4 & -1 \\ 15 & -1 \end{pmatrix}$.

Le seguenti proprietà del prodotto tra matrici si verificano usando la definizione.

1. Associativa: se (A, B) e (B, C) sono coppie moltiplicabili, allora $A(BC) = (AB)C$.
2. Distributiva a sinistra rispetto alla somma: se (A, B) e (A, C) sono coppie moltiplicabili, allora $A(B + C) = AB + AC$.
3. Distributiva a destra rispetto alla somma: se (A, C) e (B, C) sono coppie moltiplicabili, allora $(A + B)C = AC + BC$.
4. Esistenza dell'elemento neutro a destra e a sinistra: se $A \in M_{m,n}$, $AI_n = I_m A = A$.
5. Distributiva rispetto al prodotto per scalare: se (A, B) è una coppia moltiplicabile e $a \in K$, allora $A(aB) = a(AB)$.

Sempre dalle definizioni otteniamo che, se $A \in M_{m,n}$, $AO_{n,p} = O_{m,p}$ e $O_{q,m}A = O_{q,n}$. Osserviamo che, viceversa, non è vero che se $A \in M_{m,n}$, $B \in M_{n,p}$ e $AB = O_{m,p}$ allora $A = O_{m,n}$ o $B = O_{n,p}$. Per esempio, consideriamo le seguenti matrici non nulle:

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

Allora $AB = C^2 = O$.

La relazione tra prodotto e trasposizione è la seguente:

Proposizione 2.2. *Se (A, B) è moltiplicabile, allora ${}^t(AB) = {}^tB\,{}^tA$.*

PROVA

Se $A \in M_{m,n}$ e $B \in M_{n,p}$,

$$[\,{}^t(AB)]_{i,j} = [AB]_{j,i} = \sum_{k=1}^{n} [A]_{j,k}[B]_{k,i} = \sum_{k=1}^{n} [\,{}^tB]_{i,k}[\,{}^tA]_{k,j} = [\,{}^tB\,{}^tA]_{i,j}.$$

Introduciamo ora qualche strumento per il calcolo formale tra matrici.

Vettori canonici

Fissato n, indicheremo con $e_1, e_2, \ldots, e_n$ le colonne della matrice I_n e chiameremo tali vettori *n-vettori colonna canonici*. Quindi, dato $i = 1, \ldots, n$,

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

dove 1 si è in posizione i.

È immediato allora che i vettori riga ${}^te_1$, ${}^te_2, \ldots,$ ${}^te_n$ trasposti dei precedenti sono le righe di I_n; chiameremo dunque questi ultimi n*-vettori riga canonici*. Analogamente, per $i = 1, \ldots, n$,

$${}^te_i = (0, \ldots, 1, \ldots, 0).$$

con 1 in posizione i.

Esempio

In K^3, i vettori riga canonici sono ${}^te_1 = (1, 0, 0)$, ${}^te_2 = (0, 1, 0)$ e ${}^te_3 = (0, 0, 1)$, mentre in K^2 sono ${}^te_1 = (1, 0)$, ${}^te_2 = (0, 1)$.

Osservazione

> Le notazione introdotta, non essendo specificato l'ordine della matrice identica, è ambigua: per esempio ${}^te_1$ può indicare la prima riga di qualsiasi matrice identica, ${}^te_1 = (1, 0)$, ${}^te_1 = (1, 0, 0)$ e così via. Comunque tale notazione è largamente accettata in quanto è di solito chiara dal contesto ed evita simboli troppo complessi (tipo «e_i^n »).

Possiamo esprimere gli elementi, le righe e le colonne di una matrice A come prodotti di A con vettori canonici.

Proposizione 2.3. *Sia* $A \in M_{m,n}$. *Allora, per* $1 \leq i \leq m$ *e* $1 \leq j \leq n$ *abbiamo* $[A]_i = {}^te_iA$, $[A]^j = Ae_j$ *e* $[A]_{i,j} = {}^te_iAe_j$.

PROVA

Posto $B = {}^te_iA$, B è una matrice $1 \times n$, cioè un vettore riga, e $[B]^j = {}^te_iA_j = [A]_{i,j}$ per $j = 1, \ldots, n$; dunque $B = A_i$.

Analogamente, $C = Ae_j$ è una matrice $m \times 1$ e $[C]_i = A_ie_j = [A]_{i,j}$ per $i = 1, \ldots, m$; dunque $C = A^j$.

Infine, ${}^te_iAe_j = {}^te_iA^j = [A]_{i,j}$ per quanto sopra.

Applicando la Proposizione 2.3 alla matrice I_n stessa, otteniamo che

$$^{t}e_i e_j = \begin{cases} 1, & \text{se } i = j; \\ 0, & \text{se } i \neq j \end{cases}$$

Prodotto a blocchi

Un'altro strumento di calcolo formale per le matrici si ottiene tramite le rappresentazioni a blocchi.

Data una coppia moltiplicabile di matrici (A, B), supponiamo che A e B siano rappresentate con blocchi $A_{i,j}$ e $B_{h,k}$ rispettivamente, in modo che $j,\ h = 1, \ldots, p$ e che, per ogni $i,\ k$, la coppia $(A_{i,j}, B_{j,k})$ sia moltiplicabile. Allora le matrici

$$(AB)_{i,k} = \sum_{j=1}^{p} A_{i,j} B_{j,k}$$

danno una rappresentazione a blocchi di AB detta *prodotto a blocchi* di A e B.

Possiamo osservare che il prodotto a blocchi si effettua con la stessa regola formale del prodotto tra matrici, moltiplicando «righe» per «colonne» costituite da blocchi invece che da numeri.

Nel caso di 4 blocchi, se

$$M = \begin{pmatrix} A & | & B \\ - & - & - \\ C & | & D \end{pmatrix} \quad \text{e} \quad M' = \begin{pmatrix} A' & | & B' \\ - & - & - \\ C' & | & D' \end{pmatrix},$$

allora

$$MM' = \begin{pmatrix} AA' + BC' & | & AC' + BD' \\ - & - & - \\ CA' + DC' & | & CB' + DD' \end{pmatrix}.$$

Esempio

Siano

$$M = \begin{pmatrix} 0 & | & 1 & 1 \\ - & - & - & - \\ 1 & | & 2 & 2 \\ 0 & | & 2 & 2 \end{pmatrix}, \quad M' = \begin{pmatrix} 1 & | & 1 & 0 \\ - & - & - & - \\ 2 & | & 1 & 1 \\ -2 & | & -1 & -1 \end{pmatrix}.$$

Allora

$$MM' = \begin{pmatrix} 0 & | & 0 & 0 \\ - & - & - & - \\ 1 & | & 0 & 0 \\ 0 & | & 0 & 0 \end{pmatrix}.$$

2.4. Prodotto tra matrici quadrate e invertibilità

Una coppia (A, B) di matrici quadrate dello stesso ordine n è evidentemente moltiplicabile e il prodotto AB è ancora una matrice quadrata di ordine n. Quindi il prodotto tra matrici in M_n è un'operazione interna che, come già provato, gode delle proprietà associativa, distributive rispetto alla somma e prodotto per scalare e esistenza dell'elemento neutro I_n.

Riguardo a quest'ultimo, va sottolineato che I_n è l'unica matrice tale che $AI_n = I_nA = A$ per ogni matrice $A \in M_n$: se infatti $B \in M_n$ è tale che $AB = A$ per ogni $A \in M_n$, allora, prendendo $A = I_n$, si ha $B = I_nB = I_n$.

Se ora compariamo la somma e il prodotto tra matrici quadrate con le omonime operazioni tra numeri (che sono il caso particolare delle prime quando si considerino gli elementi di K come matrici 1×1), vediamo che mentre la somma tra matrici soddisfa alle stesse proprietà formali di quella numerica, il prodotto non si comporta nello stesso modo.

Intanto, non vale in generale la commutatività.

Esempio

Se $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, cerchiamo $M = \begin{pmatrix} x & y \\ z & t \end{pmatrix}$ tale che $AM = MA$. Ora

$$AM = \begin{pmatrix} x+z & y+t \\ z & t \end{pmatrix} \quad \text{e} \quad MA = \begin{pmatrix} x & x+y \\ z & z+t \end{pmatrix}.$$

Dunque $AM = MA$ se e solo se $z = 0$ e $x = t$, cioè se M è del tipo $\begin{pmatrix} x & y \\ 0 & x \end{pmatrix}$ per $x, y \in K$.

Pertanto

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

Inoltre, una proprietà fondamentale del prodotto numerico è l'esistenza dell'inverso: se $a \in K$ e $a \neq 0$ allora esiste unico $a^{-1} \in K$ tale che $aa^{-1} = 1$. Il seguente esempio mostra che questo non è più vero per le matrici quadrate.

Esempio

Sia $A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$; cerchiamo una matrice $M = \begin{pmatrix} x & y \\ z & t \end{pmatrix}$ tale che

$$\begin{pmatrix}1 & 1\\ 2 & 2\end{pmatrix}\begin{pmatrix}x & y\\ z & t\end{pmatrix}=\begin{pmatrix}x+z & y+t\\ 2x+2z & 2y+2t\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

Dunque dovremmo avere contemporaneamente $x+z=1$ e $x+z=0$, il che è assurdo e esclude l'esistenza di M.

In base alla discussione precedente, ha senso la seguente

Definizione 2.4. *Una matrice quadrata $A \in M_n$ si dice invertibile se esiste $B \in M_n$ tale che $AB = BA = I_n$. Altrimenti A si dice non invertibile o singolare.*

Esempio

Consideriamo la matrice $A = \begin{pmatrix}1 & -1\\ 2 & 2\end{pmatrix}$, che differisce dall'esempio precedente solo per un «- », e cerchiamo una matrice $M = \begin{pmatrix}x & y\\ z & t\end{pmatrix}$ tale che

$$\begin{pmatrix}1 & -1\\ 2 & 2\end{pmatrix}\begin{pmatrix}x & y\\ z & t\end{pmatrix}=\begin{pmatrix}x-z & y-t\\ 2x+2z & 2y+2t\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

Questo porta a risolvere il sistema

$$\begin{cases}x-z=1\\ 2x+2z=0\\ y-t=0\\ 2y+2t=1\end{cases}$$

che ha soluzione unica $x = 1/2$, $z = -1/2$, $y = t = 1/4$. Dunque M esiste e si verifica subito che $MA = I$.

Nella Definizione 2.4 , la condizione che B commuti con A è necessaria per avere l'unicità di B (seguente Proposizione 2.5); d'altra parte proveremo in seguito che la condizione $AB = I_n$ implica $AB = BA$ e quindi assicura l'invertibilità di A (e di B).

Proposizione 2.5. *Sia $A \in M_n$ invertibile e $B \in M_n$ tale che $AB = BA = I_n$. Se C e C' sono matrici in M_n tali che $AC = I_n$ e $C'A = I_n$, allora $B = C = C'$.*

PROVA

$B = BI_n = B(AC) = (BA)C = I_nC = C$ e $B = I_nB = (C'A)B = C'(AB) = C'$.

Dalla Proposizione precedente segue che se $A \in M_n$ è invertibile esiste *una sola* matrice B tale $AB = BA = I_n$; chiamiamo tale matrice la *(matrice) inversa* di A e la denotiamo con A^{-1}.
Il sottoinsieme di M_n formato dalla matrici invertibili di ordine n viene indicato con GL_n; se è necessario specificare il campo si scrive $GL_n(R)$ oppure $GL_n(C)$.
Sia $A \in GL_n$; valgono le seguenti proprietà fondamentali:

I1. $A^{-1} \in GL_n$ e $(A^{-1})^{-1} = A$.

I2. Se $B \in GL_n$, $AB \in GL_n$ e $(AB)^{-1} = B^{-1}A^{-1}$.
Infatti $B^{-1}A^{-1}AB = ABB^{-1}A^{-1} = I_n$.

I3. ${}^tA \in GL_n$ e $({}^tA)^{-1} = {}^t(A^{-1})$.
Infatti ${}^t(A^{-1})\,{}^tA = {}^t(AA^{-1}) = I_n = {}^t(A^{-1}A) = {}^tA\,{}^t(A^{-1})$.

I4. Se $a \in K,\ a \neq 0$, allora $aA \in GL_n$ e $(aA)^{-1} = a^{-1}A^{-1}$.

In seguito studieremo metodi per riconoscere quando una matrice è o non è invertibile e, nel primo caso, per calcolarne esplicitamente l'inversa. La prossima proposizione individua una classe di matrici la cui non invertibilità è immediatamente riconoscibile:

Proposizione 2.6. *Se $A \in M_n$ ha una riga o una colonna nulla (formata solo da zeri), allora A è non invertibile.*

PROVA

Supponiamo che sia nulla la riga $[A]_i$ di A e assumiamo per assurdo che $A \in GL_n$. Per ogni matrice $B \in M_n$ e per ogni $j = 1, \ldots, n$ abbiamo $[AB]_{i,j} = A_iB^j = 0$; quindi, presi $B = A^{-1}$ e $j = i$, abbiamo $[AA^{-1}]_{i,i} = 0$, il che è assurdo perché $AA^{-1} = I_n$ e $[AA^{-1}]_{i,i} = 1$.
Supponiamo ora che A abbia una colonna nulla e sia invertibile. Allora ${}^tA \in GL_n$ per la proprietà (I3) e ha una riga nulla: ciò è assurdo per quanto provato prima.

Evidentemente O_n non è invertibile; d'altra parte vi sono matrici singolari i cui elementi sono tutti $\neq 0$, come l'esempio all'inizio di questa sottosezione.
Per quel che riguarda le matrici invertibili, evidentemente $I_n \in GL_n$ con $I_n^{-1} = I_n$ e la proprietà (I4) assicura che tutti i suoi multipli per scalari non nulli sono invertibili. Osserviamo che la somma di matrici invertibili non è in generale invertibile: per esempio, $I_n + (-I_n) = O_n$.

Per matrici di ordine > 2 risulta in generale troppo complicato calcolativamente verificare l'invertibilità e trovare l'inversa con il metodo diretto usato per gli esempi esposti finora. Più avanti presenteremo metodi più efficienti per trattare questo problema.

2.5. Matrici e sistemi

Dato un sistema

$$S = \begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2 \\ \dots \\ \dots \\ \dots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \end{cases}$$

usando il prodotto tra matrici possiamo tradurre S in una *equazione matriciale* definendo $A \in M_{m,n}$, $B \in M_{m,1}$ e $X \in M_{n,1}$ come $[A]_{i,j} = a_{i,j}$, $[B]_i = b_i$ e $[X]_i = x_i$.
Allora potremo scrivere $S : AX = B$; le equazioni di S saranno date da $[A]_i X = b_i$ per $i = 1, \dots, m$.
La matrice a blocchi $M_S = \begin{pmatrix} A & | & B \end{pmatrix} \in M_{m,n+1}$ è la *matrice completa* associata a S, A è la *matrice dei coefficienti*, B è il *vettore dei termini noti* e X il *vettore delle incognite*.

Esempio

Se

$$S = \begin{cases} x_1 + x_2 + x_3 = 1 \\ x_2 + x_3 = 0 \\ x_1 - x_2 - x_3 = 1 \end{cases}$$

$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$ è la matrice dei coefficienti, $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ è il vettore delle incognite, $B = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ è il vettore dei termini noti e

$$M_S = \begin{pmatrix} 1 & 1 & 1 & | & 1 \\ 0 & 1 & 1 & | & 1 \\ 1 & -1 & -1 & | & 1 \end{pmatrix}$$

è la matrice completa associata a S.

Chiaramente l'associazione $S \to M_S$ stabilisce una corrispondenza biunivoca tra i sistemi lineari a m equazioni e n incognite e le matrici di $M_{m+1,n}$. Quindi, data $M \in M_{m,n+1}$, indichiamo con S_M l'unico sistema lineare che ha M come matrice completa: S_M si dice *sistema associato* a M.
Un sistema omogeneo è ovviamente determinato dalla sua matrice dei coefficienti, quindi se $A \in M_{m,n}$, $S : AX = O$ si dice *sistema omogeneo associato* a A.
Inoltre, se $S : AX = B$ è un sistema, il sistema $S_0 : AX = O$ si dice *sistema omogeneo associato* a S. La relazione fra $sol(S)$ e $sol(S_0)$ è data dal seguente

Teorema 2.7. *Sia $S : AX = B$ un sistema lineare con $A \in M_{m,n}$ e sia S_0 il sistema omogeneo associato a S. Se $\overline{X}$ è una qualsiasi soluzione di S, $Y \in K^n$ è una soluzione di S se e solo se esiste $Y_0 \in sol(S_0)$ tale che $Y = Y_0 + \overline{X}$. In tal caso Y_0 è unico.*

PROVA

Se $Y \in sol(S)$, allora $Y_0 = Y - \overline{X} \in S_0$: infatti, se $S : AX = B$, $AY_0 = A(Y - \overline{X}) = AY - A\overline{X} = B - B = O$. Dunque $Y = Y_0 + \overline{X}$ e Y_0 è unico.
Viceversa, se $Y = Y_0 + \overline{X}$ con $Y_0 \in sol(S_0)$, abbiamo $AY = A(Y_0 + \overline{X}) = AY_0 + A\overline{X} = A\overline{X} = B$. Dunque $Y \in sol(S)$.

Come vedremo nella prossima sezione, lo studio dei sistemi passa attraverso lo studio delle loro matrici associate.

3. MATRICI EQUIVALENTI E ALGORITMO DI GAUSS

3.1. Operazioni elementari

Si consideri la matrice $A \in M_{m,n}$ rappresentata per righe:

$$A = \begin{pmatrix} [A]_1 \\ \vdots \\ [A]_m \end{pmatrix}.$$

Un'operazione elementare per riga su A consiste nell'associare a A una matrice $A' \in M_{m,n}$, che rappresentiamo anch'essa per righe come

$$A' = \begin{pmatrix} [A']_1 \\ \vdots \\ [A']_m \end{pmatrix},$$

in uno dei tre seguenti modi.

1. *Operazioni elementari del I tipo.*

Dati $1 \leq i \neq j \leq m$ e $a \in K$, poniamo

$$[A']_i = [A]_i + a[A]_j$$

mentre $[A']_k = [A]_k$ per $k \neq i$.

Dunque A' si ottiene da A sommando alla i-esima riga la j-esima moltiplicata per a.

Denotiamo tale operazione con $(i) + a(j)$ e la trasformazione di A in A' con $A \stackrel{(i)+a(j)}{\rightarrow} A'$. Si verifica facilmente che $A' \stackrel{(i)-a(j)}{\rightarrow} A$.

2. *Operazioni elementari del II tipo.*

Dati $1 \leq i \neq j \leq m$, poniamo

$$[A']_i = [A]_j \ , \ [A']_j = [A]_i$$

mentre $[A']_k = [A]_k$ per $k \neq i, \ j$.

A' si ottiene da A scambiando la riga i-esima con la j-esima.

Denotiamo tale operazione con $(i) \ \leftrightarrow \ (j)$ e la trasformazione di A in A' con $A \overset{(i)\leftrightarrow(j)}{\rightarrow} A'$.

Si verifica facilmente che $A' \overset{(i)\leftrightarrow(j)}{\rightarrow} A$.

3. *Operazioni elementari del III tipo.*

Dati $1 \leq i \leq m$ e $a \in K, \ a \neq 0$, poniamo $[A']_i = a[A]_i$, mentre $[A']_k = [A]_k$ per $k \neq i$.

A' si ottiene da A moltiplicando la riga i-esima per a.

Denotiamo tale operazione con $a(i)$ e la trasformazione di A in A' con $A \overset{a(i)}{\rightarrow} A'$.

Si verifica facilmente che $A' \overset{\frac{1}{a}(i)}{\rightarrow} A'$.

Evidentemente possiamo interpretare un'operazione elementare come una applicazione da $M_{m,n}$ in $M_{m,n}$: per le considerazioni fatte, tale applicazione è biunivoca e la sua inversa è un'operazione elementare del medesimo tipo.

Osserviamo che l'applicazione identica (che lascia invariata ogni matrice) può essere vista come operazione elementare del I tipo (con $a = 0$ e $i, \ j$ qualsiasi) oppure del II tipo (con $a = 1$ e i qualsiasi).

Esempio

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 2 & 1 \\ 3 & -2 & 5 & 0 \end{pmatrix} \overset{(3)+2(1)}{\rightarrow} A' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 2 & 1 \\ 5 & 0 & 5 & 0 \end{pmatrix} \overset{(3)\leftrightarrow(2)}{\rightarrow}$$

$$A'' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 5 & 0 & 5 & 0 \\ 0 & -1 & 2 & 1 \end{pmatrix} \overset{\frac{1}{5}(2)}{\rightarrow} A''' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 1 \end{pmatrix}$$

Invertendo si ha

$$A''' \overset{5(2)}{\rightarrow} A'' \overset{(3)\leftrightarrow(2)}{\rightarrow} A' \overset{(3)-2(1)}{\rightarrow} A.$$

Definizione 3.1. *Due matrici $A,\ B \in M_{m,n}$ si dicono equivalenti per righe se B si può ottenere da A con un numero finito di operazioni elementari per riga. In tal caso sciveremo $A \sim B$.*

La relazione $\sim$ è una relazione di equivalenza: infatti è evidentemente riflessiva ($A \sim A$) e transitiva ($A \sim B$ e $B \sim C$ implica $A \sim C$), quanto alla simmetria ($A \sim B \Leftrightarrow B \sim A$), questa deriva dal fatto che le inverse delle operazioni elementari sono operazioni elementari.
La seguente proposizione deriva direttamente dal fatto che le operazioni per righe lasciano invariato l'indice di colonna.

Proposizione 3.2. *Siano $A,\ B \in M_{m,n}$ e sia $A \sim B$. Se A' è la sottomatrice $m \times k$ di A formata dalle colonne di A con indici $j_1, \dots, j_k$, allora la sottomatrice $m \times k$ di B formata dalle colonne di B con indici $j_1, \dots, j_k$ si ottiene da A' tramite le stesse operazioni elementari che trasformano A in B.*

L'equivalenza tra matrici è di grande importanza nello studio dei sistemi lineari grazie al seguente teorema.

Teorema 3.3. *Se $M,\ M' \in M_{m,n+1}$ sono matrici equivalenti per righe, allora i sistemi S_M e $S_{M'}$ sono equivalenti, cioè $sol(S_M) = sol(S_{M'})$.*

PROVA

Basta provare la tesi nel caso in cui M' sia ottenuta da M con una sola operazione elementare.
Mettiamo in evidenza le colonne dei termini noti di S_M e $S_{M'}$ scrivendo $M = \begin{pmatrix} A & | & B \end{pmatrix}$ e $M' = \begin{pmatrix} A' & | & B' \end{pmatrix}$ con $[B]_i = b_i$ e $[B']_i = b'_i$.
Per la Proposizione 3.2, $A \sim A'$ e $B \sim B'$ tramite la stessa operazione elementare che trasforma M in M'.
Distinguiamo a seconda del tipo dell'operazione elementare.

I tipo. Sia $M \stackrel{(i)+a(j)}{\to} M'$ e sia $\overline{X} \in Sol(S_M)$. Allora $A\overline{X} = B$, in particolare $[A]_i\overline{X} = b_i$ e $[A]_j\overline{X} = b_j$ da cui $([A]_i + a[A]_j)\overline{X} = b_i + ab_j$.
Questo significa che $\overline{X}$ soddisfa alla i-esima equazione di $S_{M'}$. Poiché tutte le altre equazioni di $S_{M'}$ sone le stesse di S_M, $\overline{X} \in sol(S_{M'})$.
Viceversa, se $\overline{X} \in Sol(S_{M'})$, abbiamo $(A_i + aA_j)\overline{X} = b_i + ab_j$ e $A_j\overline{X} = b_j$ da cui segue subito $A_i\overline{X} = b_i$ e quindi $\overline{X} \in Sol(S_M)$.

II tipo. Ovvio

III tipo. Sia $M \overset{a(i)}{\to} M'$. Poiché $a \neq 0$, $[A]_i X = b_i \Leftrightarrow a[A]_i X = ab_i$, dunque $sol(S_M) = sol(S_{M'})$.

Consideriamo ora un sistema S: se siamo in grado di trasformare tramite operazioni elementari per righe la matrice M_S in una matrice M' di forma più «semplice » (e.g.con molti zeri), allora, per il teorema precedente, il sistema $S_{M'}$ avrà le stesse soluzioni di S ma sarà più facile da studiare.

Esempio

Sia

$$S : \begin{cases} x + y = 0 \\ -y + 2z = 1 \\ 3x - 2y + 5z = 0 \end{cases} .$$

Allora $M_S = \begin{pmatrix} 1 & 1 & 0 & | & 0 \\ 0 & -1 & 2 & | & 1 \\ 3 & -2 & 5 & | & 0 \end{pmatrix}$ e effettuando le operazioni per riga

$$M_S \overset{(3)-3(1)}{\to} \begin{pmatrix} 1 & 1 & 0 & | & 0 \\ 0 & -1 & 2 & | & 1 \\ 0 & -5 & 5 & | & 0 \end{pmatrix} \overset{(3)-5(2)}{\to} \begin{pmatrix} 1 & 1 & 0 & | & 0 \\ 0 & -1 & 2 & | & 1 \\ 0 & 0 & -5 & | & -5 \end{pmatrix} \overset{-\frac{1}{5}(3)}{\to}$$

$$\begin{pmatrix} 1 & 1 & 0 & | & 0 \\ 0 & -1 & 2 & | & 1 \\ 0 & 0 & 1 & | & 1 \end{pmatrix} \overset{(2)-2(3)}{\to} \begin{pmatrix} 1 & 1 & 0 & | & 0 \\ 0 & -1 & 0 & | & -1 \\ 0 & 0 & 1 & | & 1 \end{pmatrix} = M'$$

abbiamo che

$$S_{M'} = \begin{cases} x + y = 0 \\ -y = -1 \\ z = 1 \end{cases}$$

da cui si ricava subito che S è determinato con unica soluzione $(-1, 1, 1)$.

Osservazione

È possibile in maniera del tutto analoga definire operazioni elementari e equivalenza per colonne, ma il corrispettivo del Teorema 3.3 non è più valido. Per esempio, sia

$$S = \begin{cases} x + y = 2 \\ x - y = 0 \end{cases} .$$

Allora $M' = \begin{pmatrix} 1 & 2 & | & 2 \\ 1 & 0 & | & 0 \end{pmatrix}$ si ottiene da $M_S = \begin{pmatrix} 1 & 1 & | & 2 \\ 1 & -1 & | & 0 \end{pmatrix}$ sommando la prima colonna alla seconda ma $sol(S) = \{(1,1)\}$ mentre $sol(S_{M'}) = \{(0,1)\}$.

3.2. Algoritmo di Gauss

In questa sottosezione introduciamo un algoritmo, cioè una procedura operativa che può essere tradotta in un programma, che data un matrice A permette di determinare una matrice A' equivalente per righe a A ma di forma semplificata.
Se $A \in M_{m,n}$ e se $[A]_i$ è una riga $\neq O$ di A, indichiamo con $j(i)$ l'indice di colonna del primo elemento non nullo di A_i. Quindi $[A]_{i,j(i)} \neq 0$, mentre $[A]_{i,k} = 0$ per $k < j(i)$.
Ad esempio, se $[A]_2 = (0, 0, -4, 1, 0)$, allora $j(2) = 3$.

Definizione 3.4. *Una matrice $A \in M_{m,n}$ si dice matrice a scala se $A = O$ oppure se valgono le seguenti condizioni:*

a. Esiste un intero r, $1 \leq r \leq m$, tale che $[A]_i \neq O$ per $i \leq r$ e $[A]_i = O$ per $i > r$.
b. Se $1 \leq i < k \leq m$, allora $j(i) < j(k)$.

Una matrice a scala A si dice ridotta se soddisfa a

c. Se $1 \leq i \leq r$ e $1 \leq k < i$, allora $[A]_{k,j(i)} = 0$.

Una matrice a scala A si dice normalizzata se soddisfa a

d. Se $1 \leq i \leq r$, allora $[A]_{i,j(i)} = 1$.

Per brevità, scriveremo «srn » per «a scala ridotta normalizzata ».
Gli elementi $p_i = [A]_{i,j(i)}$ di una matrice a scala A sono detti i pivots di A. Il numero dei pivots coincide con il numero r delle righe $\neq O$.

Prima di fare alcuni esempi, osserviamo che la condizione (b) implica $i \leq j(i)$ e $r \leq n$: quindi $r \leq min\{m,\ n\}$.
Inoltre, se A è srn, le colonne che contengono i pivots sono i vettori colonna canonici $e_1, \dots, e_r$ (cioè $[A]^{j(i)} = e_i$ per $i = 1, \dots, r$).

Esempio

1. Le seguenti matrici sono rispettivamente a scala, a scala ridotta e srn.

$$\begin{pmatrix} 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 0 & 3 & -5 \\ 0 & 0 & 0 & 0 & -6 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & -6 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

2. L'unica matrice $m \times n$ srn con $r = n$ è la matrice espressa a blocchi come $\begin{pmatrix} I_n \\ O_{m-n,n} \end{pmatrix}$. In particolare, se $m = n$, abbiamo I_n.
3. Se $A \in M_{m,n}$ è srn e $j(i) = i$ per $i = 1, \ldots, r$, allora A si rappresenta a blocchi come

$$A = \begin{pmatrix} I_r & | & A_0 \\ - & - & - \\ 0 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix}$$

dove A_0 è un blocco $r \times n - r$.
Per esempio, se $m = n = 2$ e $r = 2$,

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

Dimostriamo ora il teorema principale di questa parte del corso.

Teorema 3.5 (Algoritmo di Gauss). *Ogni matrice $A \in M_{m,n}$ è equivalente per righe a una matrice srn.*

PROVA

Il caso $A = O$ è ovvio, quindi sia $A \neq O$. Mostriamo dapprima che A è equivalente per righe a una matrice a scala. I passaggi teorici saranno accompagnati da un esempio numerico.
Sia A rappresentata per colonne, $A = \begin{pmatrix} [A]^1 & \ldots & [A]^n \end{pmatrix}$, e sia $[A]^j$ la prima colonna non nulla. Sia $[A]_{i,j}$ il primo elemento $\neq 0$ di $[A]^j$.

Esempio

Se

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

$j = 2$ e $[A]_{2,2} = 2$ è il primo elemento $\neq 0$ di $[A]^2$.

Se poniamo per semplificare $[A]_{h,k} = a_{h,k}$ e trasformiamo A con le $m - i$ operazioni elementari del I tipo date da

$$A \overset{(k)-\frac{a_{k,j}}{a_{i,j}}(i)}{\longrightarrow} A'$$

per $k = i+1, \ldots, m$, otteniamo una matrice a blocchi

$$A' = \begin{pmatrix} 0 & \ldots & 0 & 0 & * & \ldots & * \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & * & \ldots & * \\ 0 & \ldots & 0 & a_{i,j} & * & \ldots & * \\ 0 & \ldots & 0 & 0 & * & \ldots & * \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & * & \ldots & * \end{pmatrix}$$

dove $*$ indica un generico elemento.

Esempio

Se effettuiamo $(3) + \frac{1}{2}(2)$ e $(4) - (2)$, otteniamo

$$A' = \begin{pmatrix} 0 & 0 & 0 & 1 & 3 \\ 0 & 2 & 2 & 2 & 4 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & -5 \end{pmatrix}.$$

Nel caso in cui $i \neq 1$, effettuiamo un'operazione del II tipo, $A' \overset{(1)\leftrightarrow(i)}{\longrightarrow} A''$, in modo da ottenere una matrice a blocchi

$$A'' = \begin{pmatrix} 0 & \ldots & 0 & a_{i,j} & * & \ldots & * \\ 0 & \ldots & 0 & 0 & & & \\ \vdots & \ddots & \vdots & \vdots & & B & \\ 0 & \ldots & 0 & 0 & & & \end{pmatrix}$$

dove B è un blocco $(m-1) \times (n-j)$.

Poiché $[A'']_{1,j} = a_{i,j} \neq 0$, poniamo $j(1) = j$; allora $a_{i,j}$ è il pivot p_1.

Esempio

Se effettuiamo $(1) \leftrightarrow (2)$, otteniamo

$$A'' = \begin{pmatrix} 0 & 2 & 2 & 2 & 4 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & -5 \end{pmatrix},$$

$B = \begin{pmatrix} 0 & 1 & 3 \\ 0 & 1 & 3 \\ 0 & 1 & -5 \end{pmatrix}$, $j(1) = 2$ e $p_1 = 2$.

Se ora B è già a scala, anche A'' lo è, altrimenti riapplichiamo il procedimento descritto alla matrice a blocchi $\begin{pmatrix} O_{m-1,j} & | & B \end{pmatrix} \in M_{m-1,n}$, ottenuta da A'' eliminando la prima riga. Considerando le operazioni effettuate come operazioni su A'', siccome la prima riga di A'' e il blocco nullo restano invariati, otteniamo un indice di colonna $j(2) > j(1)$ e una matrice a blocchi

$$A''' = \begin{pmatrix} 0 & \dots & 0 & p_1 & * & \dots & * \dots & * & \\ 0 & \dots & 0 & 0 & 0 & \dots & p_2 & \dots & * \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & & \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & B' & \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & & \end{pmatrix}$$

dove p_2 è in posizione $2,\ j(2)$ e B' è un blocco $(m-2) \times (m - j(2))$.

Esempio

Se effettuiamo $(3) - (2)$ e $(4) - (2)$, otteniamo

$$A''' = \begin{pmatrix} 0 & 2 & 2 & 2 & 4 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -8 \end{pmatrix},$$

$B' = \begin{pmatrix} 0 \\ -8 \end{pmatrix}$, $j(2) = 4$ e $p_2 = 1$.

Quindi dopo un numero finito di applicazioni del procedimento (ciascuna consistente in un numero finito di operazioni elementari) otteniamo una matrice a scala.

Esempio

Se effettuiamo $(3) \leftrightarrow (4)$, otteniamo

$$\begin{pmatrix} 0 & 2 & 2 & 2 & 4 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & -8 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

che è una matrice a scala con $r = 3$, $j(1) = 2$, $j(2) = 4$, $j(3) = 5$, $p_1 = 2$, $p_2 = 1$, $p_3 = -8$.

Completiamo ora la dimostrazione. Coma abbiamo visto, A è equivalente a una matrice a scala. Dunque, per la transitività della relazione $\sim$, possiamo assumere che A sia a scala e provare che essa è equivalente a una matrice srn.

Se p_i per $i = 1, \ldots, r$ sono i pivots di A e trasformiamo A con le r operazioni del III tipo

$$A \stackrel{\frac{1}{p_i}(i)}{\longrightarrow} A'$$

dove $i = 1, \ldots, r$, otteniamo una matrice A' a scala normalizzata.

Esempio

Se

$$A = \begin{pmatrix} 0 & 2 & 2 & 2 & 4 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & -8 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

e se effettuiamo $\frac{1}{2}(1)$ e $-\frac{1}{8}(3)$, otteniamo

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

che è a scala normalizzata.

Infine, se $r > 1$, effettuando per ogni $i = r, \dots, 2$ le $i-1$ operazioni del I tipo

$$A' \overset{(k)-a_{k,j(i)}(i)}{\longrightarrow} A''$$

($k = 1, \dots, i-1$) abbiamo che A'' è srn.

Esempio

Se effettuiamo $(1)-2(3)$, $(2)-3(3)$ e in seguito effettuiamo sulla matrice così trasformata $(1)-(2)$, otteniamo

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

che è srn.

3.3. Studio e risoluzione di sistemi

Sia ora $S : AX = B$ un sistema a m equazioni e n incognite. Abbiamo visto che la matrice $A \in M_{m,n}$ può essere trasformata in una matrice srn A' con un numero finito di operazioni elementari. Se applichiamo ora la medesima sequenza di operazioni elementari alla matrice completa $M_S = \begin{pmatrix} A & | & B \end{pmatrix}$, otteniamo, per la Proposizione 3.2, una matrice $M' = \begin{pmatrix} A' & | & B' \end{pmatrix}$.

Dunque M' è la matrice completa associata al sistema $S' : A'X = B'$, equivalente a S per il Teorema 3.3.

Studiamo $sol(S)$ per mezzo di S': per semplificare la notazione poniamo $[A']_{i,j} = a_{i,j}$ e $[B']_i = b_i$ per $i = 1, \dots, m$, $j = 1, \dots, n$.

Scrivendo esplicitamente il sistema S' abbiamo

$$\begin{cases} x_{j(1)} + \sum\limits_{\substack{k>j(1)\\ k\neq j(i)}} a_{1,k}x_k = b_1 \\ \\ x_{j(2)} + \sum\limits_{\substack{k>j(2)\\ k\neq j(i)}} a_{2,k}x_k = b_2 \\ \vdots \\ x_{j(r)} + \sum\limits_{k>j(r)} a_{r,k}x_k = b_r \\ 0 = b_{r+1} \\ \vdots \\ 0 = b_m \end{cases}$$

dove $k \neq j(i)$ significa che k non deve essere un indice di colonna di pivots, quindi nella prima equazione $k \neq j(2), j(3), \ldots, j(r)$ e così via; ovviamente per l'equazione r-esima questa condizione è verificata d'ufficio. Per quanto detto k può assumere $n - r$ valori.

Esempio

Se

$$M' = \begin{pmatrix} 1 & 0 & 2 & 0 & 1 & | & -1 \\ 0 & 1 & -1 & 0 & 4 & | & 3 \\ 0 & 0 & 0 & 1 & 2 & | & 1 \\ 0 & 0 & 0 & 0 & 0 & | & b \end{pmatrix}$$

abbiamo

$$S' = \begin{cases} x_1 + 2x_3 + x_5 = -1 \\ x_2 - x_3 + 4x_5 = 3 \\ x_4 + 2x_5 = 1 \\ 0 = b \end{cases}$$

che sarà risolubile solo per $b = 0$.

Distinguiamo vari casi:

1. Se $b_i \neq 0$ per qualche $i > r$, allora il sistema è impossibile ($sol(S) = \emptyset$).
2. Se per ogni $i > r$ abbiamo $b_i = 0$, allora vi sono due possibilità.

2a. $r < n$: in questo caso il sistema è indeterminato e $X = (x_1, \ldots, x_n) \in sol(S)$ se e solo se sono verificate le equazioni

$$(*) = \begin{cases} x_{j(1)} = -\sum\limits_{\substack{k>j(1)\\ k\neq j(i)}} a_{1,k}x_k + b_1 \\ \\ x_{j(2)} = -\sum\limits_{\substack{k>j(2)\\ k\neq j(i)}} a_{2,k}x_k + b_2 \\ \vdots \\ x_{j(r)} = -\sum\limits_{k>j(r)} a_{r,k}x_k + b_r \end{cases}$$

Le variabili $\{x_{j(1)}, \ldots, x_{j(r)}\}$ si dicono *variabili dipendenti*, mentre le restanti $n - r$ variabili si dicono *variabili indipendenti*. Dunque, se assegnamo un valore a ciascuna delle varibili indipendenti (cioè scegliamo un elemento di K^{n-r}), possiamo determinare un elemento di $sol(S)$ tramite le equazioni $(*)$: per questo si dice che S ha ∞^{n-r} soluzioni.

Esempio

Se nell'esempio precedente poniamo $b = 0$, abbiamo

$$(*) = \begin{cases} x_1 = -2x_3 - x_5 - 1 \\ x_2 = x_3 - 4x_5 + 3 \\ x_4 = -2x_5 + 1 \end{cases}$$

con $x_1,\ x_2,\ x_4$ variabili dipendenti e $x_3,\ x_5$ variabili indipendenti.

2b. $r = n$: il sistema è determinato poiché, ricordando che l'unica matrice $m \times n$ srn con $r = n$ è $\begin{pmatrix} I_n \\ O_{m-n,n} \end{pmatrix}$, le equazioni $(*)$ diventano

$$\begin{cases} x_1 = b_1 \\ \vdots \\ x_n = b_n \end{cases}.$$

Dunque $sol(S) = \{(b_1, \ldots, b_n)\}$.

Il caso II corrisponde evidentemente al caso in cui S è risolubile: quindi diciamo che le equazioni $(*)$ sono le *equazioni risolutive* di S.

Osservazione

Se $m < n$, S può solo essere impossibile o indeterminato, in quanto $r \leq m$. In particolare, se S è omogeneo, e $m < n$, S è indeterminato (in questo caso $B = B'$).

Esempio

Si consideri la famiglia di sistemi

$$S_{a,b} = \begin{cases} x_1 - x_3 = -b \\ 2x_1 + x_2 + ax_3 = 1 \\ x_1 + x_2 + x_3 = 1 \end{cases}$$

al variare dei parametri $a,\ b$ in K. Famiglie di questo tipo vengono usualmente chiamate *sistemi con parametri*. In questo caso abbiamo una famiglia di matrici associate

$$M_{a,b} = \begin{pmatrix} 1 & 0 & -1 & | & -1 \\ 2 & 1 & a & | & b \\ 1 & 1 & 1 & | & 1 \end{pmatrix}.$$

Comunque fissati a e b e applicando l'algoritmo di Gauss, si può vedere $M_{a,b}$ è equivalente alla matrice a scala

$$\begin{pmatrix} 1 & 1 & 1 & | & 1 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & a & | & b \end{pmatrix}.$$

Ora, per $a \neq 0$ e b qualsiasi, la matrice dei coefficienti è equivalente alla matrice I: dunque abbiamo un sistema determinato con soluzione $(-b + b/a, 1 + b - 2b/a, b/a)$.

Se $a = 0$, la matrice completa è equivalente alla matrice

$$\begin{pmatrix} 1 & 0 & -1 & | & 1 \\ 0 & 1 & 2 & | & 0 \\ 0 & 0 & 0 & | & b \end{pmatrix}.$$

Se $b \neq 0$, il sistema è evidentemente impossibile, mentre se $b = 0$ il sistema è indeterminato con ∞^1 soluzioni e abbiamo le equazioni risolutive

$$\begin{cases} x_1 = x_3 + 1 \\ x_2 = -2x_3 \end{cases}$$

con x_3 variabile indipendente e $x_1,\ x_2$ variabili indipendenti.

Dunque dando un valore arbitrario a x_3 determino una soluzione: ad esempio per $x_3 = -1$ ho la soluzione $(0, 2, -1)$.

3.4. Rango di una matrice

Utilizziamo il lavoro svolto per introdurre il concetto di rango di una matrice.

Teorema 3.6. *Se $A,\ B \in M_{m,n}$ sono srn e $A \sim B$, allora $A = B$.*

PROVA

Non è restrittivo supporre che le colonne dei pivots di A siano le prime r, cioè che gli indici dei pivots di A siano $1, \dots, r$.
Siano $j(1), \dots, j(\rho)$ gli indici di colonna dei pivots di B.
Allora $j(1) = 1$, poiché $j(1)$ è l'indice della prima colonna $\neq O$, che non varia per operazioni di riga. Supponiamo ora che $j(2) \neq 2$, quindi $j(2) > 2$. Le sottomatrici A' e B' di A e B rispettivamente costituite dalle prime 2 colonne sono della forma

$$A' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}$$

$$B' = \begin{pmatrix} 1 & * \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}$$

quindi non possono essere equivalenti per riga, il che è assurdo per la Proposizione 3.2. Dunque dev'essere $j(2) = 2$. Iterando questo ragionamento possiamo provare che $\rho = r$ e $j(i) = i$ per $i = 1, \dots, r$.
Dunque le matrici A e B si rappresentano a blocchi come

$$A = \begin{pmatrix} I_r & | & A_0 \\ - & - & - \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix}, \quad B = \begin{pmatrix} I_r & | & B_0 \\ - & - & - \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix}$$

dove A_0 e B_0 sono blocchi $r \times n - r$.
I sistemi omogenei S_A e S_B associati a A e B rispettivamente sono equivalenti per il Teorema 3.3 e, posto $[A]_{i,j} = a_{i,j}$ e $[B]_{i,j} = b_{i,j}$, hanno equazioni risolutive

$$\begin{cases} x_1 = -\sum\limits_{k=r+1}^{n} a_{1,k}x_k \\ {}_2 = -\sum\limits_{k=r+1}^{n} a_{2,k}x_k \\ \vdots \\ x_r = -\sum\limits_{k=r+1}^{n} a_{r,k}x_k \end{cases} \quad \text{e} \quad \begin{cases} x_1 = -\sum\limits_{k=r+1}^{n} b_{1,k}x_k \\ x_2 = -\sum\limits_{k=r+1}^{n} b_{2,k}x_k \\ \vdots \\ x_r = -\sum\limits_{k=r+1}^{n} b_{r,k}x_k \end{cases}$$

Fissato ora $i = r+1, \ldots, n$, se nelle equazioni risolutive di S_A diamo alle variabili indipendenti $x_{r+1}, \ldots, x_n$ i valori $x_j = 0$ per $j \neq i$ e $x_i = -1$, abbiamo che $\overline{X} = (a_{1,i}, \ldots, a_{r,i}, 0, \ldots, 1, \ldots, 0)$ è soluzione di S_A e quindi di S_B.
Questo vuol dire che $\overline{X}$ verifica le equazioni risolutive di S_B: sostituendo otteniamo $a_{k,i} = b_{k,i}$ per $k = 1, \ldots, r$.
Siccome questo vale per ogni i, abbiamo provato che $A = B$.

Per i Teoremi 3.5 e 3.6 possiamo affermare che, per ogni matrice A, esiste un'unica matrice a scala ridotta normalizzata equivalente per righe a A. Indicheremo tale matrice con A_{srn} e diremo che è la *forma srn* di A. Due matrici sono equivalenti per righe se e solo se hanno la stessa forma srn: infatti, se $A \sim B$ allora $A_{srn} \sim B_{srn}$ per transitività, quindi $A_{srn} = B_{srn}$ per il Teorema 3.6. Viceversa $A \sim A_{srn} = B_{srn} \sim B$ implica $A \sim B$.
Diamo ora una importante definizione.

Definizione 3.7. *Se $A \in M_{m,n}$, il numero di pivots di A_{srn} si dice rango di A e si denota con $r(A)$. Per convenzione, $r(O) = 0$.*

Per quanto detto prima, se due matrici sono equivalenti allora hanno lo stesso rango. Inoltre, poiché il numero di pivots (che coincide con il numero r di righe $\neq O$ di A_{srn}) non varia passando da una matrice a scala a una srn, possiamo calcolare il rango di una matrice A determinando una matrice a scala equivalente per righe a A (come nella prova del Teorema 3.5) e contando i pivots (o le righe non nulle).
Osserviamo che $r(A) = 0 \Leftrightarrow A = O$ e ricordiamo che, se $A \in M_{m,n}$ allora $r(A) \leq min\{m, n\}$.
ATTENZIONE!

Il rango di una matrice A NON è il numero di righe non nulle di A, a meno che A non sia a scala. Se così non è, bisogna PRIMA trasformare A in una matrice a scala con operazioni di riga e POI contare le righe non nulle (o i pivots) della matrice trasformata.

Esempio

$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix}$ ha forma srn $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, dunque ha rango 1, anche se non ha righe nulle.

Ci sarà utile in seguito il

Lemma 3.8. *Se $A \in M_{m,n}$ e B è una sottomatrice $p \times n$ di A (quindi formata da p righe di A), allora $r(B) \leq r(A)$. Inoltre, se $r(A) = m$, allora $r(B) = p$.*

Applicando il concetto di rango alla teoria dei sistemi otteniamo il

Teorema di Rouchè-Capelli. *Sia $S : AX = B$ un sistema con m equazioni e n incognite. Allora S è risolubile se e solo se il rango della matrice dei coefficienti $A \in M_{m,n}$ è uguale al rango della matrice completa $M_S = \begin{pmatrix} A & | & B \end{pmatrix}$.*
Inoltre, nel caso in cui S sia indeterminato, S ha $\infty^{n-r(A)}$ soluzioni.

PROVA

Lo studio della risolubilità dei sistemi può essere sinteticamente espresso nel modo seguente: con le operazioni di riga che trasformano A in A_{srn} ottengo una matrice $M' = \begin{pmatrix} A_{srn} & | & B' \end{pmatrix}$ equivalente per righe a M_S; dunque il sistema $S_{M'}$ è equivalente a S.
Ora S sarà risolubile se e solo se $b_i = [B']_i = 0$ per $i = r+1, \ldots, n$. Questa condizione equivale a sua volta al fatto che $r(M') = r(M_S) = r = r(A)$.

Osservazione

> Dato un sistema $S : AX = B$, dallo studio fatto in precedenza risulta che $r(M_S) = r(A)$ oppure che $r(M_S) = r(A) + 1$.
> Infatti, posto $r(A) = r$, le ultime $n - r$ righe di $M' = \begin{pmatrix} A_{srn} & | \, B' \end{pmatrix}$ sono del tipo $(0, \ldots, 0, b_i)$ con $i = r+1, \ldots, n$. Dunque, se $b_i \neq 0$ per qualche $i > r$ possiamo annullare tutte queste righe con operazioni del I tipo e $r(M_S) = r(M') = r + 1$.

3.5. Matrici elementari e matrici invertibili

Una matrice $E \in M_n$ si dice *elementare* se è ottenuta da I_n con una operazione elementare di riga. E è del I , II o III tipo a seconda del tipo dell'operazione effettuata.

Usando le notazioni introdotte, poniamo:

1. $E = E_a(i,j)$ se $I_n \overset{(i)+a(j)}{\to} E$; dunque $[E]_i = {}^te_i + a\,{}^te_j$ e $[E]_k = {}^te_k$ per $k \neq i$.
2. $E = E(i,j)$ se $I_n \overset{(i)\leftrightarrow(j)}{\to} E$; dunque $[E]_i = {}^te_j$, $[E]_j = {}^te_i$ e $[E]_k = {}^te_k$ per $k \neq i,\ j$.
3. $E = E_a(i)$ se $I_n \overset{a(i)}{\to} E$; dunque $[E]_i = a\,{}^te_i$ e $[E]_k = {}^te_k$ per $k \neq i$.

Esempio

Per $n = 2$ abbiamo

$$E_a(1,2) = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \qquad E(1,2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad E_a(2) = \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}.$$

Proposizione 3.9. *Siano $A,\ A' \in M_{m,n}$ e supponiamo che A' sia ottenuta da A con una operazione elementare. Se E è la matrice elementare ottenuta da I_n con la medesima operazione elementare, allora $A' = EA$.*

Non diamo la prova della Proposizione 3.9, ma facciamo degli esempi.

Esempi

Se $A = \begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix}$ abbiamo:

I tipo: $E_2(1,2)A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 8 & -1 \\ 0 & 1 \end{pmatrix} = A'$.

II tipo: $E(1,2)A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 2 & -1 \end{pmatrix} = A'$.

III tipo: $E_{-4}(1)A = \begin{pmatrix} -4 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} -8 & 4 \\ 3 & 0 \end{pmatrix} = A'$.

In base alla Proposizione precedente, le matrici elementari sono invertibili. Infatti

1. $E_{-a}(i,j)E_a(i,j) = E_a(i,j)E_{-a}(i,j) = I_n$;
2. $E(i,j)E(i,j) = I_n$;
3. $E_{1/a}(i)E_a(i) = E_{1/a}(i)E_a(i) = I_n$.

È evidente che per trasformare una specifica matrice tramite una data operazione per riga conviene eseguire l'operazione direttamente piuttosto che applicare la Proposizione 3.9. Comunque le matrici elementari sono interessanti per lo studio generale di proprietà delle matrici, per esempio l'invertibilità.

In base alla Proposizione 3.9, se $A,\ A' \in M_{m,n}$ e $A \sim A'$, esistono $E_1, \ldots, E_q$ matrici elementari tali che $A' = E_q E_{q-1} \ldots E_1 A$. Dunque se $A \in GL_n$, anche $A' \in GL_n$ in quanto prodotto di matrici invertibili.

Teorema 3.10. *Sia $A \in M_n$. Allora le seguenti affermazioni sono equivalenti:*

a. $A \in GL_n$;

b. $r(A) = n$;

c. $A_{srn} = I_n$;

d. *A è prodotto di matrici elementari.*

PROVA

(a) $\Rightarrow$ (b): se $r(A) < n$, A_{srn} (o una qualsiasi matrice a scala equivalente a A) avrebbe almeno una riga nulla e quindi, per la Proposizione 2.6, sarebbe singolare, mentre ogni matrice equivalente per righe a A appartiene a GL_n.
(b) $\Rightarrow$ (c): A_{srn} è quadrata $n \times n$ e ha n pivots tutti uguali a 1. Quindi la sue colonne sono e_1, e_2, ... , e_n, da cui $A_{srn} = I_n$.
(c) $\Rightarrow$ (d): per ipotesi esistono matrici elementari $E_1, \ldots, E_k$ tali che $I_n = E_k \cdots E_1 A$; dunque $A = E_1^{-1} \cdots E_k^{-1}$ e (d) è provata poiché le inverse di matrici elementari sono ancora matrici elementari.
(d) $\Rightarrow$ (a): $A \in GL_n$ in quanto prodotto di matrici invertibili.

Il Teorema 3.10 permette di riformulare più sinteticamente il concetto di equivalenza per righe:

Corollario 3.11. *Se $A,\ B \in M_{m,n}$, A è equivalente per righe a B se e solo se esiste $P \in GL_m$ tale che $B = PA$. (ovviamente P è ottenuta come prodotto di matrici elementari).*

Vediamo ora alcune applicazioni del Teorema 3.10.

1. Dimostriamo quanto asserito dopo la definizione di matrice invertibile:

Teorema 3.12. *Se $A,\ B \in M_n$ sono tali che $AB = I_n$, allora $A,\ B \in GL_n$ e $B = A^{-1}$ (quindi $BA = I_n$).*

PROVA

Proviamo dapprima che $A \in GL_n$; se così non fosse, $A' = A_{srn}$ avrebbe almeno una riga nulla (per esempio $[A']_n$) per la Proposizione 2.6.
Ora $A' = E_k \cdots E_1 A$, con $E_1, \dots, E_k$ matrici elementari, quindi per ipotesi $A'B = E_k \cdots E_1 AB = E_k \cdots E_1 I_n = C$. Questo è assurdo perché $A'B$ ha almeno una riga nulla (per esempio $[A'B]_n$) mentre C è prodotto di matrici invertibili, e quindi invertibile.
Poiché dunque $A \in GL_n$ e $AB = I_n$, abbiamo $A^{-1}AB = A^{-1}I_n$, da cui $B = A^{-1}$.

2. Usando le operazioni per righe è possibile verificare se una matrice $A \in M_n$ è invertibile e, in caso affermativo, calcolarne nel contempo l'inversa.
Per il Teorema 3.10 , $A \in GL_n$ se e solo $A_{srn} = I_n$; in tal caso, se $E_1, \dots, E_k$ sono matrici elementari tali che $A_{srn} = E_k \cdots E_1 A$, abbiamo che $A^{-1} = E_k \cdots E_1$.
Consideriamo la matrice a blocchi $\overline{A} = \begin{pmatrix} A & | & I_n \end{pmatrix} \in M_{2n}$. Allora
$$E_k \cdots E_1 \overline{A} = \begin{pmatrix} E_k \cdots E_1 A & | & E_k \cdots E_1 \end{pmatrix} = \begin{pmatrix} I_n & | & A^{-1} \end{pmatrix}.$$
Tradotta in pratica, l'uguaglianza precedente ci dice che, operando per righe su $\overline{A}$, se riusciamo a trasformare il primo blocco in I_n, allora $A \in GL_n$ e, nel contempo il secondo blocco diventa A^{-1}; se invece compare una riga nulla nel primo blocco, allora A è singolare.

Esempio

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

$$\begin{pmatrix} 1 & 0 & -1 & | & 1 & 0 & 0 \\ 2 & 1 & 1 & | & 0 & 1 & 0 \\ 1 & 1 & 1 & | & 0 & 0 & 1 \end{pmatrix} \overset{(2)-2(1),(3)-(1)}{\longrightarrow} \begin{pmatrix} 1 & 0 & -1 & | & 1 & 0 & 0 \\ 0 & 1 & 3 & | & -2 & 1 & 0 \\ 0 & 1 & 2 & | & -1 & 0 & 1 \end{pmatrix} \overset{(3)-(2)}{\longrightarrow}$$

$$\begin{pmatrix} 1 & 0 & -1 & | & 1 & 0 & 0 \\ 0 & 1 & 3 & | & -2 & 1 & 0 \\ 0 & 0 & -1 & | & 1 & -1 & 1 \end{pmatrix} \overset{(1)-(3),(2)+3(3)}{\longrightarrow} \begin{pmatrix} 1 & 0 & 0 & | & 0 & 1 & -1 \\ 0 & 1 & 0 & | & 1 & -2 & 3 \\ 0 & 0 & 1 & | & -1 & 1 & -1 \end{pmatrix}.$$

Dunque $A \in GL_n$ e $A^{-1} = \begin{pmatrix} 0 & 1 & -1 \\ 1 & -2 & 3 \\ -1 & 1 & -1 \end{pmatrix}$.

3. Infine proviamo un teorema sui sistemi quadrati.

Teorema di Cramer. *Sia $A \in M_n$. $A \in GL_n$ se e solo se per ogni $B \in K^n$ il sistema $AX = B$ è determinato.*

PROVA

Sia $A \in GL_n$. Se $\overline{X} \in sol(S)$, allora $A\overline{X} = B$, dunque $\overline{X} = A^{-1}B$. Quindi $A^{-1}B$ è l'unica soluzione.
Viceversa, se per ogni B il sistema è determinato, per $i = 1, \dots, n$ ogni sistema $AX = e_i$ ha un'unica soluzione $\overline{X}_i$.
Se $P = \begin{pmatrix} \overline{X}_1 & \dots & \overline{X}_n \end{pmatrix}$ è la matrice che ha come colonne i vettori $\overline{X}_1, \dots, \overline{X}_n$, facendo il prodotto a blocchi abbiamo $AP = I_n$, quindi $A \in GL_n$ e $P = A^{-1}$ per il Teorema 3.12.

Esempio

Il sistema

$$S = \begin{cases} x_1 - x_3 = 1 \\ 2x_1 + x_2 + x_3 = 0 \\ x_1 + x_2 + x_3 = 2 \end{cases}$$

ha matrice dei coefficienti $A = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

Per l'esempio precedente e il Teorema di Cramer, il sistema è determinato e l'unica soluzione è data da

$$\begin{pmatrix} 0 & 1 & -1 \\ 1 & -2 & 3 \\ -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ 7 \\ -3 \end{pmatrix}.$$

4. DETERMINANTI

4.1. Definizione e proprietà fondamentali

Se consideriamo una matrice A quadrata di ordine 2, il suo *determinante* è il numero $D(A)$ espresso dalla formula

$$D(A) = [A]_{1,1}[A]_{2,2} - [A]_{1,2}[A]_{2,1}.$$

Per esempio, se $D(\begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}) = 1 \cdot (-1) - 2 \cdot 3 = -7$.
Vogliamo estendere questa nozione alle matrici quadrate di ordine qualsiasi. Ovviamente, se $A \in M_1 = K$, A è un numero ed è naturale porre $D(A) = A$.
Se $n > 1$ e se A è una matrice quadrata $n \times n$, indicheremo d'ora in poi con $A_{i,j}$ la sottomatrice di A ottenuta sopprimendo la i-esima riga e la j-esima colonna; dunque $A_{i,j} \in M_{n-1}$.
Per esempio, se $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ -2 & 3 & -1 \end{pmatrix}$, abbiamo

$$A_{1,1} = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}, \qquad A_{2,3} = \begin{pmatrix} 1 & 1 \\ -2 & 3 \end{pmatrix}$$

e così via.
Osserviamo che se $A \in M_2$, $A_{1,1} = [A]_{2,2,}$, $A_{1,2} = [A]_{2,1}$, $A_{2,1} = [A]_{1,2}$ e $A_{2,2} = [A]_{1,1}$. Dunque, posto $[A]_{i,j} = a_{i,j}$, abbiamo

$$D(A) = a_{1,1}D(A_{1,1}) - a_{1,2}D(A_{1,2}).$$

Questa formula ci dà uno spunto per generalizzare: se $A \in M_3$, possiamo porre

$$D(A) = a_{1,1}D(A_{1,1}) - a_{1,2}D(A_{1,2}) + a_{1,3}D(A_{1,3}).$$

Per esempio, se $A = \begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 2 \\ -2 & 3 & -1 \end{pmatrix}$,

$$D(A) = 2D(A_{1,1}) - D(A_{1,2}) - 3D(A_{1,3}) = -24.$$

Nel caso di una matrice A quadrata di ordine n qualsiasi, definiamo allora il *determinante* $D(A) \in K$ di A usando il seguente processo ricorsivo su n:

1. se $n = 1$, $A \in K$ e $D(A) = A$;
2. se $n > 1$, supponiamo di aver definito $D(B)$ per ogni $B \in M_{n-1}$ e poniamo

   $$\begin{aligned} D(A) =& a_{1,1}D(A_{1,1}) - a_{1,2}D(A_{1,2}) + \dots + (-1)^{j+1}a_{1,j}D(A_{1,j}) \cdots \\ &+ (-1)^{n+1}a_{1,n}D(A_{1,n}) \end{aligned}$$

   cioè

   $D(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{1,j} D(A_{1,j}).$

La definizione precedente può sembrare piuttosto arbitraria, ma mostreremo che il determinante ha notevole relazione con quanto sviluppato in precedenza.
Enunciamo qui di seguito senza dimostrazione le proprietà fondamentali del determinante. Data $A \in M_n$ valgono:

D1. Per $1 \leq i \leq n$

$$D(A) = \sum_{k=1}^{n} (-1)^{k+i} a_{i,k} D(A_{i,k}).$$

(*sviluppo* di $D(A)$ rispetto alla i-esima riga: nel caso $i = 1$ coincide con la definizione).

Per esempio, usando lo sviluppo rispetto alla seconda riga, otteniamo

$$D(\begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 2 \\ -2 & 3 & -1 \end{pmatrix}) = D(\begin{pmatrix} 2 & -3 \\ -2 & -1 \end{pmatrix}) - 2D(\begin{pmatrix} 2 & 1 \\ -2 & 3 \end{pmatrix}) = -24.$$

D2. Se $B \in M_n$, allora $D(AB) = D(A)D(B)$ (formula di Binet).

(D2) implica evidentemente che $D(AB) = D(BA)$. Per esempio, se $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ e $B = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$,

$$\begin{aligned} D(AB) = & D(\begin{pmatrix} 4 & 2 \\ 5 & 1 \end{pmatrix}) = -6 = D(A)D(B) = D(\begin{pmatrix} 2 & 4 \\ 3 & 3 \end{pmatrix}) \\ & = D(BA). \end{aligned}$$

Osserviamo che, in generale, $D(A+B) \neq D(A) + D(B)$; per esempio $D(I_2) = D(-I_2) = 1$, quindi $D(I_2) + D(-I_2) = 2$, mentre $D(I_2 + (-I_2)) = D(O_2) = 0$.

D3. $D(A) = D({}^tA)$.

Come conseguenza di (D1) e (D3), abbiamo le ulteriori proprietà

D4. Per $1 \leq j \leq n$

$$D(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{k,j} D(A_{k,j}).$$

(sviluppo di $D(A)$ rispetto alla j-esima colonna). Per esempio, usando lo sviluppo rispetto alla prima colonna, otteniamo

$$D(\begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 2 \\ -2 & 3 & -1 \end{pmatrix}) = 2 \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix} - 2 \begin{pmatrix} 1 & -3 \\ 1 & 2 \end{pmatrix} = -24.$$

D5. Se $[A]_i$ è somma di due vettori riga B e C, allora

$$D(A) = D(\begin{pmatrix} [A]_1 \\ \vdots \\ B \\ \vdots \\ [A]_n \end{pmatrix}) + D(\begin{pmatrix} [A]_1 \\ \vdots \\ C \\ \vdots \\ [A]_n \end{pmatrix}),$$

dove B e C occupano la i-esima riga.

La formula analoga vale nel caso in cui B e C siano vettori colonna e $[A]^j = B + C$.

Per esempio

$$D(\begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 2 \\ -2 & 3 & -1 \end{pmatrix}) = D(\begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 0 \\ -2 & 3 & -1 \end{pmatrix}) + D(\begin{pmatrix} 2 & 1 & -3 \\ 0 & 0 & 2 \\ -2 & 3 & -1 \end{pmatrix}) =$$

$$D(\begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 2 \\ 0 & 3 & -1 \end{pmatrix}) + D(\begin{pmatrix} 0 & 1 & -3 \\ 0 & 0 & 2 \\ -2 & 3 & -1 \end{pmatrix}) = -24.$$

Evidentemente (D1), (D4) e (D5) ci permettono di semplificare il calcolo del determinante: in particolare, dagli esempi si vede che conviene scegliere gli sviluppi rispetto alle righe o colonne che contengono il maggior numero di zeri, eliminando così addendi nella sommatoria. Queste considerazioni ci permettono di affermare che
Se $A \in M_n$ ha una riga o una colonna nulla, allora $D(A) = 0$.
Per alcuni tipi di matrici il determinante è esprimibile facilmente in qualsiasi ordine.

Definizione 4.1. *Sia $A \in M_n$.*

1. *A si dice triangolare superiore se $[A]_{i,j} = 0$ per $i > j$.*
2. *A si dice triangolare inferiore se $[A]_{i,j} = 0$ per $i < j$.*
3. *A si dice diagonale se è sia triangolare superiore che inferiore (quindi se $[A]_{i,j} = 0$ per $i \neq j$).*

Nei casi (1) e (2) diciamo genericamente che A è triangolare.

Trasponendo una matrice triangolare superiore si ottiene una triangolare inferiore e viceversa; ovviamente una matrice diagonale è simmetrica.

Esempi

1. Se
$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$
A è triangolare superiore, $B = {}^tA$ è triangolare inferiore e C è diagonale.
2. Una matrice elementare del I tipo $E_a(i,j)$ è triangolare superiore se $i < j$, inferiore se $i > j$.

3. Una matrice quadrata A di ordine n a scala è triangolare superiore con sulla diagonale i pivots $p_1, \ldots, p_r$ seguiti da $n-r$ zeri; se A è a scala ridotta e non vi sono righe nulle, A è diagonale.
4. O_n, I_n e le matrici elementari del III tipo sono diagonali.

Proposizione 4.2. *Se $A \in M_n$ è triangolare allora $D(A)$ è il prodotto degli elementi della diagonale principale.*

PROVA

Per induzione su n. Per $n = 1$ è ovvio. Sia $n > 1$ e supponiamo che la tesi valga per matrici $(n-1) \times (n-1)$.
Se A è triangolare superiore, si consideri lo sviluppo rispetto alla prima colonna. Allora $D(A) = [A]_{1,1} D(A_{1,1})$; poiché $A_{1,1}$ è $(n-1) \times (n-1)$ e triangolare superiore, per ipotesi induttiva abbiamo che $D(A_{1,1})$ è il prodotto degli elementi $[A]_{i,i}$ per $i \geq 2$, da cui la tesi.
Se A è triangolare inferiore, ${}^tA$ sarà triangolare superiore, dunque il ragionamento precedente e (D3) danno il risultato.

Esempio

$D(I_n) = 1$, $D(O_n) = 0$.

Riguardo alle rappresentazioni a blocchi, è spesso utile la seguente Proposizione di cui omettiamo la prova.

Proposizione 4.3. *Se $M \in M_n$ è rappresentata con 4 blocchi quadrati nel modo seguente*

$$M = \begin{pmatrix} A & | & B \\ - & - & - \\ O & | & C \end{pmatrix},$$

allora $D(M) = D(A)D(C)$.

Esempio

$$D\left(\begin{pmatrix} 0 & 1 & | & 0 & 2 \\ -2 & 3 & | & -1 & 1 \\ - & - & - & - & - \\ 0 & 0 & | & 5 & -4 \\ 0 & 0 & | & 1 & 1 \end{pmatrix}\right) = D\left(\begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}\right) D\left(\begin{pmatrix} 5 & -4 \\ 1 & 1 \end{pmatrix}\right) = 18.$$

4.2. Determinante e operazioni elementari

Proposizione 4.4. *Sia $E \in M_n$ una matrice elementare e sia $A \in M_n$. Allora, posto $A' = EA$,*

1. *Se E è del I tipo, $D(E) = 1$ e $D(A') = D(A)$.*
2. *Se E è del II tipo, $D(E) = -1$ e $D(A') = -D(A)$.*
3. *Se E è del III tipo e $E = E_a(i)$, $D(E) = a$ e $D(A') = aD(A)$.*

PROVA

1. Se E è del I tipo, E è triangolare con 1 sulla diagonale, quindi $D(E) = 1$ per la Proposizione 4.3.
2. Sia E è del II tipo e supponiamo per semplicità che $E = E(n-1, n)$. Dunque

$$E = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 1 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

Allora $D(E) = D\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = -1$

3. Se $E = E_a(i)$, E è diagonale con $[E]_{i,i} = a$ e $[E]_{k,k} = 1$ se $k \neq i$. Quindi $D(E) = a$.

Ricordando ora (D2), possiamo dire che se A' è ottenuta da A scambiando due righe o colonne, allora $D(A') = -D(A)$, mentre se è ottenuta da A moltiplicando una riga o colonna per $a \neq 0$, $D(A') = aD(A)$. In particolare se A ha due righe o colonne uguali

o una multiplo dell'altra, allora $D(A) = 0$.
Effettuando d scambi di righe o colonne, si ha $D(A') = (-1)^d D(A)$, mentre moltiplicando k righe o colonne di A per scalari non nulli $a_1, \ldots, a_k$ si ha $D(A') = a_1 \cdots a_k D(A)$: per esempio, se $A' = aA$, $D(A') = a^n D(A)$.
Dunque le operazioni del I tipo lasciano invariato il determinante, mentre quelle del II tipo ne cambiano solo il segno. Poiché una matrice può essere trasformata in una matrice a scala utilizzando solo operazioni per riga del I e II tipo, queste forniscono un metodo per semplificare il calcolo del determinante e ci permettono di provare il seguente

Teorema 4.5. *Se $A \in M_n$, A è invertibile se e solo se $D(A) \neq 0$.*

PROVA

Supponiamo che $A' = E_k \cdots E_1 A$ dove A' è a scala e le E_i sono matrici elementari del I o del II tipo.
(D2) implica che $D(A) = (-1)^d D(A')$, dove d è il numero di operazioni del II tipo effettuate.
Se $A \in GL_n$, allora $r(A) = n$ per il Teorema 3.10 e quindi A' non ha righe nulle. Dunque A' è diagonale e $D(A) = (-1)^d p_1 \cdots p_n \neq 0$.
Viceversa, se $D(A) \neq 0$, anche $D(A') \neq 0$ per la formula precedente e questo può essere solo se A' non ha righe nulle, cioè se $r(A) = n$: quindi, sempre per il Teorema 3.10, $A \in GL_n$.

Esempio

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

$$A \overset{(3)+(1)}{\to} \begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 2 \\ 0 & 4 & -4 \end{pmatrix} \overset{(3)-4(2)}{\to} \begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & -12 \end{pmatrix}$$

quindi $D(A) = D(\begin{pmatrix} 2 & 1 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & -12 \end{pmatrix}) = -24$.

Riassumendo, abbiamo che se $A \in M_n$, allora

$$A \in GL_n \Leftrightarrow r(A) = n \Leftrightarrow D(A) \neq 0 \quad \text{e} \quad A \notin GL_n \Leftrightarrow r(A) < n \Leftrightarrow D(A) = 0.$$

Diamo un'applicazione allo studio dei sistemi con parametri.

Esempio

Si consideri il sistema con parametri $a,\ b \in R$

$$S = \begin{cases} x_1 - 2x_3 = -b \\ 2x_1 + x_2 + ax_3 = 1 \\ x_1 + x_2 + x_3 = 1 \end{cases}.$$

La matrice dei coefficienti è $A = \begin{pmatrix} 1 & 0 & -2 \\ 2 & 1 & a \\ 1 & 1 & 1 \end{pmatrix}$ e $D(A) = -(a+1)$. Quindi, per il Teorema 4.5 e per il Teorema di Cramer, se $a \neq -1$, S è determinato.

Il caso $a = -1$ si studia separatamente applicando l'algoritmo di Gauss.

4.3. Formula dell'inversa e regola di Cramer

Possiamo esprimere l'inversa di una matrice $A \in GL_n$ utilizzando il determinante e le sue proprietà.

Formula dell'inversa. Se $A \in GL_n$ abbiamo

$$[A^{-1}]_{i,j} = (-1)^{i+j} \frac{D(A_{j,i})}{D(A)}.$$

Esempio

Sia $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ una generica matrice 2×2: se $D(M) = ad - bc \neq 0$, applicando la formula dell'inversa abbiamo

$$M^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Ricordiamo ora che, per un sistema quadrato $S : AX = B$ con matrice dei coefficienti invertibile, il Teorema di Cramer assicura l'esistenza e l'unicità della soluzione $\overline{X}$ esprimendola come prodotto di matrici $A^{-1}B$. Applicando la formula dell'inversa e lo sviluppo del determinante, possiamo scrivere esplicitamente tale soluzione.

Regola di Cramer. Sia $S : AX = B$ un sistema quadrato. Se $A \in GL_n$, l'unica soluzione $\overline{X} = (\overline{x}_1, \ldots, \overline{x}_n)$ di S è data data

$$\overline{x}_i = \frac{D(\begin{pmatrix} [A]^1 & \ldots & B & \ldots & [A]^n \end{pmatrix})}{D(A)}$$

Esempio

Sia

$$S : \begin{cases} x + y = 0 \\ -y + 2z = 1 \\ 3x - 2y + 5z = 0 \end{cases} .$$

Allora $D(\begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 2 \\ 3 & -2 & 5 \end{pmatrix}) = 5$, quindi possiamo applicare la regola di Cramer trovando

$$x = \frac{D(\begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 2 \\ 0 & -2 & 5 \end{pmatrix})}{5} = -1, \quad y = \frac{D(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 3 & 0 & 5 \end{pmatrix})}{5} = 1, \quad z = \frac{D(\begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 3 & -2 & 0 \end{pmatrix})}{5} = 1.$$

Osserviamo che il sistema precedente è quello risolto alla fine della sottosezione *Operazioni elementari* (sezione 3): si confrontino i due metodi di risoluzione per sistemi quadrati al crescere del numero delle incognite.

4.4. Determinante e rango

Concludiamo questa parte esponendo un criterio di calcolo del rango di una matrice per mezzo di determinanti di cui omettiamo la prova.

Teorema 4.6. *Se $A \in M_{m,n}$, allora*

a. *$r(A)$ è uguale al massimo valore intero k per cui esiste una sottomatrice quadrata B di A di ordine k tale che $D(B) \neq 0$.*

b. *Sia B sottomatrice quadrata di A di ordine k con $D(B) \neq 0$. Se per ogni sottomatrice quadrata B' quadrata di ordine $(k+1)$ ottenuta da B aggiungendo una riga e una colonna abbiamo $D(B') = 0$, allora $r(A) = k$.*

Usualmente, il determinante di una sottomatrice quadrata di A di ordine k si dice *minore* di ordine k di A. Nel caso in cui $A \in M_n$, si riottiene come conseguenza del Teorema 4.6 che $r(A) = n$ se e solo se $D(A) \neq 0$ (Teorema 3.10).

Esempio

Se

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

e se $B = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$, $D(B) = 1$ mentre i 2 minori di ordine 3 ottenuti aggiungendo una riga e una colonna a B sono entrambi nulli: quindi $r(A) = 2$ per il Teorema 4.6b.

5. STRUTTURA LINEARE DI K^n

5.1. VETTORI

Nella sezione dedicata alle matrici abbiamo visto che K^n è identificabile con $M_{1,n}$ direttamente e con $M_{n,1}$ tramite trasposizione. Poiché le operazioni di somma e prodotto per scalare sono compatibili con la trasposizione (Proposizione 2.1), possiamo parlare di «somma» e «prodotto per scalare» per vettori di K^n senza specificare se stiamo considerando vettori riga o colonna.
Le operazioni di somma e prodotto per scalare determinano quella che chiamiamo *struttura lineare* (o *struttura vettoriale*) di K^n. In generale, si parla di K-*struttura vettoriale* su un insieme V quando sono assegnate su V due operazioni «somma» e «prodotto per scalare» che soddisfano alle proprietà formali S1, S2, S3, S4 e P1, P2, P3, P4 della sezione II. In questo caso V viene detto K-*spazio vettoriale*: per esempio l'insieme $M_{m,n}$ con le citate operazioni è uno spazio vettoriale. Qui ci limitiamo a studiare il caso particolare dato da K^n.

5.2. SOTTOSPAZI LINEARI

La geometria analitica in due o tre dimensioni consiste essenzialmente nell'identificare il piano o lo spazio con R^2 o R^3 associando a ogni punto una coppia o terna di numeri e utilizzare metodi di algebra lineare per studiare i vari problemi geometrici. In questo contesto, a enti geometrici fondamentali come rette e piani corrispondono insiemi di soluzioni di sistemi lineari, beninteso indeterminati (per completezza, si possono considerare i singoli punti come soluzioni di sistemi determinati).
Per esempio, l'equazione lineare $2x - y = 1$ descrive una retta l nel piano, mentre $x + y + z = 0$ descrive un piano α nello spazio: di fatto identifichiamo l con

$\{(x,y) \in R^2 |\ 2x+y=1\}$ e α con $\{(x,y,z) \in R^3 |\ x+y+z=0\}$, che sono insiemi soluzioni di sistemi lineari. Se ora β è il piano di equazione $x-y+z=1$, l'intersezione tra α e β è una retta nello spazio, che è descrivibile come soluzioni del sistema

$$\begin{cases} x+y+z=0 \\ x-y+z=1 \end{cases}$$

Evidentemente i sistemi omogenei descrivono le rette e i piani per l'origine del sistema di coordinate. Mentre dal punto di vista geometrico non vi è particolare ragione di distinguere questi ultimi dalle altre rette e dagli altri piani, in algebra lineare gli insiemi di soluzioni di sistemi omogenei hanno un ruolo di primo piano. Tale importanza deriva dal fatto che, sommando due soluzioni di un sistema omogeneo o moltiplicandone una per uno scalare otteniamo ancora una soluzione dello stesso sistema: questo non avviene per sistemi non omogenei, come si verifica facilmente. Piú formalmente abbiamo

Proposizione 5.1. *Se $S: AX=O$ è un sistema omogeneo, allora*

1. *Se $X_1,\ X_2 \in sol(S)$, allora $X_1+X_2 \in sol(S)$.*
2. *Se $a \in K$ e $X_0 \in sol(S)$, allora $aX_0 \in sol(S)$.*

PROVA

Se $AX_1 = AX_2 = O$, allora $A(X_1+X_2) = AX_1 + AX_2 = O$ e $A(aX_0) = aAX_0 = O$.

Diamo quindi la seguente

Definizione 5.2. *Un sottoinsieme L non vuoto di K^n si dice sottospazio lineare di K^n se esiste un sistema omogeneo S con n incognite tale che $L = sol(S)$.*

Per brevità, scriveremo «sl» per «sottospazio lineare».
Osserviamo che un sl L non è in generale determinato da un solo sistema S: infatti qualsiasi sistema equivalente a S ha come soluzioni L.
Sia K^n, visto come insieme delle soluzioni del sistema $O_nX = O_{n,1}$, sia $\{O\}$, visto come l'unica soluzione di un qualsiasi sistema quadrato omogeneo con matrice associata invertibile (teorema di Cramer), sono sl di K^n; essi sono detti *sottospazi banali* e $\{O\}$ si dice *sottospazio nullo*. Se $L \neq \{O\}$, L si dice *non nullo*.
Se L è un sl di K^n diverso da K^n e O, L si dice *sottospazio proprio*.

È immediato osservare che se L e L' sono sl di K^n, anche l'intersezione $L \cap L'$ è un sl: infatti se $L = sol(S)$ e $L' = sol(L')$, allora $L \cap L'$ è l'insieme delle soluzioni del sistema omogeneo formato unendo le equazioni di S e S'.

Definizione 5.3. *Se $X_1, X_2, \ldots, X_k$ sono vettori di K^n e $c_1, c_2, \ldots, c_k$ sono elementi di K, il vettore $c_1X_1+c_2X_2+\cdots+c_kX_k$ si dice la combinazione lineare di $X_1, \ldots, X_k$ con coefficienti $c_1, \ldots, c_k$.*

Per brevità, scriveremo «cl» per «combinazione lineare».
Se $X_1, \ldots, X_k$ sono vettori di K^n, indicheremo $M(X_1, \ldots, X_k)$ la matrice di $M_{n,k}$ che ha come colonne i vettori $X_1, \ldots, X_k$: quindi
$M(X_1, \ldots, X_k) = \begin{pmatrix} X_1 & X_2 & \ldots & X_k \end{pmatrix}$.
Ovviamente ${}^tM(X_1, \ldots, X_k) = \begin{pmatrix} {}^tX_1 \\ \vdots \\ {}^tX_k \end{pmatrix}$ è la matrice che ha come righe i vettori $X_1, \ldots, X_k$.
Con le notazioni introdotte, la cl di $X_1, \ldots, X_k$ con coefficienti $c_1, \ldots, c_k$ è uguale, scritta come vettore colonna, al prodotto di $M(X_1, \ldots, X_k)$ per il vettore dei coefficienti $C = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$. Questo fatto ci sarà utile nel seguito.

Esempio

Se consideriamo in R^2 i vettori $(1, -3),\ (-2, 2),\ (0, 1)$ e i numeri reali $2,\ \frac{1}{2},\ -1$, otteniamo la cl $2(1, -3) + \frac{1}{2}(-2, 2) - 1(0, 1) = (1, -6)$, che si può anche scrivere come

$$\begin{pmatrix} 1 & -2 & 0 \\ -3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ \frac{1}{2} \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -6 \end{pmatrix}.$$

Indichiamo con $L(X_1, \ldots, X_k)$ l'insieme dei vettori di K^n che sono cl di $X_1, \ldots, X_k$.

Quindi $X \in L(X_1, \ldots, X_k)$ se e solo se esistono $c_1, \ldots, c_k$ tali che $X = c_1X_1 + \cdots + c_kX_k$.

Prendendo $c_1 = \cdots = c_k = 0$, vediamo che $O \in L(X_1, \ldots, X_k)$ per ogni insieme $\{X_1, \ldots, X_k\}$.

Dalla Proposizione 5.1 otteniamo che se $L \subseteq K^n$ è un sl e se $X_1, \ldots, X_k$ sono vettori di L, allora ogni cl di $X_1, \ldots, X_k$ appartiene a L, cioè $L(X_1, \ldots, X_k) \subseteq L$.
Osserviamo che uno stesso vettore può essere cl di $X_1, \ldots, X_k$ con insiemi diversi di coefficienti. Inoltre insiemi diversi di vettori possono dare origine allo stesso insieme di cl, cioè si può avere $L(X_1, \ldots, X_k) = L(Y_1, \ldots, Y_h)$ per $\{X_1, \ldots, X_k\} \neq \{Y_1, \ldots, Y_h\}$. In tal caso gli Y_i saranno cl degli X_j e viceversa.

Esempi

1. Con gli stessi vettori di R^2 dell'esempio abbiamo anche $(1, -6) = 1(1, -3) + 0(-2, 2) - 3(0, 1)$.
2. Se $X_1 = (1, 1, 1)$, $X_2 = (1, 0, 1)$ e $Y_1 = (2, 1, 2)$, $Y_2 = (0, -1, 0)$, $Y_3 = (2, 0, 2)$ allora $L(X_1, X_2) = L(Y_1, Y_2, Y_3)$; infatti $Y_1 = X_1 + X_2$, $Y_2 = -X_1 + X_2$ e $Y_3 = 2X_2$.

Applicando la teoria dei sistemi, possiamo dare un criterio per stabilire se un vettore è cl di assegnati elementi di K^n.

Proposizione 5.4. *Siano $X_1, \ldots, X_k$ vettori di K^n. Un vettore $\overline{X} \in K^n$ appartiene a $L(X_1, \ldots, X_k)$ se e solo se $r(M(X_1, \ldots, X_k)) = r(M(X_1, \ldots, X_k, X))$.*

PROVA

$\overline{X}$ appartiene a $L(X_1, \ldots, X_k)$ se e solo se esistono $c_1, \ldots, c_k \in K$ tali che $\overline{X} = c_1X_1 + \cdots + c_kX_k$.
Se $M = M(X_1, \ldots, X_k)$, l'affermazione precedente equivale a dire che il sistema $\overline{S}$: $MX = \overline{X}$ con matrice completa associata $M_{\overline{S}} = \begin{pmatrix} M & | & \overline{X} \end{pmatrix}$ è risolubile. La tesi quindi deriva dal Teorema di Rouché-Capelli.

Evidentemente, risolvendo il sistema $MX = \overline{X}$ della dimostrazione precedente possiamo trovare tutti i coefficienti che danno $\overline{X}$ come cl di $X_1, \ldots, X_k$.

Esempi

I coefficienti $c_1,\ c_2,\ c_3$ tali che $c_1(1, -3) + c_2(-2, 2) + c_3(0, 1) = (1, -6)$ si trovano risolvendo il sistema

$$\begin{cases} c_1 - 2c_2 = 1 \\ -3c_1 + 2c_2 + c_3 = -6 \end{cases}$$

che ha equazioni risolventi
$$\begin{cases} c_1 = 2c_2 + 1 \\ c_3 = 4c_2 - 3 \end{cases}$$

La relazione con i sl è data dal seguente teorema:

Teorema 5.5.

a. Se L di K^n è un sl, allora esiste un insieme di vettori $\{X_1, \dots, X_k\}$ tale che $L = L(X_1, \dots, X_k)$.

b. Se $L = L(X_1, \dots, X_k)$, L è un sl di K^n. In particolare, se $\rho = r(M(X_1, \dots, X_k))$, esiste un sistema S omogeneo con $n - \rho$ equazioni, n incognite e ∞^ρ soluzioni tale che $L = sol(S)$.

PROVA

a. Supponiamo che L sia un sl di K^n, e determiniamo un insieme $X_1, \dots, X_k$ tale che $L = L(X_1, \dots, X_k)$.
Per definizione, $L = sol(S)$, dove $S : AX = O$. Sia $A' = A_{srn}$ e supponiamo per semplicità che in A' si abbia $j(i) = i$ per $i = 1, \dots, r = r(A)$ (le colonne dei pivots sono le prime r). Allora, posto $a_{i,j} = [A']_{i,j}$, le equazioni risolutive di S sono:
$$\begin{cases} x_1 = -\sum\limits_{j=r+1}^{n} a_{1,j}x_j \\ x_2 = -\sum\limits_{j=r+1}^{n} a_{2,j}x_j \\ \vdots \\ x_r = -\sum\limits_{j=r+1}^{n} a_{r,j}x_j \end{cases} .$$
Dunque un vettore soluzione di S ha i primi r elementi determinati un volta fissati gli $n - r$ valori $x_{r+1}, \dots, x_n$. Effettuando al variare di j tra $r + 1$ e n le $n - r$ sostituzioni $x_j = 1$ e $x_k = 0$ per $k \neq i$, otteniamo i vettori soluzione

$$X_{j-r} = \begin{pmatrix} -a_{1,j} \\ \vdots \\ -a_{r,j} \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},$$

dove 1 si trova in posizione j.

A questo punto è immediato verificare che, dato $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$, $X \in sol(S)$ se e solo se $X = x_{r+1}X_1 + \cdots + x_n X_{n-r}$.

Quindi $L = sol(S) = L(X_1, \ldots, X_{n-r})$.

b. Per la Proposizione 5.4 , $\overline{X} \in L = L(X_1, \ldots, X_k)$ se e solo se il sistema $\overline{S} : MX = \overline{X}$ è risolubile, dove $M = M(X_1, \ldots, X_k)$.

Per il Corollario 3.11, sappiamo che esiste $P \in GL_n$, ottenuta come prodotto di matrici elementari, tale che $PM = M_{srn}$.

Se rappresentiamo P a blocchi come $P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}$ con P_1 formata dalle prime ρ righe e P_2 dalle ultime $n - \rho$, passando al sistema equivalente abbiamo

$$PM_{\overline{S}} = \begin{pmatrix} PM & | & P\overline{X} \end{pmatrix} = \begin{pmatrix} P_1 M & | & P_1\overline{X} \\ - & - & - \\ O_{n-\rho,k} & | & P_2\overline{X} \end{pmatrix}.$$

Dunque $\overline{S}$ è risolubile se e solo se $P_2\overline{X} = O$, quindi $L = sol(P_2 X = O)$, per il Lemma 3.8 $r(P_2) = n - \rho$ e la tesi è provata.

Dal teorema precedente ricaviamo che i sottospazi lineari possono essere presentati in due forme differenti, come soluzioni di sistemi omogenei (per definizione), e allora parliamo di *forma implicita*, oppure come insieme di cl, e in questo caso parliamo di *forma esplicita* o *parametrica*. Entrambe queste forme non sono uniche, come abbiamo gà osservato.

Sia $L \subset K^n$ un sl. Se $L = sol(S)$, le equazioni di S vengono dette *equazioni* di L; se $L = L(X_1, \ldots, X_k)$, $X_1, \ldots, X_k$ vengono detti *generatori* di L e L viene detto *sottospazio lineare generato da* $X_1, \ldots, X_k$.

Proposizione 5.6. *Se L un sl di K^n generato da $X_1, \ldots, X_k$, ogni sottoinsieme finito di L contenente $\{X_1, \ldots, X_k\}$ è un insieme di generatori di L.*

PROVA

Consideriamo un insieme del tipo $\{X_1, \ldots, X_k, X_{k+1}, \ldots, X_h\}$. Allora

$X = c_1X_1 + \cdots + c_kX_k$ implica $X = c_1X_1 + \cdots + c_kX_k + 0X_{k+1} + \cdots + 0X_h$.

Per quel che riguarda i sottospazi banali, abbiamo che $\{O\} = L(O)$ e che $K^n = L(e_1, \ldots, e_n)$. Infatti, se $X \in K^n$,

$$\begin{pmatrix} [X]_1 \\ \vdots \\ [X]_n \end{pmatrix} = [X]_1 e_1 + \cdots + [X]_n e_n.$$

Un fatto importante è che la dimostrazione del Teorema 5.5 fornisce un metodo esplicito per passare da una forma implicita a una esplicita e viceversa di un sl.

Esempi

1. Siano $X_1 = (1,1,1,1)$, $X_2 = (1,0,1,0)$, $X_3 = (0,-1,0,-1)$ vettori di R^4. Per determinare una forma implicita di $L = L(X_1, \ldots, X_4)$, portiamo il primo blocco della matrice

$$M = \begin{pmatrix} 1 & 1 & 0 & | & x_1 \\ 1 & 0 & -1 & | & x_2 \\ 1 & 1 & 0 & | & x_3 \\ 1 & 0 & -1 & | & x_4 \end{pmatrix}$$

in forma a scala. Allora

$$M \sim \begin{pmatrix} 1 & 1 & 0 & | & x_1 \\ 0 & 1 & 1 & | & x_1 - x_2 \\ 0 & 0 & 0 & | & x_3 - x_1 \\ 0 & 0 & 0 & | & x_4 - x_2 \end{pmatrix}$$

e $L = sol(S)$ con

$$S : \begin{cases} -x_1 + x_3 = 0 \\ -x_2 + x_4 = 0 \end{cases}$$

Osserviamo che, partendo dal sistema S trovato, si ottengono le equazioni risolventi $x_1 = x_3$ e $x_2 = x_4$, da cui

$L = L((1,0,1,0),(0,1,0,1)) = L(X_2, -X_3)$.

2. Sia $L = sol(S) \subseteq R^4$ con

$$S : \begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ x_1 + x_2 - 2x_3 - x_4 = 0 \end{cases} .$$

Le equazioni risolventi sono $x_1 = -x_2 - \frac{1}{3}x_4$ e $x_3 = -\frac{2}{3}x_4$ dunque $L = L((-1,1,0,0),(-\frac{1}{3},0,-\frac{2}{3},1))$.

5.3. Insiemi liberi e basi

Definizione 5.7. *Un insieme finito $\{X_1,\dots,X_k\}$ di vettori di K^n si dice insieme libero se l'unica cl di $X_1,\dots,X_k$ nulla è quella data dai coefficienti $c_1 = c_2 = \cdots = c_k = 0$.*
In tal caso i vettori $X_1,\dots,X_k$ si dicono linearmente indipendenti.
Inversamente, $\{X_1,\dots,X_k\}$ non è libero se esistono coefficienti $c_1,\dots,c_k$ non tutti nulli tali che $c_1X_1 + \cdots + c_kX_k = 0$.
In tal caso i vettori $X_1,\dots,X_k$ si dicono linearmente dipendenti.

Proposizione 5.8. *Se $X_1,\dots,X_k$ è un insieme libero di K^n, allora ogni suo sottoinsieme $\neq \emptyset$ è libero.*

PROVA

Consideriamo per semplicità un sottoinsieme del tipo $\{X_1,\dots,X_h\}$, con $0 < h < k$. Se fosse $c_1X_1+\cdots+c_hX_h = O$ con i c_i non tutti nulli, avremmo $c_1X_1+\cdots+c_hX_h+0X_{h+1}+\cdots+0X_k = O$, assurdo poiché $\{X_1,\dots,X_k\}$ è libero.

Osserviamo che O non può mai appartenere a un insieme libero: se $X_1 = O$ e $X_2,\dots,X_k$ sono qualsiasi, allora $1X_1 + 0X_2 + \cdots + 0X_n = O$ con coefficienti non tutti nulli.
D'altra parte, se $X \neq O$, allora $\{X\}$ è libero. Nel caso di 2 vettori $X,\ Y \in K^n$, essi sono linearmente dipendenti se e solo se esistono a e b in K non entrambi nulli e tali che $aX + bY = O$: se per esempio $a \neq 0$, abbiamo $Y = -\frac{b}{a}X$. Dunque due vettori sono linearmente dipendenti (indipendenti) se e solo se sono (non sono) uno multiplo dell'altro.
Possiamo generalizzare questo fatto con la seguente

Proposizione 5.9. *Un insieme $\{X_1, \ldots, X_k\}$ di vettori in K^n non è libero se e solo se per almeno un indice $i = 1, \ldots, k$ il vettore X_i è cl dei restanti vettori X_j, $j \neq i$.*

PROVA

Se $\{X_1, \ldots, X_k\}$ non è libero, esistono $c_1, \ldots, c_k$ non tutti nulli tali che $c_1 X_1 + \cdots + c_k X_k = O$. Se per comodità assumiamo $c_1 \neq 0$, possiamo scrivere

$$X_1 = -\frac{c_2}{c_1} X_2 - \cdots - \frac{c_k}{c_1} X_k,$$

quindi $X_1 \in L(X_2, \ldots, X_k)$.
Viceversa, se per esempio $X_1 = c_2 X_2 + \cdots + c_k X_k$, allora $X_1 - c_2 X_2 - \cdots - c_k X_k = O$; dunque abbiamo una cl di $X_1, \ldots, X_k$ con almeno il primo coefficiente $\neq 0$, quindi l'insieme non è libero.

Vediamo in generale come si può verificare l'indipendenza lineare.

Proposizione 5.10. *Un insieme $\{X_1, \ldots, X_k\}$ di k vettori in K^n è libero se e solo se $r(M(X_1, \ldots, X_k)) = k$. In particolare $k \leq n$.*

PROVA

Sia $M = M(X_1, \ldots, X_k)$. L'insieme in questione è libero se e solo il sistema omogeneo $MX = O$ ammette solo la soluzione nulla, cioè è determinato. Questo significa che M_{srn} ha k righe non nulle, quindi $k = r(M)$.

Esempio

L'insieme $X_1 = (1, 1, 1, 1)$, $X_2 = (1, 0, 1, 0)$, $X_3 = (0, -1, 0, -1)$ di vettori di R^4 non è libero: si verifichi infatti che $r(M(X_1, X_2, X_3)) = 2$.
Sostituendo a X_3 il vettore $Y = (0, -1, 0, 2)$ si ottiene un insieme libero.

Definizione 5.11. *Sia L un sl di K^n, L non nullo. Una base di L è un insieme libero e ordinato di generatori di L.*

La condizione «ordinato» significa che lo stesso insieme libero di generatori individua basi differenti, ottenute permutando gli elementi dell'insieme.

Esempio

L'insieme ordinato $\{e_1, \ldots, e_n\}$ è una base di K^n. Infatti tale insieme è libero per la Proposizione 5.10 ($M(e_1, \ldots, e_n) = I_n$) e genera K^n. Tale base si dice *base canonica*

di K^n.
D'altra parte $\{e_2, e_1, \ldots, e_n\}$ è una base di K^n differente da quella canonica.

Proposizione 5.12. *Sia L un sl non nullo di K^n. Allora $\{X_1, \ldots, X_k\} \subset L$ è una base di L se e solo se per ogni $\overline{X} \in L$ esistono unici $c_1, \ldots, c_k$ tali che $\overline{X} = c_1X_1 + \cdots + c_kX_k$.*

PROVA

Se $M = M(X_1, \ldots, X_k)$, $\overline{X} \in L$ se e solo se il sistema con n equazioni e k incognite $S : MX = \overline{X}$ è risolubile.
D'altra parte, $X_1, \ldots, X_k$ sono linearmente indipendenti se e solo se $r(M) = k$. Dunque $X_1, \ldots, X_k$ sono una base se e solo se S è determinato; in tal caso l'unica soluzione è il vettore $(c_1, \ldots, c_k) \in K^k$ dei coefficienti cercati.

I coefficienti $c_1, \ldots, c_k$ determinati nella Proposizione 5.12 si dicono *coordinate* di $\overline{X}$ rispetto alla base $\{X_1, \ldots, X_k\}$ di L e il vettore $(c_1, \ldots, c_k) \in K^k$ si dice *vettore delle coordinate* (spesso, per comodità, per «coordinate» intenderemo «vettore delle coordinate»).
La base canonica di K^n si caratterizza come l'unica base tale che, per ogni $X \in K^n$, le coordinate rispetto a tale base sono X stesso: infatti, $X = [X]_1e_1 + \cdots + [X]_ne_n$. Questo fatto giustifica l'aggettivo «canonica».

Esempi

1. Le coordinate di $(1, 2, -1)$ rispetto alla base canonica di R^3 sono $(1, 2, -1)$ mentre rispetto alla base $\{e_2,\ e_3,\ e_1\}$ sono $(2, -1, 1)$.
2. Sia $L \subseteq R^4$ il sl definito dal sistema
$$S : \begin{cases} x_1 + x_2 + 2x_3 = 0 \\ x_1 - x_2 + 2x_4 = 0 \end{cases} .$$
S ha come equazioni risolventi $x_1 = -x_3 + x_4$ e $x_2 = -x_3 - x_4$, quindi $L = L(X_1, X_2)$ con $X_1 = (-1, -1, 1, 0)$ e $X_2 = (-1, 1, 0, 1)$ (Teorema 5.5). X_1 e X_2 sono linearmente indipendenti (perché?), dunque sono una base di L. Dato ora $X = (2, 0, -1, -1)$, si verifica che $X \in L$ e che le sue coordinate rispetto alla base trovata sono $(-1, -1)$.
3. Il vettore nullo O ha coordinate nulle rispetto a qualsiasi base.

4. Se $\{X_1, \ldots, X_k\}$ è una base di un sl L, un elemento della base X_i ha coordinate ${}^t e_i = (0, \ldots, 1, \ldots 0)$ rispetto alla base stessa, dove 1 sta in posizione i e e_i è considerato in K^k.

Supporremo d'ora in poi che L sia non nullo. Proviamo il fondamentale

Teorema 5.13. *Se L è un sl di K^n, allora tutte le sue basi hanno lo stesso numero di elementi.*

PROVA

Siano $X_1, \ldots, X_k$ e $Y_1, \ldots, Y_h$ basi di L. Poiché le basi sono insiemi di generatori, esse danno luogo a due forme esplicite di L, cioè $L = L(X_1, \ldots, X_k) = L(Y_1, \ldots, Y_h)$. In base al Teorema 5.5 , possiamo passare alle forme implicite corrispondenti, dunque $L = sol(S) = sol(S')$.
Ora, i sistemi omogenei S e S' hanno ∞^ρ e $\infty^{\rho'}$ soluzioni, dove $\rho = r(M(X_1, \ldots, X_k)) = k$ e $\rho' = r(M(Y_1, \ldots, Y_h)) = h$ in quanto gli insiemi in questione sono liberi.
Infine, poiché S e S' sono equivalenti, dev'essere necessariamente $\rho = \rho'$, quindi $k = h$.

Il teorema precedente ci permette di dare la seguente definizione.

Definizione 5.14. *Se L è un sl di K^n, il numero di elementi di una sua base viene chiamato dimensione di L e indicato con $dimL$. Per convenzione, $dim\{O\} = 0$.*

La dimensione di K^n è n, poiché abbiamo la base canonica. In generale, utilizzando la dimostrazione del Teorema 5.5, possiamo calcolare direttamente la dimensione di un sl determinandone una base.

Proposizione 5.15. *Sia L un sl di K^n.*

a. Se $L = sol(S)$ e $S : AX = O$, allora $dimL = n - r(A)$.

b. Se $L = L(X_1, \ldots, X_k)$, allora $dimL = r(M(X_1, \ldots, X_k))$.

c. Nelle ipotesi di (a) e (b), vale $n - r(A) = r(M(X_1, \ldots, X_k))$.

PROVA

(c) è ovvia conseguenza di (a) e (b).

a. Nella dimostrazione del Teorema 5.5 a, abbiamo ricavato un insieme $X_1, \ldots, X_{n-r}$ di generatori per L, della forma

$$X_{j-r} = \begin{pmatrix} -a_{1,j} \\ \vdots \\ -a_{r,j} \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},$$

dove $r = r(A)$ e $j = r+1, \ldots, n$.
Ora la matrice $M(X_1, \ldots, X_{n-r}) \in M_{n,n-r}$ ha evidentemente rango $n - r$, dunque tale insieme di generatori è libero e forma una base.

b. Per il Teorema 5.5b, se $\rho = r(M(X_1, \ldots, X_k))$, $L = sol(S)$, dove S è omogeneo con ∞^ρ soluzioni. Dunque $r(M_S) = n - \rho$ e per la parte (a) di questa Proposizione, abbiamo $dimL = \rho$.

Osserviamo che la parte (a) della precedente dimostrazione ci garantisce che l'insieme di generatori trovato nel Teorema 5.5 è una base.

Teorema 5.16. *Sia L un sl di K^n con $dimL = d$ e sia $\mathcal{L} = \{X_1, \ldots, X_k\} \subseteq L$ un insieme finito di vettori di L.*

a. Se $\mathcal{L}$ è libero, allora $\mathcal{L}$ è contenuto in una base di L.
In particolare $k \leq d$ e, se $k = d$, allora $\mathcal{L}$ è una base.

b. Se $\mathcal{L}$ genera L, allora $\mathcal{L}$ contiene una base di L.
In particolare $k \geq d$ e, se $k = d$, allora $\mathcal{L}$ è una base.

PROVA

a. Se $X_1, \ldots, X_k$ non generano L, esiste $X_{k+1} \in L$ tale che $X_{k+1} \notin L(X_1, \ldots, X_k)$. Dunque, per la Proposizione 5.9 , $X_1, \ldots, X_{k+1}$ sono linearmente indipendenti e $L(X_1, \ldots, X_{k+1}) \subseteq L$.
Reiterando questo ragionamento, poiché il numero di elementi di un insieme libero è $\leq n$, troviamo $p > 0$ tale che $L = L(X_1, \ldots, X_{k+p})$ e $\{X_1, \ldots, X_{k+p}\}$ libero, quindi $d = k + p$ per il Teorema 5.13.

b. Se $\{X_1, \ldots, X_k\}$ non è libero, per la Proposizione 5.9 uno dei suoi elementi è cl degli altri: supponiamo per comodità che $X_k \in L(X_1, \ldots, X_{k-1})$, quindi $L = L(X_1, \ldots, X_{k-1})$.

Reiterando questo ragionamento, poiché $L \neq \{O\}$, troviamo $0 < p < k$ tale che $X_1, \ldots, X_{k-p}$ sono indipendenti e $L = L(X_1, \ldots, X_{k-p})$, quindi $d = k-p$ per il Teorema 5.13.

Corollario 5.17. *Se $L' \subseteq L$ sono sl di K^n e se $dimL = d$, allora $dimL' \leq d$ e $dimL' = d$ se e solo se $L = L'$.*
Dunque l'unico sl di K^n di dimensione n è K^n stesso e ogni altro sl di K^n ha dimensione $< n$.

PROVA

Se $\{X_1, \ldots, X_k\}$ è una base di L', $dimL' = k \leq d$ per il Teorema 5.16a. Nel caso in in cui $k = d$, abbiamo un insieme libero con d elementi contenuto in L, quindi una base di L. Dunque $L = L(X_1, \ldots, X_d) = L'$.

Esempi

1. In K non vi sono sl propri: infatti un sl $L \subseteq K$ puó avere dimensione 0 o 1: nel primo caso L é il sl nullo mentre nel secondo $L = K$.
2. I sl di K^n di dimensione $n-1$ sono tutti e soli gli insiemi di soluzioni di equazioni lineari omgenee $a_1x_1 + \cdots + a_nx_n = 0$ con gli a_i non tutti nulli.
3. In K^2 un sl L proprio ha necessariamente dimensione 1. Rappresentando L in forma esplicita abbiamo $L = L((x_0, y_0))$ con $(x_0, y_0) \neq O$: quindi
   $L = \{(tx_0, ty_0) \mid \ t \in K\}$.
   Per $K = R$ queste sono le equazioni parametriche di una retta r per l'origine in un fissato sistema di coordinate cartesiane.
   Per l'esempio (2), L ha equazione $ax + by = 0$ con $(a, b) \neq (0, 0)$. Quindi se $K = R$ una forma implicita di L è data dall'equazione cartesiana della medesima retta r.
4. Un sl non banale di K^3 ha dimensione 1 o 2.
   Se $dimL = 2$, in base all'esempio (2) L ha equazione $ax + by + cz = 0$: questa è l'equazione cartesiana di un piano per l'origine.
   Se $dimL = 1$, analogamente a (3), $L = L((x_0, y_0, z_0))$ con $(x_0, y_0, z_0)) \neq O$: quindi per $K = R$, otteniamo le equazioni parametriche di una retta per l'origine r nello spazio. D'altra parte, considerando la forma implicita di L, possiamo trovare equazioni

$$\begin{cases} ax + by + cz = 0 \\ a'x + b'y + c'z = 0 \end{cases}$$

dove $r(\begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix}) = 2$.
Dunque, se $K = R$, r viene espressa come intersezione di due piani per l'origine (equazioni cartesiane di r).
Per esempio, se $L = L((1,2,3))$, abbiamo le equazioni parametriche

$$\begin{cases} x = t \\ y = 2t \\ z = 3t \end{cases} .$$

Le equazioni cartesiane si ottengono sostituendo t:

$$\begin{cases} 2x - y = 0 \\ 3x - z = 0 \end{cases} .$$

Il Teorema 5.16 a ci dice che a un insieme libero contenuto in un sl L si possono aggiungere vettori fino ad ottenere un base di L: tale procedimento viene detto *estensione* dell'insieme a una base. D'altra parte, il Teorema 5.16 b ci dice che possiamo scegliere da un insieme di generatori di L una base: tale procedimento viene detto *estrazione* di una base. Nella prossima sottosezione daremo un esempio di tali procedimenti.
Quanto dimostrato prima porta alla seguente caratterizzazione del concetto di dimensione:
Se $L \subseteq K^n$, la dimensione di L coincide con il massimo numero di elementi di un insieme libero contenuto in L e con il minimo numero di elementi di un insieme di generatori di L.
Nella teoria generale degli spazi vettoriali si considerano i «sottospazi vettoriali» di uno K-spazio vettoriale V: $U \subseteq V$ non vuoto è un sottospazio vettoriale se, comunque dati $u, u' \in U$ e $a \in K$, si ha $u + u' \in U$ e $au \in U$.
Nel caso di K^n vediamo che i sottospazi vettoriali corrispondono esattamente ai sottospazi lineari.

Teorema 5.18. *Sia L un sottoinsieme non vuoto di K^n. Le seguenti affermazioni sono equivalenti:*

a. *L è un sl di K^n.*

b. *Comunque dati X, $Y \in L$ e $a \in K$ si ha $X + Y \in L$ e $aX \in L$.*

c. *Ogni cl formata con elementi di L è ancora un elemento di L.*

PROVA

È evidente che (b)$\Leftrightarrow$ (c) e che (a)$\Rightarrow$ (b) per la Proposizione 5.1. Proviamo che (b)$\Rightarrow$ (a). Supponiamo che $L \subseteq K^n$ sia non vuoto e valga (b). Dato $X \in L$, $0X = O \in L$. Se $L = \{O\}$, L è un sl. Altrimenti, sia $X_1 \in L$ diverso da O. Allora il sl $L(X_1)$ dei multipli di X_1 è contenuto in L.
Se $L(X_1) \neq L$, sia $X_2 \in L$ e $X_2 \notin L(X_1)$: allora $X_1,\ X_2$ sono linearmente indipendenti e $L(X_1, X_2) \subseteq L$ per (c). Reiterando questo procedimento, troviamo un insieme libero $\{X_1, \ldots, X_k\} \subset L$ tale che $L = L(X_1, \ldots, X_k)$. Quindi L è un sl di K^n.

5.4. Indipendenza lineare e rango

Data una matrice $A \in M_{m,n}$, possiamo considerare le sue righe $[A]_1, \ldots, [A]_m$ come elementi di K^n e le sue colonne $[A]^1, \ldots, [A]^n$ come elementi di K^m.
Allora indichiamo il sl di K^n generato dalle righe con $\mathcal{R}(A) = L(\,{}^t[A]_1, \ldots,\ {}^t[A]_m)$ e il sl di K^m generato dalle colonne con $\mathcal{C}(A) = L([A]^1, \ldots, [A]^n)$. $\mathcal{R}(A)$ e $\mathcal{C}(A)$ si dicono rispettivamente *spazio delle righe* e *spazio delle colonne* di A.
Va precisato che, mentre negli insiemi di generatori considerati in precedenza i vettori erano ovviamenti pensati distinti, puó ora accadere che più righe o colonne siano uguali: questo evidentemente non influenza i risultati.
Vogliamo studiare la relazione tra il $r(A)$, $dim(\mathcal{R}(A))$ e $dim(\mathcal{C}(A))$. Intanto $dim(\mathcal{C}(A)) = r(A)$ per la Proposizione 5.15b. Proviamo il seguente

Lemma 5.19. *Se $A,\ B \in M_{m,n}$ e $A \sim B$, allora $\mathcal{R}(A) = \mathcal{R}(B)$.*

PROVA

Basta verificare la tesi nel caso in cui B è ottenuta da A con una sola operazione elementare. Nel caso di operazioni del secondo tipo, le righe rimangono le stesse, negli altri due casi viene cambiata una sola riga, sia essa la i-esima.
Allora abbiamo $[B]_i = [A]_i + a[A]_j$ per $a \in K$ e $j \neq i$ oppure $[B]_i = a[A]_i$ per $a \neq 0$. In entrambi i casi $[B]_i \in \mathcal{R}(A)$, quindi $\mathcal{R}(B) \subseteq \mathcal{R}(A)$; poiché $\sim$ è simmetrica vale anche l'inclusione opposta e quindi la tesi.

Lemma 5.20. *Se $A \in M_{m,n}$ è $\neq O$, $r = r(A)$ e $A' = A_{srn}$, allora*

a. $[A']_1, \ldots, [A']_r$ è una base di $\mathcal{R}(A)$, quindi $dim\mathcal{R}(A) = r$.

b. Se $j(1), \ldots, j(r)$ sono gli indici di colonna dei pivots di A', allora $[A]^{j(1)}, \ldots, [A]^{j(r)}$ è una base di $\mathcal{C}(A)$.

PROVA

a. Poiché $[A']_i = O$ per $i > r$, $\mathcal{R}(A') = L({}^t[A']_1 \ldots, {}^t[A']_r)$. Inoltre, se M è la matrice $n \times r$ che ha per colonne le prime r righe di A', allora $r(M) = r$. Quindi $[A']_1, \ldots, [A']_r$ sono linearmente indipendenti e danno una base di $\mathcal{R}(A') = \mathcal{R}(A)$ (Lemma 5.19).

b. Sia $\overline{A} = M([A]^{j(1)}, \ldots, [A]^{j(r)}) \in M_{m,r}$. Poiché le operazioni elementari lasciano invariati gli indici di colonna, $\overline{A}_{srn} = M(e_1, \ldots, e_r)$ e $r(\overline{A}) = r$; per la Proposizione 5.10, $\{[A]^{j(1)}, \ldots, [A]^{j(r)}\}$ è libero e dunque è una base per il Teorema 5.16a.

Teorema 5.21. *Se $A \in M_{m,n}$, $r(A) = dim\mathcal{R}(A) = dim\mathcal{C}(A)$.*

Corollario 5.22. *Se $A \in M_{m,n}$, $r(A) = r({}^tA)$.*

Il rango di un matrice può dunque essere caratterizzato come *il massimo numero di righe o colonne linearmente indipendenti.*

Da queste considerazioni sul rango otteniamo un metodo effettivo per estrarre una base da un insieme di generatori o estendere a una base un insieme libero, metodo che illustriamo con un esempio.

Esempio

Siano $X_1 = (1, -1, 0, 1)$, $X_2 = (0, -3, -1, 2)$, $X_3 = (2, 1, 1, 0)$, $X_4 = (3, 0, 1, 1)$ e sia $L = L(X_1, \ldots, X_4)$. Vogliamo estrarre da $\{X_1, \ldots, X_4\}$ una base e estendere tale base a una base di R^4.

Se consideriamo $\{X_1, \ldots, X_4,\ e_1, \ldots, e_4\}$, questo insieme contiene la base canonica, dunque è un insieme di generatori di R^4 (Proposizione 5.6). Se estraiamo da tale insieme una base di R^4 contenente una base di L abbiamo risolto il nostro problema. Poiché $M(e_1, \ldots, e_4) = I$, abbiamo $M = M(X_1, \ldots, X_4,\ e_1, \ldots, e_4) = \Big(M(X_1, \ldots, X_4) \quad | \quad I\Big)$. Dunque

$$M = \begin{pmatrix} 1 & 0 & 2 & 3 & |\,1 & 0 & 0 & 0 \\ -1 & -3 & 1 & 0 & |\,0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 1 & |\,0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 1 & |\,0 & 0 & 0 & 1 \end{pmatrix}.$$

M è equivalente alla matrice a scala

$$M' = \begin{pmatrix} 1 & 0 & 2 & 3 & | & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & | & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & | & -1 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & | & 0 & -1 & 5 & 1 \end{pmatrix}.$$

Le colonne di M' contenenti i pivots sono la prima, la seconda, la quinta e la sesta: dunque, applicando il Lemma 5.20, $X_1 = [M]^1,\ X_2 = [M]^2,\ e_1 = [M]^5,\ e_2 = [M]^6$ formano una base di R^4 e $X_1,\ X_2$ una base di L.

Evidentemente, l'esempio appena esposto può essere generalizzato a un insieme qualsiasi di vettori $\{X_1, \dots, X_k\}$ in K^n, ottenendo così una procedura per estrarre una base da un sl L di K^n e estendere tale base a una base di K^n.
Riguardo a tale procedura possiamo fare le seguenti osservazioni:

1. È possibile utilizzare al posto della base canonica una qualsiasi base di K^n.
2. Se $\{X_1, \dots, X_k\}$ è un insieme libero, la procedura fornisce un'estensione a una base di K^n. Se d'altra parte vogliamo solo estrarre una base di $L(X_1, \dots, X_k)$ da $\{X_1, \dots, X_k\}$, basta utilizzare nella procedura la matrice $M(X_1, \dots, X_k)$ al posto di M.
3. Se $L \subseteq K^n$ è un sl e $\{X_1, \dots, X_k\} \subset L$ è un insieme libero, possiamo estendere tale insieme a una base di L con la stessa procedura pur di utilizzare una base di L al posto di una base di K^n.

Terminiamo questa sottosezione con un'ulteriore caratterizzazione delle matrici invertibili.

Proposizione 5.23. *Se $A \in M_n$, $A \in GL_n$ se e solo se le righe e/o le colonne formano una base di K^n.*

PROVA

$A \in GL_n \Leftrightarrow r(A) = n \Leftrightarrow dim\mathcal{R}(A) = dim\mathcal{C}(A) = n \Leftrightarrow \mathcal{R}(A) = \mathcal{C}(A) = K^n$.
L'ultima uguaglianza dice che le n righe (colonne) generano K^n, quindi sono una base di K^n per il Teorema 5.16.

5.5. Coordinate in K^n e cambiamenti di base

Consideriamo ora il caso delle basi di K^n. Indichiamo con $\mathcal{B}$ una base $\{X_1, \dots, X_n\}$ di K^n. Ricordiamo che, dato $X \in K^n$, per la Proposizione 5.12, esistono unici $c_1, \dots, c_n$ tali che $X = c_1X_1 + \cdots + c_nX_n$: tali numeri sono detti coordinate del vettore X rispetto alla base $\mathcal{B}$. In questo caso, il vettore delle coordinate di X rispetto $\mathcal{B}$ è ancora un vettore di K^n che indichiamo come colonna con

$$\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = [X]_{\mathcal{B}}.$$

Denotiamo con $\mathcal{C}_n$ (o più brevemente $\mathcal{C}$) la base canonica $\{e_1, \dots, e_n\}$ di K^n e ricordiamo che $X = [X]_{\mathcal{C}}$ per ogni $X \in K^n$.
D'altra parte, se $\mathcal{B} = \{X_1, \dots, X_n\}$ è una base di K^n, per ogni $i = 1, \dots, n$ abbiamo $X_i = 0X_1 + \cdots + 1X_i + \cdots + 0X_n$, dunque $[X_i]_{\mathcal{B}} = e_i$.
Consideriamo ora due basi di K^n, $\mathcal{B} = \{X_1, \dots, X_n\}$ e $\mathcal{B}' = \{Y_1, \dots, Y_n\}$ e consideriamo per ogni elemento X_i di $\mathcal{B}$ il suo vettore delle coordinate $[X_i]_{\mathcal{B}'}$ rispetto a $\mathcal{B}'$.
Se indichiamo con $M^{\mathcal{B}}_{\mathcal{B}'} = M([X_1]_{\mathcal{B}'}, \dots, [X_n]_{\mathcal{B}'})$ la matrice ottenuta mettendo in colonna le coordinate rispetto a $\mathcal{B}'$ dei vettori di $\mathcal{B}$, vale la

Formula di trasformazione delle coordinate. Se $X \in K^n$, il vettore delle coordinate di X rispetto a $\mathcal{B}'$ è uguale al prodotto di $M^{\mathcal{B}}_{\mathcal{B}'}$ con il vettore delle coordinate di X rispetto a $\mathcal{B}$, cioè

$$[X]_{\mathcal{B}'} = M^{\mathcal{B}}_{\mathcal{B}'}[X]_{\mathcal{B}}.$$

La matrice $M^{\mathcal{B}}_{\mathcal{B}'}$ si dice *matrice di trasformazione* da $\mathcal{B}$ a $\mathcal{B}'$.
Non proviamo tale formula ma la illustriamo con un esempio.

Esempio

$\mathcal{B} = \{(1,1), (1,-1)\}$ e $\mathcal{B}' = \{(0,1), (1,1)\}$ sono basi di R^2. Calcolando le coordinate, si ottiene

$$[(1,1)]_{\mathcal{B}'} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad [(1,-1)]_{\mathcal{B}'} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \quad \text{e} \quad M^{\mathcal{B}}_{\mathcal{B}'} = \begin{pmatrix} 0 & 1 \\ -2 & 1 \end{pmatrix}.$$

Se ora $X = (2,0)$, abbiamo $[X]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e quindi $[X]_{\mathcal{B}'} = \begin{pmatrix} 0 & 1 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$.
Calcolando direttamente le coordinate di X rispetto a $\mathcal{B}'$ si ottiene il medesimo risultato.

Possiamo qui osservare che il calcolo diretto è più conveniente rispetto all'uso della matrice di trasformazione: tali matrici sono utili nello studio della similitudine, come vedremo in seguito.

Vediamo ora le principali proprietà delle matrici di trasformazione.

Proposizione 5.24. *Se $\mathcal{B}$, $\mathcal{B}'$ e $\mathcal{B}''$ sono basi di K^n, allora*

a. $M_{\mathcal{B}}^{\mathcal{B}} = I_n$.

b. $M_{\mathcal{B}''}^{\mathcal{B}'} M_{\mathcal{B}'}^{\mathcal{B}} = M_{\mathcal{B}''}^{\mathcal{B}}$.

c. $M_{\mathcal{B}'}^{\mathcal{B}} \in GL_n$ *e* $(M_{\mathcal{B}'}^{\mathcal{B}})^{-1} = M_{\mathcal{B}}^{\mathcal{B}'}$.

d. Se $A \in GL_n$, $A = M_{\mathcal{C}}^{\mathcal{A}}$, dove $\mathcal{A}$ è la base di K^n formata dalle colonne $[A]^1, \ldots, [A]^n$ di A (Proposizione 5.23) e $\mathcal{C}$ è la base canonica di K^n.

PROVA

a. Se $X_i \in \mathcal{B}$, allora $[X_i]_{\mathcal{B}} = e_i$.

b. Verifica usando le definizioni.

c. Basta ricordare che, per ogni j, il vettore dell coordinate di $[A]^j$ rispetto a $\mathcal{C}$ è $[A]^j$ stesso, cioè $[[A]^j]_{\mathcal{C}} = [A]^j$.

6. SIMILITUDINE E DIAGONALIZZAZIONE

6.1. MATRICI E APPLICAZIONI

Se $A \in M_{m,n}$ e se $X \in K^n$, sappiamo che AX è un elemento di K^m. In questo modo possiamo definire un'applicazione $f_A : K^n \to K^m$ ponendo $f_A(X) = AX$. L'applicazione f_A viene detta *applicazione associata* a A e gode delle seguenti proprietà derivanti da quelle del prodotto tra matrici.

1. Se $X,\ Y \in K^n$, $f_A(X + Y) = f_A(X) + f_A(Y)$.
2. Se $X \in K^n$ e $a \in K$, $f_A(aX) = af_A(X)$.

In generale un'applicazione $f : V \to W$ tra due K-spazi vettoriali si dice *lineare* se valgono le precedenti proprietà (1) e (2), beninteso con le somme e prodotti per scalare definiti su V e W.
Per esempio, se $V = M_{m,n}$ e $W = M_{n,m}$, la trasposizione $T : M_{m,n} \to M_{n,m}$ è lineare per la Proposizione 2.1.
Le applicazioni associate a matrici $m \times n$ sono applicazioni lineari da K^n a K^m; di fatto si può provare che ogni applicazione lineare da K^n a K^m è di questo tipo.
Osserviamo che se $A \in M_{m,n}$ e $B \in M_{n,p}$, è definita l'applicazione composta $f_A \circ f_B : K^p \to K^m$ e $f_A \circ f_B = f_{AB}$.

Esempio

1. Se $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, f_A è l'applicazione da R^2 in R^2 che manda il generico vettore (x, y) nel vettore $(x + y, x - y)$ determinato da $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$.
2. f_{I_n} è l'identità su K^n.

Se $A \in M_{m,n}$, ricordiamo che l'*immagine* dell'applicazione f_A è l'insieme
$Im(A) = \{Y \in K^m | \quad \exists X \in K^n \quad \text{tale che} \quad Y = f_A(X) = AX\}$.

Allora $Im(A) = \mathcal{C}(A)$: infatti $Y \in \mathcal{C}(A)$ se e solo se esistono $c_1, \ldots, c_n$ tali che $Y = c_1[A]^1 + \cdots + c_n[A]^n = AX$, dove $[X]_i = c_i$.

Dunque $Im(A)$ è un sl di K^n e $dimIm(A) = r(A)$ per il Teorema 5.21.

Il sl definito dal sistema $AX = O$ viene detto *nucleo* di f_A e denotato con $Ker(A)$: è l'insieme dei vettori di K^n che viene trasformato in O_m da f_A. Per la Proposizione 5.15, $dimKer(A) = n - r(A)$, quindi vale il

Teorema 6.1. *Se* $A \in M_{m,n}$, $dimIm(A) + dimKer(A) = n$.

Ricordiamo che l'applicazione f_A è detta *surgettiva* se $Im(A) = K^m$, è detta *iniettiva* se $AX = AX'$ implica $X = X'$ per $X,\ X' \in K^n$ e è detta *biunivoca* se è iniettiva e surgettiva.

Se $X,\ X' \in K^n$ sono tali che $AX = AX'$, abbiamo $AX - AX' = A(X - X') = O$. Questo vuol dire che $X - X' \in Ker(A)$ e abbiamo che f_A è iniettiva se e solo se $Ker(A) = \{O\}$. Allora

i. f_A è surgettiva se e solo se $r(A) = m$ (Corollario 5.17); quindi f_A surgettiva implica $n \geq m$.
ii. f_A è iniettiva se e solo se $r(A) = n$ (Teorema 6.1); quindi f_A iniettiva implica $n \leq m$.
iii. f_A è biunivoca se e solo se $r(A) = n = m$, quindi se e solo $A \in GL_n$.
Riguardo a (iii),osserviamo che, se $A \in GL_n$, $(f_A)^{-1} = f_{A^{-1}}$. Vale la seguente

Proposizione 6.2. *Sia* $A \in GL_n$ *e sia* $\{X_1, \ldots, X_k\} \subset K^n$. *Se* $\{X_1, \ldots, X_k\}$ *è libero (è una base) , allora* $\{AX_1, \ldots, AX_k\}$ *è libero (è una base).*

PROVA

Se $c_1AX_1 + \cdots + c_kAX_k = O$, abbiamo $A(c_1X_1 + \cdots + c_kX_k) = O$, quindi $c_1X_1 + \cdots + c_kX_k \in Ker(A)$.

Siccome $A \in GL_n$, $Ker(A) = \{O\}$, dunque $c_1AX_1+\cdots+c_kAX_k = O$ e $c_1 = \cdots = c_k = 0$ per indipendenza lineare.

Nel caso di una base, $k = n$, da cui la tesi segue per li Teorema 5.16.

6.2. Similitudine

Definiamo una relazione d'equivalenza tra matrici quadrate detta di *similitudine*.

Definizione 6.3. *Se A e B sono matrici quadrate $n \times n$, B si dice simile a A, e si scrive $A \simeq B$, se esiste $N \in GL_n$ tale che $B = N^{-1}AN$.*
La matrice N si dice matrice di similitudine. Si dice anche che B è la coniugata di A tramite N.

La relazione di similitudine è una relazione di equivalenza, infatti è

1. riflessiva: $A \simeq A$ con $N = I_n$
2. simmetrica: $A \simeq B \Rightarrow B \simeq A$, poiché $B = N^{-1}AN$ implica $A = NBN^{-1}$.
3. transitiva: $A \simeq B,\ B \simeq C \Rightarrow A \simeq C$, poiché se $B = N^{-1}AN$ e $C = M^{-1}BM$, allora $C = (NM)^{-1}A(NM)$.

Esempi

1. Le matrici $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ e $B = \begin{pmatrix} 0 & -3 \\ 0 & 2 \end{pmatrix}$ sono simili: infatti, se $N = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix}$, $B = N^{-1}AN$.
2. Consideriamo le matrici in M_n del tipo aI_n per $a \in K$. Tali matrici sono particolari matrici diagonali dette *matrici scalari*: O_n e I_n sono scalari. Comunque data $N \in GL_n$, $N^{-1}aI_nN = aI_n$, perciò ogni matrice scalare è simile solo a stessa.

Alcune quantità legate alle matrici quadrate, come il rango e il deteminante, sono uguali per due matrici simili, cioè sono invarianti per similitudine.
Un'altra quantità invariante è la somma degli elementi sulla diagonale di una matrice $A \in M_n$. Tale numero si dice *traccia* di A e si indica con $tr(A)$: dunque $tr(A) = [A]_{1,1} + [A]_{2,2} + \cdots + [A]_{n,n}$.
Si può verificare che la traccia gode delle seguenti proprietà: se $A,\ B \in M_n$ e $a \in K$, allora $tr(A + B) = tr(A) + tr(B)$, $tr(aA) = atr(A)$ e $tr(AB) = tr(BA)$.
Proviamo ora che le quantità citate sono invarianti.

Proposizione 6.4. *Se $A,\ B \in M_n$ sono simili, allora*

a. $r(A) = r(B)$, $D(A) = D(B)$ *e* $tr(A) = tr(B)$.

b. $A \in GL_n$ *se e solo se* $B \in GL_n$.

PROVA

$B = N^{-1}AN$ implica che $NB = AN$. Per il Corollario 3.11 e il Corollario 5.22 abbiamo

$$r(B) = r(NB) = r(AN) = r({}^tN{}^tA) = r({}^tA) = r(A).$$

Inoltre $D(NB) = D(N)D(B) = D(A)D(N)$ implica $D(A) = D(B)$ in quanto $D(N) \neq 0$.
Infine, usando le proprietà della traccia,
$tr(B) = tr(N^{-1}AN) = tr(ANN^{-1}) = tr(A)$.

Dunque (a) è provata e (b) deriva direttamente da $r(A) = r(B)$ e dal Teorema 3.10.

ATTENZIONE!
Le condizioni (a) della Proposizione 6.5 sono necessarie ma NON sufficienti perché due matrici siano simili. Questo vuol dire che, anche se vale $r(A) = r(B)$, $D(A) = D(B)$ e $tr(A) = tr(B)$, NON NECESSARIAMENTE A e B sono simili.
Per esempio, sia $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$: allora $r(A) = r(I_2) = 2$, $D(A) = D(I_2) = 1$ e $tr(A) = tr(I_2) = 2$, ma A e I_2 non sono simili perché I_2 è scalare e quindi simile solo a se stessa.

Il problema di stabilire se due matrici sono simili o no è in generale difficile e non lo affronteremo se non nel caso particolare delle matrici diagonalizzabili; comunque la Proposizione 6.4 ci assicura che due matrici che abbiano rango, determinante oppure traccia differenti NON sono simili.
Se interpretiamo una matrice $A \in M_n$ come applicazione da K^n in K^n, possiamo domandarci che relazione sussiste tra A e una sua coniugata, anch'essa interpretata come applicazione.

Proposizione 6.5. *Siano $A \in M_n$, $N \in GL_n$ e $B = N^{-1}AN$. Se $N = M_{\mathcal{C}}^{\mathcal{B}}$, con $\mathcal{B} = \{X_1, \ldots, X_n\}$ base di K^n formata dalle colonne di N e $\mathcal{C}$ base canonica di K^n, allora per ogni $X \in K^n$ si ha*

$$[AX]_{\mathcal{B}} = B[X]_{\mathcal{B}}$$

cioè B come applicazione trasforma le coordinate di X rispetto a $\mathcal{B}$ nelle coordinate di AX rispetto a $\mathcal{B}$.
In particolare la matrice B ha come j-esima colonna il vettore delle coordinate di AX_j rispetto alla base $\mathcal{B}$, cioè

$$B = M([AX_1]_{\mathcal{B}}, \ldots, [AX_n]_{\mathcal{B}}).$$

PROVA

Per la Proposizione 5.24, $N^{-1} = M^{\mathcal{C}}_{\mathcal{B}}$, quindi se $X \in K^n$,

$$B[X]_{\mathcal{B}} = M^{\mathcal{C}}_{\mathcal{B}} A M^{\mathcal{B}}_{\mathcal{C}} [X]_{\mathcal{B}} = M^{\mathcal{C}}_{\mathcal{B}} AX = [AX]_{\mathcal{B}}.$$

Se ora prendiamo un elemento X_j di $\mathcal{B}$, $[X_j]_{\mathcal{B}} = e_j$, dunque

$$B^j = Be_j = B[X_j]_{\mathcal{B}} = [AX_j]_{\mathcal{B}}.$$

Esempio

Consideriamo le matrici $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ e $N = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix}$.
$N \in GL_2$ e, posto $\mathcal{B} = \{\, {}^tX_1 = (1, -1),\ {}^tX_2 = (2, -1)\}$, $N = M^{\mathcal{B}}_{\mathcal{C}}$.
Ora $AX_1 = O$ e $AX_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, quindi $[AX_1]_{\mathcal{B}} = O$, $[AX_2]_{\mathcal{B}} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$ e $B = \begin{pmatrix} 0 & -3 \\ 0 & 2 \end{pmatrix}$.

Definizione 6.6. *$A \in M_n$ si dice diagonalizzabile se è simile a una matrice diagonale, cioè se esistono una matrice $D \in M_n$ diagonale e $N \in GL_n$ tali che $D = N^{-1}AN$.*

Esempi

1. Se A è diagonale allora A è ovviamente diagonalizzabile prendendo $D = A$ e $N = I_n$.
2. Sia $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. A è diagonalizzabile: infatti, se $N = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, $N^{-1}AN = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$. Sia invece $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$: poiché $r(B) = 1$ e $D(B) = 0$, se B fosse simile a una matrice 2×2 diagonale C dovremmo avere, per la Proposizione 6.4, $D(C) = tr(C) = 0$ e $r(C) = 1$. Ma per le prime due condizioni deve necessariamente essere $C = O$ il che è assurdo per la terza, quindi B non è diagonalizzabile.

6.3. Autovettori e autovalori

Introduciamo dei concetti che sono essenziali per lo studio della diagonalizzabilità delle matrici.

Definizione 6.7. *Sia $A \in M_n$. Un vettore $X \in K^n$ si dice autovettore di A se $X \neq O$ e se esiste $\lambda \in K$ tale che $AX = \lambda X$. Lo scalare λ si dice autovalore di A relativo a X.*

Osserviamo che se $X = O$, $AX = \lambda X$ per ogni λ, quindi la condizione $X \neq O$ è indispensabile perché abbia senso il concetto di autovalore.
Se consideriamo A come applicazione di K^n in sé, un autovettore X di A è un vettore che viene trasformato in un multiplo di se stesso. In particolare, se $K = R$, la retta $L(X)$ viene trasformata in sé da A.

Esempi

1. Sia $A \in M_n$; poiché $Ae_j = [A]^j$, se la j-esima colonna di A è un multiplo di e_j, cioè se $[A]^j = ae_j$, allora e_j è un autovettore di A con autovalore a: per esempio
$$\begin{pmatrix} 1 & 0 & -2 \\ -1 & 3 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$
ha e_2 come autovettore con autovalore 3.
In particolare, se A è diagonale e se poniamo $[A]_{i,i} = a_i$, allora
$Ae_i = a_i e_i$ per ogni i: dunque la base canonica è costituita da autovettori. Nel caso di una matrice scalare $A = aI_n$, tali autovettori avranno tutti lo stesso autovalore a.
2. Se $A \in M_n$ e $Ker(A) \neq \{O\}$, ogni vettore X non nullo del nucleo è un autovettore di A con autovalore 0: infatti $AX = O = 0X$.
3. Sia $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. $Ker(A)$ è dato dalle soluzioni di $x + y = 0$, dunque $(1, -1)$ è un autovettore di A con autovalore 0.
D'altra parte, $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Dunque $(1, 1)$ è autovettore di A con autovalore 2.
Osserviamo che i due autovettori trovati formano una base di R^2.
Sia $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. $Ker(B)$ è dato da $y = 0$, dunque gli autovettori di B con autovalore 0 sono i multipli $(1, 0)$.
D'altra parte, imponendo la condizione $BX = \lambda X$, otteniamo $y = \lambda x$ e

$\lambda y = 0$, per cui, se $X \neq O$, abbiamo $\lambda = 0,\ y = 0$. Quindi tutti gli autovettori di B sono multipli di $(1, 0)$ e hanno autovalore 0.
Diversamente dal caso precedente, non esiste una base di R^2 formata da autovettori di B.

4. Sia $X \in K^n$ è un autovettore di $A \in M_n$ con autovalore $\lambda = 1$. Allora X é lasciato fisso da A vista come applicazione, in quanto $AX = X$. In particolare, se $K = R$, la retta $L(X)$ è tutta formata da punti fissi.

Nel precedente esempio (3), A è diagonalizzabile mentre B non lo è: tale fatto deriva dal legame tra basi di autovettori e diagonalizzabilità espresso in generale dal seguente

Teorema 6.8. *$A \in M_n$ è diagonalizzabile se e solo se esiste una base di K^n formata da autovettori di A.*
In particolare, se $\mathcal{B} = \{X_1, \ldots, X_n\}$ è una tale base, se λ_i è l'autovalore relativo a X_i per $i = 1, \ldots, n$ e se $N = M(X_1, \ldots, X_n)$, allora $N^{-1}AN = D$, con

$$D = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{pmatrix}.$$

PROVA

Sia $\mathcal{B} = \{X_1, \ldots, X_n\}$ una base di K^n tale che per ogni $i = 1, \ldots, n$ si abbia $AX_i = \lambda_i X_i$ e si consideri la matrice $N = M(X_1, \ldots, X_n) = M^{\mathcal{B}}_{\mathcal{C}}$.
Poiché per la Proposizione 6.5

$$[AX_i]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{pmatrix} = \lambda_i e_i,$$

abbiamo $N^{-1}AN = M(\lambda_1 e_1, \ldots, \lambda_n e_n) = D$.
Viceversa, supponiamo che esista $N \in GL_n$ tale che $N^{-1}AN = D$ sia diagonale. Posto $[D]_{i,i} = \lambda_i$, per $i = 1, \ldots, n$ abbiamo

$$De_i = N^{-1}ANe_i = \lambda_i e_i$$

da cui

$$A[N]^i = ANe_i = \lambda_i Ne_i = \lambda_i [N]^i,$$

cioè la i-esima colonna di N (che è $\neq 0$ poiché $N \in GL_n$) è autovettore di A con autovalore λ_i.
Ponendo $[N]^i = X_i$, $\mathcal{B} = \{X_1, \dots, X_n\}$ è una base di K^n formata da autovettori di A.

Il Teorema 6.8 ci porta a approfondire lo studio degli autovettori di una matrice.
Se $A \in M_n$ e $\lambda \in K$, consideriamo il problema di determinare tutti i vettori $X \in K^n$ tali che $AX = \lambda X$. Se scriviamo $X = I_n X$, i vettori in questione saranno le soluzioni del sistema omogeneo $S_\lambda : (A - \lambda I_n)X = O$.
Evidentemente, λ è un autovalore di A se e solo se S_λ ha soluzioni non nulle. In tal caso denotiamo il sl non nullo $sol(S_\lambda)$ con $V_A(\lambda)$ e lo chiamiamo *autospazio* di A relativo a λ.
Gli elementi di $V_A(\lambda)$ sono tutti gli autovettori di A relativi a λ più lo O (che è l'unico elemento di $sol(S_\lambda)$ se λ non è autovalore). Dunque la somma, il prodotto per scalare e quindi le cl tra autovettori con autovalore λ o sono nulli o danno autovettori con autovalore λ (Teorema 5.18).
Per la Proposizione 5.15, $dimV_A(\lambda) = n - r(A - \lambda I_n)$.
Come vedremo, una matrice di $M_n(C)$ ha sempre autovalori, mentre questo non è in generale vero per matrici di $M_n(R)$.
Se consideriamo due autovettori X_1 e X_2 di un matrice A con autovalori $\lambda_1 \neq \lambda_2$ distinti, vediamo che X_1 e X_2 devono essere linearmente indipendenti. In caso contrario, dovremmo avere $X_2 = aX_1$ per un opportuno $a \in K$, $a \neq 0$. Moltiplicando per A entrambi i membri dell'uguaglianza si avrebbe $\lambda_2 X_2 = \lambda_1 a X_1$, quindi $\lambda_2 a X_1 = \lambda_1 a X_1$ e $\lambda_1 = \lambda_2$ contro l'ipotesi. Generalizziamo questo fatto.

Lemma 6.9. *Sia $A \in M_n$ e siano $X_1, \dots, X_k$ autovettori di A tali che i relativi autovalori siano differenti. Allora $\{X_1, \dots, X_k\}$ è un insieme libero.*

PROVA

Usiamo induzione su k. Per $k = 1$ è ovvio. Supponiamo $k > 1$ e che valga la tesi per $k - 1$ autovettori con autovalori distinti.
Sia λ_i l'autovalore relativo a X_i, per $i = 1, \dots, k$ e supponiamo che una cl di $X_1, \dots, X_k$ sia nulla:

$$c_1 X_1 + c_2 X_2 + \cdots + c_k X_k = O \tag{6.1}$$

Moltiplicando la 6.1 per A e per λ_1 otteniamo

$$c_1 \lambda_1 X_1 + c_2 \lambda_2 X_2 + \cdots + c_k \lambda_k X_k = O, \tag{6.2}$$

e

$$c_1\lambda_1 X_1 + c_2\lambda_1 X_2 + \cdots + c_k\lambda_1 X_k = O. \qquad 6.3$$

Sottraendo dalla 6.2 la 6.3 otteniamo

$$c_2(\lambda_2 - \lambda_1)X_2 + \cdots + c_k(\lambda_k - \lambda_1)X_k = O. \qquad 6.4$$

Questa è una cl nulla tra $k-1$ autovettori con autovalori distinti, dunque $c_2 = \cdots = c_k = 0$ per l'ipotesi induttiva. Sostituendo in 6.1 otteniamo che anche $c_1 = 0$.

Teorema 6.10. *Siano $A \in M_n$ e $\lambda_1, \ldots, \lambda_k$ autovalori di A. Data per ogni autospazio $V_A(\lambda_i)$ una base $\mathcal{B}_i$, il sottoinsieme $\mathcal{L}$ di K^n ottenuto come unione delle $\mathcal{B}_i$ ($\mathcal{L} = \bigcup_{i=1}^{n} \mathcal{B}_i$) è un insieme libero.*

PROVA

Per semplificare, diamo un'esposizione informale della prova. Consideriamo una generica cl di elementi di $\mathcal{L}$, indicata con C. Dobbiamo provare che $C = O$ implica che tutti i coefficienti di C sono nulli. Raggruppando nella sommatoria che esprime C gli addendi che appartengono al medesimo autospazio, possiamo scrivere C come somma di cl C_i, dove C_i è cl di elementi di $\mathcal{B}_i$. Dunque C_i o è O o è un autovettore con autovalore λ_i. Ma se qualcuna delle C_i fosse $\neq O$, $C = C_1 + C_2 + \cdots + C_k = O$ sarebbe un cl nulla di autovettori con autovalori distinti e con coefficienti non tutti nulli: questo è assurdo per il Lemma 6.9.
Pertanto $C_i = O$ per ogni i; poiché le $\mathcal{B}_i$ sono basi, tutti i coefficienti delle C_i, e pertanto tutti i coefficienti di C, devono essere 0.

Sia $A \in M_n$ e supponiamo ora che $\lambda_1, \ldots, \lambda_k$ siano tutti gli autovalori di A. Scelta per ogni autospazio $V_A(\lambda_i)$ una base $\mathcal{B}_i$ e posto $d_i = dimV_A(\lambda_i)$, l'unione di tali basi $\mathcal{L}$ è un insieme libero di autovettori di A con $d = d_1 + \cdots + d_k$ elementi (le basi sono necessariamente disgiunte). Quindi $\mathcal{L}$ genera un sl di K^n la cui dimensione $d \leq n$ è la somma delle dimensioni degli autospazi: se fosse $d = n$, $\mathcal{L}$ sarebbe una base di K^n e A sarebbe diagonalizzabile per il Teorema 6.8. Per determinare le condizioni perché ciò avvenga dobbiamo analizzare meglio la natura degli autovalori.
Come abbiamo visto, λ è un autovalore di $A \in M_n$ se e solo se il sistema quadrato $S_\lambda : (A - \lambda)I_n = O$ ha soluzioni non nulle, cioè non è determinato. Questo, per il Teorema di Cramer, equivale a dire che la matrice $A - \lambda I_n$ è singolare, e quindi (Teorema 4.5) che $D(A - \lambda I_n) = 0$.

Possiamo dunque asserire che la ricerca degli autovalori di una matrice coincide con la ricerca dei $\lambda \in K$ tali che $D(A - \lambda I_n) = 0$. Per questo motivo siamo interessati all'espressione nell'incognita t definita da $p_A(t) = D(A - tI_n)$. Indichiamo con $K[t]$, $R[t]$, $C[t]$ i polinomi in t a coefficienti rispettivamente in K, R o C. Vale il seguente teorema.

Teorema 6.11. *Se $A \in M_n$, allora $p_A(t) = D(A - tI_n)$ è un polinomio di $K[t]$ di grado n il cui coefficiente direttivo (coefficiente del monomio di grado n) è $(-1)^n$.*

Il polinomio $p_A(t)$ viene detto *polinomio caratteristico* della matrice A. Dalla discussione precedente abbiamo che:

Proposizione 6.12. *$\lambda \in K$ è un autovalore di $A \in M_n$ se e solo se $p_A(\lambda) = 0$, cioè se e solo se λ è una radice di $p_A(t)$.*

Ricordiamo ora il

Teorema di Ruffini. *Se $p(t) \in K[t]$ e $\lambda \in K$ è una radice di $p(t)$, allora esistono un intero positivo $m(\lambda)$ e $q(t) \in K[t]$ tali che $q(\lambda) \neq 0$ e $p(t) = (t - \lambda)^{m(\lambda)} q(t)$.*

L'intero $m(\lambda)$ si dice *molteplicità* di λ. Se $m(\lambda) > 1$, la radice λ si dice *multipla*, mentre se $m(\lambda) = 1$ si dice *semplice*.
A questo punto, a ogni autovalore λ di una matrice $A \in M_n$ risultano associati due interi positivi, la dimensione dell'autospazio relativo $V_A(\lambda)$ e la sua molteplicità $m(\lambda)$ come radice del polinomio caratteristico $p_A(t)$. In questo senso parleremo di autovalori semplici o autovalori multipli.

Esempi

1. Sia $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} \in M_3(R)$. Allora

$$p_A(t) = D(A - tI_3) = D(\begin{pmatrix} 1-t & 2 & 3 \\ 2 & 4-t & 6 \\ 3 & 6 & 9-t \end{pmatrix}) = -t^3 + 14t^2.$$

Dunque A ha autovalori 0 e $-\frac{1}{4}$ con molteplicità $m(0) = 2$ e $m(-\frac{1}{4}) = 1$.

2. Sia $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 6 \\ 0 & 0 & 2 \end{pmatrix} \in M_3(R)$. Allora

$$p_A(t) = (t-1)^2(t-2)$$

In generale, se $A \in M_n$ è un matrice triangolare, si ha

$$p_A(t) = (t-a_1)^{m_1} \cdot (t-a_2)^{m_2} \cdot \dots \cdot (t-a_k)^{m_k}$$

dove $a_1, \dots, a_k$ sono gli elementi distinti sulla diagonale principale di A e m_i è il numero di volte che a_i compare: dunque gli a_i sono gli autovalori di A e gli m_i sono le loro molteplicità.

Un fatto importante è l'invarianza di $p_A(t)$ per similitudine.

Proposizione 6.13. *Se $A,\ B \in M_n$ sono matrici simili, allora $p_A(t) = p_B(t)$, quindi A e B hanno gli stessi autovalori con la stessa molteplicità. Inoltre, se λ è un autovalore di A e B, allora $dimV_A(\lambda) = dimV_B(\lambda)$.*

PROVA

Se $B = N^{-1}AN$, $N^{-1}(A - \lambda I_n)N = N^{-1}AN - \lambda N^{-1}I_nN = B - \lambda I_n$. Quindi le matrici $A - \lambda I_n$ e $B - \lambda I_n$ sono simili (con stessa matrice di similitudine N) e, per la Proposizione 6.4, hanno stesso determinante, da cui quindi $p_A(t) = p_B(t)$, e lo stesso rango. Pertanto $dimV_A(\lambda) = n - r(A - \lambda I_n) = n - r(B - \lambda I_n) = dimV_B(\lambda)$.

Vediamo ora che relazione sussiste tra molteplicità di un autovalore e dimensione dell'autospazio relativo.

Proposizione 6.14. *Sia $A \in M_n$ e sia λ autovalore di A. Allora $dimV_A(\lambda) \leq m(\lambda)$.*

PROVA

Sia $d = dimV_A(\lambda)$ e sia $X_1, \dots, X_d$ una base di $V_A(\lambda)$. Per il Teorema 5.16 a, posso trovare vettori $X_{d+1}, \dots, X_n$ in modo che $\mathcal{B} = \{X_1, \dots, X_n\}$ sia una base di K^n. Posto allora $N = M_{\mathcal{C}}^{\mathcal{B}}$, per la Proposizione 6.5, la matrice $B = N^{-1}AN$ è della forma a blocchi

$$B = \begin{pmatrix} \lambda I_d & | & B_1 \\ - & - & - \\ O_{n-d} & | & B_2 \end{pmatrix},$$

dove $B_1 \in M_{d,n-d}$ e $B_2 \in M_{n-d}$. Dunque

$$B - tI_n = \begin{pmatrix} (\lambda - t)I_d & | & B_1 \\ - & - & - \\ O_{n-d} & | & B_2 - tI_{n-d} \end{pmatrix},$$

e $p_A(t) = p_B(t) = (\lambda - t)^d p_{B_2}(t)$ per la Proposizione 4.3: questo implica che $d \leq m(\lambda)$.

Diamo ora il criterio generale di diagonalizzabilità.

Teorema 6.15. *Supponiamo che $A \in M_n$ con autovalori $\lambda_1, \ldots, \lambda_k$. Allora A è diagonalizzabile se e solo se*

a. $\sum_{i=1}^{k} m(\lambda_i) = n$.

b. Per ogni $i = 1, \ldots, k$, $dimV_A(\lambda_i) = m(\lambda_i)$.

PROVA

Siano vere (a) e (b) e sia $d = \sum_{i=1}^{k} dimV_A(\lambda_i)$. Per il Teorema 6.10 esiste allora un insieme libero $\mathcal{L} = \{X_1, \ldots, X_d\}$ formato da autovettori di A mentre per (a) e (b) $d = n$. Dunque $\mathcal{L}$ è una base di K^n e A è diagonalizzabile per il Teorema 6.8.
Viceversa, supponiamo che A sia diagonalizzabile. Allora $A \simeq D$, dove D è diagonale con gli autovalori di A sulla diagonale e $p_A(t) = p_D(t) = (t - \lambda_1)^{m_1} \cdots (t - \lambda_k)^{m_k}$, dove $m_i = m(\lambda_i)$: poiché $\sum_{i=1}^{k} m_i = n$, questo prova (a). Riguardo a (b) , dato $i = 1, \ldots, k$, $r(D - \lambda_i I_n) = n - m_i$, dunque $dimV_A(\lambda_i) = m_i$.

Se λ è un autovalore di A tale che $dimV_A(\lambda) = m(\lambda)$, diciamo che λ è *regolare*.
Se $A \in M_n$ è diagonalizzabile, una matrice diagonale che ha gli autovalori di A sulla diagonale principale si dice *forma diagonale* di A. Le forme diagonali di A si ottengono l'una dall'altra permutando gli autovalori. Se $\mathcal{B}$ è una base di K^n formata da autovettori di A e $N = M_{\mathcal{C}}^{\mathcal{B}}$, $N^{-1}AN = D$ è una forma diagonale: permutando gli elementi $\mathcal{B}$ si ottengono, tramite le relative matrici di similitudine, le altre forme diagonali.

Corollario 6.16. *Se $A \in M_n$ ha n autovalori semplici $\lambda_1, \ldots, \lambda_n$, allora A è diagonalizzabile.*

PROVA

$1 \leq dimV_A(\lambda_i) \leq m(\lambda_i) = 1$ per la Proposizione 6.14. Quindi (a) e (b) del Teorema 6.15 sono verificate.

ATTENZIONE! La condizione del Corollario 6.16 è sufficiente ma NON necessaria. Non si può dire che una matrice non è diagonalizzabile perché ha autovalori multipli: altrimenti I_n non sarebbe diagonalizzabile!

Corollario 6.17. *Se A e B sono diagonalizzabili e se $p_A(t) = p_B(t)$, allora $A \simeq B$.*

PROVA

A e B hanno gli stessi autovalori e quindi sono simili alla stessa matrice diagonale.

ATTENZIONE! La condizione $p_A(t) = p_B(t)$ NON implica in generale che $A \simeq B$. Per esempio, se $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ e $B = I_2$, abbiamo $p_A(t) = p_B(t) = (t-1)^2$ ma, come si è visto, A non è simile a B.
Osserviamo che se $A \in M_n$ è diagonalizzabile e se $p_A(t) = (t-\lambda)^n$ (cioè A ha un solo autovalore), allora necessariamente $A = \lambda I_n$: infatti A è simile a λI_n che , come già osservato, è simile solo a se stessa. Questo fatto può essere utilizzato per provare che una matrice non è diagonalizzabile: per esempio $\left(\begin{smallmatrix} 1 & -1 \\ 1 & 3 \end{smallmatrix}\right)$ ha autovalore 2 con moteplicità 2, quindi se fosse diagonalizzabile sarebbe simile a $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ che è simile solo a se stessa.
Riguardo allo studio della similitudine, possiamo delineare il seguente schema: siano $A,\ B \in M_n$ con $p_A(t) = p_B(t)$.

1. Se $A,\ B$ sono diagonalizzabili, allora sono simili (Corollario 6.17).
2. Se una delle due matrici è diagonalizzabile e l'altra no, allora non sono simili.
3. Se $A,\ B$ non sono diagonalizzabili, non abbiamo fornito in questi appunti un criterio per dare una risposta.

Finora abbiamo enunciato asserzioni che assumevano come ipotesi l'esistenza di autovalori, cioè di radici di un dato polinomio: ora vogliamo dire qualcosa riguardo a questo problema. L'esistenza di radici è un fatto importante che differenzia R da C.

Teorema fondamentale dell'Algebra. *Se $p(t) \in C[t]$, allora $p(t)$ ammette radici. Inoltre se n è il grado di $p(t)$ e se $\lambda_1, \ldots, \lambda_k$ sono le radici di $p(t)$, allora*

$$\sum_{i=1}^{k} m(\lambda_i) = n.$$

La seconda parte dell'enunciato significa che $p(t)$ ha n radici se «contiamo le radici con la loro molteplicità ».

D'altra parte, un polinomio di $R[t]$ può non avere affatto radici reali, per esempio t^2+1, o averne ma con somma delle molteplicità minore del grado, per esempio t^3-t^2+t-1. Poiché $R[t] \subset C[t]$, un polinomio $p(t) \in R[t]$ può sempre essere considerato come polinomio in $C[t]$. Quindi $p(t)$ ha perlomeno radici in C: riguardo a questo fatto abbiamo che se $z \in C$ è una radice di $p(t) \in R[t]$, allora anche il suo coniugato $\overline{z}$ è radice. Infatti, se $p(t) = a_n t^n + \cdots + a_1 t + a_0$ e $p(z) = a_n z^n + \cdots + a_1 t + a_0 = 0$, allora

$$0 = \overline{p(z)} = \overline{a}_n \overline{z}^n + \cdots + \overline{a}_1 \overline{z} + \overline{a}_0 = a_n \overline{z}^n + \cdots + a_1 \overline{z} + a_0 = 0,$$

poiché $a_i \in R$ per $i = 1, \ldots, n$.

Possiamo concludere dalla discussione precedente che, dato un polinomio $p(t) \in R[t]$, il numero di radici non reali di $p(t)$ considerato come elemento di $C[t]$ è pari. Quindi, se il grado di $p(t)$ è dispari, $p(t)$ avrà almeno una radice reale.

In ogni caso, determinare esattamente la radici di un polinomio di grado > 2 è un problema che non è risolubile in generale: esistono vari metodi validi per classi speciali di polinomi, oltre che metodi di calcolo approssimato.

La discussione precedente si riflette sulle matrici: se $A \in M_n(C)$, la condizione (a) del Teorema 6.15 è sempre verificata, quindi A può non essere diagonalizzabile solo in presenza di autovalori non regolari.

Invece se $A \in M(R)$ (a) può non valere, come abbiamo visto. Se A è diagonalizzabile considerata come matrice in $M_n(C)$, diciamo che A è *diagonalizzabile su* C. In questo caso, mentre $p_A(t) \in R[t]$, gli autovalori non sono in generale reali e gli autovettori sono elementi di C^n.

Esempio

Sia $A = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$; allora $D(A - \lambda I_n) = \lambda^2 + 1 \neq 0$ per ogni $\lambda \in R$, quindi $A - \lambda I_n$ è sempre invertibile e $sol(S_\lambda) = \{O\}$ per ogni $\lambda \in R$. Invece, considerando $A \in M_n(C)$, vediamo che ha autovalori $i,\ -i$, quindi è diagonalizzabile su C e è simile alla matrice $\left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right)$. Una base di C^2 formata da autovettori di A è $\{(1, i), (1, -i)\}$.

Si può provare che la traccia e il determinante di una matrice complessa sono rispettivamente la somma e il prodotto dei suoi autovalori. Noi lo dimostriamo nel caso 2×2.

Proposizione 6.18. *Se $A \in M_n(C)$ e se $\lambda_1, \ldots, \lambda_k$ sono gli autovalori di A, allora*

$$tr(A) = m(\lambda_1) + m(\lambda_2) + \cdots + m(\lambda_k) \qquad e \qquad D(A) = m(\lambda_1) \cdot m(\lambda_2) \cdot \cdots m(\lambda_k).$$

PROVA

($n = 2$). Se $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$,

$$p_A(t) = D(\begin{pmatrix} a-t & b \\ c & d-t \end{pmatrix}) = t^2 - (a+d)t + ad - bc = t^2 - tr(A)t + D(A).$$

Quindi la somma delle radici di $p_A(t)$ è $tr(A)$ mentre il prodotto è $D(A)$.

Terminiamo questa sezione con un esempio che mostra come applicare quanto fatto allo studio della diagonalizzabilità di matrici.

Esempio

Per ogni $a \in R$ si consideri la matrice

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

Studiamo la diagonalizzabilità di A su R e su C al variare di a e determiniamo in alcuni casi delle basi di autovettori. Riguardo a quest'ultimo problema, possiamo già subito osservare che, per ogni a, e_1 è autovettore di A con autovalore 3, in quanto $Ae_1 = [A]^1 = 3e_1$.

$$p_A(t) = D(\begin{pmatrix} 3-t & 2 & 1 \\ 0 & 1-t & a \\ 0 & a-3 & 1-t \end{pmatrix}) = (t-3)(t^2 - 2t + 1 - a(a-3))$$

(abbiamo usato lo sviluppo rispetto alla prima colonna: si osservi che il polinomio è già decomposto in due fattori, quindi basta risolvere una equazione di II grado).

Al variare di a gli autovalori di A sono 3 e $1 \pm \sqrt{a(a-3)}$.

1. Se $0 < a < 3$, allora il discriminante è < 0, quindi vi sono due autovalori complessi (non reali) coniugati.
   A non è diagonalizzabile su R mentre lo è su C essendoci in C tre autovalori distinti (Corollario 6.16).
   A titolo di esempio, determiniamo una base di autovettori per $a = 1$. In questo caso gli autovalori sono 3 e $1 \pm i\sqrt{2}$. Dunque, per determinare gli autospazi $V_A(1+i\sqrt{2})$ e $V_A(1-i\sqrt{2})$ risolviamo i sistemi $A-(1+i\sqrt{2})I_3)X = O$ e $A - (1-i\sqrt{2})I_3)X = O$: riducendo troviamo

$$\begin{cases} (2-i\sqrt{2})x+2y+z=0 \\ i\sqrt{2}y-z=0 \end{cases} \quad , \quad \begin{cases} (2+i\sqrt{2})x+2y+z=0 \\ i\sqrt{2}y+z=0 \end{cases}$$

quindi una base di autovettori è

$$(1,0,0),(-2-i\sqrt{2},2-i\sqrt{2},2(1+i\sqrt{2})),(-2+i\sqrt{2},2+i\sqrt{2},2(1-i\sqrt{2})).$$

2. Per $a \leq 0$ o $a \geq 3$ gli autovalori sono tutti reali.
Vi sono autovalori multipli per gli a tali che il discriminante è 0, cioè per $a=0$ e $a=3$, e per gli a tali che $1 \pm \sqrt{a(a-3)} = 3$. Questi ultimi sono $a=-1$ e $a=4$.
Sempre per il Corollario 6.16 , possiamo allora affermare che per $a > 3$, $a < 0$ e $a \neq -1,\ 4$, A è diagonalizzabile su R.
Esaminiamo ora in dettaglio i casi $a = 0,\ 3,\ 4,\ -1$.
   - a=0. Abbiamo 1 come autovalore con $m(1)=2$. Poiché

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

     ha rango 2, $dimV_A(1) = 1$, quindi A non è diagonalizzabile per il Teorema 6.15.
   - a=3. Analogo al precedente.
   - a=4. Abbiamo 3 come autovalore con $m(3)=2$. Poiché

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

     ha rango 2, $dimV_A(3) = 1$, quindi A non è diagonalizzabile per il Teorema 6.15.
   - a=-1. Abbiamo 3 come autovalore con $m(3)=2$. Poiché

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

     ha rango 1, $dimV_A(3) = 2$, quindi A è diagonalizzabile per il Teorema 6.15.

Troviamo una base di autovettori: $(A-3I)X = O$ si riduce all'equazione $2y+z=0$, dunque una base di $V_A(3)$ è $(1,0,0),(0,1,-2)$. L'altro autovalore è -1 (sicuramente regolare in quanto $m(-1) = 1$) e $V_A(-1)$ è dato dal sistema

$$\begin{cases} x+2y+z=0 \\ 2y-z=0 \end{cases}$$

Dunque possiamo completare la base precedente con $(-1,1,2)$.

7. STRUTTURA METRICA DI R^N

7.1. Prodotto scalare e distanza

Se X e Y sono vettori di R^n, il numero reale

$$^tXY = [X]_1[Y]_1 + [X]_2[Y]_2 + \cdots + [X]_n[Y]_n$$

si dice *prodotto scalare canonico* tra X e Y e si indica con $X \cdot Y$.

L'aggettivo «canonico» presuppone una definizione più generale di prodotto scalare. Poiché noi ci occuperemo solo del prodotto scalare canonico, chiameremo quest'ultimo semplicemente «prodotto scalare».

Esempio

Se $X = (1, 2, -1)$ e $Y = (0, 1, 1)$, $X \cdot Y = 1$.

Il prodotto scalare ha le seguenti proprietà: dati $X,\ Y,\ Z \in R^n$ e $a \in R$

PS1. Bilinearità: $(X + Y) \cdot Z = X \cdot Z + Y \cdot Z$, $X \cdot (Y + Z) = X \cdot Y + X \cdot Z$ e $(aX) \cdot Y = X \cdot (aY) = aX \cdot Y$.
(PS1) deriva direttamente dalle proprietà del prodotto tra matrici.

PS2. Simmetria: $X \cdot Y = Y \cdot X$.
Infatti $X \cdot Y = {}^t({}^tXY) = {}^tYX = Y \cdot X$.

PS3. Positività: $X \cdot X \geq 0$ e $X \cdot X = 0$ se e solo se $X = O$.
Infatti $X \cdot X = [X]_1^2 + [X]_2^2 + \cdots + [X]_n^2 \geq 0$ e, in quanto somma di quadrati, è $= 0$ se e solo $[X]_i = 0$ per ogni i, cioè $X = O$.

Osserviamo che (PS3) ha senso per il fatto che stiamo considerando elementi di R^n. Inoltre (PS3) implica

PS4. $X \cdot X' = 0$ per ogni $X' \in R^n$ se e solo se $X = O$.

Se $X \in R^n$, il numero reale $\geq$ dato da $\sqrt{X \cdot X}$ si dice *norma* di X e si denota con $\|X\|$. Evidentemente $\|aX\| = |a|\|X\|$ per ogni $a \in R$ e $\|X\| = 0$ se e solo se $X = O$. Vale inoltre la

Disuguaglianza di Schwarz. *Dati* $X,\ Y \in R^n$, *abbiamo*

$$|X \cdot Y| \leq \|X\|\|Y\|.$$

PROVA

La disuguaglianza è ovvia se $Y = 0$, quindi possiamo supporre $Y \neq O$. (PS3) implica che $(aX + bY) \cdot (aX + bY) \geq 0$ comunque dati $a,\ b \in R$. Dunque, per (PS1) e (PS2),

$$a^2\|X\|^2 + b^2\|Y\|^2 + 2abX \cdot Y \geq 0. \tag{7.1}$$

Se prendiamo ora $a = \|Y\|^2$ e $b = -X \cdot Y$, sostituendo in (1) otteniamo

$$\|Y\|^4\|X\|^2 + (X \cdot Y)^2\|Y\|^2 - 2(X \cdot Y)^2\|Y\|^2 \geq 0. \tag{7.2}$$

poiché $\|Y\| \neq 0$ posso semplificare la (2)

$$\|Y\|^2\|X\|^2 - (X \cdot Y)^2 \geq 0. \tag{7.3}$$

e, portando a destra $(X \cdot Y)^2$ e passando alle radici quadrate

$$X \cdot Y \leq \|Y\|\|X\|.$$

Dalla disuguaglianza di Schwarz otteniamo che, se $X,\ Y \in R^n$, risulta

$$\|X + Y\| \leq \|X\| + \|Y\|.$$

Infatti, per Schwarz,

$$\|X+Y\|^2 = \|X\|^2 + \|Y\|^2 + 2X \cdot Y \leq \|X\|^2 + \|Y\|^2 + 2\|X\|\|Y\| = (\|X\| + \|Y\|)^2.$$

Sostituendo in questa disuguaglianza X con $X_1 - X_2$ e Y con $X_2 - X_3$, X_1, X_2 e X_3 vettori di R^n, otteniamo la

Disuguaglianza triangolare.

$$\|X_1 - X_3\| \leq \|X_1 - X_2\| + \|X_2 - X_3\|.$$

Se X e Y sono elementi di R^n, il numero $\|X - Y\|$ viene chiamato *distanza* tra X e Y e indicato con $d(X, Y)$. In base a quanto detto prima, dati comunque $X,\ Y,\ Z \in R^n$ valgono le seguenti proprietà della distanza:

1. $d(X, Y) \geq 0$ e $d(X, Y) = 0 \Leftrightarrow X = Y$.
2. $d(X, Y) = d(Y, X)$.
3. $d(X, Z) \leq d(X, Y) + d(Y, Z)$.

Se consideriamo i casi $n = 2$ o $n = 3$, vediamo che la nozione qui introdotta coincide con quella di distanza euclidea nel piano o spazio cartesiano.

Per esempio, per $n = 2$, se $X = (x_1, x_2)$ e $Y = (y_1, y_2)$, allora
$X - Y = (x_1 - y_1, x_2 - y_2)$ e $d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$.

7.2. ORTOGONALITÀ E BASI ORTONORMALI

Per la disuguaglianza di Schwarz, dati X e Y in R^n non nulli, si ha

$$-1 \leq \frac{X \cdot Y}{\|X\|\|Y\|} \leq 1.$$

Quindi esiste un unico angolo θ compreso tra 0 e π tale che

$$X \cdot Y = \|X\|\|Y\| cos\theta.$$

θ viene detto l'*angolo* tra X e Y. Osserviamo che anche in questo caso tale nozione coincide con quella data nel piano o nello spazio cartesiano.

Esempio

Se $X = (1, 1, 0)$ e $Y = (0, 1, 0)$, $X \cdot Y = 1$, $\|X\| = \sqrt{2}$, $\|Y\| = 1$, quindi $cos(\theta) = \frac{1}{\sqrt{2}}$ e $\theta = \frac{\pi}{4}$.

Considerando i casi estremi $\theta = 0$ e $\theta = \pi$, otteniamo che $Y = aX$, con $a > 0$ nel primo caso e $a < 0$ nel secondo. Infatti, supponiamo dapprima che $\|Y\| = \|X\|$. Per $\theta = 0$ abbiamo

$$\|X - Y\|^2 = \|X\|^2 + \|Y\|^2 - 2\|X\|\|Y\| = (\|Y\| - \|X\|)^2 = 0$$

dunque $d(X, Y) = 0$ e $X = Y$.
Analogamente, per $\theta = \pi$, otteniamo $Y = -X$.
Se X e Y hanno norme differenti basta applicare quanto detto sopra a $\frac{\|Y\|}{\|X\|}X$ e Y.
Il caso in cui $\theta = \pi/2$ corrisponde invece a $X \cdot Y = 0$. Dunque possiamo dare la seguente definizione:

Definizione 7.1. *Due vettori X e Y di R si dicono ortogonali se $X \cdot Y = 0$. In tal caso scriviamo $X \perp Y$.*

Un vettore $X \in R^n$ si dice *versore* se $\|X\| = 1$. Se $X \neq O$, possiamo associare a X il versore $\nu(X) = \frac{1}{\|X\|}X$ detto *normalizzato* di X.

Esempi

1. $(1, -1) \perp (1, 1)$ e $(1, -1, -1) \perp (2, 1, 1)$.
   $\nu((1, -1)) = \frac{1}{\sqrt{2}}(1, -1)$ e $\nu((2, 1, 1)) = \frac{1}{\sqrt{6}}(2, 1, 1)$.
2. $O \perp X$ per ogni $X \in R^n$.
3. I vettori canonici $e_1, \ldots, e_n$ di R^n sono versori. Se $i \neq j$, $e_i \perp e_j$.

Definizione 7.2. *Un insieme $\mathcal{H} = \{X_1, \ldots, X_k\}$ di vettori di R^n si dice ortogonale se $X_i \perp X_j$ per $1 \leq i \neq j \leq k$. Un insieme ortogonale $\mathcal{H}$ si dice ortonormale se ogni X_i è un versore. Se una base $\mathcal{B}$ è un insieme ortogonale (ortonormale), $\mathcal{B}$ si dice base ortogonale (ortonormale).*

Dunque $\mathcal{H}$ è ortonormale se $X_i \cdot X_j = 0$ per $i \neq j$ e $X_i \cdot X_i = 1$ per ogni i. Ovviamente, ogni permutazione di una base ortogonale è ancora una base ortogonale.

Esempi

1. La base canonica $e_1, \ldots, e_n$ di R^n (e ogni sua permutazione) è una base ortonormale di R^n.
2. $\frac{1}{\sqrt{2}}(1, -1)$ e $\frac{1}{\sqrt{2}}(1, 1)$ formano una base ortonormale di R^2.

Proposizione 7.3. *Se $\mathcal{H} = \{X_1, \dots, X_k\}$ è un insieme ortogonale di vettori non nulli di R^n, allora $\mathcal{H}$ è libero. In particolare, se $\mathcal{H}$ è ortonormale allora è libero.*

PROVA

Sia $c_1X_1 + \cdots + c_kX_k = O$ una cl nulla di elementi di $\mathcal{H}$.
Eseguendo il prodotto scalare con un qualsiasi $X_i \in \mathcal{H}$ abbiamo

$$X_i(c_1X_1 + \cdots + c_kX_k) = c_1X_i \cdot X_1 + \cdots + c_kX_i \cdot X_k = c_i\|X_i\|^2 = O.$$

Poiché $X_i \neq O$, dev'essere $c_i = 0$. Dunque $c_1 = c_2 = \cdots = c_k = 0$.

Se L è un sl di R^n, il sottoinsieme $L^\perp = \{X \in R^n | \forall Y \in L \quad X \perp Y\}$ è un sl. Infatti, se $X,\ X' \in L^\perp$ e $a \in R$, abbiamo $X + X' \in L^\perp$ e $aX \in L^\perp$ per (PS1): dunque $L^\perp$ per il Teorema 5.18. $L^\perp$ viene detto *sottospazio ortogonale* a L.

Proposizione 7.4. *Sia $L \subseteq R^n$ con $dimL = d$.*

a. Se $L = L(X_1, \dots, X_k)$, allora $L^\perp = \{X \in R^n \mid X_i \cdot X = 0,\ i = 1, \dots, k\}$, dunque $L^\perp$ è l'insieme delle soluzioni del sistema omogeneo ${}^tM(X_1, \dots, X_k)X = O$.

b. $dimL^\perp = n - d$.

c. Se $L = sol(S)$ con $S : AX = O$ e $A \in M_{m,n}(R)$, allora $L^\perp = \mathcal{R}(A)$.

PROVA

a. Se $X \in L^\perp$, allora $X_i \cdot X = 0$ per ogni i.

   Viceversa, se $X_i \cdot X = 0$ per ogni i, data un cl di $X_1, \dots, X_k$ abbiamo $X \cdot (c_1X_1 + \cdots + c_kX_k) = c_1(X \cdot X_1) + \cdots + c_k(X \cdot X_k) = 0$. Dunque $X \in L^\perp$.

b. Da (a) e dalla Proposizioner 5.15b deduciamo che
   $dimL^\perp = n - r({}^tM(X_1, \dots, X_k)) = n - d$.

c. Poiché $[A]_iX = O$ per ogni i, ${}^t[A]_i \in L^\perp$. Dunque $\mathcal{R}(A) \subseteq L^\perp$. Siccome per il Teorema 5.21, la Proposizione 5.15a e per (b) abbiamo $dim\mathcal{R}(A) = r(A) = n - d = dimL^\perp$, per il Corollario 5.17 $L^\perp = \mathcal{R}(A)$.

Corollario 7.5. *Se $L \subseteq R^n$, $(L^\perp)^\perp = L$.*

PROVA

Per definizione, $L \subseteq (L^{\perp})^{\perp}$. Poiché per la Proposizione 7.4 $dim(L^{\perp})^{\perp} = n - dimL^{\perp} = n - (n - dimL) = dimL$, la tesi segue dal Corollario 5.17.

Esempi

1. $\{O\}^{\perp} = R^n$ e $(R^n)^{\perp} = \{O\}$.
2. In R^2, se L è un sl con equazione $ax + by = 0$ (retta per O), allora $L^{\perp} = L((a, b))$ e dunque ha equazione $bx - ay = 0$ (retta ortogonale in O a L).

   In R^3, se $L = L((a, b, c))$ con $(a, b, c) \neq O$, $L^{\perp}$ ha equazione $ax+by+cz = 0$: dunque considerando L come retta per O, la precedente è l'equazione del piano ortogonale a tale retta nell'origine.
3. Se $L \subset R^4$ ha equazioni

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ -2x_1 - x_2 + x_3 - x_4 = 0 \end{cases}$$

   una base di L è $\{(2, -3, 1, 0), (0, -1, 0, 1)\}$. Dunque per la Proposizione 7.5, $L^{\perp} = L((1, 1, 1, 1), (-2, -1, 1, -1))$ e $L^{\perp}$ ha equazioni

$$\begin{cases} 2x_1 - 3x_2 + x_3 = 0 \\ -x_2 + x_4 = 0 \end{cases}$$

Enunciamo ora il teorema di esistenza delle basi ortonormali.

Teorema 7.6. *Sia L un sl di R^n con $dimL = d$. Se $\mathcal{H} = \{X_1, \ldots, X_k\} \subset L$ è un insieme ortonormale, allora esistono $X_{k+1}, \ldots, X_d$ vettori di L tali che $\{X_1, \ldots, X_d\}$ è una base di L.*
In particolare, L ammette una base ortonormale.

Non proviamo il Teorema 7.6 ma forniamo un metodo per determinare una base ortonormale di un sl $L \subseteq R^n$. Sia $L = sol(S)$, con $S : AX = O$, $A \in M_{m,n}$, e sia $dimL = d > 0$. Se Y_1 è un vettore non nullo di L, consideriamo il sistema

$$S' : \begin{cases} AX = O \\ {}^tY_1 X = 0 \end{cases}$$

Per la Proposizione 7.4 c e il Corollario 7.5 , $Y_1 \in \mathcal{R}(A)^\perp$. Dunque $Y_1 \notin \mathcal{R}(A)$, $r(M_{S'}) = r(A) + 1$ (Teorema 5.21) e $dim(sol(S')) = d - 1$ (Proposizione 5.15a).
Se $d = 1$, allora $\{\nu(Y_1)\}$ è una base ortonormale di L, altrimenti $dim(sol(S')) > 0$ e possiamo scegliere una soluzione non nulla Y_2 di S': allora $Y_2 \perp Y_1$ e $Y_2 \in L$.
Se $d > 2$, ripetendo il procedimento precedente con $L' = sol(S')$ al posto di L e Y_2 al posto di Y_1, troviamo un vettore $Y_3 \neq O$ di L ortogonale a Y_1 e Y_2.
In generale, con d passi otterremo una base ortogonale $\{Y_1, \ldots, Y_d\}$ di L e, ponendo $X_i = \nu(Y_i)$, avremo una base ortonormale.
Il caso $L = R^n$ è più semplice, in quanto non compare il sistema iniziale S e si parte dall'equazione ${}^tY_1 X = O$, con $Y_1 \neq O$ vettore di R^n.

Esempi

1. Sia $L \subset R^4$ di equazione $S : x_1 + x_2 - x_3 - x_4 = 0$ e sia $Y_1 = (1, 1, 1, 1) \in L$. Il sistema
$$S' : \begin{cases} x_1 + x_2 - x_3 - x_4 = 0 \\ x_1 + x_2 + x_3 + x_4 = 0 \end{cases} \qquad \text{è equivalente a} \qquad \begin{cases} x_1 + x_2 = 0 \\ x_3 + x_4 = 0 \end{cases}$$
Allora $Y_2 = (1, -1, 1, -1)$ è una soluzione di S', quindi $Y_2 \perp Y_1$ e $Y_2 \in L$. Se
$$S'' : \begin{cases} x_1 + x_2 = 0 \\ x_3 + x_4 = 0 \\ x_1 - x_2 + x_3 - x_4 = 0 \end{cases}$$
allora $Y_3 = (1, -1, -1, 1)$ è una soluzione di S'', quindi $Y_3 \perp Y_1,\ Y_2$ e $Y_3 \in L$. Dunque $\{Y_1, Y_2, Y_3\}$ e $\{\frac{1}{\sqrt{4}} Y_1, \frac{1}{\sqrt{4}} Y_2, \frac{1}{\sqrt{4}} Y_3\}$ sono rispettivamente una base ortogonale e una base ortonormale di L.
2. Sia $Y_1 = (1, 0, -1) \in R^3$. I vettori ortogonali a $L(Y_1)$ sono le soluzioni di $x_1 - x_3 = 0$. Scelto quindi $Y_2 = (1, 0, 1) \perp Y_1$, $L(Y_1, Y_2)^\perp$ ha equazioni
$$\begin{cases} x_1 - x_3 = 0 \\ x_1 + x_3 = 0 \end{cases} .$$
Dunque, ponendo $Y_3 = (0, 1, 0) = e_3$, abbiamo che Y_1, Y_2, Y_3 e $\{\frac{1}{\sqrt{2}} Y_1, \frac{1}{\sqrt{2}} Y_2, Y_3\}$ sono rispettivamente una base ortogonale e una base ortonormale di R^3.

7.3. Matrici ortogonali e matrici simmetriche

Sia $\mathcal{B} = \{X_1, \ldots, X_n\}$ una base ortonormale di R^n e sia $M = M_{\mathcal{C}}^{\mathcal{B}}$. Siccome $M = M(X_1, \ldots, X_n)$ ha per colonne i vettori di $\mathcal{B}$ e ${}^tM$ ha tali vettori come righe, effettuando il prodotto righe per colonne ${}^tMM$ abbiamo che

$$[\,{}^tMM]_{i,j} = {}^tX_iX_j = X_i \cdot X_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} .$$

Pertanto ${}^tMM = M\,{}^tM = I_n$ e M è invertibile con $M^{-1} = {}^tM$.
Inversamente, se una matrice $M \in M_n(R)$ è tale che ${}^tMM = I_n$, dal calcolo precedente e dal fatto che $M\,{}^tM = I_n$, abbiamo che sia le righe che le colonne formano basi ortonormali di R^n. Diamo ora la seguente definizione

Definizione 7.7. *Una matrice $A \in M_n(R)$ si dice ortogonale se ${}^tAA = I_n$ o, equivalentemente, se $A \in GL_n(R)$ e $A^{-1} = {}^tA$.*

In base alle precedenti considerazioni e al fatto evidente che se A è ortogonale anche ${}^tA = A^{-1}$ lo è, vale

Proposizione 7.8. *Se $A \in M_n$ è ortogonale, le righe e le colonne di A formano basi ortonormali di R^n.*
Viceversa, se le righe o le colonne di A formano una base ortonormale di R^n, allora A è ortogonale.

Indichiamo il sottoinsieme delle matrici ortogonali di M_n con $O(n)$. È immediato che $I_n, -I_n \in O(n)$. Inoltre, dalla definizione, vediamo subito che se $A \in O(n)$ e $B \in O(n)$, allora $AB \in O(n)$: infatti $AB\,{}^t(AB) = AB\,{}^tB\,{}^tA = I_n$. Pertanto

il prodotto di matrici ortogonali è una matrice ortogonale.

Invece né la somma né il prodotto per uno scalare sono in generale matrici ortogonali: $I_n + (-I_n) = O_n$ e $2I_n \notin O(n)$.

La seguente proposizione riassume alcune proprietà fondamentali delle matrici ortogonali.

Proposizione 7.9. *Se $A \in O(n)$ e se $X,\ Y \in R^n$, allora*

a. $D(A) = \pm 1$.

b. $AX \cdot AY = X \cdot Y$.

c. $d(AX, AY) = d(X, Y)$.

PROVA

a. Per definizione $D(A)^2 = 1$.

b. $AX \cdot AY = {}^t(AX)AY = {}^tX\,{}^tAAY = X \cdot Y$.

c. Conseguenza di (b) in quanto $\|AX - AY\|^2 = A(X - Y) \cdot A(X - Y)$.

Il punto (c) della Proposizione 7.9 dice alle matrici ortogonali sono associate applicazioni di R^n in sé che conservano le distanze (*isometrie*).

Esempi

1. In base alla Proposizione 7.8 e all'esempio precedente (parte (2)), la matrice

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

è ortogonale.

2. Sia $A \in M_n(R)$ una matrice diagonale. Poichè $A = {}^tA$, $A \in O(n)$ se e solo se $A^2 = I_n$. Ponendo $[A]_{i,i} = a_i$, abbiamo

$$[A^2]_{i,j} = \begin{cases} a_i^2 & i = j \\ 0 & i \neq j \end{cases}.$$

Dunque $A \in O(n)$ se e solo se $a_i = \pm 1$ per ogni i.

3. Analizziamo in dettaglio come si può descrivere $O(2)$.
Se $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in O(2)$, abbiamo

$${}^tAA = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix} = I_n$$

quindi

$$\begin{cases} a^2 + c^2 = 1 \\ b^2 + d^2 = 1 \\ ab + cd = 0 \end{cases}.$$

Pertanto esistono angoli $\theta,\ \phi \in [0 \quad 2\pi)$ tali che $a = cos\theta$, $c = sen\theta$, $b = cos\phi$, $d = sen\phi$ e $cos\theta cos\phi + sen\theta sen\phi = cos(\theta - \phi) = 0$.
Allora dev'essere $\phi = \theta - \frac{\pi}{2}$ o $\phi = \theta - \frac{3}{2}\pi$ e, fissato θ, otteniamo due matrici ortogonali

$$S_\theta = \begin{pmatrix} cos\theta & sen\theta \\ sen\theta & -cos\theta \end{pmatrix}$$

e

$$R_\theta = \begin{pmatrix} cos\theta & -sen\theta \\ sen\theta & cos\theta \end{pmatrix}.$$

Osserviamo che $S_\theta = R_\theta S_0$, essendo $S_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ e che $R_0 = I$, $R_\pi = -I$, $R_{\pi/2} = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $S_\pi = -S_0$, $S_{\pi/2} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$.
Se consideriamo ora R_θ come applicazione di R^2 in sé (identificandola con f_{R_θ}) e utilizziamo le coordinate polari nel piano $x = rcos\alpha,\ y = rsen\alpha$, abbiamo

$$R_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} xcos\theta - ysen\theta \\ xsen\theta + ycos\theta \end{pmatrix} = \begin{pmatrix} r(cos\alpha cos\theta - sen\alpha sen\theta) \\ r(cos\alpha sen\theta + sen\alpha cos\theta) \end{pmatrix} = \begin{pmatrix} rcos(\alpha + \theta) \\ rsen(\alpha + \theta) \end{pmatrix}.$$

Quindi R_θ corrisponde alla rotazione di angolo θ.
Gli autovalori di R_θ sono $cos\theta \pm isen\theta = e^{\pm i\theta}$, dunque R_θ è diagonalizzabile su C e, se $\theta \neq 0,\ \pi$, non su R.
Consideriamo ora le S_θ. S_0 e S_π corrispondono evidentemente alle simmetrie rispetto agli assi delle ascisse e delle ordinate.
In generale, gli autovalori di S_θ sono 1 e -1 per ogni θ, dunque sono tutte diagonalizzabili su R e $S_\theta \simeq S_0$.
Fissato θ, l'autospazio $V_{S_\theta}(1)$ di S_θ è retta r di equazione $(cos\theta - 1)x + sen\theta y = 0$.
Tale retta è formata dai punti fissi dell'isometria S_θ, che quindi rappresenta la simmetria rispetto a r.

7.4. Matrici simmetriche e Teorema Spettrale

Definizione 7.10. *Se $A,\ B \in M_n(R)$, allora A si dice congruente a B e si scrive $A \equiv B$ se esiste $N \in O(n)$ tale che $B = N^{-1}AN = {}^tNAN$.*
Dunque due matrici congruenti sono simili mentre il viceversa non è vero.

Il Teorema 6.8 e la Proposizione 7.9 implicano la

Proposizione 7.11. *$A \in M_n(R)$ è congruente a una matrice diagonale se e solo se esiste una base ortonormale di R^n formata da autovettori di A.*

Studieremo ora le matrici simmetriche reali, allo scopo di provare che sono sempre congruenti a una matrice diagonale. Diamo prima qualche esempio.

Esempio

1. Se $A = aI_n$ è una matrice scalare, qualsiasi base ortonormale di R^n è una base di autovettori per A.
2. Consideriamo il caso delle matrici simmetriche 2×2. Se $A = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$, abbiamo $p_A(t) = t^2 - (a+c)t + ac - b^2$.
   Quindi gli autovalori di A sono i numeri reali $\frac{1}{2}(a + c \pm \sqrt{(a-c)^2 + 4b^2})$.
   Tali autovalori saranno coincidenti se e solo se $a = c$ e $b = 0$, cioè se $A = aI$, pertanto A è sempre diagonalizzabile.
   Se ora consideriamo il caso degli autovalori distinti, calcolando esplicitamente troviamo un base di autovettori data da $\{X_1 = (-2b, a - c - \sqrt{\Delta}), X_2 = (-2b, a - c + \sqrt{\Delta})$, dove $\Delta = (a-c)^2 + 4b^2$. Poiché $X_1 \perp X_2$, passando ai normalizzati otteniamo una base ortonormale di R^2 di autovettori di A.

Lemma 7.12. *Se $A \in M_n(R)$ è simmetrica e $B \equiv A$, anche B è simmetrica.*

PROVA

Se $B = {}^tNAN$ con $N \in O(n)$, ${}^tB = {}^t({}^tNAN) = {}^t(AN)N = {}^tNAN = B$.

Lemma 7.13. *Se $A \in M_n(R)$ è simmetrica, i suoi autovalori sono tutti reali.*

PROVA

Supponiamo che $\lambda \in C$ sia un autovalore di A (come matrice di $M_n(C)$) e che $X \in C^n$ sia un autovettore relativo a λ.
Se $[X]_i = x_i \in C$, il vettore coniugato $\overline{X}$ di X sarà dato da $[\overline{X}]_i = \overline{x}_i$ e la quantità $\xi = {}^tX\overline{X} = {}^t\overline{X}X = \sum_{i=1}^{n} |x_i|^2$ è > 0 poiché $X \neq O$.
Evidentemente $AX = \lambda X$ implica $\overline{AX} = A\overline{X} = \overline{\lambda X}$ in quanto A è reale, dunque $\overline{X}$ è un autovettore di A con autovalore $\overline{\lambda}$. Ora

$$\lambda\xi = {}^t\overline{X}AX = {}^t({}^t\overline{X}AX) = {}^tXA\overline{X} = \overline{\lambda}\xi$$

da cui $\lambda = \overline{\lambda}$, cioè $\lambda \in R$.

Lemma 7.14. *Se $A \in M_n(R)$ è simmetrica e se $X_1,\ X_2 \in R^n$ sono autovettori di A relativi a autovalori $\lambda_1,\ \lambda_2$ distinti, allora $X_1 \perp X_2$.*

PROVA

$$\lambda_2 X_1 \cdot X_2 = {}^t X_1 A X_2 = {}^t({}^t X_1 A X_2) = {}^t X_2 A X_1 = \lambda_1 X_1 \cdot X_2.$$

Poiché $\lambda_1 \neq \lambda_2$, dev'essere $X_1 \cdot X_2 = 0$.

Teorema Spettrale. *Se $A \in M_n(R)$ è simmetrica, allora A è congruente a una matrice diagonale.*

PROVA

Dimostriamo il teorema per induzione su n. Per $n = 1$ è ovvio. Supponiamo allora che $n > 1$ e che la tesi sia vera per $n - 1$.

Per il Lemma 7.13 , sia $\lambda_1 \in R$ un autovalore di A con autovettore X_1; passando al normalizzato possiamo assumere che X_1 sia un versore. Applicando il Teorema 7.5 , estendiamo X_1 a una base ortonormale $\mathcal{B} = \{X_1, \dots, X_n\}$ di R^n.

Allora $N_1 = M(X_1, \dots, X_n) \in O(n)$ e, se $B = {}^t N_1 A N_1$, abbiamo che $[B]^1 = \lambda_1 e_1$: infatti, per la Proposizione 6.5, $[B]_{i,1} = [AX_1]_{\mathcal{B}}$ e $AX_1 = \lambda_1 X_1$.

Per il Lemma 7.12, B è simmetrica, quindi ha una forma a blocchi

$$B = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix}$$

dove $A' \in M_{n-1}(R)$ è simmetrica. Per ipotesi induttiva, esiste $N' \in O(n-1)$ tale che ${}^t N' A' N'$ è diagonale. Consideriamo allora la matrice a blocchi

$$N_2 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & N' & \\ 0 & & & \end{pmatrix}.$$

Si verifica facilmente con la regola del prodotto a blocchi che $N_2 \in O(n)$ e che

$${}^t N_2 B N_2 = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & {}^t N' A' N' & \\ 0 & & & \end{pmatrix}$$

è diagonale. Dunque, posto $N = N_1 N_2$, $N \in O(n)$ e ${}^t N A N$ è diagonale.

Esempio

Sia

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

Allora $p_A(t) = -(t+2)(t^2+t-2) = -(t+2)^2(t-1)$, quindi A ha autovalori -2 e 1 con $m(-2) = 2$ e $m(1) = 1$.
Poiché

$$A + 2I = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{e} \quad A - I = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

abbiamo che $V_A(-2)$ ha equazione $x + y + z = 0$ e $V_A(1) = L((1,1,1)) = (V_A(-2))^\perp$.
Con i metodi introdotti in precedenza si determina una base ortogonale di $V_A(-2)$: per esempio $(1,0,-1), (1,-2,1)$. Quindi

$$\{\frac{1}{\sqrt{3}}(1,1,1), \quad \frac{1}{\sqrt{2}}(1,0,-1), \quad \frac{1}{\sqrt{6}}(1,-2,1)\}$$

è una base ortonormale di R^3 formata da autovettori di A.

Questo libro fa parte dei materiali prodotti nel progetto didattica in rete del Politecnico di Torino sostenuto dal contributo della Fondazione Cassa di Risparmio di Vercelli

ISBN 88-87503-39-7

€ 5