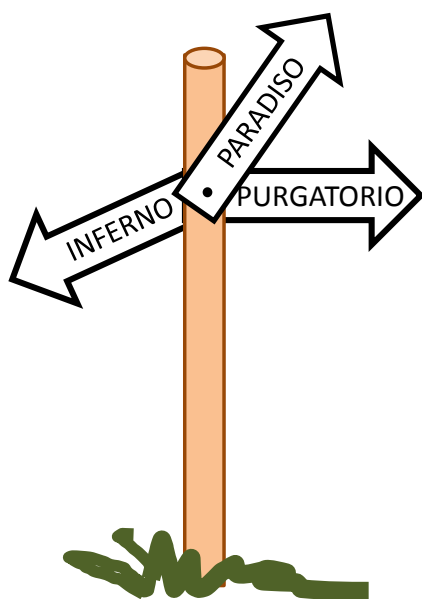


INTRODUZIONE ALLA MATEMATICA DISCRETA – PARTE II: ALGEBRA LINEARE

Giuseppe Lancia

Dipartimento di Matematica e Informatica
Università di Udine



Contents

1	Le n-ple di numeri reali	1
1.1	I punti dello spazio	1
1.2	Operazioni sulle n -ple	2
1.2.1	Somma di n -ple.	2
1.2.2	Prodotto di n -ple e scalari.	4
1.2.3	Proprietà matematiche delle operazioni.	5
1.3	Rette in \mathbb{R}^2	9
2	Matrici	11
2.1	Operazioni su matrici	14
2.2	Prodotto matriciale	15
2.3	Sottomatrici e scomposizione in blocchi	19
2.4	Matrice inversa	23
3	Spazi vettoriali	27
3.1	Dipendenza e indipendenza lineare	35
3.2	Basi e dimensione	38
3.2.1	Sistemi di coordinate	42
3.3	Somme di sottospazi e somme dirette	43
4	Applicazioni lineari	47

4.1	Introduzione	47
4.1.1	Teorema della nullità + rango	51
4.1.2	Isomorfismo tra spazi della stessa dimensione	52
4.2	Applicazioni lineari da un punto di vista geometrico	53
5	Matrici e applicazioni lineari	59
5.1	Prodotto di matrici e composizione di applicazioni	62
5.2	Cambiamenti di base e nuove coordinate	62
5.3	Matrici associate a basi diverse	66
6	Sistemi lineari	67
6.1	Sistemi lineari e matrici	68
6.2	Il metodo di Gauss per i sistemi lineari	71
6.2.1	Riduzione a scalini di un sistema	72
6.3	Il rango di una matrice	77
7	Determinanti	79
7.1	Esistenza e unicità del determinante	81
7.2	Determinante di matrici quadrate	82
7.3	Note da mettere a posto	85
7.3.1	Determinante di applicazioni lineari	85
7.3.2	Definizione ricorsiva	85
8	Autovalori e autovettori	89
8.1	Calcolo di autovalori e autovettori	93
8.1.1	Note da mettere a posto	94
9	Geometria dello spazio \mathbb{R}^n	97
9.1	Prodotto scalare e ortogonalità	97
9.2	Equazioni parametriche di rette e piani	100

9.3	Proiezioni e distanze tra insiemi	100
-----	---	-----

Chapter 1

Le n -ple di numeri reali

1.1 I punti dello spazio

Noi tutti abbiamo familiarità con la nozione di dimensione dello spazio in cui viviamo. Diciamo che il nostro spazio è tridimensionale in quanto gli oggetti nello spazio ci appaiono possedere tre caratteristiche, che possiamo chiamare altezza, larghezza e profondità. Queste tre dimensioni sono percepite a livello sensoriale. La nostra vista ci permette di riconoscere oggetti come vicini, lontani, lunghi, larghi, ecc. Anche chiudendo gli occhi possiamo, tramite il tatto, apprezzare le forme degli oggetti, che si sviluppano in tre dimensioni. Inoltre, muovendoci liberamente, abbiamo la sensazione di poter andare in alto, in basso, a destra, sinistra, avanti e indietro.

Questa percezione "fisica" dello spazio a tre dimensioni ci permette anche di comprendere spazi di dimensione inferiore, ossia spazi bidimensionali e monodimensionali. Ad esempio possiamo, con la fantasia, immaginare un mondo in cui oggetti, animali, persone, sono delle figure magnetiche (tutte dello stesso spessore, infinitesimo) che si muovono su una grande lastra metallica. Un tale mondo sarebbe bidimensionale. Un carrello montato su una rotaia rettilinea si muove in uno spazio che percepiamo come mono-dimensionale, in quanto il carrello non può sganciarsi dalla rotaia e virare verso destra o sinistra, nè verso il basso o verso l'alto. La sua esistenza è confinata nella monotonia di una linea. Quello che ci riesce difficile, invece, è riuscire a figurare nella nostra mente spazi con un numero di dimensioni maggiori di 3 in quanto, istintivamente, pensiamo alle dimensioni come caratteristiche "geometriche", e non sapremmo da che parte muoverci per andare verso la quarta dimensione. Se però, con uno sforzo di astrazione, introduciamo delle dimensioni non necessariamente geometriche nel nostro concetto di spazio allora possiamo riuscire a concepire (i.e., a farci delle immagini mentali) spazi di dimensioni superiori a tre. Ad esempio, il tempo, o la temperatura, potrebbero rappresentare ulteriori dimensioni.

Tornando all'usuale spazio tridimensionale in cui ci muoviamo, per individuarne un punto preciso utilizziamo 3 valori, che indicano come raggiungere il punto a partire da un punto di riferimento (che possiamo chiamare *origine* o *zero*). In particolare, se questi valori sono x , y e z , per raggiungere il punto dobbiamo percorrere un tratto lungo $|x|$ in una certa direzione, poi girare di 90 gradi a sinistra (a destra se $y < 0$) e percorrere un tratto di lunghezza $|y|$, e infine muoverci in verticale verso l'alto (verso il basso se $z < 0$) per un tratto di lunghezza $|z|$.

Gli spostamenti appena descritti avvengono lungo tre direzioni (tre rette) a due a due perpendicolari tra loro. Il sistema che esprime ogni punto tramite tre valori (detti *coordinate* del punto) viene detto un *sistema di assi cartesiani*, e le tre rette vengono chiamate *assi* del sistema (solitamente asse x , asse y e asse z). Nel piano, il sistema di assi cartesiani utilizza solo due assi perpendicolari, ossia l'asse x (detto anche asse delle *ascisse*) e l'asse y (o asse delle *ordinate*).

Un punto P del piano è individuato da due valori e quindi corrisponde a (di fatto, è) una coppia (x, y) con $x, y \in \mathbb{R}$. Nello spazio tridimensionale, un punto è una terna di valori (x, y, z) con $x, y, z \in \mathbb{R}$. Per quanto dal punto di vista geometrico i nostri sensi non siano in grado di percepire spazi con un numero di dimensioni maggiore di 3, niente vieta di considerare un sistema in cui i punti sono caratterizzati da più di tre coordinate. Vogliamo quindi generalizzare il concetto di spazio tridimensionale a quello di spazio n -dimensionale, uno spazio con un sistema di n coordinate e in cui ogni punto è caratterizzato da una n -pla di valori reali (x_1, \dots, x_n) .

Definizione 1: [Punti e coordinate] Chiamiamo *spazio (numerico) reale n -dimensionale* l'insieme

$$\mathbb{R}^n = \{(x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{R}\}.$$

Gli elementi di \mathbb{R}^n sono detti *n -ple* o *punti*. Le componenti x_i sono dette le *coordinate* del punto (x_1, \dots, x_n) .

Sottolineiamo (per quanto ciò dovrebbe comunque risultare ovvio) che perchè due n -ple risultino diverse non è necessario che siano diverse in ciascuna componente, ma semplicemente che esista almeno una componente in cui le due n -ple differiscono. Quindi

$$(x_1, \dots, x_n) \neq (y_1, \dots, y_n) \iff \bigvee_{i=1}^n (x_i \neq y_i)$$

In particolare $(x_1, \dots, x_n) = (0, \dots, 0) \iff x_i = 0 \forall i$.

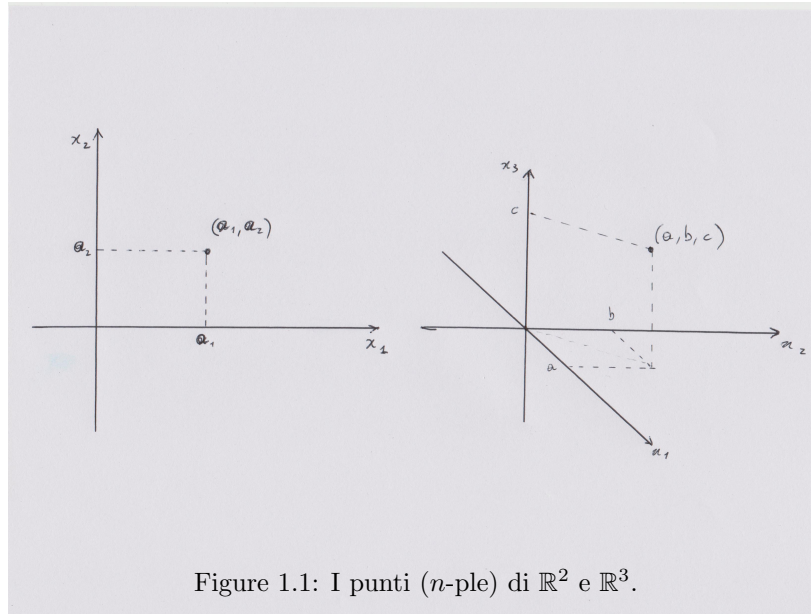
Lo spazio reale di dimensione due è detto anche il piano cartesiano, mentre lo spazio tridimensionale in cui ci muoviamo è rappresentato da \mathbb{R}^3 . Anzichè usare le lettere x, y, z per denotare le coordinate, in conformità al caso più generale possiamo indicare i punti del piano come coppie (x_1, x_2) di reali, e quelli dello spazio come terne (x_1, x_2, x_3) (si veda la fig.1.1).

1.2 Operazioni sulle n -ple

1.2.1 Somma di n -ple.

Un ragazzino deve fare la spesa per la mamma. La spesa tipicamente riguarda determinate categorie di beni, espresse nelle relative unità di misura. Ad es. c'è il latte (in litri), le uova, il prosciutto (in hg.), biscotti (in pacchi), ecc. In modo astratto, possiamo pensare alla spesa come ad un'opportuna n -pla, dove n è il numero complessivo di tipologie di beni che possono venir richiesti, ed ogni componente è relativa ad uno specifico bene. Ad esempio, supponiamo che la spesa possa riguardare 10 tipologie di beni, di cui la prima è il latte, la seconda le uova, la terza il prosciutto, ecc. La seguente è una possibile spesa:

$$(2, 0, 1.5, 0, 1, 4, 0, 0, 3.5, 2)$$

Figure 1.1: I punti (n -ple) di \mathbb{R}^2 e \mathbb{R}^3 .

In questo caso, il ragazzo deve acquistare 2 litri di latte, nessun uovo, un etto e mezzo di prosciutto, ecc. La spesa richiesta dalla mamma è quindi una n -pla

$$x = (x_1, x_2, \dots, x_n)$$

in cui ogni x_i rappresenta la quantità del bene i da acquistare. Ora, supponiamo che la zia venga a sapere che il ragazzino sta per andare a fare la spesa, e che gli chieda la cortesia di farla anche a lei. Qual'è la quantità complessiva di ognuno dei beni che il ragazzo dovrà acquistare? Se $y = (y_1, y_2, \dots, y_n)$ rappresenta la spesa richiesta dalla zia è chiaro che per ogni bene i -mo il ragazzo dovrà acquistare una quantità complessiva pari a

$$x_i + y_i$$

Possiamo allora definire una nuova spesa "virtuale", che chiameremo z , e che è una n -pla rappresentante la spesa complessiva (che possiamo anche chiamare *somma*) delle spese x e y . Scriviamo $z = x + y$, e quindi

$$(z_1, \dots, z_n) = (x_1, \dots, x_n) + (y_1, \dots, y_n)$$

se $z_i = x_i + y_i$ per ogni $i = 1, \dots, n$.

Esempio. Un imprenditore possiede 3 negozi in altrettante città. Non tutti i negozi sono in attivo, e la loro redditività dipende da vari fattori (ad esempio, un negozio in una località di mare è più redditizio nei mesi estivi che negli altri). Considerando i 12 mesi dell'anno come altrettante coordinate, il rendimento nell'anno di ogni negozio può essere rappresentato da una n -pla di uno spazio 12-dimensionale:

$$x = (x_1, \dots, x_{12})$$

dove x_1 è il guadagno (se > 0) o la perdita (se < 0) del negozio nel mese di Gennaio, e così via fino a x_{12} che è il guadagno (o perdita) del mese di Dicembre. Ai tre negozi dell'imprenditore corrispondono altrettante

n -ple, siano x , y e z . Ad esempio

$$\begin{aligned} x &= (1.5, -2.0, 1.8, 3.0, -0.5, -0.7, 0.8, 2.1, -0.3, 1.2, 4.0, 1.4) \\ y &= (1.2, -0.5, 3.8, 1.2, 0.5, -0.8, 0.9, 2.3, -0.5, -1.0, 2.1, 1.7) \\ z &= (1.1, 0.9, -1.4, 0.9, -0.5, -1.1, 0.6, 3.1, -1.4, 2.2, 1.8, -0.4) \end{aligned}$$

(dove i valori sono espressi in migliaia di euro). All'imprenditore interessa conoscere il proprio guadagno (o perdita) complessivo relativo ad ogni mese. Tale operazione corrisponde a *sommare* le tre n -ple. Ad esempio, in Marzo il guadagno complessivo è stato

$$x_3 + y_3 + z_3 = 1.8 + 3.8 - 1.4 = 4.2$$

ossia 4200 euro. Quindi, detta u la n -pla relativa alla rendita complessiva dei tre negozi, avremo $u = x + y + z$ se $u_i = x_i + y_i + z_i$ per ogni $i = 1, \dots, 12$.

1.2.2 Prodotto di n -ple e scalari.

Ogni ricetta specifica gli ingredienti necessari per la preparazione di un piatto e le relative quantità, facendo riferimento a un numero ben preciso di persone da sfamare. Ad esempio, per preparare una pasta all'amatriciana per 4 persone sono necessari

- (1) 400 g. di spaghetti
- (2) 250 g. di guanciale
- (3) 500 g. di pomodoro fresco
- (4) 150 g. di pecorino
- (5) 1 cucchiaio di strutto
- (6) 1 peperoncino rosso
- (7) 1 pugno di sale grosso

Facendo corrispondere ai singoli ingredienti le componenti di un'opportuna n -pla, questa ricetta può essere rappresentata dalla seguente 7-upla:

$$(400, 250, 500, 150, 1, 1, 1)$$

Notiamo come con 7-uple diverse sia possibile rappresentare altre ricette (la maggior parte delle quali darebbe luogo a preparati immangiabili). Ad esempio, una pasta all'arrabbiata per 4 persone è codificata dalla 7-upla

$$(400, 0, 500, 0, 0, 5, 1)$$

Supponiamo ora di voler preparare gli spaghetti all'amatriciana, ma al nostro pranzo saranno presenti 10 persone anziché 4. Come dobbiamo fare? Detta

$$x = (x_1, \dots, x_7)$$

la ricetta corrispondente alle dosi per 4 persone, e sapendo che $10 = 2.5 \times 4$, la quantità di ogni ingrediente andrà moltiplicata per 2.5. Chiamando $y = (y_1, \dots, y_7)$ la nuova ricetta, avremo quindi

$$y = (y_1, y_2, \dots, y_7) = (2.5x_1, 2.5x_2, \dots, 2.5x_7)$$

Moltiplicare il valore di ogni componente per lo stesso coefficiente λ (nel nostro caso $\lambda = 2.5$) può essere visto come un'operazione che moltiplica un'intera n -pla per il numero λ , e quindi

$$y = \lambda x$$

se $y_i = \lambda x_i$ per ogni $i = 1, \dots, n$. Nel contesto dell'algebra lineare, le grandezze numeriche vengono anche dette gli *scalari*, in quanto ognuna di esse può essere vista come un valore su una scala graduata di uno strumento di misura. Quindi il numero λ è uno scalare e il prodotto appena definito corrisponde a moltiplicare una n -pla per uno scalare.

1.2.3 Proprietà matematiche delle operazioni.

Diamo una definizione formale delle operazioni che abbiamo appena introdotto tramite esempi. Tali operazioni permettono di sommare due n -ple e di moltiplicare una n -pla per un numero reale, ottenendo in entrambi i casi come risultato una nuova n -pla. L'operazione di addizione fra n -ple è definita al seguente modo: dati due elementi di \mathbb{R}^n

$$A = (a_1, \dots, a_n)$$

e

$$B = (b_1, \dots, b_n)$$

definiamo la loro somma come la n -pla $C = A + B$ le cui coordinate sono

$$C = (a_1 + b_1, \dots, a_n + b_n). \quad (1.1)$$

Ad esempio, se $A = (3, -1, 2)$ e $B = (1, 0, -1)$ si ha $A + B = (4, -1, 1)$.

L'operazione che permette di moltiplicare un numero reale per una n -pla è definita come segue. Sia $\lambda \in \mathbb{R}$ e $A \in \mathbb{R}^n$. Definiamo λA come la n -pla le cui coordinate sono

$$(\lambda a_1, \dots, \lambda a_n). \quad (1.2)$$

Ad esempio, se $\lambda = 3$ e $A = (-1, 0, 2, \frac{1}{3})$ abbiamo $\lambda A = (-3, 0, 6, 1)$.

Si noti che valgono le seguenti proprietà:

1. La somma di n -ple è sia *associativa* che *commutativa*, i.e.

$$(A + B) + C = A + (B + C)$$

e

$$A + B = B + A$$

per ogni $A, B, C \in \mathbb{R}^n$.

2. Esiste un *elemento neutro* per la somma. In particolare, detta $\mathbf{0} = (0, \dots, 0)$ la n -pla con coordinate tutte nulle, si ha

$$A + \mathbf{0} = \mathbf{0} + A = A$$

per ogni $A \in \mathbb{R}^n$. La n -pla $\mathbf{0}$ viene chiamata *zero*, e non va confusa con il numero $0 \in \mathbb{R}$.

3. Per ogni n -pla A esiste un *inverso rispetto alla somma* (detto il suo *opposto*), ossia una n -pla \bar{A} tale che

$$A + \bar{A} = \mathbf{0}$$

L'opposto di A risulta essere $\bar{A} = (-1)A$.

4. Il prodotto è distributivo, sia rispetto a \mathbb{R} che a \mathbb{R}^n , i.e.

$$(\lambda + \delta)A = \lambda A + \delta A$$

e

$$\lambda(A + B) = \lambda A + \lambda B$$

per ogni $\lambda, \delta \in \mathbb{R}$ e $A, B \in \mathbb{R}^n$.

5. Esiste un elemento neutro per il prodotto. In particolare, il numero $1 \in \mathbb{R}$ è tale che

$$1 A = A$$

per ogni $A \in \mathbb{R}^n$.

6. La moltiplicazione soddisfa a

$$(\lambda\delta)A = \lambda(\delta A)$$

per ogni $\lambda, \delta \in \mathbb{R}$ e $A \in \mathbb{R}^n$.

Dimostriamo alcune di queste proprietà. La proprietà 3 si dimostra notando che

$$A + ((-1)A) = (a_1, \dots, a_n) + (-a_1, \dots, -a_n) = \mathbf{0}$$

Si noti che l'opposto di una n -pla è unico, perchè detto $B = (b_1, \dots, b_n)$ un opposto di A deve essere $a_i + b_i = 0$ per ogni i , ossia $b_i = -a_i$ per ogni i .

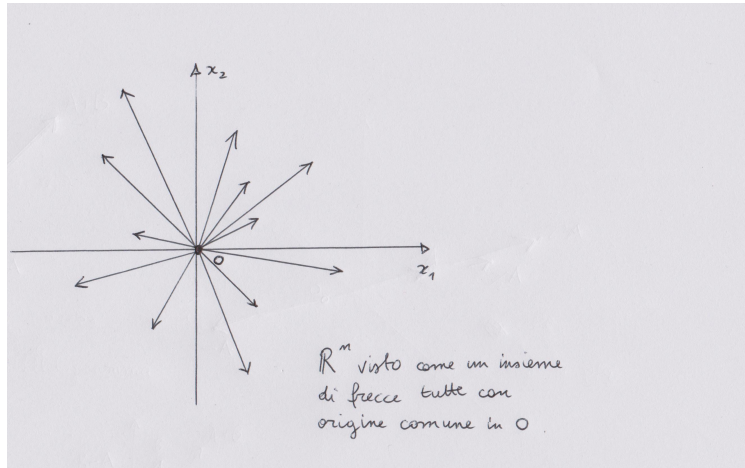
La prima delle proprietà 4 si dimostra come segue:

$$(\lambda + \delta)A = ((\lambda + \delta)a_1, \dots, (\lambda + \delta)a_n) \quad (1.3)$$

$$= (\lambda a_1 + \delta a_1, \dots, \lambda a_n + \delta a_n) \quad (1.4)$$

$$= (\lambda a_1, \dots, \lambda a_n) + (\delta a_1, \dots, \delta a_n) \quad (1.5)$$

$$= \lambda A + \delta A. \quad (1.6)$$

Figure 1.2: Le frecce (i punti) di \mathbb{R}^n .

mentre per la seconda si ha

$$\lambda(A + B) = \lambda(a_1 + b_1, \dots, a_n + b_n) \quad (1.7)$$

$$= (\lambda(a_1 + b_1), \dots, \lambda(a_n + b_n)) \quad (1.8)$$

$$= (\lambda a_1 + \lambda b_1, \dots, \lambda a_n + \lambda b_n) \quad (1.9)$$

$$= (\lambda a_1 + \dots + \lambda a_n) + (\lambda b_1 + \dots, \lambda b_n) \quad (1.10)$$

$$= \lambda A + \lambda B. \quad (1.11)$$

La dimostrazione delle altre proprietà è lasciata al lettore come esercizio.

Come osservato nell'introduzione, ogni tripla $P \in \mathbb{R}^3$ può essere vista come un punto P dello spazio tridimensionale. Chiamiamo ogni segmento orientato con coda nel punto $\mathbf{0}$ (detto anche l'*origine* di \mathbb{R}^3) e terminazione in un qualsiasi altro punto una *freccia* (Si veda Fig.1.2). Dato un punto P , resta definita la freccia che termina in P . Viceversa, ogni freccia identifica univocamente il punto del suo estremo finale. C'è quindi una corrispondenza biunivoca tra l'insieme di tutti i punti di \mathbb{R}^3 e l'insieme di tutte le frecce nello spazio (considerazioni analoghe valgono per \mathbb{R}^2 e le frecce del piano). L'uso delle frecce rende più semplice descrivere graficamente sia l'operazione di somma tra punti, tramite quella che è detta anche la *regola del parallelogramma*, che la moltiplicazione di un punto per un numero (si veda la figura 1.3).

A ogni freccia F possiamo associare in modo naturale una lunghezza, che denotiamo con $\|F\|$. Il valore $\|F\|$ è anche detto la *norma* di F . Consideriamo una freccia del piano che, partendo dall'origine $(0,0)$, termina nel punto (x,y) . In base al teorema di Pitagora abbiamo

$$\|F\| = \sqrt{x^2 + y^2}$$

Nello spazio tridimensionale, la lunghezza di una freccia che termina in (x,y,z) è

$$\|F\| = \sqrt{x^2 + y^2 + z^2}$$

Generalizziamo ora il concetto di norma alle n -ple di \mathbb{R}^n .

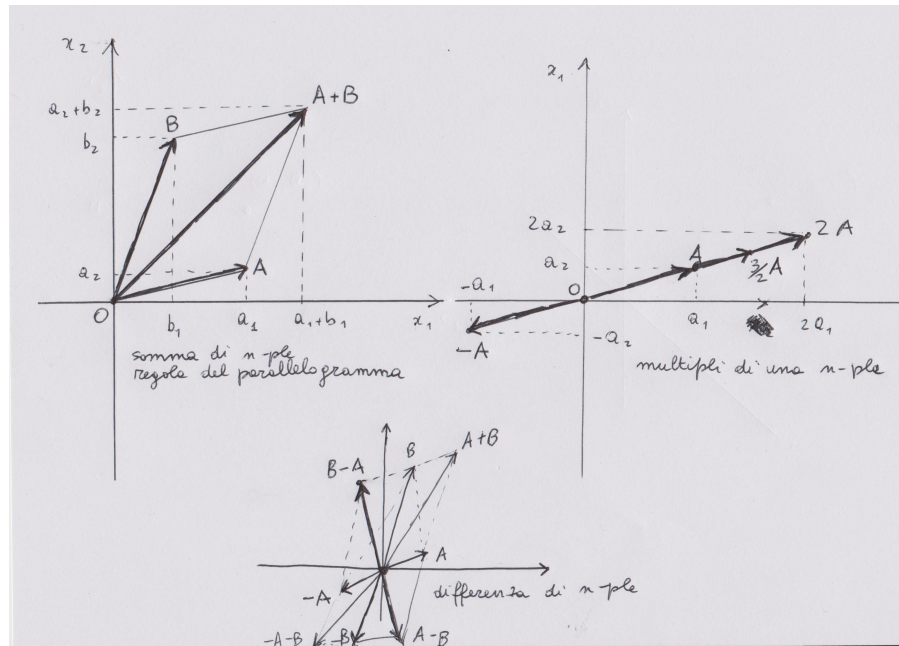


Figure 1.3: Somma di frecce e prodotto per scalari

Definizione 2: [Norma] Sia $x = (x_1, \dots, x_n)$ una generica n -ple di \mathbb{R}^n . Definiamo la norma di x , denotata come $\|x\|$, come il valore

$$\|x\| = \sqrt{x_1^2 + \dots + x_n^2}$$

Si noti che per la norma valgono le seguenti proprietà:

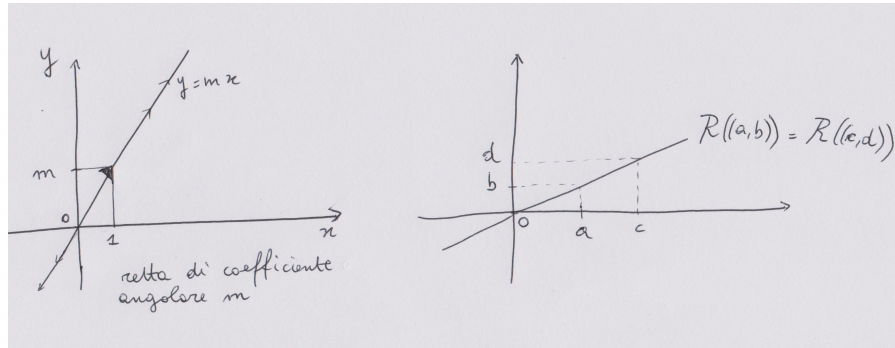
1. $\|x\| \geq 0$ per ogni $x \in \mathbb{R}^n$. Inoltre $\|x\| = 0$ se e solo se $x = \mathbf{0}$.
2. $\|\lambda x\| = |\lambda| \|x\|$ per ogni $x \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$.
3. (Disuguaglianza triangolare:) Per ogni $x, y \in \mathbb{R}^n$ si ha $\|x + y\| \leq \|x\| + \|y\|$.

La prima e la seconda proprietà sono immediate. La terza proprietà può essere dimostrata facilmente (esercizio) utilizzando la seguente disuguaglianza, di cui omettiamo la dimostrazione:

Disuguaglianza di Cauchy-Schwarz: Per ogni coppia $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ di punti in \mathbb{R}^n si ha

$$\left| \sum_{i=1}^n x_i y_i \right| \leq \|x\| \|y\|$$

-NOTA: per una F di dominio \mathbb{R}^n , anziché scrivere $F((x_1, \dots, x_n))$ scriviamo indifferentemente $F(x_1, \dots, x_n)$

Figure 1.4: rette per l'origine in \mathbb{R}^2

1.3 Rette in \mathbb{R}^2

Dalle scuole superiori siamo abituati a descrivere le rette in \mathbb{R}^2 tramite un'equazione (detta *l'equazione parametrica*) che specifica come varia l'ordinata (y) in funzione dell'ascissa (x). Se la retta passa per l'origine (consideriamo rette non parallele all'asse y) si tratta di una dipendenza *lineare*

$$y = mx$$

con m detto il *coefficiente angolare*. Se la retta non passa per l'origine l'equazione che la descrive è del tipo

$$y = mx + q$$

con $q \neq 0$. Consideriamo una retta passante per l'origine. Per $x = 1$ otteniamo $y = m$ e quindi la retta passa per il punto $(1, m)$. Ogni generico punto (x, mx) della retta è ottenuto come un multiplo di $(1, m)$ per un'opportuno scalare x . Ricordando che i punti di \mathbb{R}^2 corrispondono anche a frecce con origine nel punto $(0, 0)$, possiamo dire che la retta è individuata (o generata) dalla freccia $(1, m)$ (vedi la figura 1.4). In generale, ogni $v \in \mathbb{R}^2$ con $v = (a, b) \neq (0, 0)$ genera una retta. Dato v , definiamo *retta passante per l'origine* e v l'insieme

$$\mathcal{R}(v) = \{w \in \mathbb{R}^2 : w = \lambda v, \lambda \in \mathbb{R}\} = \{(x, y) : (x, y) = (\lambda a, \lambda b), \lambda \in \mathbb{R}\}$$

Se $a = 0$ la retta coincide con l'asse y . Se $a \neq 0$, per calcolare il coefficiente angolare della retta $\mathcal{R}(v)$ ci chiediamo per quale valore di m si ha $(1, m) \in \mathcal{R}(v)$. Deve esistere λ tale che $(1, m) = (\lambda a, \lambda b)$. Guardando la prima componente, otteniamo che $\lambda a = 1$, ossia $\lambda = 1/a$. Sostituendo nell'equazione relativa alla seconda componente si ha $m = b/a$ che è quindi il coefficiente angolare di $\mathcal{R}(v)$.

L'asse x del piano cartesiano è identificato dalla retta $\mathcal{R}((1, 0))$, mentre l'asse y è la retta $\mathcal{R}((0, 1))$.

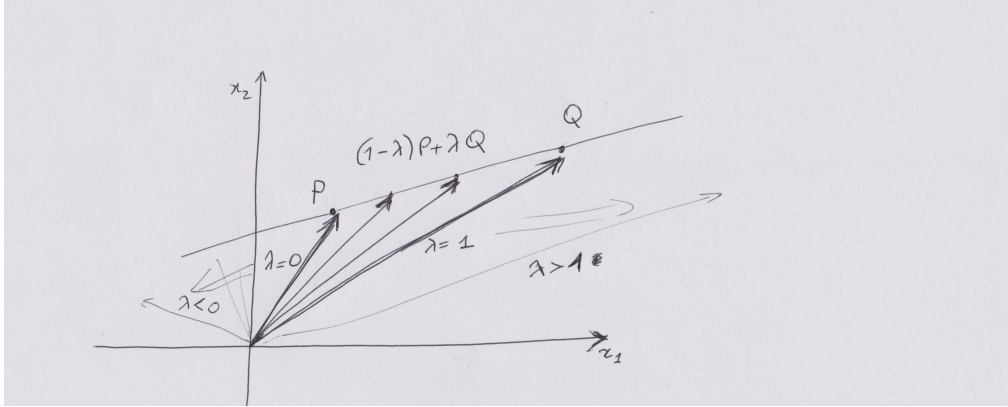
Si noti che se $w \in \mathcal{R}(v)$, $w \neq \mathbf{0}$, allora $\mathcal{R}(w) = \mathcal{R}(v)$. Infatti, esiste $\mu \neq 0$ con $w = \mu v$ e quindi

$$u \in \mathcal{R}(w) \iff \exists \lambda : u = \lambda w \iff \exists \lambda : u = \lambda \mu v \iff \exists \beta : u = \beta v \iff u \in \mathcal{R}(v).$$

Consideriamo l'insieme $\mathcal{R}(v)$ e sia $w = (c, d)$ un qualsiasi elemento di \mathbb{R}^2 . Se sommiamo w ad ogni punto di $\mathcal{R}(v)$ otteniamo l'insieme

$$\mathcal{R}(v)_w = \mathcal{R}(v) + w = \{(\lambda a + c, \lambda b + d), \lambda \in \mathbb{R}\}$$

Questo insieme descrive ancora una retta che

Figure 1.5: Retta passante per P e Q (retta affine)

1. se $w \in \mathcal{R}(v)$ coincide con $\mathcal{R}(v)$ (lo si dimostri per esercizio)
2. se $w \notin \mathcal{R}(v)$ non è passante per l'origine (altrimenti, si avrebbe $\lambda a + c = 0$, $\lambda b + d = 0$ ossia $(c, d) = -\lambda(a, b)$ contraddicendo l'ipotesi che $(c, d) \notin \mathcal{R}(v)$).

Ogni tale retta viene anche detta una *retta affine* o una *retta traslata* (di una n -pla w). In generale l'insieme di tutte le rette affini contiene anche le rette passanti per l'origine, ottenute traslando di un n -pla nulla: $\mathcal{R}(v) = \mathcal{R}(v)_{(0,0)}$. Una generica retta in \mathbb{R}^2 è quindi un insieme del tipo $\{\lambda v + w \mid \lambda \in \mathbb{R}\}$ in cui v è la *direzione* e w la *traslazione*. Due rette $\{\lambda v + w \mid \lambda \in \mathbb{R}\}$ e $\{\mu v' + w' \mid \mu \in \mathbb{R}\}$ si dicono *parallele* se esiste λ tale che $\lambda v = v'$, ossia se hanno la stessa direzione, a meno di multipli scalari.

Dati due punti $P = (p_1, p_1)$ e $Q = (q_1, q_2)$ supponiamo di voler trovare una coppia v, w tale che $\mathcal{R}(v)_w$ sia la retta passante per P e Q . Se consideriamo il punto P e sommiamo ad esso $Q - P$ arriviamo al punto Q . Ogni punto del tipo

$$P + \lambda(Q - P)$$

con $\lambda \in [0, 1]$ si trova sul tratto di linea compreso tra P e Q . Quando $\lambda = 0$ esso è il punto P , mentre per $\lambda = 1$ è il punto Q . Ogni tale punto può anche essere riscritto come

$$(1 - \lambda)P + \lambda Q$$

per $0 \leq \lambda \leq 1$, e viene detto una *combinazione convessa* di P e Q . Se consideriamo anche valori $\lambda > 1$ otteniamo i punti "a destra" di Q , e per $\lambda < 0$ quelli "a sinistra" di P (si veda la Fig.1.5). L'espressione $(1 - \lambda)P + \lambda Q$ per $\lambda \in \mathbb{R}$ viene detta una *combinazione affine* di P e Q .

Quindi la retta $\mathcal{R}(Q - P)_P = \{X : X = \lambda(Q - P) + P, \lambda \in \mathbb{R}\}$ è la retta passante per P e Q . In essa, $Q - P$ è la direzione, e P è la traslazione.

Chapter 2

Matrici

Una tabella A di numeri reali disposti su m righe e n colonne è detta una *matrice* $m \times n$ di reali. Diciamo che A ha m righe e n colonne. Denotiamo con $\mathbb{R}^{m \times n}$ l'insieme di tutte le matrici di reali con m righe e n colonne.

La generica matrice $m \times n$ è

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

che possiamo indicare anche come $A = (a_{ij})$. Se $m = n$ diciamo che A è una *matrice quadrata*. Per comodità di notazione e tipografiche, alle volte useremo indifferentemente uno tra i simboli a_{ij} , o A_{ij} , o $A[i, j]$ ad indicare la stessa cosa, i.e., *l'elemento in riga i , colonna j di A* .

Esempio. Consideriamo le seguenti matrici:

$$A = \begin{pmatrix} 2 & -\pi & 4 \\ 1 & 3 & -\sqrt{2} \end{pmatrix} \quad B = \begin{pmatrix} 0 & -\pi & 1 \\ 7 & -3 & 2 \\ 1 & 1 & 0 \end{pmatrix} \quad C = (1 \quad 0 \quad 1 \quad 1)$$

A è una matrice 2×3 , B è una matrice quadrata 3×3 , mentre C è una matrice 1×4 . ◇

Esempio. Consideriamo le città di New York, Roma, Rio de Janeiro e Stoccolma e le temperature medie in tali città nei vari mesi dell'anno. Numerando le città da 1 a 4 e i mesi da 1 a 12, possiamo rappresentare i valori di temperatura di ogni città in ogni mese con una matrice $A = (a_{ij})$ di 4 righe e 12 colonne, in cui

$$a_{ij} = \text{temperatura media nella città } i \text{ nel mese } j$$

Tale matrice (con valori in gradi Celsius) risulta la seguente:

$$\begin{pmatrix} 0.5 & 1.8 & 5.7 & 11.5 & 16.9 & 22.3 & 25.2 & 24.6 & 20.6 & 14.5 & 9.1 & 3.4 \\ 8.3 & 9.1 & 10.6 & 13.2 & 16.9 & 20.6 & 23.4 & 23.6 & 20.9 & 17.2 & 12.7 & 9.5 \\ 27.2 & 27.5 & 26.6 & 24.7 & 22.6 & 21.4 & 21.0 & 22.1 & 23.1 & 23.6 & 24.5 & 26.3 \\ -2.5 & -3.2 & 1.2 & 3.6 & 10.4 & 14.5 & 17.2 & 16.8 & 11.3 & 6.7 & 1.4 & -2.6 \end{pmatrix}$$

◇

ESERCIZIO 2.1. Consideriamo una particella che si muove casualmente tra n posizioni P_1, \dots, P_n . Ad ogni iterazione viene tirata una moneta e se esce testa la particella si sposta nella posizione successiva a quella corrente (a meno che questa non sia la P_n , nel qual caso la particella resta ferma) mentre se esce croce la particella si sposta nella posizione precedente a quella corrente (a meno che questa non sia la P_1 , nel qual caso la particella resta ferma). L'intero processo può essere descritto da una matrice quadrata A (detta *matrice di transizione*) in cui A_{ij} è la probabilità che la particella passi dalla posizione P_i alla posizione P_j . Si scriva tale matrice per $n = 4$. ◇

Definizione 3: [Vettori-riga e vettori-colonna.] Una matrice $A \in \mathbb{R}^{1 \times n}$ è detta un *vettore-riga*. Ad esempio

$$\left(3 \quad -\pi \quad \frac{1}{2} \quad 0 \quad 7 \right)$$

Una matrice $A \in \mathbb{R}^{m \times 1}$ è detta un *vettore-colonna*. Ad esempio

$$\begin{pmatrix} -3 \\ 1 \\ \sqrt{2} \\ 0 \end{pmatrix}$$

Ad ogni vettore-riga o vettore-colonna corrisponde in modo ovvio una n -pla per un opportuno n . Ad esempio le n -ple corrispondenti ai vettori di cui sopra sono $(3, -\pi, \frac{1}{2}, 0, 7)$ e $(-3, 1, \sqrt{2}, 0)$. Nel seguito, con un leggero abuso di terminologia, ci riferiremo alle volte ai vettori-riga o vettori-colonna come alle n -ple ad essi corrispondenti.

Data la matrice $A = (a_{ij}) \in \mathbb{R}^{m \times n}$, indichiamo con A_i , per $i = 1, \dots, m$, il vettore-riga corrispondente alla riga i -ma di A :

$$A_i = \left(a_{i1} \quad a_{i2} \quad \dots \quad a_{in} \right)$$

indichiamo con A^j , per $j = 1, \dots, n$, il vettore-colonna corrispondente alla colonna j -ma di A :

$$A^j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{mj} \end{pmatrix}$$

Evidenziando le righe e le colonne rispettivamente, possiamo allora pensare alla matrice A come ad un (super)vettore-colonna, le cui componenti sono le righe di A :

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \dots \\ A_m \end{pmatrix}$$

o un (super)vettore-riga,

$$A = (A^1 \quad A^2 \quad \dots \quad A^n)$$

le cui componenti sono le colonne di A .

Definizione 4: [Diagonale principale e traccia] Sia $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ una matrice quadrata. Gli elementi a_{ii} costituiscono la *diagonale* principale di A . La loro somma viene detta la *traccia* di A ,

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

Definizione 5: [Matrici diagonali e triangolari] Una matrice $A \in \mathbb{R}^{n \times n}$ si dice *diagonale* se ogni elemento che non si trova sulla diagonale principale è 0 (mentre gli elementi della diagonale principale possono essere nulli o meno). La generica matrice diagonale sarà anche indicata con

$$\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

Una particolare matrice diagonale è la matrice *identità*, denotata con I_n (o, più semplicemente con I qualora non ci sia possibilità di confusione), in cui $a_{ii} = 1$ per ogni $i = 1, \dots, n$:

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

Una matrice $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ si dice *triangolare superiore* se $a_{ij} = 0$ per ogni $1 \leq j < i \leq n$. Una matrice $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ si dice *triangolare inferiore* se $a_{ij} = 0$ per ogni $1 \leq i < j \leq n$. Una matrice si dice *triangolare* se è triangolare superiore o triangolare inferiore.

Definizione 6: [Trasposta] Sia A una matrice $A = (a_{ij}) \in \mathbb{R}^{m \times n}$. Diciamo che $B = (b_{hk}) \in \mathbb{R}^{n \times m}$ è la matrice *trasposta* di A se $b_{ij} = a_{ji}$ per ogni $1 \leq i \leq m$ e $1 \leq j \leq n$. La trasposta di A verrà anche indicata con tA .

Semplificando, possiamo dire che la trasposta di A è la matrice che si ottiene scambiando le righe di A con le sue colonne. Infatti, sia $A \in \mathbb{R}^{m \times n}$. Se $B = {}^tA$, allora $B_i = ({}^tA^i)$ per ogni $i = 1, \dots, n$, e $B^j = ({}^tA_i)$ per ogni $j = 1, \dots, m$.

Esempio. Abbiamo

$${}^t\begin{pmatrix} 2 & -1 & 4 & 0 \\ 6 & 1 & -3 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ -1 & 1 \\ 4 & 3 \\ 0 & 3 \end{pmatrix}$$

◇

Si noti che una matrice A è triangolare superiore se e solo se tA è triangolare inferiore. Inoltre una matrice A è diagonale se e solo se anche tA è diagonale.

Definizione 7: [Matrice simmetrica e antisimmetrica] Una matrice quadrata $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ si dice *simmetrica* se $a_{ij} = a_{ji}$ per ogni $1 \leq i \leq n$ e $1 \leq j \leq n$. La matrice A si dice *antisimmetrica* se $a_{ij} = -a_{ji}$ per ogni $1 \leq i \leq n$ e $1 \leq j \leq n$.

Si noti che ogni matrice diagonale è simmetrica. Inoltre, una matrice triangolare non è mai simmetrica, a meno che essa non sia anche diagonale. Infine, ogni matrice antisimmetrica ha traccia nulla.

2.1 Operazioni su matrici

Per le matrici si definiscono due operazioni principali: una *somma* di matrici, e un *prodotto di matrice per scalare*

$$(+): \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \mapsto \mathbb{R}^{m \times n}$$

$$(\cdot): \mathbb{R} \times \mathbb{R}^{m \times n} \mapsto \mathbb{R}^{m \times n}$$

Si noti che la somma è definita solo per *matrici delle stesse dimensioni* e restituisce una matrice di medesime dimensioni. Date due matrici A, B e un numero reale λ , anzichè scrivere $(+)(A, B)$ e $(\cdot)(\lambda, A)$ scriveremo la forma più familiare $A + B$ e λA .

La somma di $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ e $B = (b_{ij}) \in \mathbb{R}^{m \times n}$ è la matrice $C = (c_{ij}) \in \mathbb{R}^{m \times n}$ definita come segue:

$$c_{ij} = a_{ij} + b_{ij} \quad \forall i = 1, \dots, m, \quad j = 1, \dots, n$$

Ad esempio

$$\begin{pmatrix} 3 & 0 & -1 \\ -5 & 8 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 & -2 \\ 2 & -2 & -1 \end{pmatrix} = \begin{pmatrix} 5 & 0 & -3 \\ -3 & 6 & 0 \end{pmatrix}$$

Il prodotto tra il numero $\lambda \in \mathbb{R}$ e la matrice $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ è la matrice $B = (b_{ij}) \in \mathbb{R}^{m \times n}$ definita come segue:

$$b_{ij} = \lambda a_{ij} \quad \forall i = 1, \dots, m \quad j = 1, \dots, n$$

Ad esempio

$$-2 \begin{pmatrix} 3 & 0 & -1 \\ -5 & 8 & 1 \end{pmatrix} = \begin{pmatrix} -6 & 0 & 2 \\ 10 & -16 & -2 \end{pmatrix}$$

Per ogni $m, n \in \mathbb{N}^+$, definiamo $\mathbf{0}_{m,n}$ (o, qualora non ci sia pericolo di confusione, semplicemente $\mathbf{0}$) come la matrice $m \times n$ i cui elementi sono tutti uguali a zero. È immediato notare come $\mathbf{0}$ sia *l'elemento neutro rispetto alla somma di matrici*.

2.2 Prodotto matriciale

Date due matrici $A \in \mathbb{R}^{m \times r}$ e $B \in \mathbb{R}^{r \times n}$ resta definito il loro *prodotto*, che è una matrice $C = (c_{ij})$ di dimensioni $m \times n$, al seguente modo:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj} \quad \forall i = 1, \dots, m, \quad j = 1, \dots, n$$

Si dice anche che il prodotto è effettuato *righe per colonne*, in quanto ogni elemento di C è ottenuto prendendo una riga di A , una colonna di B e sommando i prodotti fra ogni elemento della riga di A e il corrispondente elemento della colonna di B . Sottolineiamo che per poter moltiplicare una matrice A per una matrice B , è necessario e sufficiente che *il numero di colonne di A sia uguale al numero di righe di B* . La matrice prodotto avrà tante righe quante A e tante colonne quante B .

Se C è il prodotto di A e B , scriveremo $C = AB$. Si noti che *il prodotto di matrici non è commutativo*, i.e., in generale, $AB \neq BA$. Questo è molto semplice da vedere, in quanto esistono casi dovuti alle dimensioni di A e di B in cui è possibile effettuare uno dei due prodotti ma non l'altro (ad esempio si può moltiplicare una matrice 2×3 con una 3×4 , ma non viceversa). Anche qualora rispetto alle dimensioni esistano entrambi i prodotti, può comunque risultare $AB \neq BA$. Ad esempio, siano

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

Abbiamo

$$AB = \begin{pmatrix} -1 & 1 & 2 \\ 7 & 6 & -3 \\ 2 & 5 & 2 \end{pmatrix} \quad BA = \begin{pmatrix} 0 & 3 & -3 \\ 5 & 7 & -3 \\ 5 & 9 & -2 \end{pmatrix}$$

Per il prodotto di matrici valgono invece la proprietà associativa e distributiva. Siano A , B e C matrici per le quali esistono i prodotti AB e BC . Allora

$$A(BC) = (AB)C$$

Se inoltre B e C hanno le stesse dimensioni, allora

$$A(B + C) = AB + AC$$

La dimostrazione di questi fatti è lasciata al lettore.

È immediato notare come la matrice identità I_n sia *l'elemento neutro rispetto al prodotto di matrici* in $\mathbb{R}^{n \times n}$, ossia, per ogni $A \in \mathbb{R}^{n \times n}$ vale

$$IA = AI = A.$$

Definizione 8: [Potenze di matrici] Sia A una matrice quadrata $n \times n$. Allora esiste la matrice AA , che chiameremo A^2 . In generale, definiamo $A^0 = I$ e per ogni $k > 0, k \in \mathbb{N}$, definiamo $A^k = A^{k-1}A$. Le matrici A^k si chiamano le *potenze k -me* della matrice A . Una matrice quadrata A tale che $A^2 = A$ viene detta *idempotente*. Ad esempio la seguente è una matrice idempotente: $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

ESERCIZIO 2.2. Detta M una matrice quadrata, indichiamo con M^2 la matrice MM . Analogamente allo sviluppo del quadrato di un binomio, i.e., $(a+b)^2 = a^2 + 2ab + b^2$ con $a, b \in \mathbb{R}$, si ricavi una formula per lo sviluppo di $(A+B)^2$ con $A, B \in \mathbb{R}^{n \times n}$. Si applichi tale formula al caso

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

◇

Esempio. Consideriamo un gruppo di n azionisti, che possiedono stock-options di p compagnie. Le stock-options sono azioni vincolate, che non possono essere vendute prima di un prefissato periodo di tempo (nel nostro caso, supponiamo un anno). Conoscendo l'andamento del mercato su un anno solare (i.e., conoscendo il valore medio dell'azione di ogni compagnia in ciascun mese dell'anno), vogliamo valutare l'andamento del valore delle stock options per gli azionisti, ossia il capitale virtuale posseduto da ognuno degli azionisti in ogni mese.

Per risolvere questo problema, rappresentiamo i dati come segue. Sia $A = (a_{ij})$ una matrice $m \times p$ (con $m = 12$) in cui il generico elemento a_{ij} è il valore medio di un'azione della compagnia j nel mese i . Sia poi $B = (b_{hk})$ una matrice $p \times n$ in cui il generico elemento b_{hk} è il numero di stock-options della compagnia h possedute dall'azionista k . Sia infine $C = AB = (c_{ij})$ il prodotto di A e B .

Per ogni $i = 1, \dots, m$ e $j = 1, \dots, n$ il numero $\sum_{k=1}^p a_{ik}b_{kj}$ rappresenta il valore nel mese i delle stock-options possedute dall'azionista j . Quindi la matrice C contiene esattamente il capitale virtuale posseduto da ognuno degli azionisti in ogni mese. ◇

Esempio. Siano p_1, \dots, p_m delle stazioni da cui si può partire per arrivare a una tra n stazioni di destinazione, denotate con d_1, \dots, d_n . Ogni viaggio da p_i a d_j (per $i \in \{1, \dots, m\}$ e $j \in \{1, \dots, n\}$) attraversa una tra t stazioni intermedie, siano esse s_1, \dots, s_t . Supponendo di conoscere il numero di percorsi tra p_i e s_j per ogni $i \in \{1, \dots, m\}$ e $j \in \{1, \dots, t\}$ e il numero di percorsi tra s_i e d_j per ogni $i \in \{1, \dots, t\}$ e $j \in \{1, \dots, n\}$, come possiamo calcolare il numero di percorsi tra p_i e d_j per ogni $i \in \{1, \dots, m\}$ e $j \in \{1, \dots, n\}$?

Definiamo due opportune matrici, $A = (a_{ij}) \in \mathbb{R}^{m \times t}$ e $B = (b_{ij}) \in \mathbb{R}^{t \times n}$. L'elemento a_{ij} rappresenta il numero di percorsi tra p_i e s_j , mentre b_{ij} rappresenta il numero di percorsi tra s_i e d_j . In base alle leggi combinatoriche della somma e del prodotto, per ogni $i \in \{1, \dots, m\}$ e $j \in \{1, \dots, n\}$, il numero di percorsi possibili tra p_i e d_j è

$$\sum_{h=1}^t a_{ih}b_{ht}$$

Quindi, detta $C = (c_{ij}) \in \mathbb{R}^{m \times n}$ la matrice il cui elemento generico c_{ij} è il numero di percorsi possibili tra p_i e d_j , si ha

$$C = AB.$$

◇

Un caso importante del prodotto matriciale AX si ha quando X è un vettore-colonna. Sia ad esempio $A \in \mathbb{R}^{m \times n}$ e $X \in \mathbb{R}^{n \times 1}$. In questo caso il prodotto è un vettore-colonna $m \times 1$ e si ha

$$AX = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}$$

Fissato $i \in \{1, \dots, n\}$, sia $X = \mathbf{e}_i$ definito da $x_i = 1$ e $x_j = 0$ per $j \neq i$. Allora il prodotto AX restituisce la colonna i -ma di A . Se invece $X = \mathbf{1} = {}^t(1 \ 1 \ \cdots \ 1)$ allora AX ha l'effetto di sommare gli elementi di ogni riga di A , e restituire la colonna con tutti i risultati.

Esempio. Sia $G = (V, E)$ un grafo con $V = \{1, \dots, n\}$. Definiamo la *matrice di adiacenza* di G come la matrice quadrata $A = (a_{ij})$ di dimensioni $n \times n$ tale che

$$a_{ij} = 1 \text{ se } ij \in E \text{ mentre } a_{ij} = 0 \text{ se } ij \notin E$$

si noti che A è simmetrica e la sua diagonale è fatta di soli 0. Sia $\mathbf{1} = {}^t(1 \ 1 \ \cdots \ 1)$ il vettore-colonna i cui elementi sono tutti pari a 1. Consideriamo il prodotto

$$A\mathbf{1} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} + \cdots + a_{1n} \\ \vdots \\ a_{n1} + \cdots + a_{nn} \end{pmatrix} = \begin{pmatrix} d(1) \\ \vdots \\ d(n) \end{pmatrix}$$

dove ogni $d(i)$ è il grado del nodo i in G . Infatti, la somma $\sum_{j=1}^n a_{ij}$ restituisce esattamente il numero di nodi j adiacenti al nodo i in G .

Ogni arco è ovviamente anche un cammino tra due nodi. In particolare, è l'unico cammino di lunghezza 1, ed è quindi corretto affermare che $A^1 (= A)$ è la matrice che, per ogni coppia di nodi i e j , conta quanti cammini di lunghezza 1 esistono fra i e j in G . Vogliamo ora dimostrare come per ogni $k \geq 1$, detta $A^k = (a_{ij}^k)$, si ha che

$$a_{ij}^k = \text{numero di cammini di lunghezza } k \text{ tra } i \text{ e } j \text{ in } G$$

Per vedere ciò, ragioniamo per induzione. Il caso $k = 1$ è vero. Supponiamo allora $k > 1$ e sia vero che A^{k-1} restituisca, per ogni coppia di nodi i, j , il numero di cammini di lunghezza $k - 1$ fra i e j . Ogni cammino di lunghezza k fra i e j è dato da un cammino di lunghezza $k - 1$ tra i e v , con v un nodo adiacente a j , più l'arco vj . Quindi, per ogni v adiacente a j (ossia tale che $a_{vj} = 1$) bisognerà contare tutti i cammini di lunghezza $k - 1$ per ottenere i cammini di lunghezza k passanti per v . In conclusione, il numero totale di cammini di lunghezza k fino a j sarà

$$\sum_{v=1}^n a_{iv}^{k-1} a_{vj}$$

Si noti che sulla diagonale compare il numero di cicli di lunghezza k per ogni nodo. Tali cicli, così come i cammini, non sono necessariamente elementari nè semplici. Ogni arco ij , ad esempio, dà luogo a un ciclo

di lunghezza 2, i.e., (i, j, i) . Consideriamo il grafo

$$G = (\{1, 2, 3, 4\}, \{12, 13, 14, 24, 34\})$$

la cui matrice di adiacenza è

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

Abbiamo

$$A^3 = \begin{pmatrix} 4 & 5 & 5 & 5 \\ 5 & 2 & 2 & 5 \\ 5 & 2 & 2 & 5 \\ 5 & 5 & 5 & 4 \end{pmatrix}$$

Quindi ci sono, ad esempio, 5 cammini di lunghezza 3 tra 1 e 2. Essi sono $(1, 2, 1, 2)$, $(1, 3, 1, 2)$, $(1, 4, 1, 2)$, $(1, 3, 4, 2)$ e $(1, 2, 4, 2)$. \diamond

Il seguente teorema mette in relazione il prodotto di matrici con la loro trasposizione.

TEOREMA 1: Siano $A \in \mathbb{R}^{m \times n}$ e $B \in \mathbb{R}^{n \times m}$. Allora

$${}^t(AB) = {}^tB {}^tA$$

Dim: Siano $C = (c_{ij}) = AB$ e $D = (d_{ij}) = {}^tC$. Per ogni $i = 1, \dots, m$ e $j = 1, \dots, m$ abbiamo

$$d_{ij} = c_{ji} = \sum_{k=1}^n A_{jk} B_{ki} = \sum_{k=1}^n ({}^tA)_{kj} ({}^tB)_{ik} = \sum_{k=1}^n ({}^tB)_{ik} ({}^tA)_{kj}$$

e quindi $D = {}^tB {}^tA$. ♣

Definizione 9: [Matrici di permutazione] Una *matrice di permutazione* è una matrice $P = (p_{ij}) \in \mathbb{R}^{n \times n}$ in cui ciascun elemento $p_{ij} \in \{0, 1\}$ e tale che

$$\sum_{k=1}^n p_{ik} = \sum_{k=1}^n p_{kj} = 1 \quad \text{per ogni } i = 1, \dots, n \text{ e } j = 1, \dots, n$$

(i.e., ogni riga e ogni colonna di P contengono esattamente un "1").

Il nome delle matrici di permutazione è dovuto al fatto che la moltiplicazione di una matrice $A \in \mathbb{R}^{n \times n}$ a sinistra o a destra per una matrice di permutazione ha l'effetto di permutare le righe o, rispettivamente, le colonne di A .

Esempio. Siano

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad A = \begin{pmatrix} 4 & -2 & 6 \\ 3 & -5 & -3 \\ -7 & 0 & 9 \end{pmatrix}$$

Abbiamo

$$PA = \begin{pmatrix} 3 & -5 & -3 \\ -7 & 0 & 9 \\ 4 & -2 & 6 \end{pmatrix} \quad AP = \begin{pmatrix} 6 & 4 & -2 \\ -3 & 3 & -5 \\ 9 & -7 & 0 \end{pmatrix}$$

◇

La pre-moltiplicazione per P ha avuto l'effetto di permutare le righe di A , mentre la post-moltiplicazione ne ha permutato le colonne.

Ad ogni permutazione $\pi \in S_n$ corrisponde biunivocamente una matrice di permutazione, che indicheremo con $P_{(\pi)}$. In particolare, se $\pi = (\pi(1), \pi(2), \dots, \pi(n))$ allora $P_{(\pi)}[i, j] = 1$ se $j = \pi(i)$ e $P_{(\pi)}[i, j] = 0$ altrimenti.

Ad esempio, se $\pi = (4, 1, 3, 2)$ abbiamo

$$P_{(\pi)} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Data una permutazione π , si ha

$$P_{(\pi)}A = \begin{pmatrix} A_{\pi(1)} \\ A_{\pi(2)} \\ \dots \\ A_{\pi(n)} \end{pmatrix}$$

Dimostrazione: Sia $B = P_{(\pi)}A$ e consideriamo l'elemento in riga i , colonna j di B . Tale elemento è il prodotto della riga i di $P_{(\pi)}$ e della colonna j di A . Siccome in riga i di $P_{(\pi)}$ ogni elemento vale 0 tranne quello in colonna $\pi(i)$, che vale 1, si ha $b_{ij} = a_{\pi(i), j}$.

Denotiamo ora con $\pi^{-1} = (\pi^{-1}(1), \dots, \pi^{-1}(n))$ la permutazione inversa di π , ossia la permutazione tale che $\pi^{-1}(\pi(i)) = i$ per ogni $i = 1, \dots, n$. Nel nostro esempio $\pi^{-1} = (2, 4, 3, 1)$. Abbiamo allora

$$AP_{(\pi)} = \begin{pmatrix} A^{\pi^{-1}(1)} & A^{\pi^{-1}(2)} & \dots & A^{\pi^{-1}(n)} \end{pmatrix}.$$

2.3 Sottomatrici e scomposizione in blocchi

Finora abbiamo considerato le matrici come delle tabelle rettangolari i cui elementi sono dei numeri. Astruendo questa idea ad un livello superiore, possiamo immaginare delle matrici in cui gli elementi sono a loro volta matrici.

Consideriamo ad esempio la matrice

$$A = \begin{pmatrix} 0 & 2 & 0 & 1 & 3 & 9 & 9 \\ 5 & 0 & 4 & 7 & 6 & 5 & 0 \\ 2 & 1 & 5 & 1 & 0 & 0 & 4 \\ 5 & 7 & 4 & 3 & 0 & 3 & 0 \\ 0 & 9 & 3 & 1 & 2 & 7 & 0 \\ 1 & 1 & 0 & 1 & 0 & 2 & 4 \\ 8 & 0 & 9 & 2 & 6 & 6 & 1 \end{pmatrix}$$

Evidenziamone dei blocchi in questo modo

$$A = \left(\begin{array}{cccc|ccc} 0 & 2 & 0 & 1 & 3 & 9 & 9 \\ 5 & 0 & 4 & 7 & 6 & 5 & 0 \\ \hline 2 & 1 & 5 & 1 & 0 & 0 & 4 \\ 5 & 7 & 4 & 3 & 0 & 3 & 0 \\ 0 & 9 & 3 & 1 & 2 & 7 & 0 \\ 1 & 1 & 0 & 1 & 0 & 2 & 4 \\ \hline 8 & 0 & 9 & 2 & 6 & 6 & 1 \end{array} \right)$$

La matrice A può ora essere vista come una matrice 3×2 i cui elementi sono a loro volta delle matrici B_{ij} :

$$A = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{pmatrix}$$

In questo esempio,

$$B_{11} = \begin{pmatrix} 0 & 2 & 0 & 1 \\ 5 & 0 & 4 & 7 \end{pmatrix} \quad B_{22} = \begin{pmatrix} 0 & 0 & 4 \\ 0 & 3 & 0 \\ 2 & 7 & 0 \\ 0 & 2 & 4 \end{pmatrix}$$

eccetera.

Definizione 10: [Sottomatrice] Sia $A = (a_{ij})$ una matrice $m \times n$, e siano i_1, i_2, j_1, j_2 degli indici tali che $1 \leq i_1 \leq i_2 \leq m$ e $1 \leq j_1 \leq j_2 \leq n$. La matrice $B \in \mathbb{R}^{(i_2-i_1+1) \times (j_2-j_1+1)} = (b_{hk})$ contenuta tra le righe i_1 e i_2 (comprese) di A e le colonne j_1 e j_2 (comprese) di B è detta una *sottomatrice* di A . Per ogni $k = 1, \dots, i_2 - i_1 + 1$ e $h = 1, \dots, j_2 - j_1 + 1$ si ha

$$b_{hk} = a_{i_1+h-1, j_1+k-1}$$

Definizione 11: [Scomposizione in blocchi] Una *scomposizione in blocchi* di una matrice $A \in \mathbb{R}^{m \times n}$ è definita prendendo $p \geq 1$ naturali positivi r_1, \dots, r_p tali che

$$r_1 + \dots + r_p = m$$

e q naturali positivi c_1, \dots, c_q con

$$c_1 + \dots + c_q = n$$

La scomposizione ripartisce A in $p \times q$ sottomatrici (blocchi) come segue

$$A = \begin{pmatrix} B_{11} & \dots & B_{1q} \\ \dots & \dots & \dots \\ B_{p1} & \dots & B_{pq} \end{pmatrix}$$

dove la generica sottomatrice B_{hk} ha dimensioni $r_h \times c_k$, ed è contenuta in A tra le righe $\sum_{i=1}^{h-1} r_i + 1$ e $\sum_{i=1}^h r_i$ e tra le colonne $\sum_{j=1}^{k-1} c_j + 1$ e $\sum_{j=1}^k c_j$.

Definizione 12: [Matrice diagonale a blocchi] Sia $A \in \mathbb{R}^{n \times n}$ una matrice quadrata, e siano d_1, \dots, d_k naturali positivi con $\sum_i d_i = n$. Denotiamo con $D_i \in \mathbb{R}^{d_i \times d_i}$ la sottomatrice quadrata di A la cui prima riga e prima colonna hanno indice $1 + \sum_{j=1}^{i-1} d_j$. Se $A_{uv} = 0$ per ogni elemento uv che non appartiene ad alcuna delle sottomatrici D_i , diremo che A è una matrice *diagonale a k blocchi*, con blocchi D_1, \dots, D_k . Ad esempio, la seguente è una matrice diagonale a 3 blocchi

$$A = \left(\begin{array}{ccc|cc|cc} 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 4 & 0 & 0 & 0 & 0 \\ 2 & 1 & 5 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 6 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 6 & 1 \end{array} \right)$$

Ogni operazione tra matrici può essere generalizzata ad una corrispondente operazione tra matrici scomposte in blocchi, a patto che le scomposizioni siano *compatibili con l'operazione*. Il concetto di compatibilità è definito come segue.

Compatibilità per la somma. Siano $A, B \in \mathbb{R}^{m \times n}$. Consideriamo una scomposizione di A in $p \times q$ blocchi di righe r_1, \dots, r_p e colonne c_1, \dots, c_q e una scomposizione di B in $s \times t$ blocchi di righe r'_1, \dots, r'_s e colonne c'_1, \dots, c'_t . Allora le scomposizioni sono compatibili con la somma delle matrici A e B se

1. $p = s$ e $q = t$
2. $r_i = r'_i$ e $c_j = c'_j$ per ogni $i = 1, \dots, p$ e $j = 1, \dots, q$

(in pratica A e B sono "scomposte nei medesimi blocchi"). Il risultato della somma sarà una matrice scomposta in $p \times q$ blocchi le cui lunghezze sulle righe sono r_1, \dots, r_p e sulle colonne sono c_1, \dots, c_q .

Date le scomposizioni di A e B

$$A = \begin{pmatrix} C_{11} & \cdots & C_{1q} \\ \cdots & \cdots & \cdots \\ C_{p1} & \cdots & C_{pq} \end{pmatrix} \quad B = \begin{pmatrix} D_{11} & \cdots & D_{1q} \\ \cdots & \cdots & \cdots \\ D_{p1} & \cdots & D_{pq} \end{pmatrix}$$

la somma può essere effettuata "a blocchi", ottenendo

$$A + B = \begin{pmatrix} C_{11} + D_{11} & \cdots & C_{1q} + D_{1q} \\ \cdots & \cdots & \cdots \\ C_{p1} + D_{p1} & \cdots & C_{pq} + D_{pq} \end{pmatrix}$$

La verifica di questa proprietà è immediata ed è lasciata al lettore.

Compatibilità per il prodotto matriciale. Siano $A \in \mathbb{R}^{m \times t}$ e $B \in \mathbb{R}^{t \times n}$ due matrici di cui vogliamo calcolare il prodotto.

Consideriamo una scomposizione di A in 2×1 blocchi, del tipo

$$A = \begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix}$$

Allora è immediato verificare come anche la matrice AB può essere calcolata "a blocchi", e risulta

$$AB = \begin{pmatrix} X_{11}B \\ X_{21}B \end{pmatrix}$$

Infatti, ogni elemento di AB è il prodotto di una riga di A e una colonna di B . Le righe di A prese da X_{11} danno luogo alla sottomatrice $X_{11}B$, mentre quelle prese da X_{21} danno luogo alla sottomatrice $X_{21}B$.

Se A è scomposta in $r \times 1$ blocchi, del tipo

$$A = \begin{pmatrix} X_{11} \\ \dots \\ X_{r1} \end{pmatrix}$$

allora

$$AB = \begin{pmatrix} X_{11}B \\ \dots \\ X_{r1}B \end{pmatrix}$$

Per vedere ciò, si può usare il principio di induzione su r . Il caso base è $r = 2$. Per $r > 2$ basta guardare alla scomposizione di A come se fosse fatta di due soli blocchi, di cui il secondo è X_{r1} e il primo, sia esso X , è dato da tutte le restanti righe:

$$A = \begin{pmatrix} X_{11} \\ \dots \\ X_{r1} \end{pmatrix} = \begin{pmatrix} X \\ X_{r1} \end{pmatrix}$$

Per induzione abbiamo

$$XB = \begin{pmatrix} X_{11}B \\ \dots \\ X_{r-1,1}B \end{pmatrix}$$

e quindi

$$AB = \begin{pmatrix} XB \\ X_{r1}B \end{pmatrix} = \begin{pmatrix} X_{11}B \\ \dots \\ X_{r-1,1}B \\ X_{r1}B \end{pmatrix}$$

Se non scomponiamo A , ma consideriamo una scomposizione delle colonne di B , ossia scomponiamo B in $1 \times c$ blocchi, del tipo

$$B = (Y_{11} \dots Y_{1c})$$

allora in modo analogo a quanto visto precedentemente, abbiamo che

$$AB = (AY_{11} \dots AY_{1c})$$

Consideriamo ora una scomposizione di A in $p \times q$ blocchi di righe r_1, \dots, r_p e colonne c_1, \dots, c_q e una scomposizione di B in $s \times l$ blocchi di righe r'_1, \dots, r'_s e colonne c'_1, \dots, c'_l . Allora le scomposizioni sono compatibili con il prodotto delle matrici A e B se

1. $q = s$
2. $c_j = r'_j$ per ogni $j = 1, \dots, q$

Il risultato del prodotto sarà una matrice scomposta in $p \times l$ blocchi le cui lunghezze sulle righe sono r_1, \dots, r_p e sulle colonne sono c'_1, \dots, c'_l .

In particolare, consideriamo le scomposizioni di A e B

$$A = \begin{pmatrix} C_{11} & \cdots & C_{1q} \\ \cdots & \cdots & \cdots \\ C_{p1} & \cdots & C_{pq} \end{pmatrix} \quad B = \begin{pmatrix} D_{11} & \cdots & D_{1l} \\ \cdots & \cdots & \cdots \\ D_{q1} & \cdots & D_{ql} \end{pmatrix}$$

Detta

$$C_i = (C_{i1} \ C_{i2} \ \cdots \ C_{iq})$$

la generica sottomatrice di A data dalle righe del blocco i -mo, e detta

$$D^j = \begin{pmatrix} D_{1j} \\ \cdots \\ D_{qj} \end{pmatrix}$$

la generica sottomatrice di B data dalle colonne del blocco j -mo, abbiamo

$$AB = \begin{pmatrix} C_1 \\ \cdots \\ C_p \end{pmatrix} B = \begin{pmatrix} C_1 B \\ \cdots \\ C_p B \end{pmatrix} = \begin{pmatrix} C_1(D^1 \ \cdots \ D^l) \\ \cdots \\ C_p(D^1 \ \cdots \ D^l) \end{pmatrix} = \begin{pmatrix} C_1 D^1 \ \cdots \ C_1 D^l \\ \cdots \\ C_p D^1 \ \cdots \ C_p D^l \end{pmatrix}$$

In conclusione, il prodotto AB risulta la matrice a blocchi

$$AB = \begin{pmatrix} \sum_{h=1}^q C_{1h} D_{h1} & \cdots & \sum_{h=1}^q C_{1h} D_{hl} \\ \cdots & \cdots & \cdots \\ \sum_{h=1}^q C_{ph} D_{h1} & \cdots & \sum_{h=1}^q C_{ph} D_{hl} \end{pmatrix}$$

Si noti la somiglianza di questo prodotto a blocchi con il prodotto "normale" di due matrici. In quest'ultimo il generico elemento è dato dalla somma dei prodotti degli elementi di una riga di A con gli elementi di una colonna di B , mentre qui il generico elemento è dato dalla somma dei prodotti dei blocchi di una riga (della scomposizione) di A con i blocchi di una colonna (della scomposizione) di B .

2.4 Matrice inversa

Definizione 13: [Matrice inversa] Sia $A \in \mathbb{R}^{n \times n}$. Diremo che la matrice $B \in \mathbb{R}^{n \times n}$ è l'*inversa* di A , e la denoteremo con A^{-1} , se

$$AB = BA = I$$

Se una matrice A ammette un'inversa, allora diremo che A è *invertibile*. Si noti che la condizione di essere quadrata è necessaria perchè una matrice sia invertibile. Infatti, assumendo $A \in \mathbb{R}^{m \times n}$ con $m \neq n$, il prodotto AB avrebbe m righe, mentre il prodotto BA ne avrebbe n e quindi tali prodotti non potrebbero essere uguali. Non tutte le matrici quadrate sono invertibili. Banalmente, la matrice $\mathbf{0}_{n,n}$ non ha inversa, in quanto $\mathbf{0}_{n,n}B = B\mathbf{0}_{n,n} = \mathbf{0}_{n,n}$ per ogni $B \in \mathbb{R}^{n \times n}$. Per un ulteriore esempio, consideriamo la matrice

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

e sia

$$B = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

una sua ipotetica inversa. Da $AB = I$ otteniamo

$$x - z = 1, \quad y - w = 0, \quad 2x - 2z = 0, \quad 2y - 2w = 1$$

condizioni ovviamente impossibili, in quanto implicano contemporaneamente che $x - z = 1$ e $x - z = 0$ (o, similmente, $y - w = 0$ e $y - w = 1/2$). Sia ora

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

e sia

$$B = \begin{pmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Si può verificare che $AB = BA = I$ e quindi $B = A^{-1}$.

È facile vedere che se una matrice A è invertibile allora anche A^{-1} lo è e si ha

$$(A^{-1})^{-1} = A.$$

Inoltre, se A è invertibile allora la sua inversa è *unica*. Siano infatti B e C inverse della matrice A . Allora si ha

$$B = BI = B(AC) = (BA)C = IC = C$$

Come esempio di matrice invertibile, consideriamo una qualsiasi matrice di permutazione $A = P_{(\pi)}$, dove π è una permutazione di $\{1, \dots, n\}$. Sia π^{-1} la permutazione inversa di π e sia $B = P_{(\pi^{-1})}$. Nel prodotto AB , le righe di B vengono riordinate secondo la permutazione π , ossia la riga i -ma è $B_{\pi(i)}$. Tale riga ha un unico 1, in posizione $\pi^{-1}(\pi(i))$, ossia in posizione i . Quindi

$$P_{(\pi)}P_{(\pi^{-1})} = I$$

Similmente, nel prodotto BA , le righe di A vengono riordinate secondo la permutazione π^{-1} , ossia la riga i -ma è $A_{\pi^{-1}(i)}$. Tale riga ha un unico 1, in posizione $\pi(\pi^{-1}(i)) = i$. Quindi

$$P_{(\pi^{-1})}P_{(\pi)} = I$$

In conclusione, la matrice di permutazione $P_{(\pi)}$ è sempre invertibile, e la sua inversa è la matrice di permutazione $P_{(\pi^{-1})}$.

Come esempio, consideriamo la permutazione $\pi = (3, 1, 2, 4)$. Abbiamo $\pi^{-1} = (2, 3, 1, 4)$ e le seguenti matrici di permutazione:

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

Per esercizio, il lettore verifichi che $AB = BA = I$.

Per le matrici di permutazione vale un'importante relazione tra ogni matrice e la sua inversa, ossia, per una matrice di permutazione si ha

Proprietà 2: [Inversa di una matrice di permutazione] Sia $A = P_{(\pi)}$ una matrice di permutazione. Allora

$$A^{-1} = {}^tA$$

Dim: Il generico elemento ij di $A {}^tA$ è il prodotto della riga i di A e la colonna j di tA . Quest'ultima coincide (nei suoi elementi) con la riga j di A . Tale prodotto vale 1 se e solo se $i = j$, e quindi si ottiene che

$$A {}^tA = I$$

Similmente,

$${}^tAA = I$$

e quindi ${}^tA = A^{-1}$. ♣

Dall'unicità dell'inversa si deduce anche che per una matrice di permutazione $P_{(\pi)}$ si ha

$$P_{(\pi^{-1})} = {}^tP_{(\pi)}$$

Definizione 14: [Matrici ortogonali] Sia $A \in \mathbb{R}^{n \times n}$ una matrice quadrata. Diciamo che A è una *matrice ortogonale* se A è invertibile e si ha

$$A^{-1} = {}^tA$$

In base a quanto poc'anzi osservato, concludiamo che ogni matrice di permutazione è una matrice ortogonale. Come esempio di matrice ortogonale ma non di permutazione, si consideri la seguente per $\alpha \in \mathbb{R} - \{k\pi/2, k \in \mathbb{Z}\}$

$$A_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$$

Proprietà 3: [Inversa del prodotto] Siano $A, B \in \mathbb{R}^{n \times n}$ matrici invertibili. Allora anche il prodotto AB è una matrice invertibile, e si ha

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

Dim: Abbiamo

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$$

ed anche

$$(AB)(B^{-1}A^{-1}) = A(B^{-1}B)A^{-1} = AIA^{-1} = AA^{-1} = I.$$



La Proprietà 3 può essere ricordata come: *l'inversa di un prodotto è il prodotto delle inverse, ma nell'ordine inverso!* Questa regola vale anche per il prodotto di un numero arbitrario di matrici, come è facile verificare (esercizio). Ad esempio

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}.$$

Definizione 15: [Matrici simili] Siano $A, B \in \mathbb{R}^{n \times n}$ matrici quadrate. Diciamo che B è *simile* ad A se esiste una matrice invertibile $Q \in \mathbb{R}^{n \times n}$ tale che

$$B = Q^{-1}AQ$$

La similitudine tra matrici è una relazione di equivalenza. Infatti valgono le seguenti proprietà:

1. (Riflessiva:) Essendo $A = I^{-1}AI$, si ha che A è simile ad A
2. (Simmetrica:) Sia A simile a B . Allora esiste Q tale che $A = Q^{-1}BQ$. Moltiplicando a sinistra per Q e a destra per Q^{-1} otteniamo $QAQ^{-1} = QQ^{-1}BQ^{-1}Q = B$. Si ha quindi che $(Q^{-1})^{-1}A(Q^{-1}) = B$ sicchè B è simile ad A .
3. (Transitiva) Sia A simile a B e B simile a C . Esistono matrici P e Q invertibili tali che $A = P^{-1}BP$ e $B = Q^{-1}CQ$. Abbiamo quindi

$$A = (P^{-1}Q^{-1})C(QP) = A = (QP)^{-1}C(QP) =$$

sicchè A è simile a C .

Esempio. Sia $X = \begin{pmatrix} -2 & 1 \\ 3 & 2 \end{pmatrix}$. Consideriamo la matrice ortogonale A_α per $\alpha = \pi/4$. Abbiamo $A_\alpha = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$ e $A_\alpha^{-1} = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$. Essendo

$$Y = A_\alpha^{-1} X A_\alpha = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} = \begin{pmatrix} -2 & -3 \\ -1 & 2 \end{pmatrix}$$

otteniamo che $\begin{pmatrix} -2 & 1 \\ 3 & 2 \end{pmatrix}$ e $\begin{pmatrix} -2 & -3 \\ -1 & 2 \end{pmatrix}$ sono matrici simili. ◇

Chapter 3

Spazi vettoriali

Definizione 16: [Spazio Vettoriale] Uno *spazio vettoriale sul corpo dei reali* (o, più semplicemente, uno *spazio vettoriale*) è un insieme V su cui sono definite due operazioni:

- (i) un'operazione di somma, che a ogni coppia di elementi di V associa un elemento di V :

$$(+): V \times V \mapsto V$$

Denotiamo $(+)(v, w)$ con l'usuale notazione per la somma, ossia $v + w$.

- (ii) un'operazione di prodotto fra numeri in \mathbb{R} e elementi di V , che a ogni coppia (numero, elemento di V) associa un elemento di V :

$$(\cdot): \mathbb{R} \times V \mapsto V$$

Denotiamo $(\cdot)(\lambda, v)$ con l'usuale notazione per il prodotto, ossia λv .

per le quali valgono le seguenti proprietà:

PV1: La somma tra elementi di V è sia associativa che commutativa, i.e.

$$(u + v) + w = u + (v + w)$$

e

$$u + v = v + u$$

per ogni $u, v, w \in V$.

PV2: Esiste in V un elemento neutro per la somma, denotato con $\mathbf{0}$, tale che

$$v + \mathbf{0} = v$$

per ogni $v \in V$.

PV3: Per ogni elemento $v \in V$, esiste in V un suo opposto, denotato con $< -v >$, tale che

$$v + < -v > = \mathbf{0}$$

PV4: Il prodotto è distributivo, sia rispetto a \mathbb{R} che a V , i.e.

$$(\lambda + \delta)v = \lambda v + \delta v$$

e

$$\lambda(v + w) = \lambda v + \lambda w$$

per ogni $\lambda, \delta \in \mathbb{R}$ e $v, w \in V$.

PV5: Esiste un elemento neutro per il prodotto. In particolare, il numero $1 \in \mathbb{R}$ è tale che

$$1 \cdot v = v$$

per ogni $v \in V$.

PV6: La moltiplicazione soddisfa

$$(\lambda\delta)v = \lambda(\delta v)$$

per ogni $\lambda, \delta \in \mathbb{R}$ e $v \in V$.

Se V è uno spazio vettoriale, gli elementi di V vengono detti *vettori*. Bisogna fare attenzione a non confondere il numero zero con il vettore zero (i.e., l'elemento neutro della somma in V). Per questo motivo adotteremo la seguente notazione tipografica: denotiamo il numero zero con 0, il vettore zero con $\mathbf{0}$, e, qualora nel discorso appaiano più spazi vettoriali distinti, per denotare il vettore zero dello spazio V scriveremo $\mathbf{0}_V$.

Uno spazio vettoriale può anche consistere di un singolo elemento, che deve necessariamente essere il vettore $\mathbf{0}$. Chiameremo questo caso estremo lo *spazio vettoriale nullo* (o spazio *banale*). Per ogni spazio V non banale si può dimostrare che $|V| = \infty$ (segue dalla proprietà 3 qui sotto).

Per uno spazio vettoriale valgono sempre le seguenti proprietà:

1. (unicità dello zero) Esiste un unico vettore $\mathbf{0} \in V$ tale che $v + \mathbf{0} = v$ per ogni $v \in V$.
2. $0v = \mathbf{0}$ per ogni $v \in V$
3. (legge di annullamento del prodotto) Sia $\alpha \in \mathbb{R}$ e $v \in V$. Allora $\alpha v = \mathbf{0}$ se e solo se $\alpha = 0$ OR $v = \mathbf{0}$
4. Per ogni terna di vettori $v, u, w \in V$ si ha $v + u = v + w$ se e solo se $u = w$.
5. $(-1)v = < -v >$

Dimostrazione:

1. Siano $\mathbf{0}, \mathbf{0}' \in V$ tali che $v + \mathbf{0} = v + \mathbf{0}' = v$ per ogni $v \in V$. Allora $\mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}'$.
2. Dalla proprietà PV4 si ha $0v = (0+0)v = 0v + 0v$, da cui, sommando $< -(0v) >$ ad entrambi i membri si ottiene $\mathbf{0} = 0v$.

3. (Se:) Sappiamo dal punto 2. che se $\alpha = 0$ allora $\alpha v = \mathbf{0}$. Sia ora $v = \mathbf{0}$. Abbiamo $\alpha v = \alpha(v + v) = \alpha v + \alpha v$. Sommando $< -(\alpha v) >$ a entrambi i membri dell'equazione otteniamo $\mathbf{0} = \alpha v$. (Solo se:) Sia $\alpha v = \mathbf{0}$ con $\alpha \neq 0$. Allora $v = (1/\alpha)\alpha v = (1/\alpha)\mathbf{0} = \mathbf{0}$.

Si noti che da questa proprietà segue il fatto che multipli distinti di un vettore non-nullo sono vettori distinti. Infatti, siano $\alpha \neq \beta \in \mathbb{R}$ e $v \neq \mathbf{0} \in V$. Allora $\alpha v - \beta v = (\alpha - \beta)v \neq \mathbf{0}$ in quanto $\alpha - \beta \neq 0$. Questo fatto implica anche che per qualsiasi spazio vettoriale non banale si ha $|V| = \infty$.

4. Una direzione è ovvia. Per l'altra, sia $v + u = v + w$. Sommando $< -v >$ a sinistra e destra si ha $u = w$.

Si noti che da questa proprietà segue anche la proprietà di *unicità dell'opposto*: per ogni v esiste un unico elemento u tale che $v + u = \mathbf{0}$. Infatti, per ogni altro $w \in V$ tale che $v + w = \mathbf{0}$ si ha $v + u = v + w$ da cui $u = w$.

5. Dalle proprietà PV5 e PV4 si ha $v + (-1)v = 1v + (-1)v = (1 - 1)v = 0v = \mathbf{0}$ e quindi $(-1)v = < -v >$. Nel seguito indicheremo $< -v >$ semplicemente con $-v$, e anziché scrivere $u + < -v >$ scriveremo $u - v$.

Vediamo ora alcuni esempi di spazi vettoriali:

Esempio. L'insieme \mathbb{R} con le usuali operazioni di somma e prodotto è uno spazio vettoriale, come si può immediatamente verificare. In esso il vettore $\mathbf{0}$ è il numero zero

$$\mathbf{0} = 0$$

◇

Esempio. L'insieme \mathbb{R}^n con le operazioni di somma fra n -ple e prodotto di numeri per n -ple descritte in (1.1) e (1.2) è uno spazio vettoriale, come si può immediatamente verificare. In esso il vettore $\mathbf{0}$ è la n -pla

$$\mathbf{0} = (0, \dots, 0)$$

◇

Esempio. L'insieme $\mathbb{R}^{m \times n}$ con le operazioni di somma tra matrici e prodotto di una matrice per uno scalare descritte in 2.1 è uno spazio vettoriale per ogni coppia di naturali positivi m ed n . In esso il vettore $\mathbf{0}$ è la matrice

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

◇

Esempio. Sia A un insieme e sia $\mathcal{F} = \{f \mid f : A \mapsto \mathbb{R}\}$ l'insieme di tutte le funzioni di A in \mathbb{R} . Prese due funzioni $f, g \in \mathcal{F}$ possiamo definire la loro somma $f + g$ come la funzione $h : A \mapsto \mathbb{R}$ definita da

$$h(x) = f(x) + g(x) \quad x \in \mathbb{R}$$

Inoltre, per $\lambda \in \mathbb{R}$ e $f \in \mathcal{F}$ possiamo definire il prodotto λf come la funzione $h : A \mapsto \mathbb{R}$ definita da

$$h(x) = \lambda f(x) \quad x \in \mathbb{R}$$

È facile dimostrare che rispetto a queste operazioni, l'insieme \mathcal{F} è uno spazio vettoriale. In esso il vettore $\mathbf{0}$ è la funzione $\mathbf{0} : A \mapsto \mathbb{R}$ definita da $\mathbf{0}(x) = 0$ per ogni $x \in A$. \diamond

Esempio. Sia \mathcal{P} l'insieme di tutti i polinomi di una singola variabile a valori reali. Ogni polinomio è un'espressione del tipo

$$p(x) = \sum_{i \in \mathbb{N}} a_i x^i$$

in cui $a_i \in \mathbb{R}$ per ogni $i \in \mathbb{N}$ e $a_i \neq 0$ solo per un numero finito di indici. Sia n il massimo indice per il quale $a_n \neq 0$. Allora diciamo che $p(x)$ ha grado n e possiamo scrivere $p(x)$ come

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n.$$

Denotiamo con $\deg(p(x))$ il grado di un polinomio $p(x)$.

Presi due polinomi

$$p(x) = \sum_{i \in \mathbb{N}} a_i x^i$$

e

$$q(x) = \sum_{i \in \mathbb{N}} b_i x^i$$

possiamo definire la loro somma $p(x) + q(x)$ come il polinomio $h(x)$ definito da

$$h(x) = \sum_{i \in \mathbb{N}} (a_i + b_i) x^i$$

(si noti che $\deg(p(x) + q(x)) \leq \max\{\deg(p(x)), \deg(q(x))\}$).

Inoltre, per $\lambda \in \mathbb{R}$ e $p(x) \in \mathcal{P}$ possiamo definire il prodotto $\lambda p(x)$ come il polinomio $h(x)$ definito da

$$h(x) = \sum_{i \in \mathbb{N}} (\lambda a_i) x^i$$

È facile dimostrare che rispetto a queste operazioni, l'insieme \mathcal{P} è uno spazio vettoriale. In esso il vettore $\mathbf{0}$ è il polinomio in cui $a_i = 0$ per ogni i . \diamond

Esempio. Facciamo ora un esempio meno matematico dei precedenti, che illustra come i vettori possano essere (una volta rispettate le proprietà richieste dalla loro definizione) elementi di un insieme qualsiasi. In particolare, fissato un particolare punto P del piano, consideriamo tutte le figure, che chiameremo "faccine" disegnate come segue:

1. Tracciamo una circonferenza di centro P , che delimita la faccia. Il punto P rappresenta il naso, ed è comune a ogni faccia.

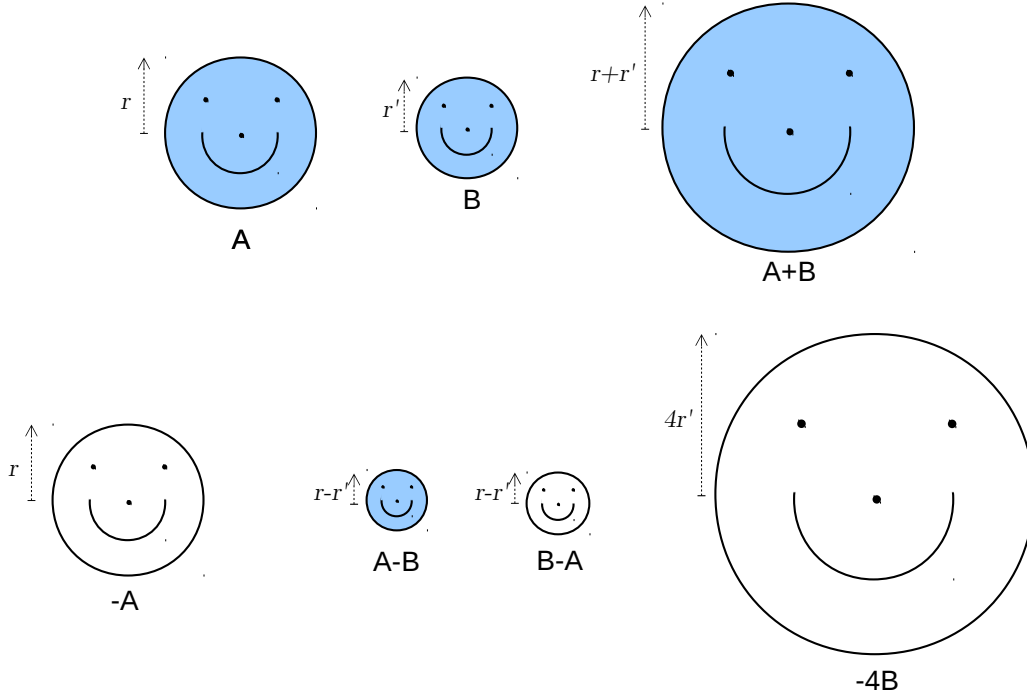


Figure 3.1: Lo spazio delle faccine bicolore.

2. Sia r il raggio della faccina. Con centro P e raggio $r/2$, tracciamo una semicirconferenza che termina in corrispondenza del diametro orizzontale della faccia. Tale semicirconferenza rappresenta una bocca sorridente
3. In corrispondenza agli estremi della bocca, posizionati in direzione verticale a una distanza di $r/2$ dagli stessi, disegniamo due punti, rappresentanti gli occhi della faccina (si veda la figura 3.1 per degli esempi).

Le faccine sono colorate e possono avere uno fra due colori, che chiameremo rosso (R) e verde (V). Tra le faccine includiamo il caso limite della faccina di raggio $r = 0$, che chiameremo faccina vuota. La faccina vuota ha area nulla e coincide con il punto P . Tale faccina, per convenzione, ha colore rosso (questa scelta è arbitraria: avremmo potuto colorare la faccina vuota verde e tutte le argomentazioni sarebbero rimaste valide).

Sia \mathcal{F} l'insieme delle faccine colorate. Per ogni faccina $f \in \mathcal{F}$ denotiamo con $r(f)$ il suo raggio e con $c(f)$ il suo colore. Inoltre, dato un colore $x \in \{R, V\}$, denotiamo con \bar{x} "l'altro" colore (ossia $\bar{R} = V$ e $\bar{V} = R$). Su \mathcal{F} definiamo la seguente operazione di somma:

Siano $f, f' \in \mathcal{F}$. Denotiamo con $g = f + f'$ la faccina somma di f e f' definita in questo modo

1. se $c(f) = c(f')$, allora $c(g) := c(f)$ e $r(g) := r(f) + r(f')$

2. Altrimenti, sia $r^+ = \max\{r(f), r(f')\}$. Poniamo $r(g) := r^+ - r^-$. Se $r(f) = r^+$ poniamo $c(g) = c(f)$, altrimenti poniamo $c(g) = c(f')$.

Informalmente, sommando ad una faccina F un'altra dello stesso colore, F “si ingrandisce”. Viceversa, sommando ad una faccina F un'altra di colore diverso e un po' più piccola di lei, F “si rimpicciolisce”. Si veda la figura 3.1 per degli esempi. Si può verificare che rispetto a questa somma la faccina vuota funge da elemento neutro. Inoltre, ogni faccina f ha un'opposto, che è la faccina h definita da $r(h) := r(f)$ e $c(h) := \overline{c(f)}$ se $r(h) > 0$, mentre $c(h) = R$ se $r(h) = 0$.

Definiamo ora il prodotto tra una faccina e un numero reale come segue

Siano $f \in \mathcal{F}$ e $\lambda \in \mathbb{R}$. Denotiamo con $g = \lambda f$ la faccina per la quale

1. $r(g) := |\lambda| r(f)$
2. Se $\lambda > 0$ allora $c(g) := c(f)$; se $\lambda < 0$ allora $c(g) := \overline{c(f)}$; altrimenti $c(g) := R$.

Informalmente, moltiplicare una faccina per λ ha l'effetto di “espanderla” (se $|\lambda| > 1$) o “contrarla” (se $|\lambda| < 1$), lasciandola dello stesso colore se λ è positivo o cambiandole colore se λ è negativo. Si veda la figura 3.1 per degli esempi.

Si può verificare che con le operazioni di somma e prodotto per scalari appena definite l'insieme \mathcal{F} risulta uno spazio vettoriale. Quindi, anche delle faccine bicolore possono essere (in un'opportuna interpretazione) dei vettori! \diamond

ESERCIZIO 3.1. Fissiamo un insieme non vuoto A e, per ogni $X \subseteq A$, $X \neq \emptyset$, definiamo $\mathcal{F}_X = \{f \mid f : X \mapsto \mathbb{R}\}$. Sia ora $\mathcal{F} = \bigcup_{X \subseteq A} \mathcal{F}_X$ l'insieme di tutte le funzioni a valori reali che hanno per dominio un sottoinsieme di A . Definiamo su \mathcal{F} la seguente operazione di somma. Siano $f : X \mapsto \mathbb{R}$ e $g : Y \mapsto \mathbb{R}$. Definiamo la loro somma $h = f + g$ come la funzione $h : X \cup Y \mapsto \mathbb{R}$ tale che

- $h(x) = f(x)$ se $x \in X - Y$
- $h(x) = g(x)$ se $x \in Y - X$
- $h(x) = 0$ se $x \in X \cap Y$

Inoltre, definiamo $(\lambda f)(x) = \lambda f(x)$.

La somma appena definita è commutativa? È associativa? È distributiva rispetto al prodotto? L'insieme \mathcal{F} con le operazioni descritte è uno spazio vettoriale? \diamond

Definizione 17: [Sottospazio] Sia V uno spazio vettoriale e $W \subseteq V$ un suo sottoinsieme. Se W risulta essere uno spazio vettoriale rispetto alla somma e al prodotto per scalari definiti su V , allora W si dice un *sottospazio* di V .

Sia W un sottoinsieme di V . Le proprietà richieste alla somma di vettori e al prodotto per scalari affinché W sia uno spazio vettoriale sono valide perchè sono valide in V . Affinchè W sia uno spazio vettoriale basterà

che W sia chiuso rispetto alla somma di vettori e al prodotto per scalari, per cui sono necessarie e sufficienti queste due condizioni:

1. Per ogni $v, w \in W$ si ha

$$v + w \in W$$

2. Per ogni $\lambda \in \mathbb{R}$ e $v \in W$ si ha

$$\lambda v \in W$$

Equivalentemente, perchè W sia un sottospazio di V è necessario e sufficiente che sia soddisfatta la seguente singola condizione:

1. Per ogni $v, w \in W$ e $\lambda, \mu \in \mathbb{R}$ e si ha

$$\lambda v + \mu w \in W.$$

Esempio. L'insieme $V = \{(x, y) : x = 3t, y = 2t, t \in \mathbb{R}\}$ è un sottospazio di \mathbb{R}^2 . Infatti, siano $v = (3p, 2p)$ e $w = (3q, 2q)$ due vettori in V , e siano $\lambda, \mu \in \mathbb{R}$. Allora si ha

$$\lambda v + \mu w = (3\lambda p, 2\lambda p) + (3\mu q, 2\mu q) = (3(\lambda p + \mu q), 2(\lambda p + \mu q))$$

e quindi $\lambda v + \mu w \in V$. Si noti che V è la retta del piano la cui direzione è data dalla freccia $u = (3, 2)$. \diamond

Esempio. In generale, i sottospazi di \mathbb{R}^2 sono di 3 tipi, di cui due "banali" e uno proprio:

1. il sottospazio $\mathbf{0}$, costituito dal solo vettore zero: $\mathbf{0} = \{(0, 0)\}$;
2. ogni retta passante per l'origine;
3. l'intero spazio \mathbb{R}^2 .

Una situazione simile si presenta per i sottospazi di \mathbb{R}^3 , che sono di 4 tipi, di cui due "banali" e due propri:

1. il sottospazio $\mathbf{0}$, costituito dal solo vettore zero: $\mathbf{0} = \{(0, 0, 0)\}$;
2. ogni retta passante per l'origine;
3. ogni piano passante per l'origine;
4. l'intero spazio \mathbb{R}^3 .

La classificazione di questi sottospazi sarà chiarita dal concetto di dimensione, che descriveremo nelle prossime sezioni. In particolare, \mathbb{R}^3 ha dimensione 3 e i sottospazi descritti hanno dimensione, rispettivamente, 0, 1, 2 e 3. \diamond

Esempio. Sia (a_1, \dots, a_n) una n -pla assegnata. L'espressione

$$a_1x_1 + \dots + a_nx_n = 0$$

è detta un'equazione lineare omogenea, a coefficienti a_1, \dots, a_n nelle variabili x_1, \dots, x_n .

Una n -pla $(v_1, \dots, v_n) \in \mathbb{R}^n$ è detta una *soluzione* della summenzionata equazione se assegnando il valore v_i alla variabile x_i per ogni $i = 1, \dots, n$, l'espressione a sinistra dell'= $\text{vale esattamente } 0$.

Consideriamo l'insieme di tutte le soluzioni dell'equazione lineare, i.e., $X = \{(v_1, \dots, v_n) \mid \sum_{i=1}^n a_i v_i = 0\}$. Vogliamo dimostrare che X è un sottospazio di \mathbb{R}^n . Infatti, siano $v, w \in X$ e $\alpha, \beta \in \mathbb{R}$ allora

$$\sum_{i=1}^n (\alpha v_i + \beta w_i) = \alpha \sum_{i=1}^n v_i + \beta \sum_{i=1}^n w_i = \alpha 0 + \beta 0 = 0$$

e quindi $\alpha v + \beta w \in X$, sicchè X è un sottospazio. \diamond

Esempio. Sia $n \in \mathbb{N}$. Chiamiamo \mathcal{P}_n l'insieme di tutti i polinomi di grado minore o uguale a n . Siccome sia la somma tra due polinomi di grado $\leq n$ che il prodotto di un polinomio di grado $\leq n$ con una costante restituiscono un polinomio di grado $\leq n$, si ottiene che \mathcal{P}_n è un sottospazio di \mathcal{P} . \diamond

Esempio. Sia $n \in \mathbb{N}$. Chiamiamo $\mathcal{P}_n^=$ l'insieme di tutti i polinomi di grado uguale a n . Si noti che la somma tra due polinomi di grado $= n$ può essere un polinomio di grado $< n$. Ad esempio, presi $p(x) = 5 + 2x + 3x^2 - x^3$ e $q(x) = -1 + 3x - 3x^2 + x^3$ si ha $(p+q)(x) = 4 + 5x$. Quindi $\mathcal{P}_n^=$ non è chiuso rispetto all'operazione di somma, e perciò non è un sottospazio di \mathcal{P} . \diamond

Intersezione di sottospazi. Siano U e W sottospazi di V . Allora anche $U \cap W$ è un sottospazio di V . Infatti, siano $v, v' \in U \cap W$ e $\alpha, \beta \in \mathbb{R}$. Dall'appartenenza di v, v' a U abbiamo $\alpha v + \beta v' \in U$, mentre Dall'appartenenza di v, v' a W abbiamo $\alpha v + \beta v' \in W$. Quindi $\alpha v + \beta v' \in U \cap W$, c.v.d.

Abbiamo visto in un precedente esempio come l'insieme di soluzioni di un'equazione lineare omogenea sia un sottospazio di \mathbb{R}^n . In virtù di quanto appena osservato, l'insieme di n -ple che soddisfano contemporaneamente *due* equazioni lineari omogenee è anch'esso un sottospazio di \mathbb{R}^n . Infatti, esso è l'intersezione fra lo spazio delle soluzioni della prima equazione e quello delle soluzioni della seconda. Più in generale, l'insieme di n -ple che soddisfano contemporaneamente m equazioni lineari omogenee è un sottospazio di \mathbb{R}^n . Ad esempio, l'insieme definito da

$$\{(x, y, z, w) \in \mathbb{R}^4 \mid (x - 2y + w = 0) \wedge (2x + y - z = 0) \wedge (x + 3w = 0)\}$$

è un sottospazio di \mathbb{R}^4 .

3.1 Dipendenza e indipendenza lineare

Definizione 18: [Combinazione lineare] Sia V uno spazio vettoriale e siano $v_1, \dots, v_n \in V$ e $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. L'espressione

$$\lambda_1 v_1 + \dots + \lambda_n v_n$$

viene detta una *combinazione lineare* dei vettori v_1, \dots, v_n con *coefficienti* $\lambda_1, \dots, \lambda_n$.

Diciamo che il vettore $w = \sum_{i=1}^n \lambda_i v_i$ è *generato* dai vettori v_1, \dots, v_n . Si noti che qualsiasi insieme non vuoto di vettori può generare il vettore $\mathbf{0}$. Infatti, se $\lambda_i = 0$ per ogni i , si ha $\sum_{i=1}^n \lambda_i v_i = \mathbf{0}$ (questa combinazione lineare viene anche detta *banale*). D'altro canto, non è detto che ogni vettore $w \neq \mathbf{0}$ possa essere generato da un particolare insieme di vettori v_1, \dots, v_n . Ad esempio consideriamo i seguenti vettori in \mathbb{R}^3 :

$$v_1 = (1, 1, 0), v_2 = (-2, 3, 0), v_3 = (0, 2, 0) \quad \text{e} \quad w = (0, 0, 1)$$

Se guardiamo alla terza componente di v_1, v_2 e v_3 notiamo che essa è 0. Quindi in ogni combinazione lineare di v_1, v_2, v_3 tale componente sarà ancora 0. Ma la terza componente di w è diversa da zero e quindi w non può essere generato da v_1, v_2, v_3 .

Un modo alternativo di indicare una combinazione lineare di v_1, \dots, v_n è

$$\sum_{v \in W} \lambda_v v$$

dove $W = \{v_1, \dots, v_n\}$ e $\lambda_v \in \mathbb{R}$ per ogni $v \in W$. Nel seguito utilizzeremo qualche volta anche questo tipo di notazione.

Definizione 19: [Spazio lineare] Sia V uno spazio vettoriale e sia $W = \{v_1, \dots, v_n\} \subseteq V$. L'insieme

$$\{\lambda_1 v_1 + \dots + \lambda_n v_n \mid \lambda_1, \dots, \lambda_n \in \mathbb{R}\}$$

dato da tutte le combinazioni lineari di v_1, \dots, v_n è detto lo *spazio lineare generato da* v_1, \dots, v_n ed è indicato con

$$\mathcal{L} \langle v_1, \dots, v_n \rangle$$

o con

$$\mathcal{L} \langle W \rangle$$

Esempio. Siano $v_1 = (1, 0, 0)$, $v_2 = (-2, 3, 0)$, e $v_3 = (0, 2, 0)$. Allora

$$\mathcal{L} \langle v_1, v_2, v_3 \rangle = \{(\alpha - 2\beta, 3\beta + 2\gamma, 0) \mid \alpha, \beta, \gamma \in \mathbb{R}\}$$

◇

Esempio. Siano $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$, e $v_3 = (0, 0, 1)$. Allora

$$\mathcal{L} \langle v_1, v_2, v_3 \rangle = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$$

Quindi lo spazio lineare generato da v_1, v_2, v_3 coincide con \mathbb{R}^3 . \diamond

È facile notare che per ogni insieme di vettori $v_1, \dots, v_n \in V$ si ha che $\mathcal{L}\langle v_1, \dots, v_n \rangle$ è un sottospazio vettoriale di V . Infatti, siano $\alpha, \beta \in \mathbb{R}$ e $u, w \in \mathcal{L}\langle v_1, \dots, v_n \rangle$ con

$$u = \sum_{i=1}^n x_i v_i, \quad w = \sum_{i=1}^n y_i v_i$$

allora

$$\alpha u + \beta w = \alpha \sum_{i=1}^n x_i v_i + \beta \sum_{i=1}^n y_i v_i = \sum_{i=1}^n (\alpha x_i + \beta y_i) v_i \in \mathcal{L}\langle v_1, \dots, v_n \rangle$$

Definizione 20: [Dipendenza e indipendenza lineare] Sia V uno spazio vettoriale e siano $v_1, \dots, v_n \in V$. Se esistono coefficienti $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ non tutti nulli tali che

$$\lambda_1 v_1 + \dots + \lambda_n v_n = \mathbf{0}$$

allora i vettori v_1, \dots, v_n vengono detti *linearmente dipendenti*, e l'insieme $W = \{v_1, \dots, v_n\}$ viene detto un *insieme dipendente* di vettori. In caso contrario, essi sono *linearmente indipendenti* e W è un *insieme indipendente* di vettori.

Consideriamo i vettori $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$, e $v_3 = (0, 0, 1)$ dell'esempio precedente. Una loro generica combinazione lineare è del tipo (x, y, z) con $x, y, z \in \mathbb{R}$. Ora, si ha $(x, y, z) = (0, 0, 0)$ se e solo se $x = 0$, $y = 0$ e $z = 0$. Quindi, i vettori v_1, v_2, v_3 sono linearmente indipendenti. Siano ora $v_1 = (1, 0, 0)$, $v_2 = (-2, 3, 0)$, e $v_3 = (0, 2, 0)$. Una loro generica combinazione lineare è del tipo $(x - 2y, 3y + 2z, 0)$ con $x, y, z \in \mathbb{R}$. Presi $x = 2$, $y = 1$ e $z = -3/2$ si ha

$$2(1, 0, 0) + 1(-2, 3, 0) - \frac{3}{2}(0, 2, 0) = (2 - 2 + 0, 0 + 3 - 3, 0 + 0 + 0) = (0, 0, 0)$$

e quindi v_1, v_2, v_3 non sono linearmente indipendenti.

Banalmente, possiamo notare che ogni insieme V tale che $\mathbf{0} \in V$ è un insieme dipendente. Infatti, la combinazione lineare che assegna il coefficiente 1 a $\mathbf{0}$ e il coefficiente 0 a tutti i vettori appartenenti a $V - \{\mathbf{0}\}$ è il vettore $\mathbf{0}$ stesso.

La proprietà di essere un insieme indipendente si propaga "verso il basso" da un insieme a tutti i suoi sottoinsiemi non vuoti. In particolare, se V è un insieme indipendente di vettori e $W \subseteq V$, allora anche W è un insieme indipendente. Per vedere ciò ragioniamo per assurdo. Se esistesse una combinazione lineare tale che

$$\sum_{v \in W} \lambda_v v = \mathbf{0}$$

con coefficienti λ_v non tutti nulli, allora ponendo $\lambda_v = 0$ per ogni $v \in V \setminus W$ avremmo

$$\sum_{v \in V} \lambda_v v = \sum_{v \in W} \lambda_v v + \sum_{v \in V \setminus W} 0v = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

sicché l'insieme V risulterebbe linearmente dipendente. D'altro canto, la proprietà di indipendenza non si propaga "verso l'alto": se W è un insieme indipendente e $W \subseteq V$, non è detto che V sia indipendente.

Ad esempio $V = W \cup \{\mathbf{0}\}$ è dipendente. Possiamo descrivere simili risultati per quel che riguarda i vettori linearmente dipendenti. La proprietà di essere un insieme dipendente si propaga "verso l'alto" da un insieme a tutti i suoi sovrainsiemi. In particolare, se W è un insieme dipendente di vettori e $W \subseteq V$, allora anche V è un insieme dipendente, come è immediato vedere. D'altro canto, la proprietà di dipendenza non si propaga "verso il basso": se V è un insieme dipendente e $W \subseteq V$, non è detto che W sia dipendente. Ad esempio sia $v \in V$ con $v \neq \mathbf{0}$. Allora $W = \{v\}$ è un insieme indipendente.

Per quanto queste proprietà siano molto semplici, vogliamo sottolineare l'importanza che lo studente familiarizzi fin da subito con esse e per comodità le ripetiamo ulteriormente nella forma di due regole:

- (i) Se ad un insieme di vettori linearmente indipendenti togliamo alcuni vettori, i vettori restanti sono ancora linearmente indipendenti. Se invece aggiungiamo alcuni vettori non è più detto che l'insieme risultante sia ancora indipendente.
- (ii) Se ad un insieme di vettori linearmente dipendenti aggiungiamo alcuni vettori, l'insieme risultante è ancora dipendente. Se invece togliamo alcuni vettori non è più detto che vettori restanti siano ancora linearmente dipendenti.

Per gli insiemi indipendenti vale un'importante caratterizzazione che è la seguente.

TEOREMA 4: I vettori v_1, \dots, v_n sono linearmente indipendenti se e solo se nessuno di essi può essere espresso come combinazione lineare degli altri.

Dim: (Se:) Supponiamo che i v_i siano linearmente indipendenti ma che uno di essi sia combinazione lineare degli altri. Senza perdita di generalità, supponiamo che v_n sia combinazione lineare di v_1, \dots, v_{n-1} , e sia

$$v_n = \alpha_1 v_1 + \dots + \alpha_{n-1} v_{n-1}$$

Ma allora

$$\alpha_1 v_1 + \dots + \alpha_{n-1} v_{n-1} - v_n = \mathbf{0}$$

ed esisterebbe una combinazione lineare nulla a coefficienti non tutti nulli dei v_i : assurdo.

(Solo se:) Siano i v_i tali che nessuno di essi può essere espresso come combinazione lineare degli altri. Supponiamo ora che siano linearmente dipendenti. In questo caso, esisterebbero coefficienti α_i non tutti nulli tali che

$$\sum_{i=1}^n \alpha_i v_i = \mathbf{0}$$

Sia j un indice per cui $\alpha_j \neq 0$. Ma allora

$$v_j = \sum_{i \neq j} \left(\frac{-\alpha_i}{\alpha_j} \right) v_i$$

sicché uno dei vettori sarebbe combinazione lineare degli altri: assurdo.



3.2 Basi e dimensione

Dato uno spazio vettoriale V non banale, esistono sempre insiemi indipendenti (ad esempio, ogni singolo vettore $v \neq \mathbf{0}$ fornisce un insieme indipendente). È legittimo chiedersi quanto grande possa essere, al massimo, un insieme indipendente. In particolare, la cardinalità massima di un insieme indipendente può risultare finita (nel qual caso diremo che V ha *dimensione finita*) o infinita (nel qual caso diremo che V ha *dimensione infinita*).

Consideriamo il caso $V = \mathbb{R}$. Sia $W = \{v, u\}$ un qualunque insieme di due numeri reali. Senza perdita di generalità, sia $u \neq 0$. Allora la combinazione lineare

$$(-1)v + \frac{v}{u} \cdot u = 0$$

è una combinazione lineare nulla a coefficienti non nulli e quindi W è linearmente dipendente. Siccome la dipendenza lineare vale per tutti i sovrainsiemi di un insieme dipendente si deduce che ogni insieme $A \subseteq \mathbb{R}$ con $|A| \geq 2$ è linearmente dipendente. Quindi \mathbb{R} ha dimensione finita.

Si consideri ora lo spazio vettoriale \mathcal{P} dei polinomi di grado arbitrario. Fissato $n \in \mathbb{N}$, si prendano i polinomi $p_0(x) = 1, p_1(x) = x, p_2(x) = x^2$ eccetera fino a $p_n(x) = x^n$. Questi $n+1$ polinomi sono linearmente indipendenti, per cui non esiste un limite superiore al numero di vettori indipendenti in \mathcal{P} che quindi ha dimensione infinita.

Nel seguito ci occuperemo esclusivamente di spazi vettoriali di dimensione finita: se non specificato altrimenti, per spazio vettoriale intenderemo sempre spazio vettoriale di dimensione finita.

Dato un insieme A e una proprietà p definita sui sottoinsiemi di A , diremo che B è *massimale* (rispetto a p) se la proprietà vale per B ma non vale per alcun insieme $B \cup \{v\}$ con $v \in A - B$. In pratica, l'aggiunta di un qualsiasi elemento a B rende falsa la proprietà. Diremo inoltre che B è *massimo* se la proprietà vale per B e non vale per alcun $C \subseteq A$ con $|C| > |B|$. Quindi, tutti gli insiemi massimi hanno la stessa cardinalità.

Definiamo la *dimensione* di uno spazio come segue:

Definizione 21: [Dimensione] Sia V uno spazio vettoriale e sia W un sottoinsieme massimo di vettori linearmente indipendenti. Sia $n = |W|$. Allora n viene detta la *dimensione* di V , indicata con $\dim(V)$.

Se $V = \{\mathbf{0}\}$ è lo spazio banale allora in V non esistono insiemi indipendenti. Il massimo insieme indipendente è vuoto e quindi $\dim(\{\mathbf{0}\}) = 0$.

Dimostriamo che in uno spazio vettoriale gli insiemi indipendenti massimi consistono di vettori che generano tutto V :

TEOREMA 5: Sia V uno spazio vettoriale e sia $W = \{v_1, \dots, v_n\}$ un insieme indipendente. Allora W è massimo se e solo se

$$\mathcal{L}\langle W \rangle = V$$

Dim: (se:) Sia W un insieme indipendente massimo. Supponiamo, per assurdo, che esista $v \in V - \mathcal{L}\langle W \rangle$. Siccome W è massimo, i vettori v, v_1, \dots, v_n sono dipendenti ed esistono coefficienti $\lambda, \lambda_1, \dots, \lambda_n$ non tutti

nulli tali che

$$\lambda v + \sum_{i=1}^n \lambda_i v_i = \mathbf{0}$$

Deve essere $\lambda \neq 0$ o si avrebbe una combinazione lineare nulla a coefficienti non nulli di v_1, \dots, v_n . Ma allora

$$v = -\frac{1}{\lambda} \sum_{i=1}^n \lambda_i v_i = \sum_{i=1}^n \left(-\frac{\lambda_i}{\lambda}\right) v_i$$

e quindi $v \in \mathcal{L}\langle W \rangle$, assurdo.

(solo se:) Sia $\mathcal{L}\langle W \rangle = V$ e siano u_1, \dots, u_k vettori linearmente indipendenti. Descriviamo una funzione iniettiva $f : \{1, \dots, k\} \mapsto \{1, \dots, n\}$ dimostrando così che $k \leq n$ e quindi W è massimo. In pratica, utilizziamo un procedimento in cui, per ogni $i = 1, \dots, k$ rimpiazziamo in W uno dei vettori v_j con u_i . Se v_t è rimpiazzato da u_i poniamo $f(i) = t$.

Esistono coefficienti $\lambda_1, \dots, \lambda_n$ tali che $u_1 = \sum_{i=1}^n \lambda_i v_i$. Siccome $u_1 \neq \mathbf{0}$, non tutti i coefficienti sono nulli ed esiste un indice t tale che $\lambda_t \neq 0$. Poniamo $f(1) = t$ e $B_1 = \{1, \dots, n\} \setminus \{t\}$. Si ha

$$v_t = \sum_{i \in B_1} \left(\frac{-\lambda_i}{\lambda_t}\right) v_i + \frac{1}{\lambda_t} u_1$$

Notiamo che l'insieme

$$W_1 = \{v_i | i \in B_1\} \cup \{u_1\}$$

ottenuto da W rimpiazzando v_t con u_1 genera ancora tutto V . Infatti, per ogni $w \in V$ esistono coefficienti $\alpha_1, \dots, \alpha_n$ tali che

$$w = \sum_{i=1}^n \alpha_i v_i \tag{3.1}$$

$$= \sum_{i \in B_1} \alpha_i v_i + \alpha_t v_t \tag{3.2}$$

$$= \sum_{i \in B_1} \alpha_i v_i + \alpha_t \left(\sum_{i \in B_1} \left(\frac{-\lambda_i}{\lambda_t}\right) v_i + \frac{1}{\lambda_t} u_1 \right) \tag{3.3}$$

$$= \sum_{i \in B_1} \left(\alpha_i - \frac{\alpha_t \lambda_i}{\lambda_t} \right) v_i + \frac{\alpha_t}{\lambda_t} u_1 \tag{3.4}$$

e quindi $w \in \mathcal{L}\langle W_1 \rangle$. Proseguendo nelle sostituzioni, supponiamo dopo $l < n$ passi di avere

$$B_l = \{1, \dots, n\} - \{f(1), \dots, f(l)\}$$

e che

$$W_l = \{v_i | i \in B_l\} \cup \{u_i | i = 1, \dots, l\}$$

generi tutto V . Allora esistono coefficienti μ_i , per $i \in B_l$, e $\lambda_1, \dots, \lambda_l$ tali che

$$u_{l+1} = \sum_{i \in B_l} \mu_i v_i + \sum_{i=1}^l \lambda_i u_i$$

Siccome u_{l+1} non può essere combinazione lineare di u_1, \dots, u_l , esiste un indice $t \in B_l$ tale che $\mu_t \neq 0$. Poniamo $f(l+1) = t$ e $B_{l+1} = B_l - \{t\}$. Si ha

$$v_t = \sum_{i \in B_{l+1}} \left(\frac{-\mu_i}{\mu_t} \right) v_i + \sum_{i=1}^l \left(\frac{-\lambda_i}{\mu_t} \right) u_i + \frac{1}{\mu_t} u_{l+1}$$

Sia $W_{l+1} = \{v_i | i \in B_{l+1}\} \cup \{u_i | i = 1, \dots, l+1\}$. Facciamo vedere che anche W_{l+1} genera tutto V . Per ogni $w \in V$ esistono coefficienti α_i con $i = 1, \dots, l$ e β_i con $i \in B_l$ tali che

$$w = \sum_{i \in B_l} \beta_i v_i + \sum_{i=1}^l \alpha_i u_i \quad (3.5)$$

$$= \sum_{i \in B_{l+1}} \beta_i v_i + \beta_t v_t + \sum_{i=1}^l \alpha_i u_i \quad (3.6)$$

$$= \sum_{i \in B_{l+1}} \beta_i v_i + \beta_t \left(\sum_{i \in B_{l+1}} \left(\frac{-\mu_i}{\mu_t} \right) v_i + \sum_{i=1}^l \left(\frac{-\lambda_i}{\mu_t} \right) u_i + \frac{1}{\mu_t} u_{l+1} \right) + \sum_{i=1}^l \alpha_i u_i \quad (3.7)$$

$$= \sum_{i \in B_{l+1}} \left(\beta_i - \frac{\beta_t \mu_i}{\mu_t} \right) v_i + \sum_{i=1}^l \left(\alpha_i - \frac{\beta_t \lambda_i}{\mu_t} \right) u_i + \frac{\beta_t}{\mu_t} u_{l+1} \quad (3.8)$$

e quindi $w \in \mathcal{L} \langle W_{l+1} \rangle$. Il processo appena descritto può essere ripetuto finchè $l = k$. A quel punto la mappa f appena definita ha associato ai vettori u_1, \dots, u_k vettori distinti $v_{f(1)}, \dots, v_{f(k)}$ sicchè $k < n$. ♣

Dal teorema appena dimostrato segue un risultato fondamentale relativo agli spazi vettoriali, ossia che tutti gli insiemi indipendenti massimali hanno la stessa cardinalità e, in particolare, essi sono tutti massimi:

COROLLARIO 6: Sia V uno spazio vettoriale. Ogni insieme $W \subseteq V$ massimale di vettori linearmente indipendenti è massimo.

Dim: Sia $W = \{v_1, \dots, v_m\}$ un insieme indipendente massimale. Per ogni vettore $w \in V \setminus W$, i vettori w, v_1, \dots, v_m sono dipendenti e quindi esiste una loro combinazione lineare nulla $\lambda_w w + \sum_{i=1}^m \lambda_i v_i$ non banale. Deve essere $\lambda_w \neq 0$ o i vettori v_i sarebbero dipendenti, e quindi

$$w = \sum_{i=1}^m \left(\frac{-\lambda_i}{\lambda_w} \right) v_i$$

Quindi $w \in \mathcal{L} \langle W \rangle$ e $\mathcal{L} \langle W \rangle = V$. Dal teorema 5 concludiamo che W è massimo. ♣

In base al corollario 6, se n è la dimensione di V e W è un qualsiasi insieme di vettori linearmente indipendenti con $|W| = k < n$ allora esistono sempre $n - k$ vettori che possono essere aggiunti a W in modo da ottenere un insieme indipendente massimo.

Definizione 22: [Base] Sia V uno spazio vettoriale. Una sequenza (v_1, \dots, v_n) di vettori è detta una *base* di V se

- (i) $\{v_1, \dots, v_n\}$ è un insieme indipendente
- (ii) $\mathcal{L}\langle v_1, \dots, v_n \rangle = V$

Dal teorema 5 segue che (v_1, \dots, v_n) è una base di V se e solo se $\{v_1, \dots, v_n\}$ è un insieme indipendente massimo. Questa è una proprietà molto importante delle basi, per cui la sottolineiamo sotto forma del seguente corollario:

COROLLARIO 7: Sia V uno spazio vettoriale. Allora, tutte le basi di V hanno la stessa cardinalità.

Lemma 8: Sia V uno spazio vettoriale e sia W un suo sottospazio non banale. Allora $\dim(W) \leq \dim(V)$ e $\dim(W) = \dim(V)$ implica $V = W$.

Dim: Sia $\dim(W) = k$ e sia $S = \{w_1, \dots, w_k\}$ un insieme indipendente massimo in W . Sia $T \subset V$ un insieme di vettori tali che $S \cup T$ è un insieme indipendente massimo in V . Essendo $\dim(V) = |S| + |T| = \dim(W) + |T|$ si ha $\dim(W) \leq \dim(T)$. Se $T = \emptyset$ allora i vettori di S sono anche una base di V . Quindi, ogni elemento $v \in V$ può essere espresso come combinazione lineare dei w_i da cui $V \subseteq W$, ed essendo anche $W \subseteq V$ ne segue $V = W$. ♣

Esempio. Abbiamo visto che $\dim(\mathbb{R}) < 2$. Siccome $\mathbb{R} \neq \{0\}$ abbiamo $\dim(\mathbb{R}) = 1$ e (v) è una base di \mathbb{R} per ogni $v \in \mathbb{R} - \{0\}$. ◇

Esempio. Abbiamo visto che in \mathbb{R}^3 i vettori $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$ sono linearmente indipendenti. Inoltre, per ogni $v = (x, y, z) \in \mathbb{R}^3$ si ha

$$v = x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1)$$

e quindi e_1, e_2, e_3 generano \mathbb{R}^3 . Possiamo concludere che (e_1, e_2, e_3) è una base di \mathbb{R}^3 e $\dim(\mathbb{R}^3) = 3$. ◇

Definizione 23: [Base canonica di \mathbb{R}^n] Nello spazio vettoriale \mathbb{R}^n si considerino le n -ple $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, \dots , $e_n = (0, \dots, 0, 1)$ (i.e., ogni e_i ha un 1 in posizione i e uno 0 in ogni altra posizione). È facile vedere che i vettori e_i sono linearmente indipendenti e che generano tutto \mathbb{R}^n . La base

$$\mathcal{C} = (e_1, \dots, e_n)$$

viene detta la *base canonica* di \mathbb{R}^n e si ha $\dim(\mathbb{R}^n) = n$.

Definizione 24: [Base canonica di $\mathbb{R}^{m \times n}$] Nello spazio vettoriale $\mathbb{R}^{m \times n}$ delle matrici $m \times n$, si considerino le matrici E_{ij} , per $i = 1, \dots, m$ e $j = 1, \dots, n$, definite da

$$E_{ij}[h, k] = \begin{cases} 1 & \text{se } h = i, k = j \\ 0 & \text{altrimenti} \end{cases}$$

Ad esempio, per $m = 2, n = 3$ tali matrici sono

$$\begin{aligned} E_{11} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & E_{12} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & E_{13} &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \\ E_{21} &= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} & E_{22} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} & E_{23} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

È facile vedere che le matrici appena definite sono linearmente indipendenti e generano tutto $\mathbb{R}^{m \times n}$. Tali matrici costituiscono la base canonica di $\mathbb{R}^{m \times n}$ e si ha $\dim(\mathbb{R}^{m \times n}) = mn$.

Esempio. Nello spazio vettoriale \mathcal{P}_n dei polinomi di grado $\leq n$ i vettori $p_i(x) = x^i$, per $i = 0, \dots, n$, sono linearmente indipendenti e generano \mathcal{P}_n . Essi ne costituiscono quindi una base e si ha $\dim(\mathcal{P}_n) = n + 1$. \diamond

Definizione 25: [Retta, piano, iperpiano] Siano V uno spazio vettoriale di dimensione n e W un suo sottospazio. Se $\dim(W) = 1$ diciamo che W è una *retta* (passante per l'origine) di V . Se $\dim(W) = 2$ diciamo che W è un *piano* (passante per l'origine) di V . Se $\dim(W) = n - 1$ diciamo che W è un *iperpiano* (passante per l'origine) di V .

3.2.1 Sistemi di coordinate

Sia V uno spazio vettoriale e $\mathcal{B} = (v_1, \dots, v_n)$ una sua base. Per ogni vettore $v \in V$ esistono coefficienti $x_1, \dots, x_n \in \mathbb{R}$ tali che $v = \sum_{i=1}^n x_i v_i$. È immediato dimostrare che tali coefficienti sono univocamente determinati da v :

TEOREMA 9: Sia V uno spazio vettoriale e $\mathcal{B} = (v_1, \dots, v_n)$ una sua base. Siano x_1, \dots, x_n e y_1, \dots, y_n coefficienti tali che

$$\sum_{i=1}^n x_i v_i = \sum_{i=1}^n y_i v_i$$

Allora $x_i = y_i$ per ogni $i = 1, \dots, n$.

Dim: Si ha $\sum_{i=1}^n (x_i - y_i) v_i = \mathbf{0}$ da cui la tesi. ♣

Consideriamo l'applicazione di V in \mathbb{R}^n che ad ogni vettore di V associa la n -pla dei coefficienti necessari per esprimerlo come combinazione lineare dei vettori della base \mathcal{B} :

$$v \mapsto (x_1, \dots, x_n) \iff v = x_1 v_1 + \dots + x_n v_n$$

È facile verificare che si tratta di un'applicazione iniettiva e suriettiva, che quindi stabilisce una biiezione tra V e \mathbb{R}^n . Denotiamo l'immagine di v con la notazione $(v)_{\mathcal{B}}$, i.e.,

$$(v)_{\mathcal{B}} = (x_1, \dots, x_n)$$

e diciamo che $(v)_{\mathcal{B}}$ è la *rappresentazione di v rispetto alla base \mathcal{B}* . I coefficienti x_i verranno anche detti le *coordinate di v rispetto alla base \mathcal{B}* . Si noti che per i vettori della base si ha sempre

$$(v_1)_{\mathcal{B}} = (1, 0, \dots, 0), \quad (v_2)_{\mathcal{B}} = (0, 1, 0, \dots, 0), \quad \dots, \quad (v_n)_{\mathcal{B}} = (0, \dots, 0, 1)$$

In alcune situazioni può essere utile considerare i coefficienti di un vettore rispetto a una base come elementi di un vettore-colonna anzichè di una n -pla. Denoteremo allora con $\mathbf{C}_{\mathcal{B}}^v$ il vettore-colonna

$$\mathbf{C}_{\mathcal{B}}^v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

tale che $v = \sum_{i=1}^n x_i v_i$. Si noti che $\mathbf{C}_{\mathcal{B}}^v = {}^t(v)_{\mathcal{B}}$.

Esempio. Consideriamo la base $\mathcal{B} = (1, x, x^2, x^3, x^4)$ dello spazio vettoriale \mathcal{P}_4 dei polinomi di grado ≤ 4 , e sia p il polinomio $p = -2x^3 + 3x^2 - 5$. Allora

$$(p)_{\mathcal{B}} = (-5, 0, 3, -2, 0)$$

◇

Esempio. Consideriamo lo spazio vettoriale \mathbb{R}^3 e la sua base canonica \mathcal{C} data dai vettori $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. È immediato verificare che la rappresentazione rispetto alla base \mathcal{C} di una qualsiasi tripla v è la tripla v stessa. Quindi in questo caso la mappa dai vettori alle loro coordinate non è altro che l'identità su \mathbb{R}^3 .

Consideriamo l'insieme \mathcal{B} dato dai vettori $v_1 = (1, 2, 0)$, $v_2 = (3, 1, 0)$, $v_3 = (0, 0, 2)$. Verifichiamo che sono linearmente indipendenti (per ora lo facciamo con delle opportune considerazioni. Più avanti, quando avremo introdotto i sistemi lineari, vedremo un metodo più generale per determinare se un insieme di vettori siano linearmente indipendenti). Notiamo che v_2 è linearmente indipendente da v_1 , in quanto non è un suo multiplo. Inoltre v_3 non appartiene allo spazio generato da v_1 e v_2 in quanto sia v_1 che v_2 – e quindi ogni loro combinazione lineare – hanno la terza componente nulla, mentre v_3 no. In conclusione, v_1, v_2, v_3 sono linearmente indipendenti e quindi \mathcal{B} è una base di \mathbb{R}^3 . Si può verificare che

$$e_1 = -\frac{1}{5}v_1 + \frac{2}{5}v_2, \quad e_2 = \frac{3}{5}v_1 - \frac{1}{5}v_2, \quad e_3 = \frac{1}{2}v_3$$

e quindi

$$(e_1)_{\mathcal{B}} = (-1/5, 2/5, 0), \quad (e_2)_{\mathcal{B}} = (3/5, -1/5, 0), \quad (e_3)_{\mathcal{B}} = (0, 0, 1/2)$$

◇

3.3 Somme di sottospazi e somme dirette

Partiamo da alcune semplici osservazioni (le cui dimostrazioni sono lasciate al lettore). Sia V uno spazio vettoriale e U un suo sottospazio. Allora $\dim(U) \leq \dim(V)$. Inoltre, $\dim(U) < \dim(V)$ se e solo se

$U \neq V$. Uno spazio vettoriale V di dimensione $n > 1$ contiene (infiniti) sottospazi di dimensione k per ogni $k = 1, \dots, n-1$. Esiste però un unico sottospazio di dimensione zero, ed un unico di dimensione n , i.e., V stesso. L'intersezione di due sottospazi U e W è ancora un sottospazio, di dimensione $\dim(U \cap W) \leq \min\{\dim(U), \dim(W)\}$. D'altro canto, l'unione di sottospazi non è necessariamente un sottospazio. Consideriamo il seguente esempio in \mathbb{R}^2 . Siano $U = \{(x, 0), x \in \mathbb{R}\}$ e $W = \{(0, y), y \in \mathbb{R}\}$. I vettori $e_1 = (1, 0)$ ed $e_2 = (0, 1)$ appartengono a $U \cup W$, ma $e_1 + e_2 = (1, 1)$ non vi appartiene, per cui $U \cup W$ non è uno spazio vettoriale. Definiremo ora un'operazione che permette, in un certo senso, di "unire" due sottospazi per crearne uno "più grande". Cerchiamo di formalizzare meglio questi concetti.

Sia V uno spazio vettoriale e siano U e W suoi sottospazi. Definiamo la *somma* di U e W come l'insieme

$$U + W = \{u + w \mid u \in U, w \in W\}$$

È facile vedere che $U + W$ è anch'esso un sottospazio di V . Infatti, siano $\alpha, \beta \in \mathbb{R}$ e $v, v' \in U + W$ (con $v = u + w$ e $v' = u' + w'$ per opportuni $u, u' \in U$ e $w, w' \in W$). Allora si ha

$$\alpha v + \beta v' = \alpha u + \alpha w + \beta u' + \beta w' = (\alpha u + \beta u') + (\alpha w + \beta w') \in U + W$$

Per quel che riguarda le dimensioni dei sottospazi, vale la seguente relazione

$$\max\{\dim(U), \dim(W)\} \leq \dim(U + W) \leq \dim(U) + \dim(W)$$

La prima disuguaglianza deriva dal fatto che sia U che W sono contenuti in $U + W$. Infatti essendo $\mathbf{0} \in U \cap W$, si ha $U = \{u + \mathbf{0} : u \in U\}$ e $W = \{\mathbf{0} + w : w \in W\}$. Per quel che riguarda la seconda, siano (u_1, \dots, u_h) una base di U e (w_1, \dots, w_k) una base di W . È facile vedere che $\{u_1, \dots, u_h, w_1, \dots, w_k\}$ genera $U + W$, e quindi $\dim(U + W) \leq h + k$, come volevasi dimostrare.

Se $\{u_1, \dots, u_h\} \cap \{w_1, \dots, w_k\} = \emptyset$ allora $\dim(U + W) = \dim(U) + \dim(W)$. D'altro canto, è immediato vedere come siano possibili situazioni in cui $\dim(U + W) < \dim(U) + \dim(W)$. Ad esempio se sommassimo due sottospazi di dimensione $> n/2$ e valesse l'uguaglianza, allora si avrebbe $\dim(U + W) > n = \dim(V)$, impossibile. In generale, se sommiamo due sottospazi la cui somma delle dimensioni è maggiore di n , necessariamente deve essere $U \cap W \neq \emptyset$.

Esempio. Siano V uno spazio vettoriale di dimensione 3, U un sottospazio di dimensione 1 (retta passante per l'origine) e W un sottospazio di dimensione 2 (piano passante per l'origine). Allora

$$U + W = \begin{cases} V & \text{se } U \not\subset W \\ W & \text{se } U \subset W. \end{cases}$$

◇

Esempio. Sia V uno spazio vettoriale di dimensione 3, e siano U e W sottospazi di dimensione 1 (rette passanti per l'origine). Allora

$$U + W = \begin{cases} \text{il piano contenente } U \text{ e } W & \text{se } U \neq W \\ U & \text{se } U = W. \end{cases}$$

◇

TEOREMA 10: (Formula di Grassmann) Siano V uno spazio vettoriale e U, W due suoi sottospazi. Allora

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$$

Dim: Se $U \cap W = \{\mathbf{0}\}$ allora è facile dimostrare la validità della formula (esercizio). Sia allora $\dim(U \cap W) = p > 1$ e sia (v_1, \dots, v_p) una base di $U \cap W$. Dette $r = \dim(U)$ e $t = \dim(W)$, esistono $r-p$ vettori u_1, \dots, u_{r-p} tali che $(v_1, \dots, v_p, u_1, \dots, u_{r-p})$ è una base di U e $t-p$ vettori w_{p+1}, \dots, w_t tali che $(v_1, \dots, v_p, w_{p+1}, \dots, w_t)$ è una base di W . Dimostriamo allora che $\mathcal{B} = (v_1, \dots, v_p, u_1, \dots, u_{r-p}, w_{p+1}, \dots, w_t)$ è una base di $U + W$, da cui segue che

$$\dim(U + W) = p + (r - p) + (t - p) = r + t - p = \dim(U) + \dim(W) - \dim(U \cap W).$$

Cominciamo col notare che i vettori in \mathcal{B} sono linearmente indipendenti. Infatti, consideriamo coefficienti α, β, γ tali che

$$\sum_{i=1}^p \alpha_i v_i + \sum_{j=1}^{r-p} \beta_j u_j + \sum_{h=1}^{t-p} \gamma_h w_h = \mathbf{0}$$

Siano $\hat{v} = \sum_{i=1}^p \alpha_i v_i$, $\hat{u} = \sum_{j=1}^{r-p} \beta_j u_j$ e $\hat{w} = \sum_{h=1}^{t-p} \gamma_h w_h$. Essendo $U \ni \hat{v} + \hat{u} = -\hat{w} \in W$, si ha $w \in U \cap W$ e quindi esistono coefficienti λ_i tali che $\sum_{i=1}^p \lambda_i v_i = \sum_{h=1}^{t-p} (-\gamma_h) w_h$, ossia

$$\sum_{i=1}^p \lambda_i v_i + \sum_{h=1}^{t-p} \gamma_h w_h = \mathbf{0}$$

Siccome $(v_1, \dots, v_p, w_{p+1}, \dots, w_t)$ sono indipendenti (base di W), tutti i coefficienti sono nulli, e, in particolare, $\gamma_h = 0$ per ogni h . Sostituendo in (3.3) si ottiene che anche $\alpha_i = 0$ e $\beta_j = 0$ per ogni i e j .

Infine, È immediato verificare che \mathcal{B} genera $U + W$ e quindi è una base di $U + W$, c.v.d. ♣

Si noti la somiglianza tra la formula di Grassmann e la formula combinatorica basata sul principio di inclusione-esclusione, in base alla quale se A e B sono sottoinsiemi di X , allora

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Definizione 26: [Somma diretta] Siano V uno spazio vettoriale e U, W due suoi sottospazi. Diremo che V è *somma diretta* di U e W , scrivendo

$$V = U \oplus W$$

se ogni vettore $v \in V$ può essere espresso, in un unico modo, da una somma $v = u + w$ con $u \in U$ e $w \in W$.

Si può dimostrare (esercizio) che $V = U \oplus W$ se e solo se

1. $V = U + W$
2. $U \cap W = \{\mathbf{0}\}$.

Inoltre, se V è somma diretta di sottospazi U e W , allora

$$\dim(V) = \dim(U) + \dim(W)$$

TEOREMA 11: Sia V uno spazio vettoriale e sia U un suo sottospazio. Allora esiste un sottospazio W tale che $V = U \oplus W$.

Dim: Sia $n = \dim(V)$ e $m = \dim(U)$. Sia u_1, \dots, u_m una base di U . Possiamo completare tale base in una base di V aggiungendo opportuni vettori w_1, \dots, w_{n-m} . A questo punto basta prendere $W = \mathcal{L}\langle w_1, \dots, w_{n-m} \rangle$ per ottenere il risultato cercato.



Si noti che, per un dato sottospazio U , possono esistere infiniti sottospazi W tali che $V = U \oplus W$. Si consideri ad esempio $V = \mathbb{R}^2$, $U = \{(\alpha, 0) | \alpha \in \mathbb{R}\}$. Per ogni vettore (a, b) con $b \neq 0$ lo spazio (la retta) $W = \{\lambda(a, b) | \lambda \in \mathbb{R}\}$ è tale che $\mathbb{R}^2 = U \oplus W$. Infatti $U \cap W = \{(0, 0)\}$ e, per ogni $(x, y) \in \mathbb{R}^2$ si ha

$$(x, y) = (x - ya/b, 0) + y/b(a, b) \in U + W.$$

Esercizio Siano $U = \{(x, y, z) \in \mathbb{R}^3 : x + 2y - z = 0\}$ e $W = \mathcal{L}\langle w \rangle$ con $w = (1, 0, 1)$.

Determinare $\dim(U + W)$ e una base di $U + W$.

Sol: Qualsiasi sia il valore di y e z abbiamo che x è univocamente determinato dall'equazione $x = -2y + z$. Quindi

$$U = \{y(-2, 1, 0) + z(1, 0, 1), y, z \in \mathbb{R}\}$$

Si ha $\dim(U) = 2$ e $((-2, 1, 0), (1, 0, 1))$ è una base di U .

Chiaramente $\dim(W) = 1$. Inoltre $U \cap W = W$ in quanto la retta W è contenuta nello spazio U . Infatti, essendo $1 + 2 \cdot 0 - 1 = 0$ il punto $w = (1, 0, 1) \in U$. Dalla formula di Grassmann otteniamo

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W) = 2 + 1 - 1 = 2.$$

Essendo $U + W = U$, la base di U è anche base di $U + W$.

Esercizio Siano $U = \{(x, y, z, t) \in \mathbb{R}^4 : x - t = 0\}$ e $W = \mathcal{L}\langle w \rangle$ con $w = (1, 0, 1, 1)$.

Determinare $\dim(U + W)$ e una base di $U + W$.

Sol: Abbiamo

$$U = \{t(1, 0, 0, 1) + y(0, 1, 0, 0) + z(0, 0, 1, 0), t, y, z \in \mathbb{R}\}$$

Quindi $\dim(U) = 3$ e $((1, 0, 0, 1), (0, 1, 0, 0), (0, 0, 1, 0))$ è una base di U .

Chiaramente $\dim(W) = 1$. Inoltre $U \cap W = W$ in quanto la retta W è contenuta nello spazio U . Infatti, essendo $1 - 1 = 0$ il punto $w = (1, 0, 1, 1) \in U$. Dalla formula di Grassmann otteniamo

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W) = 3 + 1 - 1 = 3.$$

Essendo $U + W = U$, la base di U è anche base di $U + W$.

Chapter 4

Applicazioni lineari

4.1 Introduzione

Definizione 27: [Applicazione Lineare] Siano V e W spazi vettoriali. Un'applicazione $F : V \mapsto W$ che soddisfa le seguenti proprietà

1. $F(v + w) = F(v) + F(w)$ per ogni $v, w \in V$
2. $F(\lambda v) = \lambda F(v)$ per ogni $v \in V$ e $\lambda \in \mathbb{R}$

si dice un'applicazione *lineare* (o una funzione lineare) di V in W .

Esempio. Siano $V = \mathbb{R}^3$ e $W = \mathbb{R}^2$ e sia F definita da $F(x, y, z) = (2x - z, x + y + z)$ per ogni $(x, y, z) \in \mathbb{R}^3$. Verifichiamo che F è lineare. Per ogni $(x, y, z), (x', y', z') \in \mathbb{R}^3$ si ha

$$\begin{aligned} F((x, y, z) + (x', y', z')) &= F(x + x', y + y', z + z') \\ &= (2(x + x') - (z + z'), (x + x') + (y + y') + (z + z')) \\ &= (2x - z + 2x' - z', x + y + z + x' + y' + z') \\ &= (2x - z, x + y + z) + (2x' - z', x' + y' + z') \\ &= F(x, y, z) + F(x', y', z') \end{aligned}$$

Inoltre, per ogni $\lambda \in \mathbb{R}$,

$$\begin{aligned} F(\lambda(x, y, z)) &= F(\lambda x, \lambda y, \lambda z) \\ &= (2\lambda x - \lambda z, \lambda x + \lambda y + \lambda z) \\ &= (\lambda(2x - z), \lambda(x + y + z)) \\ &= \lambda(2x - z, x + y + z) \\ &= \lambda F(x, y, z) \end{aligned}$$

e quindi F è un'applicazione lineare. \diamond

Esempio. Le seguenti sono applicazioni lineari notevoli:

- **(L'identità.)** L'applicazione $\text{id}_V : V \mapsto V$ tale che $\text{id}_V(v) = v$ per ogni $v \in V$. (Qualora lo spazio V risulti chiaro dal contesto, ci limiteremo a scrivere id anzichè id_V)
- **(L'applicazione nulla.)** L'applicazione $0 : V \mapsto V$ tale che $0(v) = \mathbf{0}$ per ogni $v \in V$

La dimostrazione (banale) che si tratti di applicazioni lineari è lasciata al lettore. \diamond

Esempio. Sia $w \in W$ un vettore non nullo di W , e sia $F : V \mapsto W$ tale che $F(v) = w$ per ogni $v \in V$. Allora F non è un'applicazione lineare. Infatti, $F(u + v) = w \neq F(u) + F(v) = 2w$ per ogni $u, v \in V$. \diamond

Si noti che per ogni applicazione lineare $F : V \mapsto W$ si ha $F(\mathbf{0}_V) = \mathbf{0}_W$. Infatti

$$F(\mathbf{0}_V) = F(\mathbf{0}_V + \mathbf{0}_V) = 2F(\mathbf{0}_V)$$

da cui, sommando $(-F(\mathbf{0}_V))$ ad entrambi i termini si ottiene $\mathbf{0}_W = F(\mathbf{0}_V)$.

Definizione 28: [Morfismi] Un'applicazione lineare tra due spazi vettoriali V e W è anche detta un *omomorfismo* tra V e W . Se $V = W$ l'applicazione è detta *endomorfismo*. Se l'applicazione lineare è iniettiva viene detta un *monomorfismo*. Se essa è suriettiva viene detta un *epimorfismo*. Se l'applicazione lineare è invertibile (i.e., è sia un monomorfismo che un epimorfismo) allora viene detta un *isomorfismo* tra V e W . In quest'ultimo caso si dice anche che V e W sono spazi vettoriali *isomorfi*.

ESERCIZIO 4.1. Siano V e W spazi vettoriali, e sia \mathcal{F} l'insieme di tutte le funzioni di dominio V e codominio W . Sia $\text{Hom}(V, W)$ l'insieme di tutti gli omomorfismi tra V e W . Si dimostri che $\text{Hom}(V, W)$ è un sottospazio di \mathcal{F} . Si dimostri infine che l'insieme di tutti gli isomorfismi tra V e W è un sottospazio vettoriale di $\text{Hom}(V, W)$. \diamond

Definizione 29: [Nucleo e Immagine]. Sia F un'applicazione lineare di V in W . Il *nucleo* di F , denotato con $\ker(F)$, è il sottoinsieme di V definito da

$$\ker(F) = \{v \in V \mid F(v) = \mathbf{0}_W\}.$$

L'*immagine* di F , denotata con $\text{im}(F)$ (o con $F(V)$), è il sottoinsieme di W definito da

$$\text{im}(F) = \{w \in W \mid \exists v \in V \text{ con } w = F(v)\}.$$

Sia il nucleo che l'immagine della F sono spazi vettoriali. In particolare

- $\ker(F)$ è un sottospazio di V

- $\text{im}(F)$ è un sottospazio di W .

Infatti, siano $v, w \in \ker(F)$ e $\lambda, \mu \in \mathbb{R}$. Allora

$$\begin{aligned} F(\lambda v + \mu w) &= \lambda F(v) + \mu F(w) \\ &= \lambda \mathbf{0} + \mu \mathbf{0} \\ &= \mathbf{0} \end{aligned}$$

il che dimostra che $\lambda v + \mu w \in \ker(F)$ e quindi $\ker(F)$ è un sottospazio di V . Inoltre, siano $u, y \in \text{im}(F)$. Allora esistono $v, w \in V$ tali che $u = F(v)$ e $y = F(w)$. Si ha allora

$$\begin{aligned} \lambda u + \mu y &= \lambda F(v) + \mu F(w) \\ &= F(\lambda v + \mu w) \end{aligned}$$

il che dimostra che $\lambda u + \mu y \in \text{im}(F)$ e quindi $\text{im}(F)$ è un sottospazio di W .

Definizione 30: [Nullità e rango] Chiamiamo *nullità* di F la dimensione di $\ker(F)$. Chiamiamo *rango* di F , denotato con $\text{rk}(F)$, la dimensione di $\text{im}(F)$.

Esempio. Se F è un'applicazione lineare di codominio \mathbb{R}^3 allora l'immagine di F può essere un sottospazio di dimensione 0, 1, 2, o 3. Quindi l'immagine di uno spazio vettoriale tramite l'applicazione F può essere

1. L'origine (in questo caso F è l'applicazione nulla)
2. Una retta passante per l'origine
3. Un piano passante per l'origine
4. Tutto \mathbb{R}^3 .

◇

Dalla definizione di applicazione lineare ricaviamo immediatamente che F è lineare se e solo se

$$F(\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n) = \lambda_1 F(v_1) + \lambda_2 F(v_2) + \cdots + \lambda_n F(v_n)$$

per ogni $v_1, \dots, v_n \in V$ e $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Quindi *le applicazioni lineari mappano ogni combinazione lineare di un insieme di vettori nella combinazione lineare, con i medesimi coefficienti, delle immagini di tali vettori.*

Consideriamo un insieme v_1, \dots, v_k di vettori in V . È facile dimostrare che l'immagine dello spazio lineare $\mathcal{L}\langle v_1, \dots, v_k \rangle$ tramite la F è lo spazio lineare generato dai vettori $F(v_1), \dots, F(v_k)$, ossia

$$F(\mathcal{L}\langle v_1, \dots, v_k \rangle) = \mathcal{L}\langle F(v_1), \dots, F(v_k) \rangle$$

Per conoscere il valore della funzione in ognuno degli infiniti vettori $v \in \mathcal{L}\langle v_1, \dots, v_k \rangle$ basta conoscere il valore della funzione in k vettori, ovvero nei generatori di $\mathcal{L}\langle v_1, \dots, v_k \rangle$. Infatti, se $v = \sum_{i=1}^k \lambda_i v_i$ si ha $F(v) = \sum_{i=1}^k \lambda_i F(v_i)$.

Se v_1, \dots, v_n sono una base di V allora $\mathcal{L}\langle v_1, \dots, v_n \rangle = V$. Quindi possiamo fare la seguente importante osservazione:

Per conoscere il valore di F in ogni vettore $v \in V$ basta conoscere il suo valore negli n vettori di una base.

Qualora v_1, \dots, v_n siano una base di V , i vettori $F(v_1), \dots, F(v_n)$ generano $\text{im}(F)$. Si ricava perciò che $\dim(\text{im}(F)) \leq n$ e $\dim(\text{im}(F)) = n$ se e solo se i vettori $F(v_i)$ sono linearmente indipendenti.

TEOREMA 12: Sia $F : V \mapsto W$ un'applicazione lineare. Allora F è iniettiva se e solo se $\ker(F) = \{\mathbf{0}\}$

Dim: (i) Sia F iniettiva e sia $v \in \ker(F)$. Essendo $F(v) = \mathbf{0}_W = F(\mathbf{0})$, dall'iniettività di F segue che $v = \mathbf{0}$.

(ii) Sia $\ker(F) = \{\mathbf{0}\}$ e siano $v_1, v_2 \in V$ tali che $F(v_1) = F(v_2)$. Abbiamo $F(v_1) - F(v_2) = \mathbf{0} = F(v_1 - v_2)$ e perciò $v_1 - v_2 \in \ker(F)$. Quindi $v_1 - v_2 = \mathbf{0}$ da cui $v_1 = v_2$ e F è iniettiva. ♣

ESERCIZIO 4.2. Sia

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

e consideriamo l'applicazione $L_A : \mathbb{R}^3 \mapsto \mathbb{R}^3$ definita da $L_A(X) = AX$. Vogliamo determinare $\text{rk}(L_A)$ e una base di $\text{im}(L_A)$. *

Sol: Consideriamo la base canonica di \mathbb{R}^3 fatta dai vettori-colonna $e_1 = {}^t(1, 0, 0)$, $e_2 = {}^t(0, 1, 0)$ ed $e_3 = {}^t(0, 0, 1)$. Abbiamo $L_A(e_1) = {}^t(2, 4, 0)$, $L_A(e_2) = {}^t(0, 1, -1)$ ed $L_A(e_3) = {}^t(3, 2, 4)$. Le immagini della base di \mathbb{R}^3 sono le colonne di A , le quali quindi generano $\text{im}(L_A)$. Le prime due colonne sono chiaramente linearmente indipendenti. Per vedere se la terza colonna dipende dalle prime due, ci chiediamo se esistono coefficienti $\alpha, \beta \in \mathbb{R}$ tali che

$$\begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix} = \alpha \begin{pmatrix} 2 \\ 4 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

Dalla prima componente ricaviamo che deve essere $\alpha = 3/2$. Sostituendo nell'equazione relativa alla seconda componente, abbiamo che

$$2 = 4\alpha + \beta = 6 + \beta$$

e quindi deve essere $\beta = -4$. Sostituendo i valori di α e β nell'equazione della terza componente, si ha che tale equazione è effettivamente soddisfatta, e quindi la terza colonna è combinazione lineare delle prime due con coefficienti $\alpha = 3/2$ e $\beta = -4$.

In conclusione $\text{rk}(L_A) = \dim(\mathcal{L}\langle A^1, A^2, A^3 \rangle) = 2$ e le prime due colonne di A costituiscono una base di $\text{im}(L_A)$. ◇

*In questo esercizio consideriamo \mathbb{R}^3 come uno spazio di triple rappresentate da vettori-colonna. Più correttamente, avremmo dovuto riferirci a $\mathbb{R}^{3 \times 1}$ visto che tipicamente, abbiamo sempre considerato le triple come vettori-riga. In generale non ci preoccuperemo mai di distinguere fra \mathbb{R}^n e $\mathbb{R}^{n \times 1}$ o $\mathbb{R}^{1 \times n}$ a meno che questo non sia assolutamente necessario, il che è molto raro.

4.1.1 Teorema della nullità + rango

TEOREMA 13: Sia $F : V \mapsto W$ un'applicazione lineare. Allora

$$\dim(\ker(F)) + \dim(\operatorname{im}(F)) = \dim V$$

In altre parole, la nullità di F più il rango di F sono uguali alla dimensione di V , i.e., del dominio di F .

Dim: Siano $n = \dim V$ e $k = \dim(\ker(F))$. Sia (v_1, \dots, v_k) una base di $\ker(F)$ e sia $(v_1, \dots, v_k, v_{k+1}, \dots, v_n)$ una qualsiasi sua estensione ad una base di V .

Sappiamo che

$$\operatorname{im}(F) = \mathcal{L} \langle F(v_1), \dots, F(v_n) \rangle$$

ma, essendo $F(v_1) = \dots = F(v_k) = \mathbf{0}_W$, si ha

$$\operatorname{im}(F) = \mathcal{L} \langle F(v_{k+1}), \dots, F(v_n) \rangle$$

Per ottenere la tesi sarà allora sufficiente dimostrare che $F(v_{k+1}), \dots, F(v_n)$ sono linearmente indipendenti. Siano $\lambda_{k+1}, \dots, \lambda_n \in \mathbb{R}$ tali che

$$\sum_{i=k+1}^n \lambda_i F(v_i) = \mathbf{0}_W$$

Dalla linearità di F segue che $F(\sum_{i=k+1}^n \lambda_i v_i) = \mathbf{0}$ e quindi $\sum_{i=k+1}^n \lambda_i v_i \in \ker(F)$. Siccome v_1, \dots, v_k è una base di $\ker(F)$, esistono coefficienti $\alpha_1, \dots, \alpha_k$ tali che

$$\sum_{i=k+1}^n \lambda_i v_i = \sum_{i=1}^k \alpha_i v_i$$

e quindi

$$-\alpha_1 v_1 + \dots - \alpha_k v_k + \lambda_{k+1} v_{k+1} + \dots + \lambda_n v_n = \mathbf{0}$$

è una combinazione lineare nulla di vettori di una base di V . Ne segue che tutti i coefficienti sono nulli e, in particolare, $\lambda_{k+1}, \dots, \lambda_n = 0$. Quindi i vettori $F(v_{k+1}), \dots, F(v_n)$ sono indipendenti e sono una base di $\operatorname{im}(F)$, da cui

$$\dim(\operatorname{im}(F)) = n - k = n - \dim(\ker(F)).$$



ESERCIZIO 4.3. Sia $W = \{A \in \mathbb{R}^{n \times n} : \operatorname{tr}(A) = 0\}$. Determinare $\dim W$ e una base di W .

Sol: L'applicazione

$$\operatorname{tr} : \mathbb{R}^{n \times n} \mapsto \mathbb{R}$$

è lineare. In particolare, $W = \ker(\operatorname{tr})$ e $\operatorname{im}(\operatorname{tr}) = \mathbb{R}$. Dal teorema della nullità e del rango abbiamo

$$\dim W = \dim(\mathbb{R}^{n \times n}) - \dim(\mathbb{R}) = n^2 - 1$$

Per quel che riguarda la base di W , consideriamo le $n^2 - n$ matrici E_{ij} definite, per ogni $1 \leq i \leq n$ e $1 \leq j \leq n$, con $i \neq j$, da

$$E_{ij}[k, h] = \begin{cases} 1 & \text{se } k = i, h = j \\ 0 & \text{altrimenti} \end{cases}$$

A queste aggiungiamo le $n - 1$ matrici T_{ii} , per $1 \leq i \leq n - 1$, definite da

$$T_{ii}[k, h] = \begin{cases} 1 & \text{se } k = i, h = i \\ -1 & \text{se } k = i + 1, h = i + 1 \\ 0 & \text{altrimenti} \end{cases}$$

È facile vedere che le matrici appena definite sono tutte linearmente indipendenti fra loro. Essendo in totale $n^2 - n + n - 1 = n^2 - 1$, formano una base di $\ker(\text{tr})$.

Ad esempio, per $n = 3$ tale base è

$$\begin{aligned} E_{12} &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & E_{13} &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & E_{21} &= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & E_{23} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \\ E_{31} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} & E_{32} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} & T_{11} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & T_{22} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \end{aligned}$$

◇

Nella dimostrazione del Teorema 13 abbiamo visto che per i vettori linearmente indipendenti v_{k+1}, \dots, v_n , anche le rispettive immagini $F(v_{k+1}), \dots, F(v_n)$, erano linearmente indipendenti. Questo fatto *non è vero in generale*, ossia, vettori linearmente indipendenti di V possono essere mappati su immagini linearmente dipendenti in W . Ad esempio, se F è la proiezione sull'asse x , i punti $v_1 = (3, 1, 0)$ e $v_2 = (6, 2, 1)$ vengono mappati in $F(v_1) = (3, 0, 0)$ e $F(v_2) = (6, 0, 0)$. In questo esempio, $F(v_2)$ è multiplo di $F(v_1)$, ma v_2 non è multiplo di v_1 . Quello che però è sempre vero per un'applicazione lineare è che vettori linearmente indipendenti in $\text{im}(F)$ provengono da vettori linearmente indipendenti in V . Abbiamo infatti

Lemma 14: Sia $F : V \mapsto W$ un'applicazione lineare, e siano $F(v_1), \dots, F(v_p)$ vettori linearmente indipendenti in $\text{im}(F)$. Allora i vettori v_1, \dots, v_p sono linearmente indipendenti in V .

Dim: Siano $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ degli scalari tali che

$$\sum_{i=1}^p \lambda_i v_i = \mathbf{0}$$

Essendo

$$\sum_{i=1}^p \lambda_i F(v_i) = F\left(\sum_{i=1}^p \lambda_i v_i\right) = F(\mathbf{0}) = \mathbf{0}_W$$

dall'indipendenza lineare degli $F(v_i)$ segue che $\lambda_1 = \dots = \lambda_p = 0$. Quindi anche v_1, \dots, v_p sono linearmente indipendenti. ♣

4.1.2 Isomorfismo tra spazi della stessa dimensione

In questa sezione dimostreremo che *tutti gli spazi della medesima dimensione sono isomorfi tra loro*. Per giungere a questo risultato cominciamo con l'osservare come l'isomorfismo definisca una relazione di equivalenza tra spazi vettoriali. Infatti, valgono le seguenti proprietà:

1. **Riflessiva:** ogni spazio vettoriale è banalmente isomorfo a se stesso. L'isomorfismo più semplice è in questo caso l'identità.
2. **Commutativa:** supponiamo che lo spazio vettoriale V sia isomorfo a W . Allora esiste un'applicazione lineare invertibile F da V a W . L'applicazione F^{-1} è un'applicazione lineare invertibile da W a V e quindi anche W è isomorfo a V .
3. **Transitiva:** supponiamo che V sia isomorfo a W e W sia isomorfo a U . Allora esistono isomorfismi $F : V \mapsto W$ e $G : W \mapsto U$. È facile verificare che l'applicazione composta GF è un isomorfismo tra V e U .

Sia ora V un qualsiasi spazio vettoriale di dimensione n . Dimostreremo come V sia isomorfo a \mathbb{R}^n . Per la proprietà transitiva dell'isomorfismo, questo implica anche che V è isomorfo a qualsiasi altro spazio vettoriale di dimensione n .

TEOREMA 15: Sia V uno spazio vettoriale di dimensione n . Allora V è isomorfo a \mathbb{R}^n .

Dim: Sia $\mathcal{B} = (v_1, \dots, v_n)$ una base di V . Consideriamo l'applicazione $\pi_{\mathcal{B}} : V \mapsto \mathbb{R}^n$ che ad ogni vettore $v \in V$ associa le sue coordinate rispetto alla base \mathcal{B} , i.e., se $v = x_1 v_1 + \dots + x_n v_n$, si ha

$$\pi_{\mathcal{B}}(x_1 v_1 + \dots + x_n v_n) = (x_1, \dots, x_n)$$

L'applicazione $\pi_{\mathcal{B}}$ è chiaramente lineare. Dimostriamo che è sia iniettiva che suriettiva.

Iniettività: siano $v, w \in V$ tali che $\pi_{\mathcal{B}}(v) = \pi_{\mathcal{B}}(w) = (x_1, \dots, x_n)$. Allora $v = \sum_{i=1}^n x_i v_i$ ed anche $w = \sum_{i=1}^n x_i v_i$, sicché $v = w$.

Suriettività: Sia $w = (x_1, \dots, x_n) \in \mathbb{R}^n$. Allora, detto v il vettore $x_1 v_1 + \dots + x_n v_n$ si ha che $\pi_{\mathcal{B}}(v) = w$ e quindi $\pi_{\mathcal{B}}$ è suriettiva.

In conclusione, essendo sia iniettiva che suriettiva, $\pi_{\mathcal{B}}$ è un isomorfismo tra V e \mathbb{R}^n . ♣

COROLLARIO 16: Siano V e W spazi vettoriali tali che $\dim(V) = \dim(W)$. Allora V e W sono isomorfi.

4.2 Applicazioni lineari da un punto di vista geometrico

Abbiamo visto nella sezione 1.3 le rette in \mathbb{R}^2 . Abbiamo poi ricordato nella definizione 3.2, come nomi quali "retta", "piano" ecc. siano individuati dalla dimensione dei corrispondenti spazi vettoriali. A costo di sembrare noiosi, richiamiamo ancora una volta questi concetti basilari.

Definizione 31: [Retta] Sia V uno spazio vettoriale e sia $v \in V$ con $v \neq \mathbf{0}$. L'insieme

$$\mathcal{R}(v) = \{w \in V : w = \lambda v, \lambda \in \mathbb{R}\}$$

è detto una *retta* di V .

Si noti che $\mathbf{0} \in \mathcal{R}(v)$ per ogni $v \in V$. Alle volte, per sottolineare questo fatto, parliamo di *retta passante per l'origine*. In questo modo possiamo anche definire il concetto di rette non passanti per l'origine (ottenute *traslando* una retta per l'origine, ossia sommando un vettore non-nullo a tutti i suoi elementi) e rette parallele (ottenute traslando una medesima retta due volte, con due traslazioni diverse).

Ogni applicazione lineare mappa una retta del dominio o nella singola origine o in una retta del codominio, in quanto si ha

Lemma 17: Sia $F : V \mapsto W$ un'applicazione lineare, e sia $v \neq \mathbf{0}_V \in V$. Supponiamo inoltre che $F(v) \neq \mathbf{0}_W$. Allora si ha

$$F(\mathcal{R}(v)) = \mathcal{R}(F(v))$$

Dim: Abbiamo che $p \in \mathcal{R}(F(v))$ se e solo se esiste λ tale che $p = \lambda F(v)$ e quindi se e solo se $p = F(\lambda v)$, i.e., $p \in F(\mathcal{R}(v))$. ♣

Definizione 32: [**Piano**] Sia V uno spazio vettoriale e siano $v, w \in V$ vettori linearmente indipendenti. L'insieme

$$\mathcal{P}(v, w) = \{u \in V : u = \lambda v + \mu w, \lambda, \mu \in \mathbb{R}\}$$

è detto un *piano* di V .

Si noti che $\mathbf{0} \in \mathcal{P}(v, w)$. Alle volte, per sottolineare questo fatto, parliamo di *piano passante per l'origine*. In questo modo possiamo anche definire il concetto di piani paralleli e piani non passanti per l'origine.

Ogni applicazione lineare mappa un piano del dominio o nella singola origine, o in una retta del codominio, o in un piano del codominio. Quest'ultimo caso si ha quando $F(v)$ e $F(w)$ sono linearmente indipendenti.

Lemma 18: Sia $F : V \mapsto W$ un'applicazione lineare, e siano $v, w \in V$, linearmente indipendenti. Supponiamo inoltre che $F(v), F(w)$ siano linearmente indipendenti in W . Allora si ha

$$F(\mathcal{P}(v, w)) = \mathcal{P}(F(v), F(w))$$

La dimostrazione del lemma è lasciata al lettore.

Definizione 33: [**Combinazione convessa**] Sia V uno spazio vettoriale e siano $v, w \in V$. Per ogni $\lambda \in [0, 1]$ il vettore

$$u = \lambda v + (1 - \lambda)w$$

è detto una *combinazione convessa* di v e w .

Si noti che la combinazione convessa può anche essere espressa come $p = w + \lambda(v - w)$ e quindi possiamo vedere come il punto p si sposta, al variare di λ dal punto w (per $\lambda = 0$) al punto v (per $\lambda = 1$), lungo la direzione $v - w$, i.e., seguendo la retta che passa per v e w .

Definizione 34: [**Segmento**] Sia V uno spazio vettoriale e siano $v, w \in V$. Chiamiamo *segmento* di estremi v e w l'insieme

$$\mathcal{S}(v, w) = \{p : p = \lambda v + (1 - \lambda)w, \lambda \in [0, 1]\}$$

Un segmento si dice *degenere* se $v = w$, in quanto in questo caso il segmento degenera in un singolo punto. Ogni applicazione lineare mappa segmenti in segmenti, in quanto si ha

Lemma 19: Sia $F : V \mapsto W$ un'applicazione lineare, e siano $v, w \in V$. Allora si ha

$$F(\mathcal{S}(v, w)) = \mathcal{S}(F(v), F(w))$$

Dim: Abbiamo che $p \in \mathcal{S}(F(v), F(w))$ se e solo se esiste $\lambda \in [0, 1]$ tale che $p = \lambda F(v) + (1 - \lambda)F(w)$ e quindi se e solo se $p = F(\lambda v + (1 - \lambda)w)$, i.e., $p \in F(\mathcal{S}(v, w))$. ♣

Il concetto di combinazione convessa può essere esteso a un insieme di n vettori.

Definizione 35: [Combinazione convessa di n vettori] Sia V uno spazio vettoriale e siano $v_1, \dots, v_n \in V$. Siano inoltre $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ coefficienti tali che $\lambda_i \geq 0$ per ogni i e $\sum_{i=1}^n \lambda_i = 1$. Allora il vettore

$$u = \sum_{i=1}^n \lambda_i v_i$$

è detto una *combinazione convessa* di v_1, \dots, v_n .

Definizione 36: [Insieme convesso.] Sia V uno spazio vettoriale e sia $A \subseteq V$. Si dice che A è un insieme *convesso* se per ogni $u, v \in A$ e $\lambda \in [0, 1]$ si ha $\lambda u + (1 - \lambda)v \in A$. (ossia, informalmente, è possibile andare da ogni punto dell'insieme ad ogni altro punto senza mai uscire dall'insieme stesso)

Come esempio di oggetti convessi tridimensionali nella vita reale, consideriamo la sfera piena, una moneta o un dado da gioco. Tra gli oggetti non convessi troviamo, ad esempio, il bicchiere, il cappello, l'ombrello. Si può facilmente dimostrare che ogni applicazione lineare mappa insiemi convessi del dominio in insiemi convessi del codominio, e insiemi non convessi del dominio in insiemi non convessi del codominio.

Definizione 37: [Triangolo.] Sia V uno spazio vettoriale e siano $u, v, w \in V$. L'insieme

$$\mathcal{T}(u, v, w) = \cup_{p \in \mathcal{S}(u, v)} \mathcal{S}(p, w)$$

è detto *triangolo* di vertici u, v, w .

Si può dimostrare (esercizio) che

$$\mathcal{T}(u, v, w) = \{p : p = \lambda_1 u + \lambda_2 v + \lambda_3 w, \lambda_1, \lambda_2, \lambda_3 \geq 0, \lambda_1 + \lambda_2 + \lambda_3 = 1\}.$$

Se nessuno dei tre vertici è combinazione convessa degli altri due, allora il triangolo si dice *non degenere*. Se esattamente uno dei tre è combinazione convessa degli altri due, il triangolo degenera in un segmento. Se ognuno è combinazione convessa degli altri due, i tre vertici coincidono e il triangolo degenera in un punto.

Ogni applicazione lineare mappa triangoli del dominio in triangoli del codominio. Infatti vale il

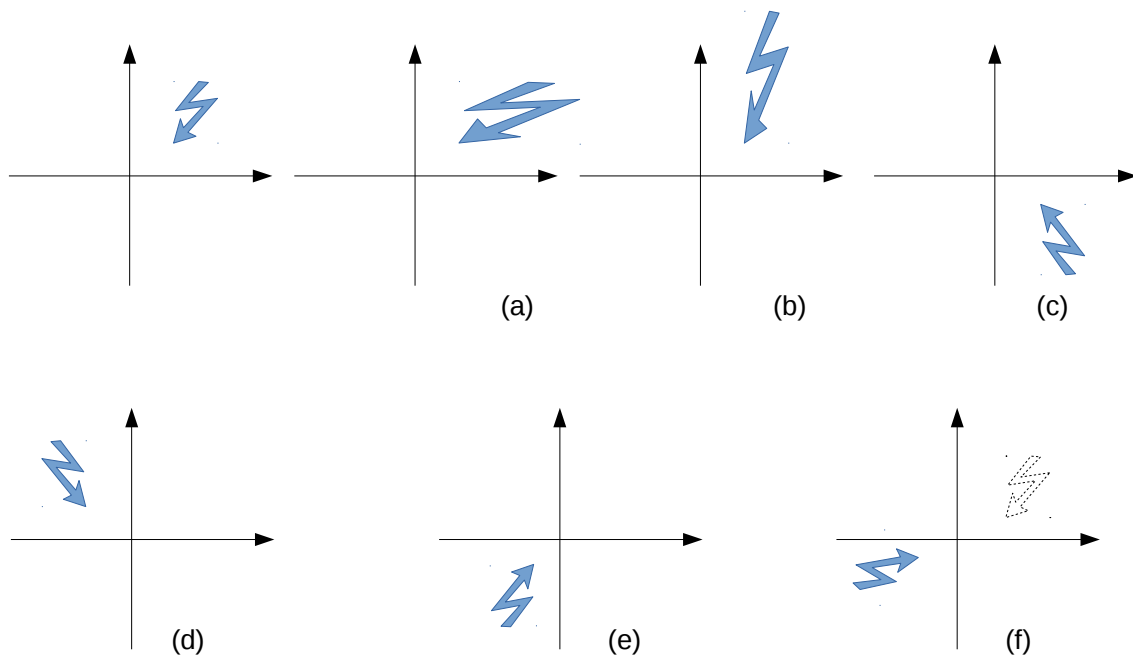


Figure 4.1: Trasformazioni lineari elementari nel piano.

Lemma 20: Sia $F : V \mapsto W$ un'applicazione lineare, e siano $u, v, w \in V$. Allora

$$F(\mathcal{T}(u, v, w)) = \mathcal{T}(F(u), F(v), F(w))$$

La dimostrazione è lasciata per esercizio.

Alla luce delle proprietà appena ricordate delle trasformazioni lineari possiamo interpretare queste mappe in modo "geometrico" notando come esse allungano e/o accorciano e/o ruotano le varie forme senza però snaturarle (nel senso che una linea retta non diventerà mai curva a seguito di una trasformazione lineare) ma, al massimo, proiettandole su spazi di dimensione minore. Consideriamo alcune semplici applicazioni lineari di \mathbb{R}^2 in \mathbb{R}^2 che hanno un'immediata interpretazione geometrica, e che sono esemplificate in figura 4.1:

1. *Cambiamento di scala sull'asse x:* $F(x, y) = (\lambda x, y)$, con $\lambda > 0$. Qui, se $\lambda > 1$ ogni figura viene "allungata" nella sua dimensione orizzontale, mentre se $\lambda < 1$ essa viene "accorciata". (Vedi fig.4.1(a))
2. *Cambiamento di scala sull'asse y:* $F(x, y) = (x, \lambda y)$, con $\lambda > 0$. Qui, se $\lambda > 1$ ogni figura viene "allungata" nella sua dimensione verticale, mentre se $\lambda < 1$ essa viene "accorciata". (Vedi fig.4.1(b))

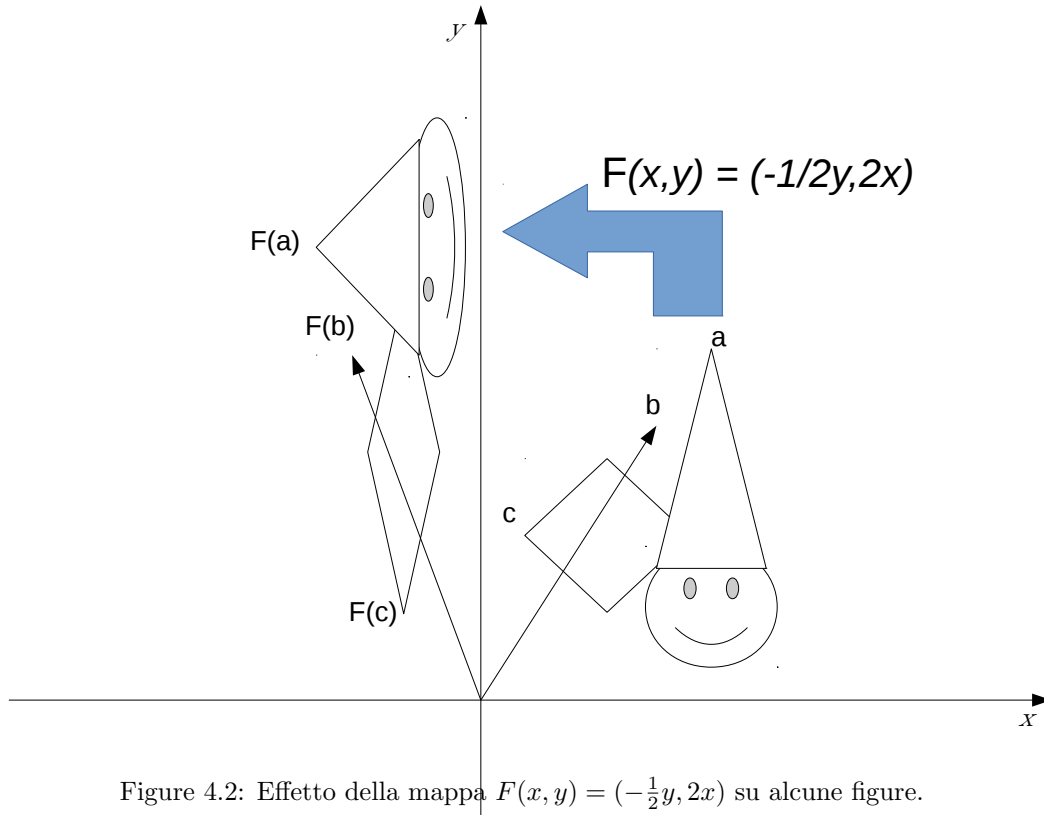


Figure 4.2: Effetto della mappa $F(x, y) = (-\frac{1}{2}y, 2x)$ su alcune figure.

3. *Proiezione sull'asse x*: $F(x, y) = (x, 0)$. Ogni figura viene "schiacciata" verticalmente sull'asse x .
4. *Proiezione sull'asse y*: $F(x, y) = (0, y)$. Ogni figura viene "schiacciata" orizzontalmente sull'asse y .
5. *Simmetria rispetto all'asse x*: $F(x, y) = (x, -y)$. Ogni figura viene "specchiata" verticalmente dalla parte opposta rispetto all'asse x . (Vedi fig.4.1(c))
6. *Simmetria rispetto all'asse y*: $F(x, y) = (-x, y)$. Ogni figura viene "specchiata" orizzontalmente dalla parte opposta rispetto all'asse y . (Vedi fig.4.1(d))
7. *Simmetria rispetto all'origine*: $F(x, y) = (-x, -y)$. Ogni figura viene "specchiata" sia orizzontalmente che verticalmente. (Vedi fig.4.1(e))
8. *Rotazione di un angolo θ* : $F_\theta(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$. Ogni figura viene ruotata di un angolo di θ radianti in senso antiorario. In particolare, per $\theta = k\pi/2$ con $k = 1, 2, 3$ abbiamo $F_{\pi/2}(x, y) = (-y, x)$; $F_\pi(x, y) = (-x, -y)$; $F_{3\pi/2}(x, y) = (y, -x)$; (Vedi fig.4.1(f) con $\theta \simeq 170^\circ$)

Quando un'applicazione lineare è la composta di alcune di queste mappe elementari, l'effetto finale è quello ottenuto applicando in sequenza tali trasformazioni. Ad esempio, se allarghiamo una figura in

orizzontale raddoppiandone le dimensioni, poi la dimezziamo nella dimensione verticale, e infine la ruotiamo di $\pi/2$ l'effetto è quello di comporre tre mappe:

$F(x, y) = (2x, y)$; $G(x, y) = (x, 1/2y)$; $H(x, y) = (-y, x)$. Abbiamo

$$H(G(F(x, y))) = H(G(2x, y)) = H(2x, 1/2y) = (-1/2y, 2x)$$

Il risultato di tale mappa su alcune forme nel piano è visibile in Figura 4.2.

Chapter 5

Matrici e applicazioni lineari

Siano V e W spazi vettoriali di dimensione finita, con $n = \dim V$ e $m = \dim W$. In questa sezione vedremo come l'insieme delle applicazioni lineari di V in W possa essere messo in corrispondenza biunivoca con $\mathbb{R}^{m \times n}$, l'insieme delle matrici di m righe e n colonne. In base a tale corrispondenza, ad ogni applicazione lineare $L : V \mapsto W$ corrisponderà univocamente una matrice $A \in \mathbb{R}^{m \times n}$, che diremo *rappresenta* L . Inoltre, data una qualsiasi matrice $A \in \mathbb{R}^{m \times n}$ sarà possibile definire un'applicazione lineare $L : V \mapsto W$ tale che A è la matrice che rappresenta L .

Sia \mathcal{L} l'insieme di tutte le applicazioni lineari di V in W . Definiamo una mappa $f : \mathcal{L} \mapsto \mathbb{R}^{m \times n}$ nel modo seguente. Innanzitutto, fissiamo una base $\mathcal{B} = (v_1, \dots, v_n)$ di V e una base $\mathcal{B}' = (w_1, \dots, w_m)$ di W (nel seguito ci riferiremo a V e \mathcal{B} come lo "spazio di partenza" e la "base di partenza", e a W e \mathcal{B}' come lo "spazio di arrivo" e la "base di arrivo"). Data un'applicazione lineare $L \in \mathcal{L}$ definiamo $f(L)$ come la matrice $A = (a_{ij})$ per la quale

$$a_{ij} = i\text{-mo coefficiente del vettore } L(v_j) \text{ rispetto alla base } \mathcal{B}', \text{ per ogni } i = 1, \dots, m \text{ e } j = 1, \dots, n$$

In altre parole, la colonna A^j coincide con $\mathbf{C}_{\mathcal{B}'}^{L(v_j)}$ (si veda la sezione 3.2.1) e quindi

$$L(v_j) = a_{1j}w_1 + a_{2j}w_2 + \dots + a_{mj}w_m, \quad \forall j = 1, \dots, n.$$

TEOREMA 21: Siano V e W spazi vettoriali. Sia \mathcal{L} l'insieme di tutte le applicazioni lineari di V in W . Allora esiste una corrispondenza biunivoca tra \mathcal{L} e $\mathbb{R}^{m \times n}$, dove $n = \dim(V)$ e $m = \dim(W)$.

Dim: Siano $\mathcal{B} = (v_1, \dots, v_n)$ e $\mathcal{B}' = (w_1, \dots, w_m)$ basi, rispettivamente, di V e W . Consideriamo la funzione $f : \mathcal{L} \mapsto \mathbb{R}^{m \times n}$ che associa ad ogni applicazione lineare L la matrice dei coefficienti dei vettori $L(v_1), \dots, L(v_n)$ rispetto alla base \mathcal{B}' e dimostriamo che f è una funzione biiettiva.

(Iniettività:) Siano F e G applicazioni lineari. Per ogni $v \in V$, detti (x_1, \dots, x_n) i coefficienti di v rispetto a \mathcal{B} si ha $F(v) = \sum_{j=1}^n x_j F(v_j)$ e $G(v) = \sum_{j=1}^n x_j G(v_j)$. Se $F \neq G$, deve esistere almeno un $j \in \{1, \dots, n\}$ tale che $F(v_j) \neq G(v_j)$. Quindi la colonna j -ma di $f(F)$ è diversa dalla colonna j -ma di $f(G)$ sicchè $F \neq G \implies f(F) \neq f(G)$.

(Suriattività:) Sia $A \in \mathbb{R}^{m \times n}$. Definiamo un'applicazione L in questo modo: per ogni $v \in V$, detti (x_1, \dots, x_n) i coefficienti di v rispetto a \mathcal{B} poniamo

$$L(v) := \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j \right) w_i$$

Non è difficile verificare che L è un'applicazione lineare. Inoltre, considerato che per $j = 1, \dots, n$ il vettore delle componenti di v_j è tale che $x_j = 1$ e $x_h = 0$ per $h \neq j$ si ottiene che la j -ma colonna di A contiene le coordinate di $L(v_j)$ rispetto a \mathcal{B}' . Quindi $A = f(L)$. ♣

È importante notare come la corrispondenza biunivoca f appena descritta dipenda dalla scelta delle basi \mathcal{B} e \mathcal{B}' . In particolare, quindi, data una matrice A non è possibile dire quale sia l'applicazione lineare rappresentata da A se non decidiamo una particolare base di partenza e una particolare base di arrivo. Analogamente, non possiamo descrivere la matrice che rappresenta un'applicazione lineare senza fissare una particolare base di partenza e una particolare base di arrivo. Rispetto a scelte diverse per le basi di riferimento, la stessa matrice rappresenta applicazioni lineari diverse, e la stessa applicazione lineare è rappresentata da matrici diverse.

Definizione 38: [Matrice rappresentativa di un'applicazione lineare.] Per denotare la matrice che riporta i coefficienti, rispetto a una base \mathcal{W} di W , delle immagini di un'applicazione lineare F sui vettori di una specifica base \mathcal{V} di V useremo la seguente notazione:

$$\mathcal{M}[F, \mathcal{V}, \mathcal{W}]$$

Esempio. Consideriamo lo spazio \mathbb{R}^2 e l'applicazione L di rotazione di un angolo di $\pi/3$ in senso antiorario

$$L(x, y) = (1/2x - \sqrt{3}/2y, \sqrt{3}/2x + 1/2y)$$

Fissiamo sia come base di partenza che come base di arrivo la base canonica $\mathcal{C} = (e_1, e_2)$ di \mathbb{R}^2 . Il vettore-colonna immagine di e_1 è

$$L(1, 0) = \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}$$

mentre il vettore-colonna immagine di e_2 è

$$L(0, 1) = \begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix}$$

Quindi, la matrice che rappresenta la rotazione di $\pi/3$ rispetto alla base canonica sia di partenza che d'arrivo è

$$\mathcal{M}[L, \mathcal{C}, \mathcal{C}] = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}$$

I vettori $L(e_1)$ e $L(e_2)$ sono linearmente indipendenti, come è facile verificare. Essi quindi costituiscono una base di \mathbb{R}^2 , sia essa \mathcal{B}'' . Supponiamo ora di considerare la stessa applicazione ma di prendere come base di partenza la base canonica e come base di arrivo \mathcal{B}'' . Rispetto alla base di arrivo, le coordinate di $L(e_1)$ sono $(1, 0)$, in quanto

$$L(e_1) = 1 \times L(e_1) + 0 \times L(e_2)$$

similmente

$$L(e_2) = 0 \times L(e_1) + 1 \times L(e_2)$$

Quindi, la matrice che rappresenta la rotazione di $\pi/3$ rispetto alla base canonica di partenza e a \mathcal{B}'' come base d'arrivo è

$$\mathcal{M}[L, \mathcal{C}, \mathcal{B}''] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

ossia, l'identità. Da questo esempio si capisce come conoscere la sola matrice associata all'applicazione non sia sufficiente a conoscere l'applicazione stessa, ma è fondamentale conoscere anche le basi di partenza e di arrivo utilizzate nel calcolo della matrice. \diamond

Siano

- (a_{ij}) le coordinate dei vettori $L(v_j)$ rispetto ai vettori w_i della base \mathcal{W} , i.e.,

$$L(v_j) = \sum_{i=1}^m a_{ij} w_i \quad \forall j = 1, \dots, n$$

- (x_1, \dots, x_n) le coordinate di un generico $v \in V$ rispetto alla base \mathcal{V}
- (y_1, \dots, y_m) le coordinate di $L(v) \in W$ rispetto alla base \mathcal{W} .

Si ha

$$L(v) = \sum_{j=1}^n x_j L(v_j) = \sum_{j=1}^n x_j \left(\sum_{i=1}^m a_{ij} w_i \right) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} x_j \right) w_i$$

da cui notiamo che

$$y_i = \sum_{j=1}^n a_{ij} x_j \quad \text{per } i = 1, \dots, m$$

Diamo ora un'interpretazione matriciale all'equazione appena descritta. Siano

- (i) $A = \mathcal{M}[F, \mathcal{V}, \mathcal{W}]$ la matrice $m \times n$ le cui colonne sono le coordinate, rispetto alla base \mathcal{W} , dei vettori $L(v_1), \dots, L(v_n)$, ossia, delle immagini dei vettori della base \mathcal{V} .
- (ii) $X = \mathbf{C}_{\mathcal{V}}^v$ il vettore-colonna $\in \mathbb{R}^{n \times 1}$ delle coordinate di v rispetto alla base \mathcal{V}
- (iii) $Y = \mathbf{C}_{\mathcal{W}}^{L(v)}$ il vettore-colonna $\in \mathbb{R}^{m \times 1}$ delle coordinate di $L(v)$ rispetto alla base \mathcal{W}

Allora abbiamo

$$Y = AX$$

e quindi

le coordinate dell'immagine di un vettore di V vengono calcolate tramite il prodotto matriciale tra la matrice associata all'applicazione lineare e la n -pla delle coordinate del vettore (NOTA BENE: devono essere utilizzate le stesse basi sia per il calcolo della n -pla che per la definizione della matrice associata all'applicazione)

5.1 Prodotto di matrici e composizione di applicazioni

Siano V, U, W spazi vettoriali con $\dim(V) = n$, $\dim(U) = r$ e $\dim(W) = m$. Siano $\mathcal{V} = (v_1, \dots, v_n)$, $\mathcal{U} = (u_1, \dots, u_r)$ e $\mathcal{W} = (w_1, \dots, w_m)$ basi di, rispettivamente, V , U e W . Siano $F : V \mapsto U$ e $G : U \mapsto W$ applicazioni lineari e sia $H = GF : V \mapsto W$ l'applicazione composta. Ci chiediamo che relazione c'è tra $\mathcal{M}[H, \mathcal{V}, \mathcal{W}]$ e le matrici $A = \mathcal{M}[F, \mathcal{V}, \mathcal{U}]$ e $B = \mathcal{M}[G, \mathcal{U}, \mathcal{W}]$.

Dato $v \in V$ abbiamo

$$\mathbf{C}_{\mathcal{U}}^{F(v)} = A\mathbf{C}_{\mathcal{V}}^v$$

Inoltre

$$\mathbf{C}_{\mathcal{W}}^{G(F(v))} = B\mathbf{C}_{\mathcal{U}}^{F(v)} = BAC_{\mathcal{V}}^v$$

Dalla definizione di $\mathcal{M}[H, \mathcal{V}, \mathcal{W}]$ si ha anche

$$\mathbf{C}_{\mathcal{W}}^{G(F(v))} = \mathbf{C}_{\mathcal{W}}^{H(v)} = \mathcal{M}[H, \mathcal{V}, \mathcal{W}]\mathbf{C}_{\mathcal{V}}^v$$

e quindi

$$\mathcal{M}[H, \mathcal{V}, \mathcal{W}]\mathbf{C}_{\mathcal{V}}^v = BAC_{\mathcal{V}}^v$$

per ogni $v \in V$. In conclusione si ha quindi che

$$\mathcal{M}[H, \mathcal{V}, \mathcal{W}] = BA$$

ossia

la matrice che rappresenta la composta di due applicazioni lineari è data dal prodotto delle matrici che rappresentano tali applicazioni, nell'ordine inverso.

Tale regola si può estendere alla composizione di un numero arbitrario di applicazioni. Per ottenere la matrice della composta basterà fare il prodotto delle matrici delle singole applicazioni, dove le matrici più a destra rappresentano le funzioni applicate per prime e quelle più a sinistra le funzioni applicate per ultime. Inoltre, bisogna che la base di arrivo di ogni singola applicazione coincida con la base di partenza dell'applicazione successiva.

5.2 Cambiamenti di base e nuove coordinate

Sia V uno spazio vettoriale di dimensione n e $\mathcal{B} = (v_1, \dots, v_n)$ una sua base. Chiamiamo $(\cdot)_{\mathcal{B}} : V \mapsto \mathbb{R}^n$ l'applicazione che associa a ogni vettore di V la n -pla delle sue coordinate rispetto alla base \mathcal{B} :

$$(v)_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

laddove $v = x_1v_1 + x_2v_2 + \dots + x_nv_n$. È immediato vedere che si tratta di un'applicazione lineare di V in \mathbb{R}^n .

Sia ora $\mathcal{B}' = (v'_1, \dots, v'_n)$ un'altra base di V . Come possiamo calcolare le coordinate di v rispetto alla base \mathcal{B}' a partire dalle sue coordinate rispetto alla base \mathcal{B} ?

Dalla relazione tra matrici e applicazioni lineari sappiamo che

$$\mathcal{M}[\text{id}, \mathcal{B}, \mathcal{B}'](v)_{\mathcal{B}} = (\text{id}(v))_{\mathcal{B}'} = (v)_{\mathcal{B}'}$$

e quindi per ottenere le nuove coordinate di v basta moltiplicare il vettore delle sue vecchie coordinate per la matrice

$${}_{\mathcal{B}}M_{\mathcal{B}'} := \mathcal{M}[\text{id}, \mathcal{B}, \mathcal{B}']$$

detta la *matrice del cambiamento di base da \mathcal{B} a \mathcal{B}'* .

Abbiamo che

$${}_{\mathcal{B}}M_{\mathcal{B}'} = \begin{pmatrix} (v_1)_{\mathcal{B}'} & (v_2)_{\mathcal{B}'} & \dots & (v_n)_{\mathcal{B}'} \end{pmatrix}$$

Sia ora ${}_{\mathcal{B}'}M_{\mathcal{B}} = \mathcal{M}[\text{id}, \mathcal{B}', \mathcal{B}]$ la matrice del cambiamento di base da \mathcal{B}' a \mathcal{B} . Essendo

$${}_{\mathcal{B}'}M_{\mathcal{B}} {}_{\mathcal{B}}M_{\mathcal{B}'} = {}_{\mathcal{B}'}M_{\mathcal{B}'} = I \quad \text{ed anche} \quad {}_{\mathcal{B}}M_{\mathcal{B}'} {}_{\mathcal{B}'}M_{\mathcal{B}} = {}_{\mathcal{B}}M_{\mathcal{B}} = I$$

abbiamo che ${}_{\mathcal{B}}M_{\mathcal{B}'} = ({}_{\mathcal{B}'}M_{\mathcal{B}})^{-1}$, i.e.,

$${}_{\mathcal{B}}M_{\mathcal{B}'} = \begin{pmatrix} (v'_1)_{\mathcal{B}} & (v'_2)_{\mathcal{B}} & \dots & (v'_n)_{\mathcal{B}} \end{pmatrix}^{-1}.$$

Cambiamento di base nel piano

Sia $V = \mathbb{R}^2$ e sia $\mathcal{B} = (e_1, e_2)$ la base canonica di \mathbb{R}^2 , con $e_1 = (1, 0)$ ed $e_2 = (0, 1)$. Sia $\mathcal{B}' = (v_1, v_2)$ una qualsiasi base di \mathbb{R}^2 , con $v_1 = (a, c)$ ed $v_2 = (b, d)$. Siccome \mathcal{B}' è una base, deve essere $ad \neq bc$. Infatti: (i) almeno uno fra c e d deve essere $\neq 0$, o v_1 e v_2 non sarebbero indipendenti. Sia, s.p.d.g., $c \neq 0$; (ii) se fosse $ad = bc$, posto $\lambda = \frac{d}{c}$ si avrebbe $b = \lambda a$ e $d = \lambda c$, sicchè $v_2 = \lambda v_1$, contraddicendo l'indipendenza lineare di v_1 e v_2 .

Il valore $\Delta = ad - bc$ è detto il *determinante* della matrice 2×2 . In particolare, nel caso di nostro interesse in cui le colonne della matrice sono linearmente indipendenti si ha sempre $\Delta \neq 0$.

Abbiamo

$${}_{\mathcal{B}'}M_{\mathcal{B}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

e quindi

$$(v_1)_{\mathcal{B}} = \begin{pmatrix} a \\ c \end{pmatrix} \quad \text{e} \quad (v_2)_{\mathcal{B}} = \begin{pmatrix} b \\ d \end{pmatrix}$$

Consideriamo le due equazioni

$$v_1 = ae_1 + ce_2 \tag{5.1}$$

e

$$v_2 = be_1 + de_2 \tag{5.2}$$

Moltiplichiamo la prima equazione per d , la seconda per c e sottraiamo la seconda dalla prima, ottenendo

$$dv_1 - cv_2 = (ad - bc)e_1$$

e quindi abbiamo che il vettore delle coordinate di e_1 rispetto alla base \mathcal{B}' è

$$(e_1)_{\mathcal{B}'} = \begin{pmatrix} d/\Delta \\ -c/\Delta \end{pmatrix}$$

Se ora moltiplichiamo l'equazione (5.1) per b , l'equazione (5.2) per a e sottraiamo la prima dalla seconda, otteniamo

$$-bv_1 + av_2 = (ad - bc)e_2$$

e quindi abbiamo che il vettore delle coordinate di e_2 rispetto alla base \mathcal{B}' è

$$(e_2)_{\mathcal{B}'} = \begin{pmatrix} -b/\Delta \\ a/\Delta \end{pmatrix}$$

In conclusione

$${}_{\mathcal{B}}M_{\mathcal{B}'} = \begin{pmatrix} d/\Delta & -b/\Delta \\ -c/\Delta & a/\Delta \end{pmatrix}$$

e le coordinate (x', y') rispetto alla base \mathcal{B}' di un generico punto $(x, y) \in \mathbb{R}^2$ saranno date da

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (5.3)$$

Sottolineiamo che il procedimento da noi seguito ci ha portato a trovare l'inversa di una generica matrice reale 2×2 con determinante $\Delta \neq 0$. La formula che definisce l'inversa è pertanto

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

Esempio. [Rotazione.] Consideriamo il cambiamento di base dalla base canonica $\mathcal{B} = (e_1, e_2)$ alla base $\mathcal{B}' = (e'_1, e'_2)$ in cui e'_i corrisponde ad e_i ruotato di un angolo θ in senso antiorario. Abbiamo $e'_1 = (\cos \theta, \sin \theta)$ e $e'_2 = (-\sin \theta, \cos \theta)$. La matrice del cambiamento di base da \mathcal{B}' a \mathcal{B} è

$${}_{\mathcal{B}'}M_{\mathcal{B}} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

Dalla formula (5.3) ricaviamo che la matrice del cambiamento di base da \mathcal{B} a \mathcal{B}' è

$${}_{\mathcal{B}}M_{\mathcal{B}'} = \frac{1}{\cos^2 \theta + \sin^2 \theta} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

Sia $v = (x, y)$ un generico punto del piano. Abbiamo

$$(v)_{\mathcal{B}} = \begin{pmatrix} x \\ y \end{pmatrix}$$

da cui

$$(v)_{\mathcal{B}'} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \cos \theta + y \sin \theta \\ -x \sin \theta + y \cos \theta \end{pmatrix}$$

◇

Esempio. [Cambiamento di scala sugli assi.] Consideriamo il cambiamento di base dalla base canonica $\mathcal{B} = (e_1, e_2)$ alla base $\mathcal{B}' = (e'_1, e'_2)$ in cui e'_i è un multiplo di e_i , per $i = 1, 2$. Abbiamo $e'_1 = (\alpha_1, 0)$ e $e'_2 = (0, \alpha_2)$ con $\alpha_1, \alpha_2 > 0$. Si noti che nel nuovo sistema di riferimento, se $\alpha_i < 1$ l'unità di misura sull'asse i -mo è stata accorciata, mentre se $\alpha_i > 1$ essa è stata allungata. La matrice del cambiamento di base da \mathcal{B}' a \mathcal{B} è

$${}_{\mathcal{B}'}M_{\mathcal{B}} = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix}$$

la cui inversa è

$${}_{\mathcal{B}}M_{\mathcal{B}'} = \begin{pmatrix} 1/\alpha_1 & 0 \\ 0 & 1/\alpha_2 \end{pmatrix}$$

Nel nuovo sistema di coordinate, le coordinate di generico punto (x, y) del piano diventano

$$\begin{pmatrix} 1/\alpha_1 & 0 \\ 0 & 1/\alpha_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x/\alpha_1 \\ y/\alpha_2 \end{pmatrix}$$

Ad esempio, se $\alpha_1 = 2$ e $\alpha_2 = 1/3$, le nuove coordinate del punto $(4, 3)$ sono $(2, 9)$.

◇

Esempio. Sia $\mathcal{B}' = \{(3, 1), (-1, 1)\}$ e sia $v = (0, 3)$. Vogliamo determinare $(v)_{\mathcal{B}'}$. Abbiamo

$${}_{\mathcal{B}'}M_{\mathcal{B}} = \begin{pmatrix} 3 & -1 \\ 1 & 1 \end{pmatrix}$$

Da (5.3) otteniamo

$${}_{\mathcal{B}}M_{\mathcal{B}'} = ({}_{\mathcal{B}'}M_{\mathcal{B}})^{-1} = \begin{pmatrix} 1/4 & 1/4 \\ -1/4 & 3/4 \end{pmatrix}$$

da cui

$$(v)_{\mathcal{B}'} = \begin{pmatrix} 1/4 & 1/4 \\ -1/4 & 3/4 \end{pmatrix} (v)_{\mathcal{B}} = \begin{pmatrix} 1/4 & 1/4 \\ -1/4 & 3/4 \end{pmatrix} \begin{pmatrix} 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 3/4 \\ 9/4 \end{pmatrix}$$

◇

ESERCIZIO 5.1. Stabilire quali tra le seguenti matrici sono invertibili. Per ogni tale matrice, se ne calcoli l'inversa.

$$A = \begin{pmatrix} 2 & 0 \\ -2 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 1 \\ -6 & -3 \end{pmatrix} \quad C = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$$

◇

ESERCIZIO 5.2. Stabilire quali valori del parametro reale k la matrice A è invertibile. In corrispondenza a tali valori, si calcoli A^{-1} .

$$A = \begin{pmatrix} 2 & k \\ 3 & 5 \end{pmatrix}$$

◇

5.3 Matrici associate a basi diverse

Siano V e W spazi vettoriali con basi \mathcal{V} e, rispettivamente, \mathcal{W} . Siano inoltre \mathcal{V}' e \mathcal{W}' ulteriori basi di V e, rispettivamente, W . Data un'applicazione lineare $F : V \mapsto W$ e supponendo nota $\mathcal{M}[F, \mathcal{V}, \mathcal{W}]$, ci chiediamo come fare ad ottenere $\mathcal{M}[F, \mathcal{V}', \mathcal{W}']$. In base alla legge per la composizione di applicazioni abbiamo

$$\mathcal{M}[F, \mathcal{V}', \mathcal{W}'] = \mathcal{M}[\text{id}_W, \mathcal{W}, \mathcal{W}'] \mathcal{M}[F, \mathcal{V}, \mathcal{W}] \mathcal{M}[\text{id}_V, \mathcal{V}', \mathcal{V}]$$

(Possiamo, intuitivamente, pensare a questa composizione di matrici come ad un processo di traduzione, dove le varie basi rappresentano dei "linguaggi". Avendo a disposizione un traduttore dal linguaggio \mathcal{V} a \mathcal{W} (la matrice $\mathcal{M}[F, \mathcal{V}, \mathcal{W}]$) ma dovendo tradurre da \mathcal{V}' a \mathcal{W}' procediamo componendo tre fasi: prima la traduzione da \mathcal{V}' a \mathcal{V} (la matrice $\mathcal{M}[\text{id}_V, \mathcal{V}', \mathcal{V}]$), poi la traduzione da \mathcal{V} a \mathcal{W} e infine a traduzione da \mathcal{W} a \mathcal{W}' (la matrice $\mathcal{M}[\text{id}_W, \mathcal{W}, \mathcal{W}']$))

Poniamoci ora nel caso in cui $V = W$ e prendiamo due basi distinte di V , siano esse \mathcal{B} e \mathcal{B}' . Sia $P = {}_{\mathcal{B}'}M_{\mathcal{B}}$ la matrice del cambiamento di base da \mathcal{B}' a \mathcal{B} . Siano inoltre $A = \mathcal{M}[F, \mathcal{B}, \mathcal{B}]$, e $A' = \mathcal{M}[F, \mathcal{B}', \mathcal{B}']$ abbiamo

$$A' = P^{-1}AP$$

Come si può notare, tra la matrice di F rispetto alla base \mathcal{B} e quella rispetto alla base \mathcal{B}' vale una relazione che abbiamo già definito come di *similitudine* (si veda la sezione 2.4). In particolare si ha che

TEOREMA 22: Due matrici quadrate $A, A' \in \mathbb{R}^{n \times n}$ sono simili se e solo se rappresentano lo stesso endomorfismo, i.e., se solo se esistono basi \mathcal{B} e \mathcal{B}' di \mathbb{R}^n e un'applicazione lineare $F : \mathbb{R}^n \mapsto \mathbb{R}^n$ tali che

$$A = \mathcal{M}[F, \mathcal{B}, \mathcal{B}] \text{ e } A' = \mathcal{M}[F, \mathcal{B}', \mathcal{B}']$$

Chapter 6

Sistemi lineari

Noi tutti abbiamo familiarità con la nozione di *equazione di primo grado*, una formula del tipo

$$ax = b \tag{6.1}$$

in cui $a, b \in \mathbb{R}$ e x è detta una *variabile* o *incognita*. *Risolvere* l'equazione consiste nel determinare tutti i valori v che, sostituiti a x , rendono vera l'uguaglianza (6.1), ossia tali che

$$av = b$$

Ogni valore v siffatto viene detto una *soluzione* dell'equazione. L'equazione (6.1) può avere una, nessuna o infinite soluzioni. Il primo caso si verifica quando $a \neq 0$. In tale caso, l'unica soluzione è b/a . Il secondo caso si verifica quando $a = 0$ ma $b \neq 0$. Infine, si hanno infinite soluzioni quando $a = b = 0$.

Un'equazione di primo grado viene anche detta *equazione lineare* in quanto ax è una funzione lineare della variabile x (i.e., un'applicazione lineare di \mathbb{R} in \mathbb{R}). La nozione di equazione lineare può essere generalizzata al caso di più di una variabile. Siano $a_1, \dots, a_n \in \mathbb{R}$ e $b \in \mathbb{R}$. L'espressione

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

è detta un'*equazione lineare* nelle *variabili* x_1, \dots, x_n , a *coefficienti* a_1, \dots, a_n e *termine noto* b . Il termine noto viene anche detto *parte destra* dell'equazione, mentre $\sum_{j=1}^n a_jx_j$ è la *parte sinistra*. Se $b = 0$ l'equazione lineare viene detta *omogenea*. Una *soluzione* di un'equazione lineare è una n -pla (v_1, \dots, v_n) tale che vale l'uguaglianza $\sum_{i=1}^n a_i v_i = b$.

Un *sistema di equazioni lineari* o, più semplicemente, un *sistema lineare* è un insieme di equazioni lineari che devono essere contemporaneamente soddisfatte dalla medesima soluzione. Consideriamo il caso di m equazioni in n variabili. Indichiamo i coefficienti della i -ma equazione con a_{i1}, \dots, a_{in} e il termine noto con b_i . Risolvere il sistema consiste nel determinare tutte le n -ple (x_1, \dots, x_n) per le quali

$$\left(\sum_{j=1}^n a_{1j}x_j = b_1 \right) \wedge \left(\sum_{j=1}^n a_{2j}x_j = b_2 \right) \wedge \dots \wedge \left(\sum_{j=1}^n a_{mj}x_j = b_m \right)$$

Per comodità tipografiche, un sistema lineare viene abitualmente rappresentato elencando le equazioni una dopo l'altra in verticale, e racchiudendole tutte con una parentesi graffa aperta:

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array} \right. \quad (6.2)$$

Qualora le variabili siano poche (ad esempio quattro o meno) spesso esse vengono sostituite dalle ultime lettere dell'alfabeto, tradizionalmente utilizzate in matematica per indicare quantità indeterminate. Quindi, anzichè x_1, x_2, x_3, x_4 , spesso incontreremo sistemi lineari nelle variabili x, y, z, w .

Esempio. Il seguente è un sistema lineare di 3 equazioni in 3 incognite:

$$\left\{ \begin{array}{l} 2x + 4y + 6z = 2 \\ x - y + z = 0 \\ 3x - 2z = -1 \end{array} \right.$$

Si può verificare per sostituzione che $(x, y, z) = (-1/7, 1/7, 2/7)$ è una soluzione del sistema. Si può inoltre dimostrare (e vedremo in seguito come farlo) che si tratta dell'unica soluzione. \diamond

Esempio. Si consideri il seguente sistema lineare di 3 equazioni in 4 incognite:

$$\left\{ \begin{array}{l} x + 4y + 3z - w = 2 \\ -x - 2y + z + 2w = 3 \\ 2y + 4z + w = 4 \end{array} \right.$$

Questo sistema non ha alcuna soluzione. Infatti, i coefficienti della terza equazione sono la somma dei coefficienti della prima e della seconda. Per ogni soluzione della prima e della seconda equazione, il valore della parte sinistra della terza equazione dovrebbe essere pari alla somma del primo e del secondo termine noto, e quindi dovrebbe essere pari a 5. Ma $b_3 = 4 \neq b_1 + b_2$ per cui non esistono soluzioni. \diamond

6.1 Sistemi lineari e matrici

Un sistema lineare può essere rappresentato da un'unica equazione in cui termine noto, variabile e coefficienti sono delle matrici. In particolare, consideriamo la matrice $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ in cui la generica riga i -ma contiene i coefficienti della i -ma equazione. Siano inoltre $X = {}^t(x_1 \dots x_n) \in \mathbb{R}^{n \times 1}$ il vettore-colonna delle variabili, e $\mathbf{b} = {}^t(b_1 \dots b_m) \in \mathbb{R}^{m \times 1}$. Il sistema (6.2) può allora essere riscritto come la seguente equazione matriciale:

$$AX = \mathbf{b}$$

Infatti

$$AX = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = \mathbf{b}$$

se e solo se X è soluzione del sistema lineare.

Ricordando che A_i denota la riga i -ma di A , vediamo che l'equazione i -ma è

$$A_i X = b_i$$

Inoltre, ricordando che A^j denota la colonna j -ma di A , vediamo che l'intero sistema può essere riscritto come

$$x_1 A^1 + x_2 A^2 + \cdots + x_n A^n = \mathbf{b}$$

Quest'ultima scrittura dà luogo ad un'importante interpretazione di un sistema lineare. In pratica, il sistema chiede di determinare tutti i modi (i.e., i coefficienti) di esprimere \mathbf{b} come combinazione lineare delle colonne A^j di A . Quindi il sistema ha soluzione se e solo se \mathbf{b} appartiene allo spazio generato da A^1, \dots, A^n . In caso positivo, la soluzione è unica se e solo se le colonne A^j sono linearmente indipendenti. Possiamo esprimere la condizione di esistenza di una soluzione anche in termini di dimensioni di opportuni spazi (condizione di Rouché-Cappelli), in questo modo:

Un sistema $AX = \mathbf{b}$ è ammissibile se e solo se

$$\dim \mathcal{L} \langle A^1, \dots, A^n \rangle = \dim \mathcal{L} \langle A^1, \dots, A^n, \mathbf{b} \rangle$$

Consideriamo l'applicazione lineare $F_A : R^n \mapsto R^m$ definita da $F_A(x_1, \dots, x_n) = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. La matrice

A rappresenta F_A rispetto alle basi canoniche di R^n e R^m . Lo spazio $\text{im}(F_A)$ è generato dalle colonne di A e quindi il sistema lineare ha soluzione se e solo se $\mathbf{b} \in \text{im}(F_A)$.

Definizione 39: [Sistema omogeneo.] Sia $A \in \mathbb{R}^{m \times n}$. Il sistema

$$AX = \mathbf{0}$$

dove $\mathbf{0} \in \mathbb{R}^{m \times 1}$ è un vettore colonna con tutte le componenti uguali a 0 è detto un *sistema omogeneo*. Si noti che un sistema omogeneo è sempre ammissibile, in quanto la n -pla $(0, \dots, 0)$ ne è una soluzione. Tale soluzione viene detta *soluzione banale*, mentre ogni altra soluzione è detta *non banale*. Si noti che l'esistenza di una soluzione non banale può essere interpretata come una relazione di dipendenza lineare tra le colonne di A , i.e., esistono $(x_1, \dots, x_n) \neq (0, \dots, 0)$ tali che

$$x_1 A^1 + \cdots + x_n A^n = \mathbf{0} \quad ?$$

Questa osservazione ci permette di dimostrare il seguente risultato:

TEOREMA 23: Sia $AX = \mathbf{0}$ un sistema omogeneo di m equazioni in n variabili con $n > m$. Allora tale sistema possiede soluzioni non banali.

Dim: Il sistema ha n colonne ognuna delle quali è un elemento di \mathbb{R}^m . Essendo $\dim(\mathbb{R}^m) = m$ e $n > m$, le colonne sono linearmente dipendenti. ♣

Definizione 40: [Sistema omogeneo associato.] Sia $A \in \mathbb{R}^{m \times n}$ e $\mathbf{b} \in \mathbb{R}^{m \times 1}$. Chiameremo *sistema omogeneo* associato al sistema lineare $AX = \mathbf{b}$ il sistema

$$AX = \mathbf{0}$$

Definizione 41: [Spazio affine.] Siano U uno spazio vettoriale, V un suo sottospazio e $v \in U$ un vettore. L'insieme $S = V + v := \{w : w = u + v, u \in V\}$ viene detto uno *spazio affine* di V . Si dice che V e S sono *paralleli* e che V è lo spazio *associato* ad S . Il concetto di dimensione viene esteso ad S ponendo $\dim(S) := \dim(V)$. Si dice infine che S è un *traslato* di V e che v è la *traslazione*.

Definizione 42: [Soluzioni e sistema ammissibile.] Dato un sistema lineare $AX = \mathbf{b}$ indichiamo con

$$\text{Sol}(A, \mathbf{b})$$

l'insieme di tutte le sue soluzioni. Se $\text{Sol}(A, \mathbf{b}) \neq \emptyset$ allora il sistema si dice *ammissibile*, altrimenti si dice *non ammissibile*.

TEOREMA 24: Sia $AX = \mathbf{b}$ un sistema lineare di m equazioni in n variabili e sia $\text{Sol}(A, \mathbf{0})$ l'insieme delle soluzioni del sistema lineare omogeneo associato. Allora $\text{Sol}(A, \mathbf{0})$ è un sottospazio di \mathbb{R}^n . Inoltre, se $\text{Sol}(A, \mathbf{b}) \neq \emptyset$, allora $\text{Sol}(A, \mathbf{b})$ è un traslato di $\text{Sol}(A, \mathbf{0})$ e, detto \bar{X} un qualsiasi elemento di $\text{Sol}(A, \mathbf{b})$, si ha $\text{Sol}(A, \mathbf{b}) = \text{Sol}(A, \mathbf{0}) + \bar{X}$.

Dim: Detta $F_A : \mathbb{R}^n \mapsto \mathbb{R}^m$ l'applicazione $F_A : X \mapsto AX$ abbiamo che $\text{Sol}(A, \mathbf{0}) = \ker(F_A)$ e quindi $\text{Sol}(A, \mathbf{0})$ è un sottospazio di \mathbb{R}^n . Supponiamo ora che $\text{Sol}(A, \mathbf{b}) \neq \emptyset$ e sia $\bar{X} \in \text{Sol}(A, \mathbf{b})$. Preso $Y \in \text{Sol}(A, \mathbf{0}) + \bar{X}$, esiste $X^0 \in \text{Sol}(A, \mathbf{0})$ con $Y = X^0 + \bar{X}$ e si ha

$$F_A(Y) = F_A(X^0 + \bar{X}) = AX^0 + A\bar{X} = \mathbf{0} + \mathbf{b} = \mathbf{b}$$

e quindi $Y \in \text{Sol}(A, \mathbf{b})$.

Di converso, sia $Y \in \text{Sol}(A, \mathbf{b})$. Poniamo $X^0 = Y - \bar{X}$. Abbiamo

$$F_A(X^0) = F_A(Y - \bar{X}) = AY + A\bar{X} = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

e quindi $X^0 \in \text{Sol}(A, \mathbf{0})$ da cui $Y = X^0 + \bar{X} \in \text{Sol}(A, \mathbf{0}) + \bar{X}$. ♣

Il teorema 24 ci fa capire come la risoluzione del sistema omogeneo associato sia strettamente legata alla risoluzione di un qualsiasi sistema. In particolare, se un sistema ammette soluzioni, allora o ne ha esattamente una o ne ha infinite. Infatti l'insieme delle soluzioni è un traslato di $\ker(F_A)$ e $\ker(F_A)$ ha un unico elemento se e solo se $\ker(F_A) = \{\mathbf{0}\}$ e altrimenti ne ha infiniti. Ricordando che un'applicazione lineare è iniettiva se e solo se il suo nucleo è nullo, abbiamo che in questo caso F_A risulta iniettiva e quindi, se $\mathbf{b} \in \text{im}(F_A)$ allora la controimmagine di \mathbf{b} è unica (i.e., il sistema ha un'unica soluzione). Quello che scopriamo ora è che in un'applicazione lineare non iniettiva esistono infinite controimmagini di ogni elemento dello spazio immagine.

6.2 Il metodo di Gauss per i sistemi lineari

Occupiamoci ora di come risolvere un sistema di m equazioni in n incognite. Introduciamo un metodo di eliminazione progressiva di alcune variabili da alcune equazioni, il cui scopo è di rendere il sistema "sempre più semplice" da risolvere. Questo processo è detto il *metodo di eliminazione di Gauss*.

Definizione 43: [Sistemi equivalenti.] Siano $AX = \mathbf{b}$ e $A'X = \mathbf{b}'$ due sistemi di m equazioni nelle stesse n incognite. Allora i sistemi si dicono *equivalenti* se

$$\text{Sol}(A, \mathbf{b}) = \text{Sol}(A', \mathbf{b}')$$

Ad esempio i sistemi

$$\begin{cases} 2x - y = 1 \\ -2x + 2y = 6 \end{cases}$$

e

$$\begin{cases} 2x - y = 1 \\ y = 7 \end{cases}$$

sono equivalenti. Il secondo sistema, tuttavia, ci appare più semplice da risolvere del primo. Infatti, nel secondo sistema la seconda equazione ci dà direttamente il valore della variabile y (7). Sostituendo tale valore nella prima equazione si ricava che $2x - 7 = 1$ ossia $2x = 8$ e quindi $x = 4$.

Il metodo di Gauss è una procedura che trasforma un sistema in un sistema equivalente tramite una serie di passaggi tesi ad eliminare via via alcune variabili da equazioni consecutive (come nell'esempio precedente la x è sparita dalla seconda equazione). Ad ogni passaggio il sistema viene trasformato in un sistema equivalente intermedio, tramite l'applicazione delle seguenti regole:

G1: Si scambino fra loro due qualsiasi equazioni

G2: Si prenda una qualsiasi equazione e la si moltiplichi per un numero diverso da 0

G3: Si prenda una qualsiasi equazione si sommi ad essa un'altra equazione

Consideriamo la matrice $(A|\mathbf{b}) \in \mathbb{R}^{m \times (n+1)}$, detta anche la *matrice estesa* o *matrice completa* del sistema. Allora le operazioni appena descritte possono essere riformulate come le seguenti operazioni da applicarsi alle righe della matrice estesa:

G1: Si scambino fra loro due righe i e j $(R_i \leftrightarrow R_j)$

G2: Si prenda una riga i e la si moltiplichi per un numero $\lambda \neq 0$ $(R_i \leftarrow \lambda R_i)$

G3: Si prenda una riga i e si sommi ad essa una riga j (con $j \neq i$) $(R_i \leftarrow R_i + R_j)$

TEOREMA 25: Siano $(A|\mathbf{b})$ e $(A'|\mathbf{b}')$ le matrici estese di un sistema $AX = \mathbf{b}$ prima e dopo l'applicazione di una delle regole del processo di eliminazione di Gauss. Allora $\text{Sol}(A', \mathbf{b}') = \text{Sol}(A, \mathbf{b})$.

Dim: Se $(A'|\mathbf{b}')$ è stato ottenuto applicando G1, l'asserto è ovvio. Supponiamo sia stata applicata la regola G2 all'equazione i -ma. Siccome

$$(A_i X = b_i) \iff (\lambda(A_i X) = \lambda b_i) \iff ((\lambda A_i) X = \lambda b_i) \text{ i.e. } (A'_i X = b'_i)$$

le soluzioni del sistema sono rimaste le stesse.

Supponiamo infine che sia stata applicata la regola G3 sommando all'equazione i -ma l'equazione j -ma, $j \neq i$. Sia $X \in \text{Sol}(A, \mathbf{b})$. Essendo $A_i X = b_i$ e $A_j X = b_j$ si ha $A_i X + A_j X = b_i + b_j$, i.e., $A'_i X = b'_i$ e quindi $X \in \text{Sol}(A', \mathbf{b}')$ visto che le altre equazioni di $(A'|\mathbf{b}')$ sono le stesse di $(A|\mathbf{b})$. Viceversa, se $X \in \text{Sol}(A', \mathbf{b}')$ si ha $A'_i X = A_i X + A_j X = b_i + b_j$ ed anche $A_j X = b_j$. Quindi $A'_i X - A_j X = b_i + b_j - b_j$ ossia $A_i X = b_i$ e quindi $X \in \text{Sol}(A, \mathbf{b})$. \clubsuit

Le regole G2 e G3 possono essere convenientemente sostituite da un'unica regola, che le comprende entrambe:

G4: Si prenda una riga i , la si moltiplichi per un numero $\lambda \neq 0$ e si sommi ad essa una riga j (con $j \neq i$) moltiplicata per un numero μ qualsiasi:

$$R_i \leftarrow \lambda R_i + \mu R_j$$

Si noti che quando $\mu = 0$ la nuova regola diventa G2, mentre se $\lambda = \mu = 1$ essa diventa G3.

6.2.1 Riduzione a scalini di un sistema

Definizione 44: [Matrice a scalini. Pivots, variabili dipendenti e variabili libere.] Sia $R = (A|\mathbf{b})$ la matrice estesa, di m righe e $n + 1$ colonne, di un sistema lineare $AX = \mathbf{b}$. Sia r l'indice dell'ultima riga di R in cui compare almeno un elemento diverso da zero. Per ogni $i = 1, \dots, r$ sia $p(i)$ la prima colonna in cui appare un elemento diverso da zero (detto il *pivot*) in riga i . La matrice R (e il corrispondente sistema) si dice *a scalini* se

1. $p(1) < p(2) < \dots < p(r)$
2. Per ogni $i = r + 1, \dots, m$ la riga i -ma di R è fatta di soli zeri

In un sistema ridotto a scalini, le variabili che compaiono nelle colonne dei pivot vengono dette *variabili dipendenti*, mentre le restanti vengono dette *variabili libere*.

Esempio. La matrice

$$\begin{pmatrix} 0 & 7 & 1 & 0 & 0 & -1 \\ 0 & 0 & -3 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

è a scalini. In essa. $r = 3$, mentre $p(1) = 2$, $p(2) = 3$ e $p(3) = 5$. \diamond

Il nome dato a questo tipo di matrici deriva dal fatto che, se in ogni riga scriviamo solo gli elementi nulli consecutivi, da sinistra verso destra, otteniamo una specie di scalinata:

$$\begin{pmatrix} 00 & & & & \\ 0000 & & & & \\ 000000 & & & & \\ 00000000 & & & & \\ 0000000000 & & & & \\ 000000000000 & & & & \\ 00000000000000 & & & & \\ 0000000000000000 & & & & \end{pmatrix}$$

Esempio. Si consideri il seguente sistema lineare a scalini di 3 equazioni in 5 incognite:

$$\begin{cases} x_1 + 4x_2 + 3x_3 - x_5 = 2 \\ 2x_3 + 2x_4 + 3x_5 = 3 \\ -2x_4 + 5x_5 = 4 \end{cases}$$

In esso, x_1, x_3, x_4 sono le variabili dipendenti, mentre x_2 e x_5 sono le variabili libere. \diamond

Ogni matrice (e quindi ogni sistema) può essere ridotta a scalini tramite le due regole di eliminazione di Gauss G1 e G4, i.e.

- **Scambio di righe:** $R_i \leftrightarrow R_j$
- **Somma pesata di righe:** $R_i \leftarrow \lambda R_i + \mu R_j$ per $i \neq j$ e $\lambda \neq 0$

È facile rendersi conto di come queste operazioni siano sufficienti a rendere a scalini una matrice $R = (r_{ij})$. Infatti, se R è una matrice di soli zeri, o se R ha un'unica riga, allora R è già a scalini. Altrimenti, sia c la prima colonna da sinistra in cui compare un elemento non-nullo, e sia i l'indice di una riga tale che $r_{ic} \neq 0$. A questo punto: (i) si scambia la riga i con la riga 1; (ii) nella nuova matrice (che chiameremo ancora R), per ogni riga $j = 2, \dots, m$ tale che $r_{jc} \neq 0$, si rimpiazza la riga j con $r_{jc}R_1 - r_{1c}R_j$. Si noti che la seconda operazione fa sì che in colonna c , riga $j = 2, \dots, m$, appaia il coefficiente $r_{jc}r_{1c} - r_{1c}r_{jc} = 0$. Infine, se la matrice ha più di una riga, si prosegue ricorsivamente allo stesso modo rendendo a scalini la sottomatrice formata dalle righe $2, \dots, m$.

Esempio. Rendiamo a scalini la matrice

$$\begin{pmatrix} -1 & 2 & 1 & 0 & 2 \\ 0 & 0 & -3 & -6/5 & 4 \\ 5 & 6 & 1/2 & 1 & 2 \\ 4 & 8 & 3/2 & 1 & 4 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

La riga R_1 è già tale che $r_{11} \neq 0$. Rimpiazziamo la riga R_3 con $5R_1 + R_3$, e la riga R_4 con $4R_1 + R_4$, ottenendo la matrice

$$\begin{pmatrix} -1 & 2 & 1 & 0 & 2 \\ 0 & 0 & -3 & -6/5 & 4 \\ 0 & 16 & 11/2 & 1 & 12 \\ 0 & 16 & 11/2 & 1 & 12 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$/* R_3 \leftarrow 5R_1 + R_3, R_4 \leftarrow 4R_1 + R_4 */$$

Scambiamo fra loro la riga R_5 ed R_2 in modo da ottenere il coefficiente 1 in r_{22} :

$$\begin{pmatrix} -1 & 2 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 16 & 11/2 & 1 & 12 \\ 0 & 16 & 11/2 & 1 & 12 \\ 0 & 0 & -3 & -6/5 & 4 \end{pmatrix}$$

$$/* R_5 \leftrightarrow R_2 */$$

Procediamo poi ad azzerare gli elementi in colonna 2 nelle righe 3,4,5. In particolare, per $i = 3, 4$, sostituiamo R_i con $R_i - 16R_2$, ottenendo

$$\begin{pmatrix} -1 & 2 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 11/2 & 1 & -4 \\ 0 & 0 & 11/2 & 1 & -4 \\ 0 & 0 & -3 & -6/5 & 4 \end{pmatrix}$$

$$/* R_3 \leftarrow R_3 - 16R_2, R_4 \leftarrow R_4 - 16R_2 */$$

L'elemento $r_{33} = 11/2$ è diverso da zero. Procediamo ad azzerare gli elementi sotto ad r_{33} , sostituendo R_4 con $R_4 - R_3$ ed R_5 con $(11/2)R_5 + 3R_3$, ottenendo

$$\begin{pmatrix} -1 & 2 & 1 & 0 & 2 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 11/2 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -18/5 & 10 \end{pmatrix}$$

$$/* R_4 \leftarrow R_4 - R_3, R_5 \leftarrow (11/2)R_5 + 3R_3 */$$

Il primo elemento non-nullo nelle righe 4 e 5 appare in colonna 4. Scambiamo le righe 4 e 5 ottenendo la matrice finale a scalini:

$$\begin{pmatrix} -1 & 2 & 1 & 0 & 2 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 11/2 & 1 & -4 \\ 0 & 0 & 0 & -18/5 & 10 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$/* R_5 \leftrightarrow R_4 */$$

◇

Sia $AX = \mathbf{b}$ un sistema lineare che vogliamo risolvere. Procediamo come segue:

1. Tramite una serie di applicazioni delle regole di eliminazione di Gauss, trasformiamo $(A|\mathbf{b})$ in un sistema a scalini equivalente, sia esso $(A'|\mathbf{b}')$. Supponiamo che nel sistema finale vi siano r righe non-nulle (le righe da 1 a r).
2. Se la riga r è del tipo $(0 \cdots 0|b)$ con $b \neq 0$ il sistema non ha soluzioni. Infatti, tale riga corrisponde all'equazione

$$0x_1 + \cdots + 0x_n = b$$

che ovviamente non ha alcuna soluzione se $b \neq 0$.

3. In caso contrario, il sistema ha r variabili dipendenti, i.e., le variabili $x_{p(i)}$ per $i = 1, \dots, r$, mentre le restanti $n - r$ variabili sono libere.
4. Procedendo a ritroso, dall'equazione r -ma all'equazione 1, si ricava il valore della variabile dipendente $x_{p(i)}$ in funzione delle variabili libere di indice $> p(i)$ e delle variabili dipendenti $x_{p(j)}$ con $j > i$ precedentemente calcolate.
5. Si noti che se $n - r > 0$ il sistema avrà infinite soluzioni (una per ogni arbitrario assegnamento di valori alle variabili libere), mentre se $n = r$ il sistema ha un'unica soluzione.

Esempio. Si consideri il seguente sistema di 3 equazioni in 3 incognite:

$$\begin{cases} 2x + 4y + 6z = 2 \\ x - y + z = 0 \\ 3x - 2z = -1 \end{cases}$$

La matrice estesa dei coefficienti è

$$\begin{pmatrix} 2 & 4 & 6 & 2 \\ 1 & -1 & 1 & 0 \\ 3 & 0 & -2 & -1 \end{pmatrix}$$

Procedendo con il processo di eliminazione di Gauss rendiamo il sistema a scalini:

$$\begin{pmatrix} 2 & 4 & 6 & 2 \\ 0 & 6 & 4 & 2 \\ 0 & 12 & 22 & 8 \end{pmatrix}$$

$$/* R_2 \leftarrow R_1 - 2R_2, R_3 \leftarrow 3R_1 - 2R_3 */$$

$$\begin{pmatrix} 2 & 4 & 6 & 2 \\ 0 & 6 & 4 & 2 \\ 0 & 0 & 7 & 2 \end{pmatrix}$$

$$/* R_3 \leftarrow 1/2R_3 - R_2 */$$

Il sistema finale a scalini è

$$\begin{cases} 2x + 4y + 6z = 2 \\ 6y + 4z = 2 \\ 7z = 2 \end{cases}$$

In questo sistema non ci sono variabili libere. Dall'ultima equazione ricaviamo $z = 2/7$. Sostituendo nella seconda abbiamo $y = 1/6(2 - 8/7) = 1/7$. Sostituendo i valori di y e z nella prima equazione abbiamo $x = 1/2(2 - 4/7 - 12/7) = -1/7$. Quindi il sistema ha un'unica soluzione, i.e., $(x, y, z) = (-1/7, 1/7, 2/7)$. \diamond

Esempio. Si consideri il seguente sistema di 3 equazioni in 4 incognite:

$$\begin{cases} x + 4y + 3z - w = 2 \\ -x - 2y + z + 2w = 3 \\ 2y + 4z + w = 4 \end{cases}$$

La matrice estesa dei coefficienti è

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

Procedendo con il processo di eliminazione di Gauss rendiamo il sistema a scalini:

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

$$/* R_2 \leftarrow R_2 - R_1 */$$

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

$$/* R_3 \leftarrow R_3 - R_2 */$$

Il sistema finale a scalini è

$$\begin{cases} x + 4y + 3z - w = 2 \\ -2y + 4z + w = 5 \\ 0 + 0 + 0 + 0 = -1 \end{cases}$$

Questo sistema non ha soluzioni. L'ultima equazione è in effetti $0x + 0y + 0z + 0w = -1$ che ovviamente non ha alcuna soluzione possibile. \diamond

- esempio con infinite soluzioni
- esempio con sistema omogeneo
- riduzione a scala diagonale (pivots =1 e sopra e sotto 0)
- operazioni di gauss come premoltiplicazione per una matrice
- calcolo dell'inversa tramite operazioni di gauss

6.3 Il rango di una matrice

Sia A una matrice $m \times n$. Definiamo il *rango per righe* di A come il numero massimo di righe linearmente indipendenti e il *rango per colonne* di A come il numero massimo di colonne linearmente indipendenti. In altre parole, il rango per righe è la dimensione del sottospazio di \mathbb{R}^n generato dalle righe di A , mentre il rango per colonne è la dimensione del sottospazio di \mathbb{R}^m generato dalle colonne di A . Dimostriamo ora che i due sono uguali, e quindi potremo parlare semplicemente di rango di A .

Consideriamo le applicazioni di riduzione a scalini. Ognuna di esse può essere vista come la (pre)moltiplicazione di A per un'opportuna matrice quadrata $m \times m$. Lo scambio delle righe i e j di A corrisponde a pre-moltiplicare A per una matrice B che è uguale alla matrice I_m tranne che per i 4 elementi all'incrocio delle righe $\{i, j\}$ e colonne $\{i, j\}$ dove si ha

$$B_{ii} = B_{jj} = 0, \quad B_{ij} = B_{ji} = 1$$

La sostituzione della riga i con somma di λ volte la riga i e μ volte la riga j corrisponde a pre-moltiplicare A per una matrice B che è uguale alla matrice I_m tranne che per 2 elementi sulla riga i dove si ha

$$B_{ii} = \lambda, \quad B_{ij} = \mu$$

Si noti che, in entrambi i casi, la matrice B può essere interpretata come una matrice di un cambiamento di base in \mathbb{R}^m . In particolare, esso è il cambiamento dalla base costituita dalle colonne di B , che chiameremo \mathcal{B} , alla base canonica \mathcal{C}_m .

Se interpretiamo A come la matrice di un'applicazione lineare $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ rispetto alle basi \mathcal{B} di \mathbb{R}^m e la base canonica \mathcal{C}_n di \mathbb{R}^n , sappiamo che il rango di F , ossia la dimensione di $\text{im}(F)$, è pari alla dimensione dello spazio generato dalle colonne di A , e quindi al rango per colonne di A . La matrice A' risultante dall'applicazione di una regola di Gauss può essere vista come rappresentativa della stessa funzione F ma tra la base canonica \mathcal{C}_n di \mathbb{R}^n e la base \mathcal{C}_m di \mathbb{R}^m in quanto:

$$A' = BA = \mathcal{M}[\text{id}, \mathcal{B}, \mathcal{C}_m] \mathcal{M}[F, \mathcal{C}_n, \mathcal{B}] = \mathcal{M}[F, \mathcal{C}_n, \mathcal{C}_m]$$

Siccome l'applicazione F è la stessa, il rango per colonne di A' è uguale alla dimensione dell'immagine di F che era uguale al rango per colonne di A e concludiamo che

Nell'applicare una regola di Gauss a una matrice, il rango per colonne rimane lo stesso.

Per quel che riguarda il rango per righe, notiamo che anch'esso rimane lo stesso. Infatti, dopo uno scambio di righe l'insieme delle righe (viste come vettori di \mathbb{R}^n) è lo stesso e quindi il numero di righe indipendenti è lo stesso. Se invece l'operazione era sostituire la riga i con $\lambda A_i + \mu A_j$, dove $\lambda \neq 0$ è facile dimostrare che le nuove righe generano lo stesso sottospazio di \mathbb{R}^n delle vecchie. Quindi vale anche

Nell'applicare una regola di Gauss a una matrice, il rango per righe rimane lo stesso.

In conclusione, detta A una matrice e detta A' la matrice diagonale a scalini ottenuta da A con l'applicazione di un certo numero di regole di Gauss abbiamo che

Dopo la riduzione tramite Gauss di una matrice A a una matrice diagonale a scalini A' , il rango per righe e quello per colonne di A' sono gli stessi di quelli di A .

Nella matrice ridotta a scalini, però, il calcolo di tali ranghi è immediato. Supponiamo la matrice abbia r righe non nulle. La riga $r - 1$ è indipendente dalla riga r in quanto la r ha zero nella colonna $p(r - 1)$, mentre $A_{r-1,p(r-1)} = 1$. Allo stesso modo, la riga $r - 2$ è indipendente dalle righe r e $r - 1$ in quanto entrambe hanno zero nella colonna $p(r - 2)$, mentre $A_{r-2,p(r-2)} = 1$. Proseguendo, per ogni $i \geq 1$, la riga i è indipendente dalle righe $i + 1, \dots, r$ in quanto $A_{k,p(i)} = 0$ per $k > i$ e $A_{i,p(i)} \neq 0$. Quindi le righe da 1 a r sono indipendenti e il rango per righe di A' è r .

Per quel che riguarda le colonne, consideriamo le colonne $p(1), \dots, p(r)$ contenenti i pivots di A' . Con un ragionamento analogo a quello fatto per le righe, la colonna $p(j)$ è indipendente dalle colonne $p(j + 1), \dots, p(r)$ in quanto $A_{j,p(j)} \neq 0$ mentre $A_{j,p(k)} = 0$ per $k > j$. Quindi il rango per colonne di A' è almeno r . Ma essendo tutte le altre colonne ottenibili come combinazione lineare delle colonne dei pivot, il rango è esattamente r . Concludiamo quindi che

TEOREMA 26: In una matrice $A \in \mathbb{R}^{m \times n}$ la dimensione del sottospazio di \mathbb{R}^n generato dalle righe (il rango per righe) e del sottospazio di \mathbb{R}^m generato dalle colonne (il rango per colonne) sono uguali. Il loro valore viene detto il *rango* della matrice A . Tale rango è anche uguale al rango di una qualsiasi applicazione lineare F che abbia, scelte opportune basi, A come matrice che la rappresenta.

Chapter 7

Determinanti

In questo capitolo descriveremo una funzione, detta *determinante*, che ad ogni n -pla di vettori in \mathbb{R}^n associa un numero reale. Oltre ad altre proprietà, questa funzione ha la caratteristica di mappare una n -pla di vettori $v_1, \dots, v_n \in \mathbb{R}^n$ nel valore 0 se e solo se i vettori v_i sono linearmente dipendenti.

Cominciamo col chiederci se esiste una funzione $\det(v_1, \dots, v_n)$ di $\mathbb{R}^n \times \dots \times \mathbb{R}^n \mapsto \mathbb{R}$ che goda delle seguenti proprietà:

1. **Multilinearità:** Per ogni $1 \leq i \leq n$, scalari $\lambda, \mu \in \mathbb{R}$ e vettori $v, v_1, \dots, v_n \in \mathbb{R}^n$ si ha

$$\det(v_1, \dots, v_{i-1}, \lambda v_i + \mu v, v_{i+1}, \dots) = \lambda \det(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots) + \mu \det(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots)$$

In altri termini, la funzione $\det(v_1, \dots, v_n)$ deve essere lineare rispetto a ciascuno dei suoi argomenti.

2. **Alternanza:** Per ogni $1 \leq i < j \leq n$ si ha

$$\det(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_{j-1}, v_j, v_{j+1}, \dots) = -\det(v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots)$$

In altri termini, scambiando fra loro due qualsiasi degli argomenti, il valore della funzione diventa l'opposto.

3. **Normalizzazione:** Sia (e_1, \dots, e_n) la base canonica di \mathbb{R}^n . Allora

$$\det(e_1, \dots, e_n) = 1$$

Da queste proprietà ne seguono delle altre. Ad esempio, una tale funzione deve essere *omogenea*, ossia, per ogni $\lambda \in \mathbb{R}$ e $1 \leq i \leq n$ si ha

$$\det(v_1, \dots, v_{i-1}, \lambda v_i, v_{i+1}, \dots) = \lambda \det(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots) \quad (7.1)$$

(questo segue dalla multilinearità, essendo $\lambda v_i = \lambda v_i + 0v_i$. Dall'omogeneità segue poi che

$$\det(v_1, \dots, v_n) = 0 \quad \text{se } v_i = \mathbf{0} \text{ per qualche } i \in \{1, \dots, n\} \quad (7.2)$$

Dalla proprietà di alternanza segue che

$$\det(v_1, \dots, v_n) = 0 \quad \text{se } v_i = v_j \text{ per qualche } i \neq j \quad (7.3)$$

Infatti, se $v_i = v_j$ scambiando v_i con v_j il valore della funzione deve rimanere lo stesso ma, per la proprietà dell'alternanza, deve anche cambiare di segno, e l'unico numero che è uguale al suo opposto è lo zero.

Dalla multilinearità e da (7.2) segue la proprietà di *invarianza per scorrimento*:

$$\det(v_1, \dots, v_i + \lambda v_j, \dots, v_n) = \det(v_1, \dots, v_i, \dots, v_n) \quad (7.4)$$

In altri termini, sommando a uno qualsiasi dei vettori un multiplo di un altro, il valore della funzione non cambia.

Si noti che sia il caso in (7.2) che quello in (7.3) sono esempi in cui i vettori v_1, \dots, v_n sono linearmente dipendenti. Dimostriamo ora come in tutti i casi in cui i vettori v_1, \dots, v_n sono dipendenti il valore della funzione $\det()$ deve valere zero.

Lemma 27: Siano v_1, \dots, v_n vettori linearmente dipendenti in \mathbb{R}^n . Allora $\det(v_1, \dots, v_n) = 0$.

Dim: Se i vettori v_1, \dots, v_n sono linearmente dipendenti, allora almeno uno di loro può essere espresso come combinazione lineare degli altri. Senza perdita di generalità supponiamo che v_1 possa essere espresso come combinazione lineare di v_2, \dots, v_n , ossia esistano coefficienti $\lambda_2, \dots, \lambda_n$ tali che

$$v_1 = \lambda_2 v_2 + \dots + \lambda_n v_n$$

Abbiamo allora che

$$\det(v_1, \dots, v_n) = \det\left(\sum_{i=2}^n \lambda_i v_i, v_2, \dots, v_n\right) = \sum_{i=2}^n \lambda_i \det(v_i, v_2, \dots, v_n) = 0$$

dove la seconda uguaglianza segue dalla multilinearità, mentre la terza segue dal fatto che in ogni termine $\det(v_i, v_2, \dots, v_n)$, per $i = 2, \dots, n$, ci sono due vettori v_i uguali fra loro. ♣

Dimostriamo ora come la condizione $\det(v_1, \dots, v_n) = 0$ sia non solo necessaria, ma anche sufficiente per la dipendenza lineare di v_1, \dots, v_n .

Lemma 28: Siano v_1, \dots, v_n vettori linearmente indipendenti in \mathbb{R}^n . Allora $\det(v_1, \dots, v_n) \neq 0$.

Dim: Se i vettori v_1, \dots, v_n sono linearmente indipendenti, essi sono una base di \mathbb{R}^n . Di conseguenza, ogni vettore della base canonica può essere espresso come combinazione lineare dei v_i , ed esistono coefficienti $\alpha[i, j]$ tali che

$$e_j = \sum_{i=1}^n \alpha[i, j] v_i \quad \forall 1 \leq j \leq n$$

Dalla multilinearità si ha allora

$$\det(e_1, \dots, e_n) = \det\left(\sum_{i_1=1}^n \alpha[i_1, 1] v_{i_1}, \dots, \sum_{i_n=1}^n \alpha[i_n, n] v_{i_n}\right) = \sum_{i_1=1}^n \dots \sum_{i_n=1}^n \left(\alpha[i_1, 1] \alpha[i_2, 2] \dots \alpha[i_n, n] \det(v_{i_1}, \dots, v_{i_n})\right) \quad (7.5)$$

Ora, se fosse $\det(v_1, \dots, v_n) = 0$, allora si avrebbe che

$$\det(v_{i_1}, v_{i_1}, \dots, v_{i_n}) = 0$$

per ogni scelta di i_1, i_2, \dots, i_n , in quanto o ci sono due i_k identici, oppure tutti gli i_k sono diversi e in questo caso costituiscono una permutazione di $(1, 2, \dots, n)$ e, dall'alternanza, si avrebbe

$$|\det(v_{i_1}, v_{i_1}, \dots, v_{i_n})| = |\det(v_1, v_2, \dots, v_n)| = |0| = 0$$

Quindi, se fosse $\det(v_1, \dots, v_n) = 0$, si avrebbe che il termine a destra in (7.5) vale 0. Ma, dalla normalità, il termine di sinistra vale 1, e quindi ne deriva un assurdo. ♣

Nella dimostrazione del Lemma 28 abbiamo notato come

$$|\det(v_{i_1}, v_{i_1}, \dots, v_{i_n})| = |\det(v_1, v_2, \dots, v_n)|$$

quando i_1, \dots, i_n sono una permutazione di $1, \dots, n$. Ricordiamo dal calcolo combinatorico che, data una permutazione π , il *segno* di π è $+1$ se il numero di scambi tra elementi di π necessari a trasformare π in $(1, \dots, n)$ è pari, e -1 altrimenti. Denotiamo con $\text{sgn}(\pi)$ il segno di una permutazione π . Siccome ad ogni scambio di elementi il valore di $\det()$ cambia di segno, abbiamo che, detta π la permutazione (i_1, \dots, i_n)

$$\det(v_{i_1}, v_{i_1}, \dots, v_{i_n}) = \text{sgn}(\pi) \det(v_1, v_2, \dots, v_n)$$

7.1 Esistenza e unicità del determinante

Siano $v_1, \dots, v_n \in \mathbb{R}^n$. Esistono coefficienti $a[i, j]$ tali che

$$v_j = \sum_{i=1}^n a[i, j] e_i \quad \forall 1 \leq j \leq n$$

Procedendo come nel lemma 28, dalla multilinearità segue che

$$\det(v_1, \dots, v_n) = \sum_{i_1=1}^n \cdots \sum_{i_n=1}^n \left(a[i_1, 1] a[i_2, 2] \cdots a[i_n, n] \det(e_{i_1}, \dots, e_{i_n}) \right) \quad (7.6)$$

Nella parte destra di (7.6) gli unici termini non-nulli si hanno quando (i_1, \dots, i_n) è una permutazione $\pi = (\pi(1), \dots, \pi(n))$ di $\{1, \dots, n\}$, e in questo caso $\det(e_{i_1}, \dots, e_{i_n}) = \pm 1 = \text{sgn}((i_1, \dots, i_n))$. Detto S_n l'insieme di tutte le permutazioni di $\{1, \dots, n\}$, possiamo allora riscrivere (7.6) come

$$\det(v_1, \dots, v_n) = \sum_{\pi \in S_n} \text{sgn}(\pi) a[\pi(1), 1] \cdot a[\pi(2), 2] \cdots a[\pi(n), n] \quad (7.7)$$

Siccome ogni funzione che soddisfi alle proprietà di multilinearità, alternanza e normalizzazione deve anche soddisfare la relazione (7.7) resta dimostrata l'unicità di una tale funzione (qualora esistesse).

Per dimostrarne l'esistenza, dobbiamo allora far vedere che la funzione f che mappa una n -pla di vettori v_1, \dots, v_n , dove ogni v_j è del tipo

$$v_j = \begin{pmatrix} a[1, j] \\ a[2, j] \\ \dots \\ a[n, j] \end{pmatrix}$$

nel valore

$$f(v_1, \dots, v_n) = \sum_{\pi \in S_n} \text{sgn}(\pi) a[\pi(1), 1] \cdot a[\pi(2), 2] \cdots a[\pi(n), n] \quad (7.8)$$

soddisfa la normalizzazione, è alternante ed è multilineare.

1. Normalizzazione: Se $v_j = e_j$ per ogni j , allora $a[i, j] = 0$ se $i \neq j$ e $a[i, i] = 1$. Quindi l'unico termine non-nullo in (7.8) è $a[1, 1] \cdot a[2, 2] \cdots a[n, n] = 1$. Inoltre $\text{sgn}((1, 2, \dots, n)) = 1$ per cui

$$f(e_1, \dots, e_n) = 1$$

2. Alternanza: Se due vettori v_i e v_j vengono scambiati tra loro, l'effetto è che in ognuno dei prodotti $a[\pi(1), 1] \cdot a[\pi(2), 2] \cdots a[\pi(n), n]$ due termini vengono scambiati tra loro. Chiaramente, il valore di ogni prodotto non cambia.
3. Multilinearità: ** to do **

7.2 Determinante di matrici quadrate

1. Data una matrice $A \in \mathbb{R}^{n \times n}$ definiamo il determinante della matrice, $\det(A)$, come il determinante dei vettori-colonna di A : $\det(A) = \det(A^1, \dots, A^n)$. Si noti che dalla proprietà di normalizzazione segue che

$$\det(I_n) = 1$$

dove I_n è la matrice identità di dimensioni $n \times n$.

2. Una matrice $A \in \mathbb{R}^{n \times n}$ si dice *non-singolare* se ha rango pieno, i.e., se $rk(A) = n$. Questo avviene se e solo se il suo determinante è diverso da zero. Quindi, una matrice quadrata A è invertibile se e solo se $\det(A) \neq 0$.
3. Det di matrice 2x2 calcolato con le permutazioni. Le uniche permutazioni sono $(1, 2)$ e $(2, 1)$. Dalla formula abbiamo

$$\det(A) = \text{sgn}((1, 2))a_{11}a_{22} + \text{sgn}((2, 1))a_{21}a_{12} = a_{11}a_{22} - a_{21}a_{12}$$

4. Det di matrice 3x3 calcolato con le permutazioni. Le permutazioni pari sono $(1, 2, 3)$, $(2, 3, 1)$ e $(3, 1, 2)$. Quelle dispari sono $(1, 3, 2)$, $(2, 1, 3)$ e $(3, 2, 1)$. Abbiamo quindi

$$\det(A) = (a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23}) - (a_{11}a_{32}a_{23} + a_{21}a_{12}a_{33} + a_{31}a_{22}a_{13})$$

5. *Regola di Sarrus* per matrice 3x3. Consideriamo la matrice ottenuta affiancando ad A una copia delle prime due colonne:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{pmatrix}$$

Si considerino le prime 3 diagonali principali (NO-SE) e si sommino i prodotti dei corrispondenti elementi. Si considerino le ultime 3 anti-diagonali principali (NE-SO) e si sommino i prodotti dei corrispondenti elementi. Allora il determinante di A è dato dalla prima somma meno la seconda.

PROPOSIZIONE 29: Sia $A = (a_{ij}) \in \mathbb{R}^{n \times n}$. Allora $\det({}^t A) = \det(A)$ (i.e. *il determinante della trasposta è uguale al determinante della matrice stessa*).

Dim: Sia $\pi = (\pi(1), \dots, \pi(n)) \in S_n$ e sia $\tau = \pi^{-1}$ la permutazione inversa di π , sicchè, per ogni i , si ha $\tau(\pi(i)) = i$. Indichiamo con (a'_{ij}) gli elementi di ${}^t A$. Abbiamo

$$\begin{aligned} \mathcal{P}(a, \pi) &:= \operatorname{sgn}(\pi) a[\pi(1), 1] \cdots a[\pi(n), n] \\ &= \operatorname{sgn}(\pi) a'[1, \pi(1)] \cdots a'[n, \pi(n)] \\ &= \operatorname{sgn}(\pi) a'[\tau(\pi(1)), \pi(1)] \cdots a'[\tau(\pi(n)), \pi(n)] \\ &= \dots \text{ riordiniamo mettendo prima la coppia in cui } \pi(i) = 1, \text{ poi quella con } \pi(j) = 2 \text{ ecc } \dots \\ &= \operatorname{sgn}(\pi) a'[\tau(1), 1] \cdots a'[\tau(n), n] \\ &= \operatorname{sgn}(\tau) a'[\tau(1), 1] \cdots a'[\tau(n), n] \quad (\text{essendo } \operatorname{sgn}(\pi) = \operatorname{sgn}(\pi^{-1})) \\ &= \mathcal{P}(a', \tau) \\ &= \mathcal{P}(a', \pi^{-1}) \end{aligned}$$

Si ottiene quindi che

$$\det(A) = \sum_{\pi \in S_n} \mathcal{P}(a, \pi) = \sum_{\pi \in S_n} \mathcal{P}(a', \pi^{-1}) = \sum_{\tau \in S_n} \mathcal{P}(a', \tau) = \det({}^t A)$$

in quanto, al variare di π su S_n anche π^{-1} assume tutti i valori in S_n . ♣

PROPOSIZIONE 30: Sia $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ una matrice triangolare. Allora

$$\det(A) = \prod_{i=1}^n a_{ii}$$

(i.e. *il determinante di una matrice triangolare, e quindi anche di una matrice diagonale, è uguale al prodotto degli elementi della diagonale principale*).

Dim: S.p.d.g. supponiamo A triangolare superiore. Il determinante è pari a

$$\sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{\pi(1), 1} \cdot a_{\pi(2), 2} \cdots a_{\pi(n), n} \quad (7.9)$$

Sia π una qualsiasi permutazione diversa da $(1, 2, \dots, n)$. Sia j il minimo indice i tale che $\pi(i) \neq i$. Avremo necessariamente $\pi(j) > j$ e quindi, essendo $a_{p(j),j} = 0$, il contributo di π alla somma (7.9) è 0. In conclusione, l'unica permutazione per la quale il contributo può essere non-nullo è la permutazione in cui $\pi(i) = i$ per ogni i , da cui la tesi. \clubsuit

TEOREMA 31: (Binet) Siano $A, B \in \mathbb{R}^{n \times n}$. Allora $\det(AB) = \det(A)\det(B)$ (i.e., *il determinante del prodotto è uguale al prodotto dei determinanti*).

Dim: Sia $C = AB$. Dalla definizione di prodotto matriciale si ha che, per ogni $j = 1, \dots, n$

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$$

e quindi la j -ma colonna di C è del tipo

$$C^j = \begin{pmatrix} a_{11}b_{1j} + a_{12}b_{2j} + \dots + a_{1n}b_{nj} \\ a_{21}b_{1j} + a_{22}b_{2j} + \dots + a_{2n}b_{nj} \\ \dots \\ a_{n1}b_{1j} + a_{n2}b_{2j} + \dots + a_{nn}b_{nj} \end{pmatrix} = b_{1j} \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{n1} \end{pmatrix} + \dots + b_{nj} \begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{nn} \end{pmatrix} = \sum_{k=1}^n b_{kj} A^k$$

Abbiamo allora

$$\begin{aligned} \det(AB) &= \det(C^1, \dots, C^n) \\ &= \det\left(\sum_{i_1=1}^n b_{i_1 1} A^{i_1}, \dots, \sum_{i_n=1}^n b_{i_n n} A^{i_n}\right) \\ &= \sum_{i_1=1}^n \dots \sum_{i_n=1}^n b_{i_1 1} \dots b_{i_n n} \det(A^{i_1}, \dots, A^{i_n}) \\ &= \sum_{\sigma=(i_1, \dots, i_n) \in S_n} \text{sgn}((i_1, \dots, i_n)) b_{i_1 1} \dots b_{i_n n} \det(A^1, \dots, A^n) \\ &= \sum_{\sigma \in S_n} \mathcal{P}(b, \sigma) \det(A) \\ &= \det(A) \sum_{\sigma \in S_n} \mathcal{P}(b, \sigma) \\ &= \det(A) \det(B). \end{aligned}$$

\clubsuit

Dalla proposizione 31 segue il corollario:

COROLLARIO 32: Sia A una matrice invertibile. Allora

$$\det A^{-1} = \frac{1}{\det A}$$

(i.e., *il determinante dell'inversa è l'inverso del determinante*).

Dim: Abbiamo

$$\det(AA^{-1}) = \det(A)\det(A^{-1}) = \det(I) = 1$$

da cui la tesi. ♣

7.3 Note da mettere a posto

7.3.1 Determinante di applicazioni lineari

Sia $T : V \mapsto V$ un'applicazione lineare. Data una base \mathcal{B} sia $A = \mathcal{M}[T, \mathcal{B}, \mathcal{B}]$. Definiamo il determinante di T come il determinante di A . La definizione è corretta in quanto qualsiasi scelta di A va ugualmente bene. Infatti, sia \mathcal{B}' un'altra base, sia $B = \mathcal{M}[T, \mathcal{B}', \mathcal{B}']$ e sia $C = \mathcal{M}[\text{id}, \mathcal{B}', \mathcal{B}]$ la matrice del cambiamento di base, sicchè $B = C^{-1}AC$. Abbiamo

$$\det(B) = \det(C^{-1}AC) = \det(C^{-1})\det(A)\det(C) = \frac{1}{\det(C)}\det(A)\det(C) = \det(A)$$

7.3.2 Definizione ricorsiva

1. Data matrice $A = (a_{ij})$ per ogni riga i e colonna j definiamo la *matrice aggiunta* $A(i, j)$ di a_{ij} come la matrice ottenuta da A rimuovendo la riga i e la colonna j .
2. Formula ricorsiva per il determinante: Se $n = 1$ definiamo $\text{Det}(A) = a_{11}$. Per $n > 1$ si ha

$$\text{Det}(A) = a_{11}\text{Det}(A(1, 1)) - a_{12}\text{Det}(A(1, 2)) + \cdots + (-1)^{1+n}\text{Det}(A(1, n))$$

Come sommatoria

$$\text{Det}(A) = \sum_{j=1}^n (-1)^{1+j} a_{1j} \text{Det}(A(1, j))$$

3. Si può dimostrare l'equivalenza tra il determinante definito in termini di permutazioni segnate e quello della formula ricorsiva, quindi $\text{Det}(A) = \det(A)$ e useremo solo la notazione $\det(A)$
4. Teorema di Laplace: Per ogni riga i si ha

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A(i, j)) \quad (7.10)$$

Parliamo in questo caso di determinante *sviluppato secondo la riga i* . Essendo inoltre $\det(A) = \det({}^t A)$ abbiamo anche che per ogni colonna j

$$\det(A) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A(i, j)) \quad (7.11)$$

Parliamo in questo caso di determinante *sviluppato secondo la colonna j* .

5. Il teorema di Laplace suggerisce di sviluppare il determinante secondo una riga o una colonna che contenga molti zeri. Infatti, in corrispondenza a $a_{ij} = 0$ non è necessario calcolare $\det(A(i, j))$.
6. Per ricordare velocemente il segno da usare nelle formule (7.10) e (7.11) si può usare uno schema *a scacchiera* in cui si mette un "+" in corrispondenza a $(-1)^{i+j} = +1$ e un "-" in corrispondenza a $(-1)^{i+j} = -1$:

$$\begin{pmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

7. Calcolo dell'inversa con i determinanti: Sia A una matrice invertibile e sia $B = A^{-1}$. Allora si ha

$$b_{ij} = (-1)^{i+j} \frac{\det(A(j, i))}{\det(A)}$$

8. Complemento algebrico:

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

Possiamo riscrivere le formule del determinante come

$$\det(A) = a_{i1}\bar{a}_{i1} + a_{i2}\bar{a}_{i2} + \cdots + a_{in}\bar{a}_{in}$$

e

$$\det(A) = a_{1j}\bar{a}_{1j} + a_{2j}\bar{a}_{2j} + \cdots + a_{nj}\bar{a}_{nj}$$

e quella dell'inversa

$$b_{ij} = \frac{\bar{a}_{ji}}{\det(A)}$$

sicchè

$$A^{-1} = \frac{1}{\det(A)} {}^t \bar{A}$$

Si noti come i complementi algebrici siano quelli della trasposta di A e non di A

9. Sistemi lineari e regola di Cramer. Sia $AX = B$ un sistema lineare con $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times 1}$ e $X = {}^t(x_1, \dots, x_n)$. Supponiamo che A sia invertibile, sicchè il sistema, se ammissibile, ha un'unica soluzione. Denotiamo tale soluzione con $(\bar{x}_1, \dots, \bar{x}_n)$.

Sia j una colonna di A . Indichiamo con $A(j)$ la matrice ottenuta da A rimpiazzando la colonna j -ma con il vettore B . Allora per ogni $j = 1, \dots, n$ si ha

$$\bar{x}_j = \frac{\det(A(j))}{\det(A)}$$

La soluzione di un sistema secondo questa formula viene detta il *metodo di Cramer*.

10. Calcolo del determinante con metodo di Gauss: Data una matrice A , possiamo manipolare A in modo da rendere più facile l'applicazione dello sviluppo per righe o per colonne secondo Laplace. Ad esempio, possiamo azzerare quanti più elementi possibile in una colonna. Possiamo usare le regole di riduzione a scalini di Gauss, ricordando che:

- scambiando tra loro due righe o due colonne il determinante cambia di segno
- moltiplicando una riga o una colonna per una costante $c \neq 0$ anche il determinante viene moltiplicato per c
- sommando a una riga, o a una colonna, un multiplo di un'altra il determinante rimane lo stesso

Chapter 8

Autovalori e autovettori

(Nota: In questa sezione, laddove non specificato diversamente, con il simbolo V denotiamo uno spazio vettoriale di dimensione finita n .)

Definizione 45: [Autovalori e autovettori] Sia $T : V \mapsto V$ un'applicazione lineare. Diremo che un vettore $v \in V$ non nullo è un *autovettore* di T se esiste $\lambda \in \mathbb{R}$ tale che

$$T(v) = \lambda v$$

In questo caso, diremo che λ è un *autovalore* di T . In particolare, esso è l'autovalore *relativo all'autovettore* v . Diremo anche che v è un autovettore avente λ come autovalore.

Si noti che se v è un autovettore, allora anche cv è un autovettore per ogni $c \in \mathbb{R}$, in quanto $T(cv) = cT(v) = c\lambda v = \lambda(cv)$. Quindi, ogni autovettore giace su una retta di autovettori aventi tutti lo stesso autovalore.

Esempio. - interpretazione geometrica in \mathbb{R}^2

- direzioni privilegiate
- cambio di scala in tali direzioni
- rotazione di K gradi con K non multiplo di 180: nessun autovettore. Esempio della matrice $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.
- rotazione di 180 gradi, tutti autovettori
- simmetria rispetto all'asse y : asse x fatto di autovettori
- cambio di scala di fattori diversi: autovettori sull'asse y e sull'asse x , con autovalori diversi. Nessun altro autovettore
- cambio di scala di fattori uguali: tutti autovettori con stesso autovalore
- fare disegni ed esempi, ma senza mettere le matrici, che metteremo dopo

◇

TEOREMA 33: Sia $T : V \mapsto V$ un'applicazione lineare. Sia $\lambda \in \mathbb{R}$ un autovalore di T e sia $V_\lambda := \{v \in V : T(v) = \lambda v\}$ (V_λ è l'insieme contenente il vettore nullo e tutti gli autovettori di T che hanno λ come autovalore). Allora V_λ è un sottospazio di V .

Dim: Siano $v_1, v_2 \in V_\lambda$ e $\alpha, \beta \in \mathbb{R}$. Allora si ha

$$T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2) = \alpha \lambda v_1 + \beta \lambda v_2 = \lambda(\alpha v_1 + \beta v_2)$$

sicchè $\alpha v_1 + \beta v_2 \in V_\lambda$ il che prova che V_λ è un sottospazio di V . ♣

Definizione 46: [**Autospazio**] Sia λ un autovalore dell'applicazione lineare $T : V \mapsto V$. Il sottospazio V_λ viene detto *autospazio* di V relativo all'autovalore λ .

Si noti che, detta $I_V : V \mapsto V$ l'applicazione identità, si ha che $V_\lambda = \ker(T - \lambda I_V)$, in quanto $(T - \lambda I_V)(v) = 0$ se e solo se $T(v) = \lambda I_V(v) = \lambda v$.

Definizione 47: [**Molteplicità geometrica**] Sia λ un autovalore di un'applicazione lineare $T : V \mapsto V$. Il valore $\dim(V_\lambda)$ è detto la *molteplicità geometrica* di λ , ed è indicato con $\text{MG}(\lambda)$.

TEOREMA 34: Siano λ_1 e λ_2 autovalori distinti di un'applicazione lineare T e siano v_1 e v_2 autovettori relativi il primo a λ_1 e il secondo a λ_2 . Allora v_1 e v_2 sono linearmente indipendenti.

Dim: Supponiamo per assurdo che esista $c \in \mathbb{R}$ tale che $v_1 = cv_2$. Allora $T(v_1) = \lambda_1 v_1$, ma anche $T(v_1) = T(cv_2) = c\lambda_2 v_2 = \lambda_2 cv_2 = \lambda_2 v_1$ da cui $(\lambda_1 - \lambda_2)v_1 = 0$ e quindi $\lambda_1 = \lambda_2$. ♣

Questo risultato può essere esteso a un numero arbitrario di autovettori i cui rispettivi autovalori siano tutti diversi fra loro:

TEOREMA 35: Sia T un'applicazione lineare di V in V e siano v_1, \dots, v_m , con $m \geq 2$, autovettori di T tali che, per ogni coppia di autovettori v_i e v_j , con $i \neq j$, per i relativi autovalori vale $\lambda_i \neq \lambda_j$. Allora i vettori v_1, \dots, v_m sono linearmente indipendenti.

Dim: Usiamo l'induzione su m . Per $m = 2$ sappiamo del teorema 34 che l'asserto vale. Sia ora $m > 2$ e supponiamo vero l'asserto per $m - 1$ autovettori. Sia data una combinazione lineare nulla dei v_i a coefficienti $a_i \in \mathbb{R}$,

$$\sum_{i=1}^m a_i v_i = \mathbf{0} \tag{8.1}$$

Vogliamo dimostrare che $a_i = 0$ per ogni i .

Applicando T a sinistra e destra in (8.1) abbiamo $T(\sum_{i=1}^m a_i v_i) = \sum_{i=1}^m (\lambda_i a_i) v_i = 0$. Se invece moltiplichiamo per λ_m a sinistra e a destra (8.1) otteniamo $\sum_i \lambda_m a_i v_i = 0$. Abbiamo quindi

$$\sum_{i=1}^m (\lambda_i a_i) v_i = \sum_{i=1}^m (\lambda_m a_i) v_i$$

da cui

$$\sum_{i=1}^m a_i (\lambda_i - \lambda_m) v_i = \mathbf{0} = \sum_{i=1}^{m-1} a_i (\lambda_i - \lambda_m) v_i$$

Quindi, per induzione, per ogni $i = 1, \dots, m-1$ si ha $a_i (\lambda_i - \lambda_m) = 0$ e, poichè $\lambda_i \neq \lambda_m$ per ogni $i < m$, deduciamo che $a_i = 0$ per $i = 1, \dots, m-1$. Sostituendo in (8.1) si vede che questo implica anche $a_m = 0$ e quindi i vettori v_1, \dots, v_m sono indipendenti. ♣

Si noti che se un'applicazione lineare $T : V \mapsto V$ ha n autovalori distinti, allora esistono n autovettori linearmente indipendenti che quindi formano una base di V . La condizione di avere n autovalori distinti è sufficiente, ma non necessaria, affinché V possieda una base fatta da autovettori di T . Come semplice esempio, si consideri l'applicazione identità. Per tale applicazione ogni vettore non nullo è un autovettore, e quindi qualsiasi base è fatta da autovettori, tutti con il medesimo autovalore, i.e., 1.

PROPOSIZIONE 36: Sia $T : V \mapsto V$ un'applicazione lineare e sia \mathcal{B} una base di V . Allora $\mathcal{M}[T, \mathcal{B}, \mathcal{B}]$ è una matrice diagonale se e solo se \mathcal{B} è una base di V fatta di soli autovettori di T .

Dim: Immediato. ♣

Quando $\mathcal{B} = \{v_1, \dots, v_n\}$ è una base di V fatta di autovettori di T , allora $\mathcal{M}[T, \mathcal{B}, \mathcal{B}] = \text{diag}(\lambda_1, \dots, \lambda_n)$, dove λ_i è l'autovalore relativo all'autovettore v_i , per $i = 1, \dots, n$.

Definizione 48: [Autovalori e autovettori di una matrice] Sia $A \in \mathbb{R}^{n \times n}$ è una matrice quadrata a coefficienti reali. Diremo che $v \in \mathbb{R}^n - \{\mathbf{0}\}$ è un autovettore di A se esiste $\lambda \in \mathbb{R}$ tale che

$$Av = \lambda v$$

In questo caso diremo che λ è un autovalore di A .

Si noti che se T è un'applicazione lineare e A è la matrice che rappresenta T rispetto a una qualsiasi base di V , allora $v \in V$ è un autovettore di T se e solo se $\mathbf{C}_{\mathcal{B}}^v$ è un autovettore di A . Quindi, per trovare gli autovettori di un'applicazione lineare, possiamo procedere cercando gli autovettori di una qualsiasi matrice rappresentativa di tale applicazione.

Esempio. Si consideri una matrice diagonale $A = \text{diag}(\lambda_1, \dots, \lambda_n)$. Sia e^i l' i -mo vettore della base canonica di \mathbb{R}^n . Abbiamo $Ae^i = A^i = \lambda_i e^i$, per cui ogni vettore e^i è un autovettore di A , ed è relativo all'autovalore λ_i . ◇

Definizione 49: [Applicazione diagonalizzabile] Sia $T : V \mapsto V$ un'applicazione lineare. Se esiste una base \mathcal{B} di V tale che $\mathcal{M}[T, \mathcal{B}, \mathcal{B}]$ è una matrice diagonale, allora T si dice *diagonalizzabile*.

Definizione 50: [Matrice diagonalizzabile] Sia $A \in \mathbb{R}^{n \times n}$ una matrice quadrata. Diremo che A è *diagonalizzabile* se A è simile ad una matrice diagonale, i.e., se esiste una matrice invertibile $B \in \mathbb{R}^{n \times n}$ tale che $B^{-1}AB$ è una matrice diagonale.

Data A , consideriamo l'applicazione $L_A : \mathbb{R}^n \mapsto \mathbb{R}^n$ definita da $L_A(X) = AX$. Allora è immediato notare come la matrice A è diagonalizzabile se e solo se l'applicazione L_A è diagonalizzabile.

Esempio. Si consideri la matrice

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

Sia $B = \begin{pmatrix} 1 & -4 \\ 1 & 2 \end{pmatrix}$. Abbiamo, dalla sezione 5.2, che $B^{-1} = \begin{pmatrix} 1/3 & 2/3 \\ -1/6 & 1/6 \end{pmatrix}$, e risulta

$$B^{-1}AB = \begin{pmatrix} 1/3 & 2/3 \\ -1/6 & 1/6 \end{pmatrix} \begin{pmatrix} 3/2 & -1 \\ -1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -4 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$$

sicché A è diagonalizzabile. Il modo in cui è stata trovata la matrice B verrà descritto nella prossima sezione.

◇

Esempio. Si consideri la matrice

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

Sappiamo dall'interpretazione geometrica descritta in un esempio precedente, che A rappresenta una rotazione di 90 gradi in senso antiorario di tutti i vettori del piano e che quindi non esistono autovettori per tale applicazione. Di conseguenza A non è diagonalizzabile.

Si può arrivare alla conclusione che A non è diagonalizzabile anche senza ricorrere a un'interpretazione geometrica di A , ma semplicemente per via algebrica.

Consideriamo una qualsiasi matrice 2×2 , invertibile, sia essa $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Dalla sezione 5.2 sappiamo che $B^{-1} = \frac{1}{\Delta} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, con $\Delta = ad - bc$, sicché si ha $B^{-1}AB =$

$$= \frac{1}{\Delta} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} -c & -d \\ a & b \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} -(cd + ab) & -(d^2 + b^2) \\ a^2 + c^2 & ab + cd \end{pmatrix}$$

L'elemento in riga 2, colonna 1 è $a^2 + c^2$ e vale zero se e solo se $a = c = 0$. Ma se $a = c = 0$ la matrice B avrebbe rango 1 e non potrebbe essere invertibile. Quindi $a^2 + c^2 \neq 0$ e $B^{-1}AB$ non è diagonale. ◇

8.1 Calcolo di autovalori e autovettori

Sia $T : V \mapsto V$ un'applicazione lineare di cui vogliamo trovare gli autovalori e gli autovettori. Fissata una qualsiasi base \mathcal{B} di V , sia $A = \mathcal{M}[T, \mathcal{B}, \mathcal{B}]$ la matrice che rappresenta T rispetto alla base \mathcal{B} . Un vettore $v \in V$ è un autovettore di T , di autovalore λ , se $T(v) = \lambda v$. In termini di coordinate rispetto alla base \mathcal{B} deve perciò essere

$$\mathbf{C}_{\mathcal{B}}^{T(v)} = \lambda \mathbf{C}_{\mathcal{B}}^v \quad (8.2)$$

Sia $X = \mathbf{C}_{\mathcal{B}}^v$ la n -pla delle coordinate (incognite) rispetto alla base \mathcal{B} di un generico autovettore v e sia λ (anch'essa incognita) il relativo autovalore. In forma matriciale l'equazione (8.2) diventa

$$AX = \lambda X$$

ossia

$$(A - \lambda I)X = \mathbf{0} \quad (8.3)$$

e si tratta di determinare per quali valori di λ questa equazione ha soluzioni $X \neq \mathbf{0}$.

Si noti che se per un certo valore $\bar{\lambda}$ la matrice $(A - \bar{\lambda}I)$ avesse rango pieno, allora essa rappresenterebbe un'applicazione biunivoca, e quindi l'unico vettore X per il quale $(A - \bar{\lambda}I)X = \mathbf{0}$ sarebbe $X = \mathbf{0}$. Perciò, una condizione necessaria che ogni λ deve soddisfare per poter essere un autovalore, è che $\text{rk}(A - \lambda I) < n$. Questa condizione è anche sufficiente. Infatti, $\dim(\ker(A - \lambda I)) = n - \text{rk}(A - \lambda I) > 0$ se $\text{rk}(A - \lambda I) < n$, e quindi in questo caso esistono $X \neq \mathbf{0}$, $X \in \ker(A - \lambda I)$.

Dalla precedente argomentazione segue la validità del seguente teorema:

TEOREMA 37: Sia $T : V \mapsto V$ un'applicazione lineare e sia $\bar{\lambda} \in \mathbb{R}$. Siano inoltre \mathcal{B} una base di V e $A = \mathcal{M}[T, \mathcal{B}, \mathcal{B}]$. Allora $\bar{\lambda}$ è un autovalore di T se e solo se la matrice $A - \bar{\lambda}I$ non ha rango pieno.

Ricordando che una matrice ha rango pieno se e solo se il suo determinante non è nullo, possiamo riformulare il teorema 37 come segue.

TEOREMA 38: Sia $T : V \mapsto V$ un'applicazione lineare e sia $\bar{\lambda} \in \mathbb{R}$. Siano inoltre \mathcal{B} una base di V e $A = \mathcal{M}[T, \mathcal{B}, \mathcal{B}]$. Allora $\bar{\lambda}$ è un autovalore di T se e solo

$$\det(A - \bar{\lambda}I) = 0$$

Data un'applicazione T siano A e C due matrici che rappresentano T , la prima rispetto a una base \mathcal{B} e la seconda rispetto a una base \mathcal{B}' . Allora si può vedere che per ogni $\lambda \in \mathbb{R}$

$$\det(C - \lambda I) = \det(A - \lambda I)$$

Infatti, sia B la matrice del cambiamento di base da \mathcal{B}' a \mathcal{B} . Si ha

$$\det(C - \lambda I) = \det(B^{-1}AB - \lambda I) = \det(B^{-1}AB - \lambda B^{-1}IB) = \det(B^{-1}(A - \lambda I)B) = \frac{1}{\det(B)} \det(A - \lambda I) \det(B)$$

Quindi il valore $\det(A - \lambda I)$ non dipende dalla particolare scelta della matrice di rappresentazione di T , ma solo dalla stessa T . Considerando λ come una variabile indefinita ed esplicitando la formula per $\det(A - \lambda I)$, si nota come tale valore risulta essere un polinomio nella variabile λ .

Definizione 51: [Polinomio caratteristico] Il determinante $\det(A - \lambda I)$ è un polinomio di grado n in λ . Tale polinomio è detto il polinomio caratteristico di T e viene indicato anche con $P_A(\lambda)$.

Per notare come tale determinante sia un polinomio in λ consideriamo lo sviluppo del determinante di $A' = (A - \lambda I)$. Abbiamo

$$A' = (a'_{ij}) = \begin{pmatrix} a_{11} - \lambda & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda & \end{pmatrix}$$

Abbiamo

$$\det(A') = \sum_{\pi} \operatorname{sgn}(\pi) a'_{\pi(1),1} a'_{\pi(2),2} \cdots a'_{\pi(n),n}$$

In ogni termine della somma, $a'_{\pi(i),i} = a_{\pi(i),i}$ se $\pi(i) \neq i$, mentre altrimenti $a'_{\pi(i),i} = a_{ii} - \lambda$. Detto $K(\pi) = \{i : \pi(i) = i\}$, e detto $c(\pi) = \operatorname{sgn}(\pi) \prod_{i \notin K(\pi)} a_{\pi(i),i}$, il generico termine della somma è

$$\operatorname{sgn}(\pi) a'_{\pi(1),1} a'_{\pi(2),2} \cdots a'_{\pi(n),n} = c(\pi) \prod_{i \in K(\pi)} (a_{ii} - \lambda)$$

ed è chiaro che si tratta di un polinomio in λ , di grado uguale a $|K(\pi)| \leq n$. Inoltre, esiste una permutazione (la permutazione identica) per la quale $|K(\pi)| = n$ e dunque il determinante di A' , essendo la somma di polinomi di λ di grado $\leq n$ tra cui ne compare esattamente uno di grado n , è un polinomio in λ di grado n .

Esempio. Sia $A = (a_{ij}) \in \mathbb{R}^{2 \times 2}$ la matrice rappresentativa di un'applicazione lineare di \mathbb{R}^2 in \mathbb{R}^2 rispetto alla base canonica. Dalla formula per i determinanti 2×2 abbiamo

$$\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = \lambda^2 - \lambda(a_{11} + a_{22}) + (a_{11}a_{22} - a_{12}a_{21})$$

◇

Definizione 52: [Equazione caratteristica] L'equazione

$$\det(A - \lambda I) = 0$$

nella variabile λ è detta *equazione caratteristica* di T . Ogni sua soluzione reale è un autovalore reale di T , i.e., λ' è un autovalore sse $P(\lambda') = 0$

8.1.1 Note da mettere a posto

1. definizione di *molteplicità algebrica* di un autovalore come numero di volte in cui λ' è radice di $P()$, ossia $P()$ è divisibile per $(\lambda - \lambda')^s$ ma non per $(\lambda - \lambda')^{s+1}$. Denotiamo la molteplicità algebrica di un autovalore λ' con $\operatorname{MA}(\lambda')$.

2. Quali siano gli autovalori, nonchè le rispettive MA, lo si determina trovando le radici distinte del polinomio caratteristico. Siano esse $\lambda_1, \dots, \lambda_k$. Per poi trovare le MG di ciascuno di tali autovalori bisogna risolvere, per $i = 1, \dots, k$, il sistema

$$(A - \lambda_i I)X = 0$$

e scoprire la dimensione del kernel di $A - \lambda_i I$, ossia dell'autospazio V_{λ_i}

3. riprendere gli esempi in \mathbb{R}^2 . esempio in cui non ci sono autovalori in \mathbb{R} (la rotazione di prima). In tale caso il polinomio caratteristico risulta

$$\det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1$$

che non ha radici in \mathbb{R} . (Si noti che se il campo degli scalari fosse \mathbb{C} anzichè \mathbb{R} esisterebbero gli autovalori i e $-i$).

4. esempio in cui c'è un solo autovalore con m.a. 2 ma la corrispondente m.g. è 1
 5. esempio con 2 autovalori di m.a. e m.g. 1 ciascuno

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

6. esempio con 1 autovalore di m.a. 2 e m.g. 2

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

7. Sia T un'applicazione lineare e siano $\lambda_1, \dots, \lambda_k$ gli autovalori (distinti) di T . In virtù del teorema che dice come autovettori associati a autovalori distinti sono linearmente indipendenti, abbiamo che

$$V_{\lambda_i} \cap V_{\lambda_j} = \{\mathbf{0}\}$$

per $i \neq j$. Se prendiamo una base in V_{λ_1} unita a una di V_{λ_2} unita ... a una di V_{λ_k} avremo un insieme indipendente di vettori in V . Quindi

$$\sum_{i=1}^k \text{MG}(\lambda_i) \leq n$$

sempre. Se $\sum_{i=1}^k \text{MG}(\lambda_i) = n$ allora tali vettori costituiscono una base di V e quindi T è diagonalizzabile (condizione sufficiente). Viceversa, se T è diagonalizzabile, presa una base di autovettori possiamo ripartirla negli autovettori relativi a ciascun autovalore λ_i ed ognuno di tali insiemi sarà una base di V_{λ_i} . Otteniamo quindi che

$$\sum_{i=1}^k \text{MG}(\lambda_i) = n$$

(condizione necessaria). In conclusione:

TEOREMA 39: Sia T un'applicazione lineare e siano $\lambda_1, \dots, \lambda_k$ gli autovalori (distinti) di T . Condizione necessaria e sufficiente perchè T sia diagonalizzabile è che si abbia

$$\sum_{i=1}^k \text{MG}(\lambda_i) = n$$

Esempio. Determinare gli autovalori e gli autospazi di $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ definita da

$$T(x, y, z) = (-x + y + 5z, -y + 3z, 2z)$$

e stabilire se T è diagonalizzabile.

La matrice rappresentativa di T rispetto alla base canonica di \mathbb{R}^3 è

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

Il polinomio caratteristico è

$$P_A(\lambda) = (-1 - \lambda)^2(2 - \lambda)$$

che ha le radici $\lambda_1 = -1$, con $\text{MA}(-1) = 2$, e $\lambda_2 = 2$, con $\text{MA}(2) = 1$.

Calcoliamo l'autospazio V_{-1} . Risolviamo il sistema $(A + I)X = 0$, ossia

$$\begin{array}{rrr} 0x & +y & +5z = 0 \\ 0 & 0y & +3z = 0 \\ 0 & 0 & 3z = 0 \end{array}$$

Tale sistema ha soluzione generica $(t, 0, 0)$ con $t \in \mathbb{R}$. Quindi $\dim(\ker(A + I)) = 1 = \text{MG}(-1)$.

Calcoliamo l'autospazio V_2 . Risolviamo il sistema $(A - 2I)X = 0$, ossia

$$\begin{array}{rrr} -3x & +y & +5z = 0 \\ 0 & -3y & +3z = 0 \\ 0 & 0 & 0z = 0 \end{array}$$

Tale sistema ha soluzione generica $(2t, t, t)$ con $t \in \mathbb{R}$. Quindi $\dim(\ker(A - 2I)) = 1 = \text{MG}(2)$.

Essendo $\text{MG}(-1) + \text{MG}(2) = 2 < 3 = \dim(\mathbb{R}^3)$, l'applicazione T non è diagonalizzabile. \diamond

Chapter 9

Geometria dello spazio \mathbb{R}^n

9.1 Prodotto scalare e ortogonalità

In questa sezione introduciamo un'operazione che ad ogni coppia (a, b) di vettori di \mathbb{R}^n associa uno scalare, denotato con $a \cdot b$. Questa operazione viene chiamata *prodotto scalare*. Il prodotto scalare ha una definizione algebrica ma la sua vera natura è rivelata da un'interpretazione geometrica. In particolare, come vedremo, il prodotto scalare è legato alla lunghezza della proiezione di un vettore sull'altro, e quindi all'angolo formato dai due vettori a e b . Cominciamo allora con la definizione algebrica del prodotto scalare.

Definizione 53: [Prodotto scalare] Siano $a = (a_1, \dots, a_n)$ e $b = (b_1, \dots, b_n)$ vettori di \mathbb{R}^n . Definiamo il loro *prodotto scalare* come il numero

$$a \cdot b = \sum_{i=1}^n a_i b_i$$

Si noti che interpretando a come un vettore-riga e b come un vettore-colonna, il prodotto matriciale ab è la matrice 1×1 che contiene il solo valore $a \cdot b$. In generale, se $A = (a_{ij}) \in \mathbb{R}^{m \times k}$ e $B = (b_{ij}) \in \mathbb{R}^{k \times n}$, allora $AB = C = (c_{ij})$ è tale che

$$c_{ij} = A_i \cdot B^j \quad \text{per } i = 1, \dots, m, \quad j = 1, \dots, n$$

Il prodotto scalare gode di due importanti proprietà, i.e., è una *forma bilineare* ed è *definito positivo*:

- *Bilinearità*: il prodotto scalare è lineare rispetto ad entrambi i suoi argomenti, ossia per ogni $a, b, c \in \mathbb{R}^n$ e scalari $\lambda, \mu \in \mathbb{R}$ si ha

$$(\lambda a + \mu b) \cdot c = \lambda(a \cdot c) + \mu(b \cdot c) = c \cdot (\lambda a + \mu b)$$

Dimostriamo la prima di tali uguaglianze, in quanto la seconda vale essendo banalmente il prodotto scalare commutativo (i.e. $u \cdot v = v \cdot u$ per ogni $u, v \in \mathbb{R}^n$). Abbiamo

$$(\lambda a + \mu b) \cdot c = \sum_{i=1}^n (\lambda a_i + \mu b_i) c_i = \lambda \sum_{i=1}^n a_i c_i + \mu \sum_{i=1}^n b_i c_i = \lambda(a \cdot c) + \mu(b \cdot c)$$

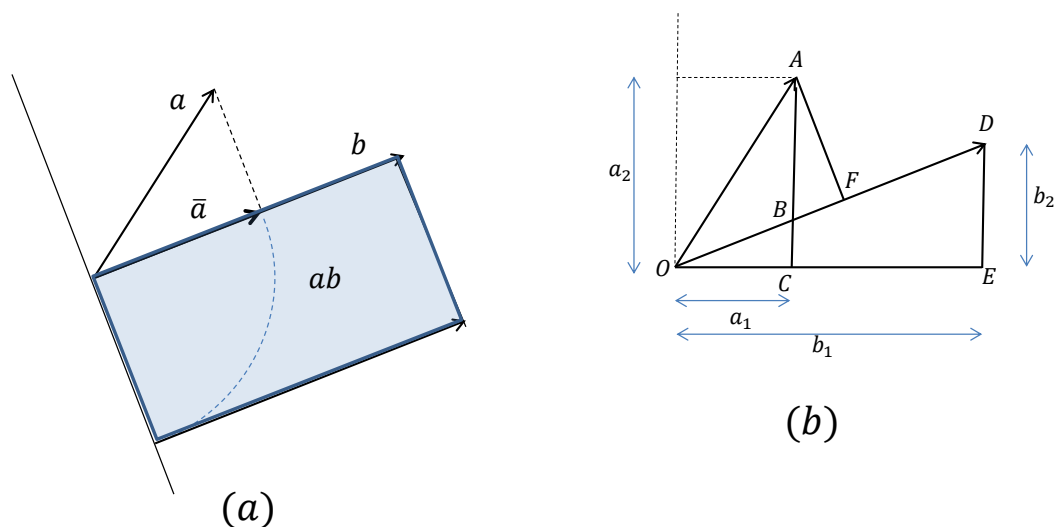


Figure 9.1: (a) xxxxxx. (b) yyyyyy.

- *Positività*: il prodotto scalare è definito positivo, ossia, per ogni $a \in \mathbb{R}^n$ si ha

$$a \cdot a \geq 0$$

e

$$a \cdot a = 0 \quad \text{se e solo se } a = \mathbf{0}$$

Per il prodotto scalare non vale la legge di annullamento del prodotto in quanto si può avere $a \cdot b = 0$ con $a \neq \mathbf{0}$ e $b \neq \mathbf{0}$. Ad esempio, se $a = (1, 2)$ e $b = (-6, 3)$ si ha $a \cdot b = -6 + 6 = 0$.

Interpretazione geometrica. Diamo ora un'interpretazione geometrica del prodotto scalare in cui facciamo uso di concetti come lunghezza, ortogonalità e proiezione ortogonale di un vettore su una retta. Questi concetti verranno formalmente definiti in seguito proprio a partire dal prodotto scalare, per cui si pone un problema di circolarità del ragionamento. A noi però basta, in questa interpretazione, supporre che i suddetti concetti siano noti, almeno informalmente, a partire dalle nozioni di geometria elementare acquisiti nelle scuole primarie.

Per semplicità, discuteremo prima del prodotto scalare in \mathbb{R}^2 . Siano a e b vettori tali che la proiezione di a su b abbia la stessa direzione di b . In pratica questo significa che a si trova "dalla stessa parte di b " rispetto alla retta ortogonale ad b . Sia \bar{a} la proiezione di a su b (si veda la Fig.9.1(a)). indichiamo con $\|v\|$ la lunghezza di un vettore v . Allora

Il prodotto scalare $a \cdot b$ è pari all'area rettangolo che ha per lati \bar{a} e b , ossia

$$a \cdot b = \|\bar{a}\| \|b\|$$

Per dimostrare la validità di questa interpretazione, facciamo riferimento alla figura 9.1(b). In questa figura abbiamo due vettori, $a = \vec{OA}$ e $b = \vec{OD}$ con $a = (a_1, a_2)$ e $b = (b_1, b_2)$. Per ogni coppia di punti X e Y indichiamo con $|\overline{XY}|$ la lunghezza del segmento \overline{XY} . Vogliamo fare vedere che

$$|\overline{OF}| |\overline{OD}| = a_1 b_1 + a_2 b_2 = a \cdot b$$

il che è lo stesso che dimostrare che

$$|\overline{OF}|^2 |\overline{OD}|^2 = (a_1 b_1 + a_2 b_2)^2 = a_1^2 b_1^2 + a_2^2 b_2^2 + 2a_1 a_2 b_1 b_2. \quad (9.1)$$

Dal teorema di Pitagora abbiamo che

$$|\overline{OD}|^2 = b_1^2 + b_2^2. \quad (9.2)$$

Per calcolare $|\overline{OF}|^2$ cominciamo con l'osservare che i triangoli $\triangle ODE$, $\triangle OBC$ e $\triangle ABF$ sono simili. Abbiamo $|\overline{BC}| : |\overline{OC}| = |\overline{DE}| : |\overline{OE}|$, da cui

$$|\overline{BC}| = \frac{a_1 b_2}{b_1}. \quad (9.3)$$

Abbiamo $|\overline{AB}| = |\overline{AC}| - |\overline{BC}|$, da cui

$$|\overline{AB}| = a_2 - \frac{a_1 b_2}{b_1} = \frac{a_2 b_1 - a_1 b_2}{b_1}.$$

Dal teorema di Pitagora si ha $|\overline{OB}|^2 = |\overline{OC}|^2 + |\overline{BC}|^2$, che da (9.3) diventa

$$|\overline{OB}|^2 = a_1^2 + \frac{a_1^2 b_2^2}{b_1^2} = \frac{a_1^2}{b_1^2} (b_1^2 + b_2^2). \quad (9.4)$$

Dalla similitudine di $\triangle ABF$ e $\triangle OBC$ si ha $|\overline{AF}|^2 : |\overline{AB}|^2 = |\overline{OC}|^2 : |\overline{OB}|^2$, e quindi

$$|\overline{AF}|^2 = \frac{a_1^2 \left(\frac{a_2 b_1 - a_1 b_2}{b_1} \right)^2}{\frac{a_1^2}{b_1^2} (b_1^2 + b_2^2)} = \frac{(a_2 b_1 - a_1 b_2)^2}{b_1^2 + b_2^2}. \quad (9.5)$$

Dal teorema di Pitagora si ha $|\overline{OF}|^2 = |\overline{OA}|^2 - |\overline{AF}|^2$, che da (9.5) diventa

$$|\overline{OF}|^2 = a_1^2 + a_2^2 - \frac{(a_2 b_1 - a_1 b_2)^2}{b_1^2 + b_2^2} = \frac{(b_1^2 + b_2^2)(a_1^2 + a_2^2) - (a_2 b_1 - a_1 b_2)^2}{b_1^2 + b_2^2} \quad (9.6)$$

In conclusione, da (9.2) e (9.6) si ha

$$|\overline{OF}|^2 |\overline{OD}|^2 = \frac{(b_1^2 + b_2^2)(a_1^2 + a_2^2) - (a_2 b_1 - a_1 b_2)^2}{b_1^2 + b_2^2} (b_1^2 + b_2^2) = a_1^2 b_1^2 + a_2^2 b_2^2 + 2a_1 a_2 b_1 b_2 = (a_1 b_1 + a_2 b_2)^2$$

e quindi vale (9.1).

-prodotto scalare e' un'area segnata

-valore assoluto=vera area

-se vettore nell'altra direzione, prendiamo il suo opposto

-piano che passa per i due vettori

L'interpretazione geometrica del prodotto scalare risulta utile a comprendere e ricordare alcune proprietà dello stesso che invece possono risultare meno intuitive nella loro forma puramente algebrica. Ne diamo di seguito alcuni esempi:

1. L'interpretazione geometrica ci dice che deve essere $\|a\|^2 = a \cdot a$, visto che la proiezione di a su a è a stesso. Quindi,

$$\|a\| = \sqrt{a \cdot a}$$

2. L'area del rettangolo di lati \bar{a} e b varia da un massimo di $\|a\| \|b\|$ (quando a giace sulla retta definita da b) ad un minimo di 0 (quando $\|\bar{a}\| = 0$). Siccome $\|\bar{a}\| = 0$ se e solo se a è perpendicolare a b , ecco che la nozione di perpendicolarità tra a e b può essere definita da $a \cdot b = 0$.
3. Possiamo definire l'angolo θ formato dai vettori a e b come l'angolo il cui coseno è

$$\cos \theta = \frac{a \cdot b}{\|a\| \|b\|}$$

in quanto $\|\bar{a}\| = \cos \theta \|a\|$.

4. La disuguaglianza di Cauchy-Schwartz, che dimostreremo formalmente in seguito, afferma

$$|a \cdot b| \leq \|a\| \|b\|$$

Nell'interpretazione geometrica, questa disuguaglianza è ovvia, perchè la parte sinistra è l'area di un rettangolo di lati \bar{a} (i.e., la proiezione di a su b) e b , mentre la parte destra è l'area di un rettangolo di lati a e b e, chiaramente, $\|\bar{a}\| \leq \|a\|$.

Le nozioni che abbiamo appena discusso intuitivamente coincidono con quelle formali, che prescindono da ogni interpretazione geometrica.

Definizione 54: [Norma di un vettore] Sia $a \in \mathbb{R}^n$. Definiamo la *norma* (o *lunghezza*) di $a = (a_1, \dots, a_n)$ come il valore

$$\|a\| = \sqrt{a \cdot a} = \sqrt{a_1^2 + \dots + a_n^2}$$

Definizione 55: [Ortogonalità] Siano $a, b \in \mathbb{R}^n$. Diciamo che a e b sono vettori *ortogonali* (o *perpendicolari*, o *normali*) se

$$a \cdot b = 0$$

9.2 Equazioni parametriche di rette e piani

9.3 Proiezioni e distanze tra insiemi