Introduzione all'Algebra Lineare

Carlo Petronio

2024

Indice

Capitolo 1

Spazi vettoriali

In questo capitolo introduciamo, a partire dagli esempi del piano cartesiano reale e più in generale dello spazio cartesiano a n componenti, la nozione di *spazio vettoriale*. Combinando le due operazioni che costituiscono una tale struttura diamo quindi la definizione di *combinazione lineare* e a partire da essa quelle di indipendenza lineare, di generazione e di base, che ci consente poi di definire la dimensione. Concludiamo il capitolo discutendo la nozione di sottospazio vettoriale e provando che se in uno stesso spazio vettoriale sono dati due sottospazi, la somma delle loro dimensioni è uguale alla somma delle dimensioni della loro intersezione e del generato della loro unione.

1.1 Definizione di spazio vettoriale

Nella meccanica di un punto materiale libero di muoversi su un piano, si rappresenta una forza agente sul punto come un *vettore applicato al punto*, ovvero come una freccia avente coda nel punto e testa in qualche altro punto del piano.

Un principio fondamentale della dinamica è che un punto soggetto a due diverse forze si muove come se fosse soggetto a una terza forza data dalla risultante delle prime due, ovvero quella ottenuta dalle prime due tramite la regola del parallelogrammo, vale a dire la diagonale del parallelogrammo che ha le prime due forze come lati. Se le forze iniziali sono date dai vettori F e G, indichiamo con $F + G$ il vettore che rappresenta la loro risultante.

Ogni forza è caratterizzata da una direzione, da un verso e da una intensità. Data una forza e un numero reale t, possiamo introdurre una nuova forza, detta multipla della prima per un fattore t, caratterizzata come segue: la sua direzione è la medesima di quella della forza iniziale; il suo verso è il medesimo di quello della forza iniziale se $t \geqslant 0$, opposto altrimenti; la sua intensità è $|t|$ volte quella della forza iniziale. Se la forza iniziale è data dal vettore F, indichiamo con $t \cdot F$ il vettore che rappresenta la multipla della prima per un fattore t.

Se nel piano introduciamo un sistema di assi cartesiano con origine nel punto materiale soggetto a una forza F, tale F è determinata dalle componenti (ascissa e ordinata) del punto in cui F ha la testa. Dunque posto

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \ x_1, x_2 \in \mathbb{R} \right\}$$

si ha che F corrisponde a un punto $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ di $\mathbb{R}^2$. Se inoltre un'altra forza G corrisponde al punto $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ di $\mathbb{R}^2$, si verifica facilmente che $F + G$ corrisponde al punto $\begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}$ di $\mathbb{R}^2$. Infine se $t \in \mathbb{R}$ la forza $t \cdot F$ corrisponde al punto $\begin{pmatrix} t \cdot x_1 \\ t \cdot x_2 \end{pmatrix}$ di $\mathbb{R}^2$. Stabilendo di indicare con $x+y$ il punto $\begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}$ e con $t \cdot x$ il punto $\begin{pmatrix} t \cdot x_1 \\ t \cdot x_2 \end{pmatrix}$ restano definite due operazioni

$$\begin{array}{ccc} \mathbb{R}^2 \times \mathbb{R}^2 & \to & \mathbb{R}^2 \\ (x, y) & \mapsto & x + y \end{array} \qquad \begin{array}{ccc} \mathbb{R} \times \mathbb{R}^2 & \to & \mathbb{R}^2 \\ (t, x) & \mapsto & t \cdot x. \end{array}$$

La costruzione si estende facilmente da $\mathbb{R}^2$ allo spazio $\mathbb{R}^n$ dei vettori colonna numerici con n componenti. Posto

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \ x_1, \ldots, x_n \in \mathbb{R} \right\}$$

e preso $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, introduciamo la notazione

$$(x)_j = x_j$$

cioè conveniamo che per qualsiasi vettore x il simbolo $(x)_j$ rappresenta la sua componente j-esima. Allora possiamo definire due operazioni

$$\begin{array}{ccc} \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R}^n \\ (x, y) & \mapsto & x + y \end{array} \qquad \begin{array}{ccc} \mathbb{R} \times \mathbb{R}^n & \to & \mathbb{R}^n \\ (t, x) & \mapsto & t \cdot x \end{array}$$

tramite le formule

$$(x+y)_j = (x)_j + (y)_j \qquad (t \cdot x)_j = t \cdot x_j.$$

È immediato verificare che sono soddisfatte le proprietà seguenti:

1. Se $0 \in \mathbb{R}^n$ è il vettore con $(0)_j = 0$ per ogni j, si ha $0 + x = x$ per ogni x in $\mathbb{R}^n$;
2. Per ogni x in $\mathbb{R}^n$, se $-x$ è il vettore tale che $(-x)_j = -(x)_j$ per ogni j, si ha $(-x) + x = 0$;
3. $x + (y + z) = (x + y) + z$ per ogni x, y, z in $\mathbb{R}^n$;
4. $x + y = y + x$ per ogni x, y in $\mathbb{R}^n$;
5. $t \cdot (x + y) = t \cdot x + t \cdot y$ per ogni t in $\mathbb{R}$ e x, y in $\mathbb{R}^n$;
6. $(t + s) \cdot x = t \cdot x + s \cdot x$ per ogni t, s in $\mathbb{R}$ e x in $\mathbb{R}^n$;
7. $(t \cdot s) \cdot x = t \cdot (s \cdot x)$ per ogni t, s in $\mathbb{R}$ e x in $\mathbb{R}^n$;
8. $1 \cdot x = x$ per ogni x in $\mathbb{R}^n$.

Nelle proprietà (5) e (6) abbiamo omesso alcune parentesi convenendo come di consueto che le operazioni "$\cdot$" di prodotto si eseguono prima delle operazioni "$+$" di somma.

Utilizzando allora il caso di $\mathbb{R}^n$ come prototipo, diamo la seguente definizione generale:

Definizione 1.1.1. Uno *spazio vettoriale reale* (o *su* $\mathbb{R}$) è un insieme V (i cui elementi sono detti *vettori*) per il quale sono definite due operazioni "$+$" e "$\cdot$" che soddisfano certi assiomi che andiamo ora a elencare. L'operazione "$+$" è binaria e interna, ovvero è una funzione da $V \times V$ in V, che associa a una coppia di vettori (v, w) un terzo vettore $v + w$, detto *somma* di v e w. L'operazione "$\cdot$" è invece una funzione da $\mathbb{R} \times V$ in V, la quale dunque associa a una coppia (t, v), costituita da un numero reale t ed un vettore v, un nuovo vettore $t \cdot v$, detto *prodotto* di v per lo *scalare* t. Le proprietà che devono essere soddisfatte sono le seguenti:

1. Esiste $0 \in V$ tale che $0 + v = v$ per ogni v in V;
2. Per ogni v in V esiste $(-v) \in V$ tale che $(-v) + v = 0$;
3. $v + (w + u) = (v + w) + u$ per ogni v, w, u in V;

4. $v + w = w + v$ per ogni v, w in V;

5. $t \cdot (v + w) = t \cdot v + t \cdot w$ per ogni t in $\mathbb{R}$ e v, w in V;

6. $(t + s) \cdot v = t \cdot v + s \cdot v$ per ogni t, s in $\mathbb{R}$ e v in V;

7. $(t \cdot s) \cdot v = t \cdot (s \cdot v)$ per ogni t, s in $\mathbb{R}$ e v in V;

8. $1 \cdot v = v$ per ogni v in V.

Di nuovo nelle proprietà (5) e (6) abbiamo convenuto che "$\cdot$" si esegue sempre prima di "+".

Matrici e polinomi Descriviamo ora altri due esempi di spazio vettoriale. Dati interi positivi m e n, chiamiamo *matrice reale* $m \times n$ una tabella rettangolare A di numeri reali con m righe ed n colonne, ovvero

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

con $a_{ij} \in \mathbb{R}$ per $i = 1, \ldots, m$ e $j = 1, \ldots, n$. Il coefficiente a_{ij} è detto di *posto* (i, j), dove i è l'indice di riga e j è l'indice di colonna. Indichiamo quindi con $\mathcal{M}_{m\times n}(\mathbb{R})$ l'insieme di tutte tali matrici, ed estendiamo la notazione precedente ponendo, per A come sopra,

$$(A)_{ij} = a_{ij}.$$

Definiamo quindi le operazioni

$$\begin{array}{ccc} \mathcal{M}_{m\times n}(\mathbb{R}) \times \mathcal{M}_{m\times n}(\mathbb{R}) & \to & \mathcal{M}_{m\times n}(\mathbb{R}) \\ (A, B) & \mapsto & A + B \end{array} \qquad \begin{array}{ccc} \mathbb{R} \times \mathcal{M}_{m\times n}(\mathbb{R}) & \to & \mathcal{M}_{m\times n}(\mathbb{R}) \\ (t, A) & \mapsto & t \cdot A \end{array}$$

ponendo

$$(A + B)_{ij} = (A)_{ij} + (B)_{ij} \qquad (t \cdot A)_{ij} = t \cdot (A)_{ij}.$$

Questo significa che si possono sommare tra loro solo matrici della stessa taglia, e che la somma si esegue posto per posto, e che per moltiplicare una matrice per un numero bisogna moltiplicare per quel numero tutti i suoi coefficienti. Chiamando 0 la matrice che ha tutti i coefficienti 0, e $-A$ la matrice definita da $(-A)_{ij} = -(A)_{ij}$ è immediato verificare che le operazioni definite soddisfano le proprietà richieste dalla definizione di spazio vettoriale.

Notiamo che $\mathbb{R}^n = \mathcal{M}_{n\times 1}(\mathbb{R})$.

Ricordiamo ora che $\mathbb{R}[x]$ rappresenta l'insieme dei polinomi a coefficienti reali nella inderminata x. Sappiamo già come sommare due polinomi e come moltiplicare un polinomio per un numero, e di nuovo è molto facile vedere che tutte le proprietà della definizione di spazio vettoriale sono soddisfatte. Notiamo che è anche possibile moltiplicare tra loro due polinomi, ma questa operazione non fa parte della struttura di spazio vettoriale.

Conseguenze della definizione Fissiamo ora fino alla fine del capitolo uno spazio vettoriale reale V. Mostriamo adesso come dalle proprietà delle operazioni elencate nella definizione ne discendano altre che appaiono del tutto naturali.

Proposizione 1.1.2. $0 \cdot v = 0$ *per ogni* v *in* V*.*

Dimostrazione. Sviluppiamo una successione di identità usando il simbolo "$\underset{(j)}{=}$" per dire che l'uguaglianza discende dalla proprietà (j) della definizione:

$$\begin{aligned} 0 &\underset{(2)}{=} -(0 \cdot v) + 0 \cdot v \\ &= -(0 \cdot v) + (0 + 0) \cdot v \\ &\underset{(6)}{=} -(0 \cdot v) + (0 \cdot v + 0 \cdot v) \\ &\underset{(3)}{=} (-(0 \cdot v) + 0 \cdot v) + 0 \cdot v \\ &\underset{(2)}{=} 0 + 0 \cdot v \\ &\underset{(1)}{=} 0 \cdot v. \end{aligned}$$

$\square$

Proposizione 1.1.3. $(-1) \cdot v = -v$ *per ogni* v *in* V*.*

Dimostrazione. Procediamo come sopra indicando con $(*)$ il contenuto della proposizione precedente:

$$\begin{aligned} -v &\underset{(1)}{=} 0 + (-v) \\ &\underset{(4)}{=} (-v) + 0 \\ &\underset{(*)}{=} (-v) + 0 \cdot v \\ &= (-v) + (1 + (-1)) \cdot v \\ &\underset{(6)}{=} (-v) + (1 \cdot v + (-1) \cdot v) \\ &\underset{(8)}{=} (-v) + (v + (-1) \cdot v) \\ &\underset{(3)}{=} ((-v) + v) + (-1) \cdot v \\ &\underset{(2)}{=} 0 + (-1) \cdot v \\ &\underset{(1)}{=} (-1) \cdot v. \end{aligned}$$

□

Analogamente possiamo vedere che $t \cdot 0 = 0$ per ogni t in $\mathbb{R}$ e che se $t \neq 0$ e $t \cdot v = 0$ allora $v = 0$. In generale, qualsiasi proprietà algebrica concernente somme di vettori e prodotti di vettori per scalari sarà vera se lo è sostituendo i vettori con numeri.

1.2 Combinazioni lineari

Dati vettori $v_1, \ldots, v_n$ di V e scalari $t_1, \ldots, t_n$, grazie alla proprietà (3) della definizione di spazio vettoriale possiamo dare significato senza mettere parentesi all'espressione

$$t_1 \cdot v_1 + \ldots + t_n \cdot v_n$$

(dove intendiamo sempre che i "$\cdot$" si eseguono prima dei "$+$"). Chiamiamo una tale espressione *combinazione lineare* dei vettori $v_1, \ldots, v_n$ con coefficenti $t_1, \ldots, t_n$, intendendo con ciò sia l'espressione stessa sia il suo risultato (un vettore di V). Possiamo ora porci due domande:

(i) L'unica combinazione lineare di $v_1, \ldots, v_n$ che ha come risultato il vettore 0 è quella con tutti i coefficienti 0?

(ii) Ogni vettore di V si può scrivere come combinazione lineare di $v_1, \ldots, v_n$ per opportuni coefficienti? Se sì, in modo unico?

Con facili calcoli possiamo rispondere a queste domande per alcuni esempi di $v_1, \dots, v_n$ nel caso $V = \mathbb{R}^2$:

- $n = 1,\ v_1 = \begin{pmatrix} 3 \\ -7 \end{pmatrix}$; (i) Sì; (ii) No;
- $n = 1,\ v_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$; (i) No; (ii) No;
- $n = 2,\ v_1 = \begin{pmatrix} 2 \\ 5 \end{pmatrix},\ v_2 = \begin{pmatrix} -3 \\ 4 \end{pmatrix}$; (i) Sì; (ii) Sì, in modo unico;
- $n = 2,\ v_1 = \begin{pmatrix} 8 \\ -20 \end{pmatrix},\ v_2 = \begin{pmatrix} -6 \\ 15 \end{pmatrix}$; (i) No; (ii) No;
- $n = 3,\ v_1 = \begin{pmatrix} 2 \\ -3 \end{pmatrix},\ v_2 = \begin{pmatrix} 5 \\ 1 \end{pmatrix},\ v_3 = \begin{pmatrix} -4 \\ 7 \end{pmatrix}$; (i) No; (ii) Sì ma non in modo unico.

Definizione 1.2.1. In generale diremo che $v_1, \dots, v_n$ sono:

(i) *Linearmente indipendenti* se l'unica combinazione lineare di $v_1, \dots, v_n$ che ha come risultato il vettore 0 è quella con tutti i coefficienti 0;

(ii) *Generatori* di V se ogni vettore di V si può scrivere come combinazione lineare di $v_1, \dots, v_n$ per opportuni coefficienti.

Osservazione 1.2.2. Per essere del tutto precisi dovremmo dire che *il sistema di vettori $v_1, \dots, v_n$ è linearmente indipendente*, oppure che *genera V*. Infatti entrambe le proprietà si riferiscono a $v_1, \dots, v_n$ collettivamente, non individualmente. Ma per semplicità diremo sempre che *sono linearmente indipendenti*, oppure che *generano V*.

Proposizione 1.2.3. *Supponiamo che $v_1, \dots, v_n$ siano generatori di V. Allora sono fatti equivalenti:*

- *$v_1, \dots, v_n$ sono linearmente indipendenti;*
- *Ogni $v \in V$ si scrive in modo unico come loro combinazione lineare.*

Dimostrazione. Supponiamo che valga la prima condizione, cioè che $v_1, \dots, v_n$ siano linearmente indipendenti. Provare la seconda significa prendere un vettore v di V e due espressioni potenzialmente diverse di v come combinazione lineare di $v_1, \dots, v_n$, concludendo che le espressioni in realtà coincidono, cioè hanno gli stessi coefficienti. Supponiamo allora che

$$v = t_1 \cdot v_1 + \ldots + t_n \cdot v_n \qquad v = s_1 \cdot v_1 + \ldots + s_n \cdot v_n.$$

Ne segue che

$$\begin{aligned} 0 &= v - v \\ &= (t_1 \cdot v_1 + \ldots + t_n \cdot v_n) - (s_1 \cdot v_1 + \ldots + s_n \cdot v_n) \\ &= (t_1 - s_1) \cdot v_1 + \ldots (t_n - s_n) \cdot v_n. \end{aligned}$$

Abbiamo trovato una espressione di 0 come combinazione lineare di $v_1, \dots, v_n$, ma per ipotesi essi sono linearmente indipendenti, dunque tutti i coefficienti di tale espressione sono nulli:

$$t_1 - s_1 = \ldots = t_n - s_n = 0 \quad \Rightarrow \quad t_1 = s_1, \ \ldots \ t_n = s_n.$$

Abbiamo provato che effettivamente le due espressioni sono uguali.

Viceversa, supponiamo che ogni $v \in V$ si scriva in modo unico come combinazione lineare di $v_1, \dots, v_n$. Dobbiamo vedere che se

$$t_1 \cdot v_1 + \ldots + t_n \cdot v_n = 0$$

allora $t_1 = \ldots = t_n = 0$. E per farlo basta notare che

$$0 \cdot v_1 + \ldots + 0 \cdot v_n = 0$$

e usare l'unicità dell'espressione di 0. $\square$

Spazi vettoriali complessi Invece dello spazio $\mathbb{R}^n$ usato come modello per la definizione di spazio vettoriale reale, possiamo considerare

$$\mathbb{C}^n = \left\{ \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} : \ z_1, \dots, z_n \in \mathbb{C} \right\}$$

su cui sono definite le operazioni

$$\begin{array}{ccc} \mathbb{C}^n \times \mathbb{C}^n & \to & \mathbb{C}^n \\ (z, w) & \mapsto & z + w \end{array} \qquad \begin{array}{ccc} \mathbb{C} \times \mathbb{C}^n & \to & \mathbb{C}^n \\ (\alpha, z) & \mapsto & \alpha \cdot z \end{array}$$

definite esattamente come in $\mathbb{R}^n$: la somma tra vettori si esegue componente per componente, uno scalare (complesso) moltiplica un vettore moltiplicando tutte le sue componenti. Si può dunque modificare la Definizione 1.1.1 sostituendo l'aggettivo "reale" con "complesso" e il simbolo $\mathbb{R}$ con $\mathbb{C}$. Più in generale, se $\mathbb{K}$ è un campo qualsiasi (dunque $\mathbb{R}$ oppure $\mathbb{C}$, ma anche $\mathbb{Q}$ o altri, che esistono) si può definire uno spazio vettoriale su $\mathbb{K}$ come un insieme V dotato di due operazioni

$$\begin{array}{rcl} V \times V & \to & V \\ (v, w) & \mapsto & v + w \end{array} \qquad \begin{array}{rcl} \mathbb{K} \times V & \to & V \\ (k, v) & \mapsto & k \cdot v \end{array}$$

che soddisfano esattamente le stesse proprietà (1)-(8) della Definizione 1.1.1, con $\mathbb{K}$ al posto di $\mathbb{R}$. Tutto quello che vedremo in questo capitolo e nei prossimi, fino al numero 4 escluso, si potrebbe enunciare nel caso di spazi vettoriali su un campo $\mathbb{K}$ qualsiasi, ma noi per semplicità e concretezza ci limiteremo al caso $\mathbb{K} = \mathbb{R}$.

Sottospazi vettoriali Abbiamo finora visto solo pochi esempi di spazio vettoriale, ma molti si possono costruire sfruttando la seguente:

Definizione 1.2.4. Se V è uno spazio vettoriale reale, un suo sottoisieme W si dice *sottospazio vettoriale* di V se valgono le proprietà seguenti:

1. W contiene 0;
2. Per ogni w_1, w_2 in W si ha che $w_1 + w_2$ è in W;
3. Per ogni t in $\mathbb{R}$ e w in W si ha che $t \cdot w$ è in W.

In altre parole, W è un sottospazio se contiene 0 e se eseguendo con i suoi elementi le operazioni di spazio vettoriale di V si trovano ancora elementi di W. Possiamo dunque dire che W *eredita* le operazioni da V. Poiché le operazioni in V soddisfano le 8 proprietà della definizione di spazio vettoriale, anche quelle che W eredita lo fanno, dunque W è esso stesso uno spazio vettoriale.

Come esempio, possiamo considerare nello spazio $\mathbb{R}^3$ il sottoinsieme W dei vettori x tali che $x_1 + x_2 + x_3 = 0$, dove $x_j = (x)_j$. Poiché il vettore 0 di $\mathbb{R}^3$ è quello con tutte le componenti 0, e $0 + 0 + 0 = 0$, abbiamo che $0 \in W$. Se $x, y \in W$, ovvero $x_1 + x_2 + x_3 = y_1 + y_2 + y_3 = 0$, abbiamo

$$\begin{aligned} (x+y)_1 + (x+y)_2 + (x+y)_3 &= (x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) \\ &= (x_1 + x_2 + x_3) + (y_1 + y_2 + y_3) \\ &= 0 + 0 \\ &= 0 \end{aligned}$$

pertanto $x+y \in W$. Analogamente se $t \in \mathbb{R}$ e $x \in W$, ovvero $x_1+x_2+x_3=0$, abbiamo

$$\begin{aligned}(t\cdot x)_1+(t\cdot x)_2+(t\cdot x)_3 &= t\cdot x_1+t\cdot x_2+t\cdot x_3\\ &= t\cdot(x_1+x_2+x_3)\\ &= t\cdot 0\\ &= 0\end{aligned}$$

pertanto $t\cdot x \in W$. Abbiamo provato che W è un sottospazio vettoriale di $\mathbb{R}^3$.

Concludiamo il paragrafo osservando che in V esistono sempre due sottospazi, detti *banali*, che sono $\{0\}$ e l'intero V.

1.3 Dimensione

È intuitivo che il piano $\mathbb{R}^2$, lo spazio $\mathbb{R}^3$ e più in generale $\mathbb{R}^n$ con $n>3$ siano oggetti intrinsecamente diversi fra loro. Tuttavia come *insiemi* puri e semplici in realtà non sono tra loro distinguibili, poiché vale il seguente fatto (non banale):

Proposizione 1.3.1. *Per ogni n,m naturali positivi esistono funzioni bigettive tra $\mathbb{R}^n$ e $\mathbb{R}^m$.*

È però vero che per $n\neq m$ si ha che $\mathbb{R}^n$ e $\mathbb{R}^m$ sono diversi fra loro come spazi vettoriali. Questa affermazione si base sulla seguente nozione:

Definizione 1.3.2. Chiamiamo *base* di uno spazio vettoriale V su $\mathbb{R}$ un insieme $\mathcal{B}=(v_1,\ldots,v_n)$ di n vettori ordinati, che siano linearmente indipendenti e generatori di V.

La Proposizione 1.2.3 comporta che se $\mathcal{B}=(v_1,\ldots,v_n)$ è una base di V, ogni $v\in V$ si scrive in modo unico come

$$v=x_1\cdot v_1+\ldots+x_n\cdot v_n$$

con $x_1,\ldots,x_n\in\mathbb{R}$. Poniamo ora $x=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\in\mathbb{R}^n$ e chiamiamo x vettore delle *coordinate* di v rispetto a $\mathcal{B}$, denotato con $[v]_{\mathcal{B}}$. Questa definizione spiega perché nella definizione di base abbiamo insistito che essa sia un insieme *ordinato* di vettori: le proprietà di essere linearmente indipendenti e generatori di V riguardano i vettori $v_1, \ldots, v_n$ senza richiedere che essi siano ordinati, ma senza l'ordine sarebbe impossibile definire il vettore delle coordinate.

Prima di procedere estendiamo leggermente un concetto già introdotto. Abbiamo detto che i vettori $v_1, \ldots, v_n$ generano V se ogni v in V si può scrivere come combinazione lineare di $v_1, \ldots, v_n$. Più in generale chiamiamo *generato* di $v_1, \ldots, v_n$, indicato con $\mathrm{Span}(v_1, \ldots, v_n)$, l'insieme delle combinazioni lineari di $v_1, \ldots, v_n$, ovvero

$$\mathrm{Span}(v_1, \ldots, v_n) = \{t_1 \cdot v_1 + \ldots + t_n \cdot v_n : \ t_1, \ldots, t_n \in \mathbb{R}\}.$$

Naturalmente $v_1, \ldots, v_n$ generano V se e solo se $\mathrm{Span}(v_1, \ldots, v_n) = V$. Conveniamo inoltre che il generato di un insieme di vettori vuoto sia $\{0\}$.

Proposizione 1.3.3. $\mathrm{Span}(v_1, \ldots, v_n)$ *è un sottospazio vettoriale di* V*.*

Dimostrazione. Se $n = 0$ abbiamo il sottospazio banale $\{0\}$. Altrimenti basta sfruttare le identità seguenti:

$$\begin{array}{rcl} 0 \cdot v_1 + \ldots + 0 \cdot v_n & = & 0 \\ (t_1 \cdot v_1 + \ldots + t_n \cdot v_n) & & \\ + (s_1 \cdot v_1 + \ldots + s_n \cdot v_n) & = & (t_1 + s_1) \cdot v_1 + \ldots + (t_n + s_n) \cdot v_n \\ s \cdot (t_1 \cdot v_1 + \ldots + t_n \cdot v_n) & = & (s \cdot t_1) \cdot v_1 + \ldots + (s \cdot t_n) \cdot v_n. \end{array}$$

□

Buona definizione di dimensione Abbiamo ora il seguente risultato di verifica non immediata:

Proposizione 1.3.4. *Siano dati* $n, m \geqslant 0$ *e vettori* $w_1, \ldots, w_m, v_1, \ldots, v_n$ *in* V *tali che:*

- $w_1, \ldots, w_m$ *sono linearmente indipendenti;*
- $w_1, \ldots, w_m \in \mathrm{Span}(v_1, \ldots, v_n)$.

Allora $m \leqslant n$*.*

Dimostrazione. Proviamo il risultato per induzione su m, introducendo la seguente affermazione $\mathcal{P}(m)$ relativa a un intero m generico: *dati* $w_1, \ldots, w_m$ *linearmente indipendenti in* V*, un intero* $n \geqslant 0$ *e vettori* $v_1, \ldots, v_n$ *tali che* $w_1, \ldots, w_m \in \mathrm{Span}(v_1, \ldots, v_n)$*, si ha* $m \leqslant n$*.*

Per il passo base potremmo iniziare dimostrando $\mathcal{P}(m)$ per $m = 0$, nel qual caso la conclusione è ovvia perché $n \geqslant 0 = m$ per ipotesi, ma dimostriamo direttamente anche $\mathcal{P}(m)$ per $m = 1$ per meglio illustrare il ragionamento. In questo caso l'insieme di vettori contenente il solo w_1 è linearmente indipendente, il che significa precisamente che w_1 è non nullo. Ora, dovendo mostrare che $n \geqslant 1$, supponiamo per assurdo che $n = 0$. In questo caso l'insieme dei v_i è vuoto, dunque il suo generato è $\{0\}$. Tuttavia per ipotesi w_1 appartiene al generato dei v_i, e questo è assurdo poiché w_1 è non nullo.

Ci rivolgiamo ora al passo induttivo, supponendo nota $\mathcal{P}(m)$ per un generico $m \geqslant 1$. Dobbiamo dimostrare $\mathcal{P}(m+1)$, dunque consideriamo $m+1$ vettori linearmente indipendenti $x_0, \ldots, x_m$ che stanno nel generato di altri vettori $y_1, \ldots, y_k$. Si noti che stiamo per comodità usando gli indici tra 0 e m per gli $m+1$ vettori $x_0, \ldots, x_m$. Stiamo inoltre usando nuovi simboli x_i, y_j e k, per avere a disposizione v_i, w_j e n per quando useremo l'ipotesi induttiva. Dobbiamo provare che $m + 1 \leqslant k$.

Poiché $x_0, \ldots, x_m \in \operatorname{Span}(y_1, \ldots, y_k)$ esistono coefficienti $t_{ij} \in \mathbb{R}$ tali che

$$x_i = \sum_{j=1}^{k} t_{ij} \cdot y_j, \qquad i = 0, \ldots m.$$

Ora osserviamo che x_0 non è il vettore nullo, altrimenti $x_0, \ldots, x_m$ non potrebbero essere linearmente indipendenti. Perciò qualcuno dei coefficienti t_{0j} è non nullo. Poiché l'ordine dei vettori $y_1, \ldots, y_k$ non ha alcuna rilevanza ai fini della proprietà da verificare, possiamo modificare questo ordine e supporre per comodità che sia t_{0k} il coefficiente non nullo. Definiamo ora:

$$w_1 = x_1 - \frac{t_{1k}}{t_{0k}} \cdot x_0, \quad \ldots \quad , w_m = x_m - \frac{t_{mk}}{t_{0k}} \cdot x_0.$$

Poniamo inoltre $n = k - 1$ e $v_1 = y_1, \ldots, v_n = y_{k-1}$, e affermiamo che l'ipotesi induttiva si applica ai vettori $w_1, \ldots, w_m$ e $v_1, \ldots, v_n$. Questo è sufficiente per concludere, dato che $\mathcal{P}(m)$ garantisce allora che $m \leqslant n$, ma $n = k - 1$ dunque $m + 1 \leqslant k$, come dovevasi dimostrare.

Resta pertanto da provare che $w_1, \ldots, w_m$ sono linearmente indipendenti e stanno nel generato di $v_1, \ldots, v_n$. Cominciamo con la prima affermazione, considerando una combinazione lineare che dia il vettore nullo:

$$\begin{aligned} 0 &= a_1 \cdot w_1 + \ldots + a_m \cdot w_m \\ &= a_1 \cdot \left(x_1 - \frac{t_{1k}}{t_{0k}} \cdot x_0 \right) + \ldots + a_m \cdot \left(x_m - \frac{t_{mk}}{t_{0k}} \cdot x_0 \right) \\ &= a_0 \cdot x_0 + a_1 \cdot x_1 + \ldots + a_m \cdot x_m \end{aligned}$$

dove a_0 è un coefficiente che dipende dagli altri a_i e dai t_{ik}. Poiché per ipotesi $x_0, \dots, x_m$ sono linearmente indipendenti, concludiamo che gli a_i, in particolare $a_1, \dots, a_m$ come dovevamo dimostrare, sono nulli.

Proviamo ora che $w_i \in \operatorname{Span}(v_1, \dots, v_n)$. Infatti

$$\begin{aligned} w_i &= x_i - \frac{t_{ik}}{t_{0k}} \cdot x_0 \\ &= \sum_{j=1}^{k} t_{ij} \cdot y_j - \frac{t_{ik}}{t_{0k}} \cdot \sum_{j=1}^{k} t_{0j} \cdot y_j \\ &= \sum_{j=1}^{k-1} \left(t_{ij} - \frac{t_{ik}}{t_{0k}} \cdot t_{0j} \right) \cdot y_j + \left(\underbrace{t_{ik} - \frac{t_{ik}}{t_{0k}} \cdot t_{0k}}_{0} \right) \cdot y_k. \end{aligned}$$

Poiché $v_1 = y_1, \dots, v_n = y_{k-1}$ la dimostrazione è conclusa. □

Teorema 1.3.5. *Se lo spazio vettoriale V ammette basi, tutte le sue basi hanno lo stesso numero di elementi.*

Dimostrazione. Supponiamo che V abbia basi

$$\mathcal{B} = (v_1, \dots, v_n) \qquad \mathcal{C} = (w_1, \dots, w_m).$$

Poiché gli elementi $w_1, \dots, w_n$ di $\mathcal{C}$ sono linearmente indipendenti e gli elementi $v_1, \dots, v_n$ di $\mathcal{B}$ generano V, possiamo applicare la proposizione precedente e dedurre che $m \leqslant n$. Scambiando i ruoli di $\mathcal{B}$ e $\mathcal{C}$ abbiamo anche che $n \leqslant m$, dunque $n = m$. □

Definizione 1.3.6. Se V ammette basi, chiamiamo *dimensione* di V, indicata con $\dim(V)$ oppure $\dim_{\mathbb{R}}(V)$, il numero di elementi di qualsiasi base di V.

Il seguente esempio mostra che la premessa nella definizione precedente è essenziale: $\mathbb{R}[x]$ non ammette basi. Infatti se $\mathbb{R}[x]$ fosse generato da un numero finito di polinomi $p_1(x), \dots, p_n(x)$, ogni altro polinomio $p(x)$ si potrebbe scrivere come

$$p(x) = t_1 \cdot p_1(x) + \ldots + t_n \cdot p_n(x),$$

ma questo è impossibile se $p(x)$ ha grado maggiore di quello di tutti i $p_i(x)$.

Se V ammette basi, dunque è possibile definire la sua dimensione, diremo che V *ha dimensione finita*.

Basi canoniche Nello spazio $\mathbb{R}^n$ poniamo

$$e_1^{(n)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad e_2^{(n)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \dots \quad e_{n-1}^{(n)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad e_n^{(n)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathcal{E}^{(n)} = (e_1^{(n)}, e_2^{(n)}, \dots, e_{n-1}^{(n)}, e_n^{(n)})$$

scrivendo per semplicità e_i invece che $e_i^{(n)}$ quando n è chiaro dal contesto. È facile verificare che $\mathcal{E}^{(n)}$ è una base di $\mathbb{R}^n$, e in particolare che $[x]_{\mathcal{E}^{(n)}} = x$ per ogni $x \in \mathbb{R}^n$, dunque $\dim_{\mathbb{R}}(\mathbb{R}^n) = n$. Chiamiamo $\mathcal{E}^{(n)}$ la *base canonica* di $\mathbb{R}^n$.

Analogamente, dati $m, n \geqslant 1$, possiamo definire la matrice

$$E_{ij}^{(m\times n)} \in \mathcal{M}_{m\times n}(\mathbb{R})$$

come quella che ha coefficiente 1 nel posto (i, j) e 0 altrove. Posto

$$\mathcal{E}^{(m\times n)} = \left(E_{ij}^{(m\times n)}\right)_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$$

abbiamo che $\mathcal{E}^{(m\times n)}$ è una base di $\mathcal{M}_{m\times n}(\mathbb{R})$, di nuovo detta *canonica*. Dunque $\dim_{\mathbb{R}}(\mathcal{M}_{m\times n}(\mathbb{R})) = m \cdot n$.

Abbiamo visto che $\mathbb{R}[x]$ non ha basi, ma se fissiamo $d \in \mathbb{N}$ e definiamo $\mathbb{R}_{\leqslant d}[x]$ come l'insieme dei polinomi di grado al più d (compreso il polinomio 0), abbiamo che esso è un sottospazio vettoriale di $\mathbb{R}[x]$. Inoltre $1, x, x^2, \dots, x^{d-1}, x^d$ è una base di $\mathbb{R}_{\leqslant d}[x]$, detta *canonica*, pertanto

$$\dim_{\mathbb{R}}(\mathbb{R}_{\leqslant d}[x]) = d + 1.$$

Estrazione e completamento Una base di V è un insieme ordinato di vettori che sono linearmente indipendenti e che generano V. Vediamo ora come da un insieme che ha una sola delle due proprietà se ne possa ottenere uno che le ha entrambe. Abbiamo intanto questa conseguenza diretta della Proposizione 1.3.4:

Proposizione 1.3.7. *Sia* $\dim_{\mathbb{R}}(V) = n$*.*

- *Se* $v_1, \ldots, v_k \in V$ *generano* V *allora* $k \geqslant n$*;*
- *Se* $v_1, \ldots, v_k \in V$ *sono linearmente indipendenti allora* $k \leqslant n$*.*

Procediamo con questo risultato più tecnico:

Lemma 1.3.8.

- *Se* $v_1, \ldots, v_k$ *sono linearmente dipendenti allora uno dei* v_i *appartiene al sottospazio generato dagli altri, e scartandolo il generato dai vettori non cambia;*
- *Se* $v_1, \ldots, v_k$ *sono linearmente indipendenti e* $v \notin \mathrm{Span}(v_1, \ldots, v_k)$*, allora* $v_1, \ldots, v_k, v$ *sono linearmente indipendenti.*

Dimostrazione. Per la prima affermazione, supponiamo che

$$t_1 \cdot v_1 + \ldots + t_k \cdot v_k = 0$$

con almeno un $t_i \neq 0$. Allora

$$v_i = \left(-\frac{t_1}{t_i}\right) \cdot v_1 + \ldots + \left(-\frac{t_{i-1}}{t_i}\right) \cdot v_{i-1} + \left(-\frac{t_{i+1}}{t_i}\right) \cdot v_{i+1} + \ldots + \left(-\frac{t_k}{t_i}\right) \cdot v_k$$

il che prova la prima parte. Se in una combinazione lineare di $v_1, \ldots, v_k$ si sostituisce l'espressione di v_i come combinazione degli altri, si ottiene una combinazione lineare dei soli altri, il che prova la seconda parte.

Per la seconda affermazione, supponiamo che

$$t_1 \cdot v_1 + \ldots + t_k \cdot v_k + t \cdot v = 0.$$

Se avessimo $t \neq 0$ avremmo (come prima) $v \in \mathrm{Span}(v_1, \ldots, v_k)$, ma l'ipotesi è il contrario. Dunque $t = 0$, ma allora essendo $v_1, \ldots, v_k$ linearmente indipendenti abbiamo $t_1 = \ldots = t_k = 0$, e abbiamo provato che $v_1, \ldots, v_k, v$ sono linearmente indipendenti. □

Il prossimo risultato descrive i procedimenti di *estrazione di una base* e di *completamento a base.*

Proposizione 1.3.9. *Sia* $\dim_{\mathbb{R}}(V) = n$*.*

- *Se* $v_1, \ldots, v_k$ *generano* V*, è possibile scartare alcuni di essi ottenendo una base di* V*;*

- *Se $v_1, \dots, v_k$ sono linearmente indipendenti, è possibile aggiungere alcuni vettori ottenendo una base di V.*

Dimostrazione. Per entrambe le affermazioni si tratta di applicare ripetutamente quella corrispondente del lemma. Per la prima è chiaro che l'iterazione finisce. Lo stesso è vero per la seconda, dato che il numero di vettori linearmente indipendenti in V non può superare la dimensione. □

L'ultimo risultato del paragrafo afferma che se un insieme di vettori ha il numero giusto di elementi per essere una base, per vedere che è davvero una base basta verificare una sola delle due proprietà della definizione (generare ed essere linearmente indipendenti) invece che entrambe. In pratica, è sempre più facile verificare l'indipendenza lineare.

Corollario 1.3.10. *Sia $\dim_{\mathbb{R}}(V) = n$. Dati $v_1, \dots, v_n \in V$ i fatti seguenti sono tra loro equivalenti:*

- *$v_1, \dots, v_n$ generano V;*
- *$v_1, \dots, v_n$ sono linearmente indipendenti.*

1.4 La formula di Grassmann

Continuiamo in questo paragrafo a mantenere fissato uno spazio vettoriale V su $\mathbb{R}$ e supponiamo che abbia dimensione finita. Abbiamo intanto la seguente:

Proposizione 1.4.1. *Se W è un sottospazio vettoriale di V, allora W ha dimensione finita e $\dim(W) \leqslant \dim(V)$. Inoltre se $\dim(W) = \dim(V)$ si ha $W = V$.*

Dimostrazione. Ammettendo che W abbia dimensione finita, la prima conclusione segue subito dalla Proposizione 1.3.4. Infatti se $(w_1, \dots, w_k)$ è una base di W e $(v_1, \dots, v_n)$ è una base di V, si ha che $w_1, \dots, w_k$ sono linearmente indipendenti e appartengono al generato di $v_1, \dots, v_n$, dunque $k \leqslant n$.

Il fatto che W ha dimensione finita si dimostra usando il secondo punto del Lemma 1.3.8, da cui segue che se W non avesse dimensione finita allora conterrebbe insiemi di vettori linearmente indipendenti con qualsiasi numero di elementi, il che contraddirebbe la Proposizione 1.3.4.

Se $\dim(W) = \dim(V)$ allora una base di W, essendo un sistema di vettori linearmente indipendenti in V, si può completare a una base di V, ma avendo già $\dim(V)$ elementi si conclude che è già una base di V, dunque $W = V$. □

Ci dedichiamo da qui alla fine del capitolo alla situazione in cui in V sono dati due sottospazi W e Z.

Lemma 1.4.2. *$W \cap Z$ è un sottospazio di V.*

Dimostrazione. A titolo di esempio dimostriamo la proprietà (2) della definizione, le altre sono ugualmente elementari. Siano $x, y \in W \cap Z$. Dobbiamo verificare che $x+y \in W \cap Z$. Poiché $x, y \in W$, si ha $x+y \in W$ per la proprietà (2) della definizione di sottospazio, che per ipotesi vale per W. Analogamente $x, y \in Z$, dunque $x+y \in Z$. Abbiamo provato che $x+y \in W$ e $x+y \in Z$, perciò $x+y \in W \cap Z$, come richiesto. □

Proposizione 1.4.3. *$W \cup Z$ è un sottospazio di V se e soltanto se $W \subset Z$ oppure $Z \subset W$.*

Dimostrazione. Se $W \subset Z$ oppure $Z \subset W$ si ha rispettivamente $W \cup Z = Z$ oppure $W \cup Z = W$, dunque $W \cup Z$ è un sottospazio.

Viceversa supponiamo che non valga né $W \subset Z$ né $Z \subset W$, dunque che esista $w \in W$ con $w \notin Z$ ed esista $z \in Z$ con $z \notin W$. Se per assurdo $W \cup Z$ fosse un sottospazio, poiché $w, z \in W \cup Z$ si dovrebbe avere, per la proprietà (2) della definizione, $w+z \in W \cup Z$. Questo significa che $w+z \in W$ oppure $w+z \in Z$, ovvero rispettivamente $w+z = w' \in W$ oppure $w+z = z' \in Z$. Ma allora si avrebbe $z = w' - w$, dunque $z \in W$ poiché $w, w' \in W$ e W è un sottospazio, il che è assurdo, oppure analogamente $w \in Z$, di nuovo assurdo. □

Il risultato precedente suggerisce di definire la *somma* di W e Z come

$$W + Z = \{w + z : \ w \in W, \ z \in Z\}$$

provando la seguente:

Proposizione 1.4.4. *$W+Z$ è un sottospazio vettoriale di V ed è il più piccolo che contiene $W \cup Z$.*

Dimostrazione. Poiché $0 \in W$ e $0 \in Z$ abbiamo $0 = 0 + 0 \in W + Z$. Se $u_1, u_2 \in W + Z$, cioè $u_1 = w_1 + z_1$ con $w_1 \in W$ e $z_1 \in Z$, e $u_2 = w_2 + z_2$ con $w_2 \in W$ e $z_2 \in Z$, abbiamo

$$u_1 + u_2 = (w_1 + z_1) + (w_2 + z_2) = (w_1 + w_2) + (z_1 + z_2)$$

e $w_1 + w_2 \in W$ e $z_1 + z_2 \in Z$ dato che W e Z sono sottospazi, dunque $u_1 + u_2 \in W + Z$. Infine se $u \in W + Z$ e $t \in \mathbb{R}$, essendo $u = w + z$ con $w \in W$ e $z \in Z$, si ha

$$t \cdot u = t \cdot (w + z) = (t \cdot w) + (t \cdot z)$$

da cui segue che $t \cdot u \in W + Z$ dato che $t \cdot w \in W$ e $t \cdot z \in Z$ essendo W e Z sottospazi.

Abbiamo dimostrato che $W + Z$ è un sottospazio di V. Ora proviamo che contiene $W \cup Z$. Infatti se $u \in W \cup Z$ abbiamo $u \in W$ oppure $u \in Z$. Nel primo caso possiamo scrivere $u = u + 0$, con $u \in W$ e $0 \in Z$, dunque $u \in W + Z$. Analogamente nel secondo caso $u = 0 + u$ con $0 \in W$ e $u \in Z$, dunque $u \in W + Z$.

Infine proviamo che se T è un sottospazio di V e contiene $W \cup Z$ allora contiene $W + Z$, il che conclude la dimostrazione. Infatti se $u \in W + Z$ si ha $u = w + z$ con $w \in W$ e $z \in Z$. Essendo $W \subset T$ e $Z \subset T$ ne segue che $w, z \in T$, dunque $u = w + z \in T$. □

Dimostriamo ora il risultato fondamentale di questo paragrafo (formula di Grassmann):

Teorema 1.4.5. $\dim(W) + \dim(Z) = \dim(W \cap Z) + \dim(W + Z)$.

Dimostrazione. Indichiamo con k la dimensione di $W \cap Z$ e notiamo che W e Z contengono $W \cap Z$, dunque hanno dimensione almeno k. Poniamo dunque $\dim(W) = k + p$ e $\dim(Z) = k + q$, dove $p, q \geqslant 0$. Dobbiamo allora provare che $Z + W$ ha dimensione $k + p + q$.

Fissiamo ora una base $(y_1, \ldots, y_k)$ di $Z \cap W$. Notando che $y_1, \ldots, y_k$ sono vettori linearmente indipendenti e stanno in W, utilizziamo il secondo punto della Proposizione 1.3.9 ottenendo una base $(y_1, \ldots, y_k, w_1, \ldots, w_p)$ di W. Similmente costruiamo una base $(y_1, \ldots, y_k, z_1, \ldots, z_q)$ di Z. Affermiamo ora che

$$(y_1, \ldots, y_k, w_1, \ldots, w_p, z_1, \ldots, z_q)$$

è una base di $W + Z$, il che comporta la conclusione desiderata.

Cominciamo col mostrare che $y_1, \ldots, y_k, w_1, \ldots, w_p, z_1, \ldots, z_q$ sono generatori di $W+Z$. Consideriamo un generico elemento di $W+Z$, che per definizione si può scrivere come $w+z$ con $w \in W$ e $z \in Z$. Dobbiamo provare che $w+z$ si può esprimere come combinazione lineare di $y_1, \ldots, y_k, w_1, \ldots, w_p, z_1, \ldots, z_q$.

Ora $y_1, \ldots, y_k, w_1, \ldots, w_p$ generano W, dunque

$$w = \alpha_1 \cdot y_1 + \ldots + \alpha_k \cdot y_k + \beta_1 \cdot w_1 + \ldots + \beta_p \cdot w_p$$

per costanti α_i e β_j opportune. Similmente

$$z = \alpha'_1 \cdot y_1 + \ldots + \alpha'_k \cdot y_k + \gamma_1 \cdot z_1 + \ldots + \gamma_q \cdot z_q.$$

Dunque

$$w+z=(\alpha_1+\alpha_1')\cdot y_1+\ldots+(\alpha_k+\alpha_k')\cdot y_k+\beta_1\cdot w_1+\ldots+\beta_p\cdot w_p+\gamma_1\cdot z_1+\ldots+\gamma_q\cdot z_q$$

e abbiamo ottenuto quanto voluto.

Passiamo a dimostrare che $y_1,\ldots,y_k,w_1,\ldots,w_p,z_1,\ldots,z_q$ sono linearmente indipendenti. Supponiamo dunque di avere una combinazione lineare con risultato nullo

$$(\heartsuit)\qquad \alpha_1\cdot y_1+\ldots+\alpha_k\cdot y_k+\beta_1\cdot w_1+\ldots+\beta_p\cdot w_p+\gamma_1\cdot z_1+\ldots+\gamma_q\cdot z_q=0.$$

Dobbiamo provare che tutti gli α_i, β_j e γ_l sono nulli. Dall'identità precedente deduciamo che

$$\alpha_1\cdot y_1+\ldots+\alpha_k\cdot y_k+\beta_1\cdot w_1+\ldots+\beta_p\cdot w_p=-\gamma_1\cdot z_1-\ldots-\gamma_q\cdot z_q.$$

Ora in questa identità il membro sinistro descrive un elemento di W, mentre il membro destro descrive un elemento di Z. L'identità comporta dunque che entrambi i membri appartengono a $W\cap Z$. In particolare il membro destro si può scrivere come combinazione lineare dei vettori $y_1,\ldots,y_k$, che costituiscono una base di $W\cap Z$:

$$\begin{aligned} &-\gamma_1\cdot z_1-\ldots-\gamma_q\cdot z_q=\alpha_1'\cdot y_1+\ldots+\alpha_k'\cdot y_k\\ \Rightarrow\quad &\alpha_1'\cdot y_1+\ldots+\alpha_k'\cdot y_k+\gamma_1\cdot z_1+\ldots+\gamma_q\cdot z_q=0.\end{aligned}$$

Ne segue che tutti i coefficienti α_i' e γ_l sono nulli, poiché $y_1,\ldots,y_k,z_1,\ldots,z_q$ sono linearmente indipendenti (costituiscono una base di Z). La $(\heartsuit)$ allora comporta che

$$\alpha_1\cdot y_1+\ldots+\alpha_k\cdot y_k+\beta_1\cdot w_1+\ldots+\beta_p\cdot w_p=0$$

dunque anche gli α_i e β_j sono nulli, poiché $y_1,\ldots,y_k,w_1,\ldots,w_q$ sono linearmente indipendenti (costituiscono una base di W). La dimostrazione è completa. $\square$

Capitolo 2

Applicazioni lineari

In questo capitolo ci occupiamo delle funzioni tra due spazi vettoriali che rispettano le operazioni di somma tra vettori e di prodotto tra scalare e vettore. Dimostriamo che le applicazioni lineari $\mathbb{R}^n \to \mathbb{R}^m$ corrispondono in modo naturale alle matrici $m \times n$, introducendo il prodotto righe per colonne tra matrici e provando che la composizione di applicazioni corrisponde al prodotto tra matrici. Quindi torniamo alla situazione generale e introduciamo nucleo e immagine, mostrando la formula che lega le loro dimensioni. Poi proviamo che se due spazi vettoriali hanno basi fissate, a un'applicazione lineare tra loro resta associata una matrice. Infine analizziamo un particolare tipo di applicazioni lineari, dette proiezioni. Concludiamo il capitolo utilizzando le nozioni introdotte per introdurre i primi rudimenti di calcolo differenziale in più variabili.

2.1 Definizione e applicazione associata a una matrice

Cominciamo introducendo le applicazioni di cui parleremo in tutto il capitolo:

Definizione 2.1.1. Dati spazi vettoriali reali V e W, diciamo che un'applicazione $f : V \to W$ è *lineare* se:

- $f(0) = 0$;
- $f(v_1 + v_2) = f(v_1) + f(v_2)$ per ogni $v_1, v_2 \in V$;
- $f(\lambda \cdot v) = \lambda \cdot f(v)$ per ogni $\lambda \in \mathbb{R}$ e $v \in V$.

Osservazione 2.1.2. In realtà la prima delle tre proprietà è conseguenza di ciascuna delle altre due, ma conviene comunque mantenerla nella definizione.

Inoltre $f : V \to W$ è lineare se e soltanto se *rispetta le combinazioni lineari*, ovvero:

- $f(t_1 \cdot v_1 + t_2 \cdot v_2) = t_1 \cdot f(v_1) + t_2 \cdot f(v_2)$ per ogni $t_1, t_2 \in \mathbb{R}$ e $v_1, v_2 \in V$.

Tra poco vedremo come le applicazioni lineari $\mathbb{R}^n \to \mathbb{R}^m$ si possano caratterizzare in modo molto concreto, ma iniziamo con alcuni esempi più generali:

- L'applicazione $\mathrm{tr} : \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathbb{R}$ data da
$$\mathrm{tr}(A) = \sum_{i=1}^{n} (A)_{ii}$$
è lineare ed è detta *traccia*; gli elementi della matrice A che vengono sommati per ottenere $\mathrm{tr}(A)$ sono quelli sulla *diagonale principale*;

- L'applicazione $\mathcal{M}_{m\times n}(\mathbb{R}) \to \mathcal{M}_{n\times m}(\mathbb{R})$ che associa ad A la matrice ${}^{\mathrm{t}}A$ data da
$$\left({}^{\mathrm{t}}A\right)_{ij} = (A)_{ji}$$
è lineare ed è detta *trasposizione*;

- L'applicazione $\mathbb{R}[x] \to \mathbb{R}[x]$ che associa al polinomio $p(x)$ il polinomio $p'(x)$, dove
$$p(x) = \sum_{k=0}^{d} a_k \cdot x^k \quad \Rightarrow \quad p'(x) = \sum_{k=1}^{d} k \cdot a_k \cdot x^{k-1}$$
è lineare ed è detta *derivazione*.

Prodotto righe per colonne Fissati tre interi positivi m, n, k definiamo un'applicazione

$$\begin{array}{rcl} \mathcal{M}_{m\times n}(\mathbb{R}) \times \mathcal{M}_{n\times k}(\mathbb{R}) & \to & \mathcal{M}_{m\times k}(\mathbb{R}) \\ (A, B) & \mapsto & A \cdot B \end{array}$$

dove

$$(A \cdot B)_{ij} = \sum_{p=1}^{n} (A)_{ip} \cdot (B)_{pj}.$$

Proposizione 2.1.3. *Per ogni* $A, A_1, A_2 \in \mathcal{M}_{m\times n}(\mathbb{R})$, *per ogni* $B, B_1, B_2 \in \mathcal{M}_{n\times k}(\mathbb{R})$ *e per ogni* $t_1, t_2 \in \mathbb{R}$, *si ha:*

$$A \cdot (t_1 \cdot B_1 + t_2 \cdot B_2) = t_1 \cdot (A \cdot B_1) + t_2 \cdot (A \cdot B_2),$$

$$(t_1 \cdot A_1 + t_2 \cdot A_2) \cdot B = t_1 \cdot (A_1 \cdot B) + t_2 \cdot (A_2 \cdot B).$$

Dimostrazione. Proviamo la prima prorietà, l'altra è analoga. L'uguaglianza da provare è tra due matrici $m \times k$, dunque per verificarla basta vedere che esse hanno lo stesso coefficiente nel posto (i, j) per ogni $i = 1, \ldots, m$ e $j = 1, \ldots, k$. Usando le definizioni delle varie operazioni coinvolte si trova infatti:

$$\begin{aligned}
&\Big(A \;\cdot\; (t_1 \cdot B_1 + t_2 \cdot B_2)\Big)_{ij} \\
&\quad= \sum_{p=1}^{n} (A)_{ip} \cdot \Big(t_1 \cdot B_1 + t_2 \cdot B_2\Big)_{pj} \\
&\quad= \sum_{p=1}^{n} (A)_{ip} \cdot (t_1 \cdot (B_1)_{pj} + t_2 \cdot (B_2)_{pj}) \\
&\quad= t_1 \cdot \sum_{p=1}^{n} (A)_{ip} \cdot (B_1)_{pj} + t_2 \cdot \sum_{p=1}^{n} (A)_{ip} \cdot (B_2)_{pj} \\
&\quad= t_1 \cdot (A \cdot B_1)_{ij} + t_2 \cdot (A \cdot B_2)_{ij} = \Big(t_1 \cdot (A \cdot B_1) + t_2 \cdot (A \cdot B_2)\Big)_{ij}.
\end{aligned}$$

□

Il risultato appena provato si interpreta come *linearità del prodotto righe per colonne in entrambi gli argomenti.* Ciò significa che fissata $B \in \mathcal{M}_{n\times k}(\mathbb{R})$, l'applicazione

$$\begin{array}{rcl} \mathcal{M}_{m\times n}(\mathbb{R}) & \to & \mathcal{M}_{m\times k}(\mathbb{R}) \\ A & \mapsto & A \cdot B \end{array}$$

è lineare. Analogamente, fissata una matrice $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, è lineare la

$$\begin{array}{rcl} \mathcal{M}_{n\times k}(\mathbb{R}) & \to & \mathcal{M}_{m\times k}(\mathbb{R}) \\ B & \mapsto & A \cdot B \end{array}$$

Questo secondo fatto si applica in particolare quando $k = 1$, consentendo di dare la seguente:

Definizione 2.1.4. Data $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, chiamiamo

$$\begin{array}{rcl} f_A : \mathbb{R}^n & \to & \mathbb{R}^m \\ x & \mapsto & A \cdot x \end{array}$$

applicazione lineare associata alla matrice A.

Oltre alla linearità in ciascun argomento, il prodotto righe per colonne ha altre proprietà algebriche naturali, ovvero ha un elemento neutro sinistro e uno destro, ed è associativo, come ora vedremo. Per ogni n definiamo la *matrice identità* di ordine n come la matrice quadrata $\mathrm{I}_n \in \mathcal{M}_{n\times n}(\mathbb{R})$ che ha coefficienti 1 sulla diagonale principale, ovvero nei posti (i,i), e 0 altrove, ovvero nei posti (i,j) con $i \neq j$.

Proposizione 2.1.5.

- *Data $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ si ha*

$$A \cdot \mathrm{I}_n = \mathrm{I}_m \cdot A = A;$$

- *Date $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $B \in \mathcal{M}_{n\times k}(\mathbb{R})$ e $C \in \mathcal{M}_{k\times h}(\mathbb{R})$ si ha*

$$A \cdot (B \cdot C) = (A \cdot B) \cdot C.$$

Dimostrazione. La prima affermazione è semplice. La seconda asserisce l'uguaglianza tra due matrici $m \times h$. Basta allora verificare che in un generico posto (i,j) le matrici hanno il medesimo coefficiente:

$$\begin{aligned} \Big(A \ \cdot \ (B \cdot C)\Big)_{ij} &= \sum_{p=1}^{n} (A)_{ip} \cdot \big(B \cdot C\big)_{pj} = \sum_{p=1}^{n} (A)_{ip} \cdot \left(\sum_{q=1}^{k} (B)_{pq} \cdot (C)_{qj} \right) \\ &= \sum_{p=1}^{n} \sum_{q=1}^{k} (A)_{ip} \cdot (B)_{pq} \cdot (C)_{qj} = \sum_{q=1}^{k} \left(\sum_{p=1}^{n} (A)_{ip} \cdot (B)_{pq} \right) \cdot (C)_{qj} \\ &= \sum_{q=1}^{k} \big(A \cdot B\big)_{iq} \cdot (C)_{qj} = \Big((A \cdot B) \cdot C\Big)_{ij}. \end{aligned}$$

□

Osservazione 2.1.6. Dalla proposizione segue che l'applicazione associata alla matrice $\mathrm{I}_n \in \mathcal{M}_{n\times n}(\mathbb{R})$ è la funzione identità $\mathrm{id}_{\mathbb{R}^n}$.

Osservazione 2.1.7. Il prodotto righe per colonne *non* è commutativo, per tre buone ragioni. Intanto se si può calcolare un prodotto $A \cdot B$ si ha che $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ e $B \in \mathcal{M}_{n\times k}(\mathbb{R})$, ma se $m \neq k$ il prodotto $B \cdot A$ non si può calcolare. Inoltre, se $m = k$, si ha $A \cdot B \in \mathcal{M}_{m\times m}(\mathbb{R})$ e $B \cdot A \in \mathcal{M}_{n\times n}(\mathbb{R})$, dunque se $n \neq m$ l'uguaguaglianza $A \cdot B = B \cdot A$ non ha senso. Infine, anche se $n = m = k$, l'uguaglianza $A \cdot B = B \cdot A$ può non valere (a meno che $n = m = k = 1$, nel qual caso naturalmente vale sempre).

Dall'associatività del prodotto righe per colonne segue che l'applicazione associata a un prodotto di due matrici è la composizione delle applicazioni associate alle due matrici:

Proposizione 2.1.8. *Date* $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ *e* $B \in \mathcal{M}_{n\times k}(\mathbb{R})$ *si ha*

$$f_{A\cdot B} = f_A \circ f_B.$$

Dimostrazione. L'uguaglianza da verificare è quella tra due applicazioni da $\mathbb{R}^k$ in $\mathbb{R}^m$. Proviamo dunque esse hanno lo stesso valore su ogni $x \in \mathbb{R}^k$:

$$f_{A\cdot B}(x) = (A \cdot B) \cdot x = A \cdot (B \cdot x) = f_A(f_B(x)) = (f_A \circ f_B)(x).$$

□

Proviamo ora che in realtà tutte le applicazioni lineari $\mathbb{R}^n \to \mathbb{R}^m$ sono associate a matrici in $\mathcal{M}_{m\times n}(\mathbb{R})$.

Proposizione 2.1.9. *Se* $g : \mathbb{R}^n \to \mathbb{R}^m$ *è lineare esiste* $A \in \mathbb{R}^n \to \mathbb{R}^m$ *tale che* $g = f_A$.

Dimostrazione. Ricordiamo che in $\mathbb{R}^n$ e $\mathbb{R}^m$ abbiamo le basi canoniche rispettivamente

$$\mathcal{E}^{(n)} = (e_1^{(n)}, e_2^{(n)}, \ldots, e_{n-1}^{(n)}, e_n^{(n)}) \qquad \mathcal{E}^{(m)} = (e_1^{(m)}, e_2^{(m)}, \ldots, e_{m-1}^{(m)}, e_m^{(m)}).$$

Poiché $e_j^{(n)} \in \mathbb{R}^n$ abbiamo $g(e_j^{(n)}) \in \mathbb{R}^m$, dunque possiamo scriverlo in coordinate rispetto a $\mathcal{E}^{(m)}$, ovvero trovare coefficienti $a_{ij} \in \mathbb{R}$ per $i = 1, \ldots, m$ tali che

$$g(e_j^{(n)}) = \sum_{i=1}^{m} a_{ij} \cdot e_i^{(m)}.$$

Abbiamo trovato una matrice $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, dove $(A)_{ij} = a_{ij}$, e affermiamo che $g = f_A$. Infatti se $x \in \mathbb{R}^n$ abbiamo

$$\begin{aligned}
x &= \sum_{j=1}^{n} x_j \cdot e_j^{(n)} \\
\Rightarrow\ g(x) &= \sum_{j=1}^{n} x_j \cdot f(e_j^{(n)}) \\
&= \sum_{j=1}^{n} x_j \cdot \sum_{i=1}^{m} a_{ij} \cdot e_i^{(m)} \\
&= \sum_{i=1}^{m} \left(\sum_{i=1}^{m} a_{ij} \cdot x_j \right) \cdot e_i^{(m)} \\
&= \sum_{i=1}^{m} (A \cdot x)_i \cdot e_i^{(m)} \\
&= A \cdot x \\
&= f_A(x).
\end{aligned}$$

□

Abbiamo provato che le applicazioni lineari $\mathbb{R}^n \to \mathbb{R}^m$ corrispondono in modo naturale alle matrici di $\mathcal{M}_{m\times n}(\mathbb{R})$, e che attraverso questa corrispondenza la composizione si traduce nel prodotto righe per colonne. Questo suggerisce di scrivere semplicemente A in luogo di f_A, cosa che faremo d'ora in poi quasi sempre.

2.2 Isomorfismi, nucleo e immagine

Ricordiamo che se $f : X \to Y$ e $g : Y \to Z$ sono applicazioni tra insiemi, la loro composizione $g \circ f : X \to Z$ è l'applicazione data da $(g \circ f)(x) = g(f(x))$. Inoltre un'applicazione $f : X \to Y$ tra insiemi è *invertibile* se esiste $g : Y \to X$, tale che $g \circ f = \mathrm{id}_X$ e $f \circ g = \mathrm{id}_Y$, ovvero $g(f(x)) = x$ per ogni $x \in X$ e $f(g(y)) = y$ per ogni $y \in Y$. Se tale g esiste, viene detta inversa di f e indicata con f^{-1}. Sappiamo che f è invertibile se e solo se è bigettiva.

Passando ora alle applicazioni lineari, abbiamo questo risultato molto semplice:

Proposizione 2.2.1.
- *Se $f : V \to W$ e $g : W \to Z$ sono applicazioni lineari, la loro composizione $g \circ f : V \to Z$ è lineare;*
- *Se $f : V \to W$ è lineare invertibile, la sua inversa $f^{-1} : W \to V$ è lineare.*

Definizione 2.2.2. Un'applicazione lineare invertibile è detta *isomorfismo* di spazi vettoriali. Se tra due spazi vettoriali V e W esiste un isomorfismo, diciamo che essi sono *isomorfi*, ovvero che sono "uguali" dal punto di vista della struttura di spazio vettoriale.

Proposizione 2.2.3. *Se V ha dimensione n e $\mathcal{B}$ è una fissata base di V, l'applicazione*

$$\begin{array}{rcl} V & \to & \mathbb{R}^n \\ v & \mapsto & [v]_{\mathcal{B}} \end{array}$$

è un isomorfismo.

Dimostrazione. Per la linearità, dobbiamo vedere che se $t, s \in \mathbb{R}$ e $v, w \in V$ si ha

$$[t \cdot v + s \cdot w]_{\mathcal{B}} = t \cdot [v]_{\mathcal{B}} + s \cdot [w]_{\mathcal{B}}.$$

Posto $x = [v]_{\mathcal{B}}$ e $y = [w]_{\mathcal{B}}$, dobbiamo dunque vedere che

$$[t \cdot v + s \cdot w]_{\mathcal{B}} = t \cdot x + s \cdot y.$$

Per definizione di coordinate, se $\mathcal{B} = (v_1, \dots, v_n)$, abbiamo

$$\begin{array}{rcl} v & = & x_1 \cdot v_1 + \ldots + x_n \cdot v_n \\ w & = & y_1 \cdot v_1 + \ldots + y_n \cdot v_n \\ \Rightarrow\ t \cdot v + s \cdot w & = & t \cdot (x_1 \cdot v_1 + \ldots + x_n \cdot v_n) + s \cdot (y_1 \cdot v_1 + \ldots + y_n \cdot v_n) \\ & = & (t \cdot x_1 + s \cdot y_1) \cdot v_1 + \ldots (t \cdot x_n + s \cdot y_n) \cdot v_n \\ & = & (t \cdot x + s \cdot y)_1 \cdot v_1 + (t \cdot x + s \cdot y)_n \cdot v_n \end{array}$$

da cui segue quanto richiesto.

L'inversa dell'applicazione $V \to \mathbb{R}^n$ che manda v nel suo vettore delle coordinate $[v]_{\mathcal{B}}$ è quella che manda $x \in \mathbb{R}^n$ in $x_1 \cdot v_1 + \ldots + x_n \cdot v_n \in V$. □

Una conseguenza immediata di questo fatto è il seguente:

Corollario 2.2.4. *Due spazi vettoriali reali aventi la medesima dimensione reale finita sono tra loro isomorfi.*

Nucleo e immagine Nel resto del paragrafo sviluppiamo una teoria che avrà come conseguenza in particolare il fatto che due spazi vettoriali reali aventi differenti dimensioni reali finite *non* sono tra loro isomorfi.

Definizione 2.2.5. Se $f : V \to W$ è lineare, introduciamo:

- Il *nucleo* di f, definito come $\mathrm{Ker}(f) = \{v \in V : \ f(v) = 0\}$;
- L'*immagine* di f, denotata con $\mathrm{Im}(f)$.

Proposizione 2.2.6. *Se $f : V \to W$ è lineare allora $\mathrm{Ker}(f)$ è un sottospazio vettoriale di V e $\mathrm{Im}(f)$ è un sottospazio vettoriale di W.*

Dimostrazione. Siano $v_1, v_2 \in \mathrm{Ker}(f)$ e $t_1, t_2 \in \mathbb{R}$; dobbiamo provare che $t_1 \cdot v_1 + t_2 \cdot v_2 \in \mathrm{Ker}(f)$. L'ipotesi su v_1 e v_2 significa che $f(v_1) = f(v_2) = 0$, e dobbiamo provare che $f(t_1 \cdot v_1 + t_2 \cdot v_2) = 0$. Ma

$$f(t_1 \cdot v_1 + t_2 \cdot v_2) = t_1 \cdot f(v_1) + t_2 \cdot f(v_2) = t_1 \cdot 0 + t_2 \cdot 0 = 0$$

e abbiamo provato che $\mathrm{Ker}(f)$ è un sottospazio di V.

Siano $w_1, w_2 \in \mathrm{Im}(f)$ e $t_1, t_2 \in \mathbb{R}$; dobbiamo provare che $t_1 \cdot w_1 + t_2 \cdot w_2$ sta in $\mathrm{Im}(f)$. L'ipotesi su w_1 e w_2 significa che esistono $v_1, v_2 \in V$ tali che $w_1 = f(v_1)$ e $w_2 = f(v_2)$, e dobbiamo esibire $v \in V$ tale che $f(v) = t_1 \cdot w_1 + t_2 \cdot w_2$. In effetti, posto $v = t_1 \cdot v_1 + t_2 \cdot v_2$ abbiamo $f(v) = t_1 \cdot f(v_1) + t_2 \cdot f(v_2) = t_1 \cdot w_1 + t_2 \cdot w_2$, e abbiamo provato che $\mathrm{Im}(f)$ è un sottospazio di W. □

Un primo facile risultato consente di dare una chiara interpretazione del fatto che il nucleo o l'immagine di un'applicazione lineare f sia banale, in termini delle proprietà di f:

Proposizione 2.2.7. *Sia $f : V \to W$ lineare. Allora:*

- *f è iniettiva se e soltanto se $\mathrm{Ker}(f) = \{0\}$;*
- *f è surgettiva se e soltanto se $\mathrm{Im}(f) = W$;*
- *f è l'applicazione identicamente nulla se e soltanto se $\mathrm{Ker}(f) = V$, oppure equivalentemente $\mathrm{Im}(f) = \{0\}$.*

Dimostrazione. Solo la prima affermazione non è evidente. Se f è iniettiva, essendo $f(0) = 0$, si ha $f^{-1}(0) = \{0\}$, dunque $\mathrm{Ker}(f) = \{0\}$. Viceversa, siano $v, v' \in V$ tali che $f(v) = f(v')$. Allora $f(v - v') = 0$, dunque $v - v' \in \mathrm{Ker}(f)$. Se quest'ultimo sottospazio è quello banale $\{0\}$, ne segue che $v = v'$, dunque f è iniettiva. □

Corollario 2.2.8. *Sia $f : V \to W$ lineare con V e W di dimensione finita. Allora:*

- *f è iniettiva se e soltanto se* $\dim(\mathrm{Ker}(f)) = 0$*;*
- *f è surgettiva se e soltanto se* $\dim(\mathrm{Im}(f)) = \dim(W)$*.*

Abbiamo ora un risultato, chiamato *formula della dimensione*, che lega tra loro la dimensione del nucleo, quella dell'immagine e quella dello spazio di partenza.

Teorema 2.2.9. *Sia $f : V \to W$ un'applicazione lineare con V di dimensione finita. Allora:*

$$\dim(\mathrm{Ker}(f)) + \dim(\mathrm{Im}(f)) = \dim(V).$$

Dimostrazione. Poiché V ha dimensione finita e $\mathrm{Ker}(f)$ ne è un sottospazio, anche quest'ultimo ha dimensione finita, che indichiamo con k. Possiamo dunque scegliere una base $(v_1, \dots, v_k)$ di $\mathrm{Ker}(f)$, e completarla a una base $(v_1, \dots, v_n)$ di V. Asseriamo ora che $(f(v_{k+1}), \dots, f(v_n))$ è una base di $\mathrm{Im}(f)$, il che comporta immediatamente la tesi.

Proviamo dunque che $f(v_{k+1}), \dots, f(v_n)$ sono linearmente indipendenti e generano $\mathrm{Im}(f)$.

Supponiamo che una combinazione lineare $\alpha_{k+1} \cdot f(v_{k+1}) + \ldots + \alpha_n \cdot f(v_n)$ sia il vettore nullo. Dalla linearità di f deduciamo che

$$f(\alpha_{k+1} \cdot v_{k+1} + \ldots + \alpha_n \cdot v_n) = 0,$$

dunque

$$\alpha_{k+1} \cdot v_{k+1} + \ldots + \alpha_n \cdot v_n \in \mathrm{Ker}(f).$$

Avendo scelto $(v_1, \dots, v_k)$ come base di $\mathrm{Ker}(f)$, ne deduciamo che esistono coefficienti $\alpha_1, \dots, \alpha_k$ tali che

$$\begin{aligned}
& \alpha_{k+1} \cdot v_{k+1} + \ldots + \alpha_n \cdot v_n = \alpha_1 \cdot v_1 + \ldots + \alpha_k \cdot v_k \\
\Rightarrow\ & (-\alpha_1) \cdot v_1 + \ldots + (-\alpha_k) \cdot v_k + \alpha_{k+1} \cdot v_{k+1} + \ldots + \alpha_n \cdot v_n = 0 \\
\Rightarrow\ & \alpha_1 = \ldots = \alpha_k = \alpha_{k+1} = \ldots = \alpha_n = 0.
\end{aligned}$$

Ciò prova l'indipendenza lineare di $f(v_{k+1}), \dots, f(v_n)$. Per dimostrare che generano $\mathrm{Im}(f)$, consideriamo un vettore $w \in \mathrm{Im}(f)$ generico. Per definizione di immagine, esiste $v \in V$ tale che $f(v) = w$. Ora, $(v_1, \dots, v_n)$ è una base di V, dunque

$$v = \alpha_1 \cdot v_1 + \ldots + \alpha_n \cdot v_n.$$

Usiamo ora la linearità di f e il fatto che $v_1, \dots, v_k \in \mathrm{Ker}(f)$, deducendo che

$$w = f(v) = \alpha_{k+1} \cdot f(v_{k+1}) + \ldots + \alpha_n \cdot f(v_n)$$

e dunque la conclusione desiderata. □

Corollario 2.2.10. *Sia $f : V \to W$ lineare con V e W di dimensione finita.*

- *Se f è iniettiva allora $\dim(V) \leqslant \dim(W)$;*
- *Se f è surgettiva allora $\dim(V) \geqslant \dim(W)$;*
- *Se f è bigettiva (cioè è un isomorfismo) allora $\dim(V) = \dim(W)$.*

Inoltre, se V e W hanno la stessa dimensione, sono fatti equivalenti:

- *f è iniettiva;*
- *f è surgettiva;*
- *f è bigettiva (cioè è un isomorfismo).*

Mostreremo nel prossimo capitolo che la conoscenza della dimensione dell'immagine di un'assegnata applicazione lineare f (un numero intero che chiameremo *rango* di f) ha cruciale importanza in molti problemi (ad esempio quello della risoluzione dei sistemi lineari). Illustreremo anche un metodo efficace di calcolo del rango.

Concludiamo questo paragrafo traducendo in termini di matrici la nozione di invertibilità e inversa per applicazioni lineari. Da quanto già provato discende quanto segue:

- L'applicazione lineare $\mathbb{R}^n \to \mathbb{R}^m$ associata a una matrice $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ può essere invertibile solo se $m = n$;
- L'applicazione lineare $\mathbb{R}^n \to \mathbb{R}^n$ associata a una matrice $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ è invertibile se solo se esiste una matrice $B \in \mathcal{M}_{n\times n}(\mathbb{R})$ tale che $B \cdot A = A \cdot B = \mathrm{I}_n$.

Se tale B esiste la chiamiamo *inversa* di A e la indichiamo con A^{-1}. Inoltre (purché A sia quadrata $n \times n$) basta conoscere una delle due relazioni $B \cdot A = \mathrm{I}_n$ oppure $A \cdot B = \mathrm{I}_n$ per concludere che vale anche l'altra.

2.3 Matrice di un'applicazione lineare rispetto a basi assegnate e cambi di base

Nella Proposizione 2.1.9 abbiamo visto come si possa associare ad ogni applicazione lineare $g : \mathbb{R}^n \to \mathbb{R}^m$ una matrice $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ tale che $g = f_A$, e nella costruzione di A abbiamo utilizzato le basi canoniche di $\mathbb{R}^n$ e $\mathbb{R}^m$. Possiamo ora estendere questa costruzione a un contesto generale.

Definizione 2.3.1. Dati spazi vettoriali V e W di dimensione finita, loro basi $\mathcal{B} = (v_1, \ldots, v_n)$ e $\mathcal{C} = (w_1, \ldots, w_m)$, e un'applicazione $f : V \to W$ lineare, chiamiamo *matrice associata a f rispetto alle basi $\mathcal{B}$ in partenza e $\mathcal{C}$ in arrivo* la matrice $A = (a_{ij}) \in \mathcal{M}_{m\times n}(\mathbb{R})$ tale che

$$f(v_j) = \sum_{i=1}^{m} a_{ij} \cdot w_i$$

per $j = 1, \ldots, n$, ovvero

$$\begin{array}{rcl} f(v_1) & = & a_{11} \cdot w_1 + \ldots + a_{m1} \cdot w_m \\ & \vdots & \\ f(v_n) & = & a_{1n} \cdot w_1 + \ldots + a_{mn} \cdot w_m \end{array} \quad \Rightarrow \quad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

Indichiamo A con $[f]^{\mathcal{C}}_{\mathcal{B}}$.

Osservazione 2.3.2. Nella colonna j-esima di $[f]^{\mathcal{C}}_{\mathcal{B}}$ si trovano le coordinate rispetto a $\mathcal{C}$ dell'immagine tramite f del j-esimo elemento di $\mathcal{B}$.

Osservazione 2.3.3. Sia $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, siano $\mathcal{E}^{(n)}, \mathcal{E}^{(m)}$ le basi canoniche di $\mathbb{R}^n$ e $\mathbb{R}^m$, e sia $f_A : \mathbb{R}^n \to \mathbb{R}^m$ l'applicazione lineare associata ad A. Allora $[f_A]^{\mathcal{E}^{(m)}}_{\mathcal{E}^{(n)}}$ è data da A stessa.

Teorema 2.3.4. $[f(v)]_{\mathcal{C}} = [f]^{\mathcal{C}}_{\mathcal{B}} \cdot [v]_{\mathcal{B}}$ *per ogni* $v \in V$.

Dimostrazione. Sia $[v]_{\mathcal{B}} = x$, dunque $v = \sum_{j=1}^{n} x_j \cdot v_j$. Allora:

$$\begin{aligned} f(v) &= f\left(\sum_{j=1}^{n} x_j \cdot v_j\right) \\ &= \sum_{j=1}^{n} x_j \cdot f(v_j) = \sum_{j=1}^{n} x_j \cdot \left(\sum_{i=1}^{m} a_{ij} \cdot w_i\right) \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} \cdot x_j\right) \cdot w_i = \sum_{i=1}^{m} \left([f]_{\mathcal{B}}^{\mathcal{C}} \cdot x\right)_i \cdot w_i \\ \Rightarrow [f(v)]_{\mathcal{C}} &= [f]_{\mathcal{B}}^{\mathcal{C}} \cdot x. \end{aligned}$$

□

Possiamo esprimere informalmente il contenuto di questo risultato nei due modi seguenti:

- L'azione di f sui vettori corrisponde esattamente all'azione della moltiplicazione a sinistra per $[f]_{\mathcal{B}}^{\mathcal{C}}$ sulle coordinate dei vettori (rispetto alle basi $\mathcal{B}$ in partenza e $\mathcal{C}$ in arrivo);
- Identificando V a $\mathbb{R}^n$ tramite la base $\mathcal{B}$ e W a $\mathbb{R}^m$ tramite $\mathcal{C}$, la f viene identificata all'applicazione associata alla matrice $[f]_{\mathcal{B}}^{\mathcal{C}}$.

Il prossimo risultato conferma, come il precedente, la naturalità dell'operazione di associazione di una matrice a un'applicazione lineare.

Proposizione 2.3.5. *Se $f : V \to W$ e $g : W \to Z$ sono applicazioni lineari e sono date basi $\mathcal{B}, \mathcal{C}, \mathcal{D}$ di V, W, Z rispettivamente, si ha:*

$$[g \circ f]_{\mathcal{B}}^{\mathcal{D}} = [g]_{\mathcal{C}}^{\mathcal{D}} \cdot [f]_{\mathcal{B}}^{\mathcal{C}}.$$

Dimostrazione. Supponiamo che

$$\mathcal{B} = (v_1, \ldots, v_n) \quad \mathcal{C} = (w_1, \ldots, w_m) \quad \mathcal{D} = (z_1, \ldots, z_k)$$

e che $[f]_{\mathcal{B}}^{\mathcal{C}} = A = (a_{ij})$ e $[g]_{\mathcal{C}}^{\mathcal{D}} = X = (x_{pi})$, ovvero che

$$f(v_j) = \sum_{i=1}^{m} a_{ij} \cdot w_i \quad g(w_i) = \sum_{p=1}^{k} x_{pi} \cdot z_p.$$

Dobbiamo vedere che $[g \circ f]_{\mathcal{B}}^{\mathcal{D}} = X \cdot A$, ovvero che

$$(g \circ f)(v_j) = \sum_{p=1}^{k} (X \cdot A)_{pj} \cdot z_p$$

e infatti

$$\begin{aligned}
(g \circ f)(v_j) &= g(f(v_j)) \\
&= g\left(\sum_{i=1}^{m} a_{ij} \cdot w_i\right) \\
&= \sum_{i=1}^{m} a_{ij} \cdot g(w_i) \\
&= \sum_{i=1}^{m} a_{ij} \cdot \left(\sum_{p=1}^{k} x_{pi} \cdot z_p\right) \\
&= \sum_{p=1}^{k} \left(\sum_{i=1}^{m} x_{pi} \cdot a_{ij}\right) \cdot z_p \\
&= \sum_{p=1}^{k} (X \cdot A)_{pj} \cdot z_p.
\end{aligned}$$

□

Affrontiamo ora le questioni seguenti: come cambiano le coordinate di un vettore cambiando la base rispetto a cui sono calcolate? Come cambia la matrice associata a un'applicazione lineare cambiando le basi rispetto a cui viene calcolata? Per rispondere introduciamo una nuova nozione:

Definizione 2.3.6. Se $\mathcal{B} = (v_1, \ldots, v_n)$ e $\mathcal{B}' = (v'_1, \ldots, v'_n)$ sono basi di uno spazio vettoriale V, chiamiamo *matrice di cambio di base* da $\mathcal{B}$ a $\mathcal{B}'$ la matrice $X = (x_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{R})$ tale che

$$v'_j = \sum_{i=1}^{n} x_{ij} \cdot v_i \qquad j = 1, \ldots, n.$$

Possiamo ora descrivere come cambiano le coordinate cambiando base:

Proposizione 2.3.7.
- *La matrice X di cambio di base da $\mathcal{B}$ a $\mathcal{B}'$ è invertibile, e la sua inversa è la matrice di cambio di base da $\mathcal{B}'$ a $\mathcal{B}$;*

- $[v]_{\mathcal{B}'} = X^{-1} \cdot [v]_{\mathcal{B}}$ *per ogni* $v \in V$.

Dimostrazione. Chiamiamo Y la matrice di cambio di base da $\mathcal{B}'$ a $\mathcal{B}$. Allora si ha per $k = 1, \ldots, n$

$$v_k = \sum_{j=1}^{n} y_{jk} \cdot v'_j$$

dunque

$$\begin{aligned} v_k &= \sum_{j=1}^{n} y_{jk} \cdot \sum_{i=1}^{n} x_{ij} \cdot v_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} x_{ij} \cdot y_{jk} \right) \cdot v_i \\ &= \sum_{i=1}^{n} (X \cdot Y)_{ik} \cdot v_i. \end{aligned}$$

Poiché il vettore delle coordinate di v_k rispetto a $\mathcal{B}$ è naturalmente il vettore e_k delle base canonica di $\mathbb{R}^n$, si ha $(X \cdot Y)_{ik}$ vale 1 per $i = k$ e 0 per $i \neq k$, dunque $X \cdot Y = \mathrm{I}_n$, il che basta a concludere che $Y = X^{-1}$.

Per il secondo punto, supponiamo che $[v]_{\mathcal{B}} = u \in \mathbb{R}^n$, ovvero

$$v = \sum_{k=1}^{n} u_k \cdot v_k.$$

Allora

$$\begin{aligned} v &= \sum_{k=1}^{n} u_k \cdot v_k \\ &= \sum_{k=1}^{n} u_k \cdot \sum_{j=1}^{n} y_{jk} \cdot v'_j \\ &= \sum_{j=1}^{n} \left(\sum_{k=1}^{n} y_{jk} \cdot u_k \right) \cdot v'_j \\ &= \sum_{j=1}^{n} (Y \cdot u)_j \cdot v'_j \end{aligned}$$

il che comporta che $[v]_{\mathcal{B}'} = Y \cdot u$, ovvero la conclusione. $\square$

Concludiamo mostrando come cambia la matrice cambiando le basi:

Proposizione 2.3.8. *Se $\mathcal{B}$ e $\mathcal{B}'$ sono basi di V con matrice di cambio X dalla prima alla seconda, $\mathcal{C}$ e $\mathcal{C}'$ sono basi di W con matrice di cambio Y dalla prima alla seconda, e $f : V \to W$ è lineare, allora*

$$[f]_{\mathcal{B}'}^{\mathcal{C}'} = Y^{-1} \cdot [f]_{\mathcal{B}}^{\mathcal{C}} \cdot X.$$

Dimostrazione. Sappiamo da sopra che la matrice di cambio di base da $\mathcal{C}'$ a $\mathcal{C}$ è la $Z = Y^{-1}$, ovvero

$$w_i = \sum_{q=1}^{m} z_{qi} \cdot w'_q.$$

Chiamiamo ora A la matrice $[f]_{\mathcal{B}}^{\mathcal{C}}$. Allora:

$$\begin{aligned}
f(v'_p) &= f\left(\sum_{j=1}^{n} x_{jp} \cdot v_j\right) \\
&= \sum_{j=1}^{n} x_{jp} \cdot f(v_j) \\
&= \sum_{j=1}^{n} x_{jp} \cdot \sum_{i=1}^{m} a_{ij} \cdot w_i \\
&= \sum_{j=1}^{n} x_{jp} \cdot \sum_{i=1}^{m} a_{ij} \cdot \sum_{q=1}^{m} z_{qi} \cdot w'_q \\
&= \sum_{q=1}^{m} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} z_{qi} \cdot a_{ij} \cdot x_{jp}\right) \cdot w'_q \\
&= \sum_{q=1}^{m} (Z \cdot A \cdot X)_{qp} \cdot w'_q
\end{aligned}$$

da cui segue che $[f]_{\mathcal{B}'}^{\mathcal{C}'} = Z \cdot A \cdot X$, ovvero la conclusione. □

2.4 Somme dirette e proiezioni

In questo paragrafo introduciamo un particolare tipo di applicazione lineare (su cui torneremo, in un contesto più geometrico, nel Capitolo 4).

Definizione 2.4.1. Se V è uno spazio vettoriale reale e $W, Z \subset V$ sono sottospazi vettoriali, diciamo che V *si decompone nella somma diretta* di W e Z, e scriviamo $V = W \oplus Z$, se $W + Z = V$ e $W \cap Z = \{0\}$.

Osservazione 2.4.2. Se V ha dimensione finita e $W, Z \subset V$ sono sottospazi, due qualsiasi delle seguenti tre condizioni implicano la terza:

- $W + Z = V$;
- $W \cap Z = \{0\}$;
- $\dim(W) + \dim(Z) = \dim(V)$.

Il prossimo risultato mostra che, data una decomposizione in somma diretta $V = W \oplus Z$, restano definite due applicazioni $p, q : V \to V$, dette *proiezioni associate* alla decomposizione.

Proposizione 2.4.3. *Se $V = W \oplus Z$ allora:*

(1) Ogni vettore v di V si scrive in uno e un solo modo come $v = w + z$ con $w \in W$ e $z \in Z$.

Definiamo ora $p, q : V \to V$ ponendo $p(v) = w$ e $q(v) = z$ se $v = w + z$ con $w \in W$ e $z \in Z$. Allora:

(2) p e q sono lineari;

(3) $p(v) + q(v) = v$ per ogni $v \in V$.

(4) $\mathrm{Im}(p) = \mathrm{Ker}(q) = W$, $\mathrm{Im}(q) = \mathrm{Ker}(p) = Z$;

(5) $p \circ q = q \circ p = 0$;

(6) $p \circ p = p$, $q \circ q = q$;

Dimostrazione. (1) Per definizione di $W + Z$, essendo $W + Z = V$, sappiamo che almeno un'espressione $v = w + z$ esiste. Supponiamo che ce ne siano due:

$$v = w_1 + z_1, \qquad v = w_2 + z_2.$$

Ne deduciamo che $w_1 - w_2 = z_2 - z_1$. L'espressione sinistra di questo vettore comporta che esso appartiene a W, mentre quella destra comporta che appartiene a Z. Tuttavia per ipotesi si ha $W \cap Z = \{0\}$, dunque $w_1 - w_2 = z_2 - z_1 = 0$. Ne segue che $w_1 = w_2$ e $z_1 = z_2$, dunque le due espressioni di v erano in realtà uguali.

(2) Verifichiamo ad esempio che $p(v_1 + v_2)$ è uguale a $p(v_1) + p(v_2)$. Per definizione di p si ha $p(v_1) = w_1$ e $p(v_2) = w_2$ se $w_1, w_2 \in W$ ed esistono $z_1, z_2 \in Z$ tali che $v_1 = w_1 + z_1$ e $v_2 = w_2 + z_2$. Da queste relazioni segue che $v_1+v_2 = (w_1+w_2)+(z_1+z_2)$. Inoltre $w_1+w_2 \in W$ e $z_1+z_2 \in Z$. Ancora dalla definizione di p segue allora che $p(v_1 + v_2) = w_1 + w_2$, ovvero la conclusione desiderata.

(3) Segue immediatamente dalla definizione.

(4) È chiaro dalla definizione che $\mathrm{Im}(p) \subset W$. Inoltre se $w \in W$ si ha $w = w+0$ e $0 \in Z$, dunque $p(w) = w$. Ne segue che $\mathrm{Im}(p) = W$. Se $z \in Z$ si ha $z = 0+z$ e $0 \in W$, dunque $p(z) = 0$, cioè $z \in \mathrm{Ker}(p)$. Viceversa se $p(v) = 0$ allora $v = p(v) + q(v) = 0 + q(v) = q(v) \in Z$. Dunque $\mathrm{Ker}(p) = Z$. Per $\mathrm{Im}(q)$ e $\mathrm{Ker}(q)$ si procede analogamente.

(5) Segue immediatamente da (3).

(6) Se $v \in V$ si ha $(p \circ p)(v) = p(p(v)) = p(v)$ poiché $p(v) \in W$ e abbiamo già osservato che $p(w) = w$ se $w \in W$. Analogamente $q \circ q = q$. □

Il seguente risultato, insieme alla Proposizione 2.4.3, caratterizza le proiezioni.

Proposizione 2.4.4. *Se V è uno spazio vettoriale e $p : V \to V$ è lineare e tale che $p \circ p = p$, posto $W = \mathrm{Im}(p)$ e $Z = \mathrm{Ker}(p)$, si ha $V = W \oplus Z$ e la proiezione su W associata a questa decomposizione è p.*

Dimostrazione. Dato $v \in V$ abbiamo $v = p(v) + (v - p(v))$, dove naturalmente $p(v) \in W$. Inoltre

$$p(v - p(v)) = p(v) - (p \circ p)(v) = p(v) - p(v) = 0$$

dunque $v - p(v) \in Z$. Resta provato che $V = W + Z$.

Sia $x \in W \cap Z$. Poiché $x \in W$ si ha $x = p(v)$. Dato che $p \circ p = p$ abbiamo allora $p(x) = p(p(v)) = p(v)$, dunque $x = p(x)$. Ma $x \in Z$, dunque $p(x) = 0$, perciò $x = 0$. Resta provato che $W \cap Z = \{0\}$.

Abbiamo verificato che $V = W \oplus Z$, e dall'argomento usato per vedere che $W + Z = V$ segue che la associata proiezione su W è la p. □

2.5 Derivate parziali di una funzione di più variabili

Per le funzioni di una variabile reale valgono questi due risultati fondamentali:

- Se $f : [a, b] \to \mathbb{R}$ è continua, allora ammette massimo e minimo assoluto;
- Se $f : (a, b) \to \mathbb{R}$ è una funzione, $x \in (a, b)$ ed esiste

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

  allora

$$f(x+t) = f(x) + f'(x) \cdot t + \mathrm{o}(t).$$

In questo paragrafo illustriamo (senza alcuna dimostrazione) gli analoghi in più variabili di questi risultati. Per farlo iniziamo ricordando che la distanza tra due punti x e y di $\mathbb{R}^n$ si calcola come

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

(su questo torneremo ampiamente nel Capitolo 4). Definiamo ora la palla aperta di centro $x \in \mathbb{R}^n$ e raggio $r > 0$ come

$$B(x, r) = \{y \in \mathbb{R}^n : \ d(x, y) < r\}.$$

Definizione 2.5.1. Diciamo che $A \subset \mathbb{R}^n$ è *aperto* se per ogni $x \in A$ esiste $r > 0$ tale che $B(x, r) \subset A$. Per $A \subset \mathbb{R}^n$ qualsiasi chiamiamo *chiusura* di A l'insieme

$$\overline{A} = \{y \in \mathbb{R}^n : \ \text{per ogni } r > 0 \text{ si ha } B(y, r) \cap A \neq \emptyset\}.$$

Diciamo che A è *chiuso* se $\overline{A} = A$. Inoltre diciamo che A è *limitato* se esiste $R > 0$ tale che $A \subset B(0, R)$. Dati $A \subset \mathbb{R}^n$, una funzione $f : A \to \mathbb{R}$ e un punto $x \in A$ diciamo che f è continua in x se per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che $d(f(y), f(x)) < \varepsilon$ per ogni $y \in A$ con $d(y, x) < \delta$, e che f è continua in A se lo è in ogni $x \in A$. Infine se $A \subset \mathbb{R}^n$, data una funzione $f : A \to \mathbb{R}$, diciamo che $x \in A$ è un *punto di massimo assoluto* (oppure *punto di minimo assoluto*) di f se $f(y) \leqslant f(x)$ per ogni $y \in A$ (oppure $f(y) \geqslant f(x)$ per ogni $y \in A$) e in tal caso che $f(x)$ è il *massimo assoluto* (oppure *minimo assoluto*) di f.

Per prima cosa abbiamo il seguente risultato:

Teorema 2.5.2. *Se $A \in \mathbb{R}^n$ è chiuso e limitato e $f : A \to \mathbb{R}$ è continua, allora f ha massimo e minimo assoluto.*

Introduciamo ora una nuova nozione:

Definizione 2.5.3. Dati $A \subset \mathbb{R}^n$ aperto, una funzione $f : A \to \mathbb{R}$ e $x \in A$, chiamiamo *derivata parziale* j-esima di f in x derivata in x di f (se esiste) ottenuta considerandola come funzione della sola x_j, con le altre variabili viste come parametri, ovvero

$$\frac{\partial f}{\partial x_j}(x) = \lim_{h \to 0} \frac{f(x + he_j) - f(x)}{h}.$$

E abbiamo il seguente risultato (le cui ipotesi potrebbero essere indebolite):

Teorema 2.5.4. *Siano dati $A \subset \mathbb{R}^n$ aperto e una funzione $f : A \to \mathbb{R}$ tale che ogni $\frac{\partial f}{\partial x_j}$ esista in ogni punto di A e sia continua come funzione da A a $\mathbb{R}$. Allora in ogni punto di x di A si ha*

$$f(x + v) = f(x) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) \cdot v_j + \mathrm{o}(d(0, v))$$

dove v varia in una palla $B(0, r)$ di $\mathbb{R}^n$ con r abbastanza piccolo perché $B(x, r)$ sia contenuta in A, e l'ultimo termine $\mathrm{o}(d(0, v))$ *rappresenta una funzione $g(v)$ tale che per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che $|g(v)| < \varepsilon \cdot d(0, v)$ se $d(0, v) < \delta$.*

Osservazione 2.5.5. Sia $A \subset \mathbb{R}^n$ un insieme aperto e supponiamo che valga la proprietà seguente: per ogni coppia di punti $x, y \in A$ esiste una successione di segmenti $s_0, \ldots, s_q$ di $\mathbb{R}^n$ contenuti in A e ciascuno parallelo a uno degli assi coordinati, tali che il primo estremo di s_0 è x, per $k = 1, \ldots, q$ il secondo estremo di s_{k-1} coincide con il primo estremo di s_k, e il secondo estremo di s_k è y. Allora date $f, g : A \to \mathbb{R}$ tali che $\frac{\partial f}{\partial x_j}(x) = \frac{\partial g}{\partial x_j}(x)$ per $j = 1, \ldots, n$ e per ogni $x \in A$, si ha necessariamente che f e g differiscono per una costante.

Chiudiamo il paragrafo enunciando l'analogo in più variabili della regola di derivazione della funzione composta. Ricordiamo che in una variabile per funzioni $f : (a, b) \to (c, d)$ e $g : (c, d) \to \mathbb{R}$, se $x \in (a, b)$ ed esistono $f'(x)$ e $g'(f(x))$ si ha che $g \circ f$ è derivabile in x e

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x).$$

Definizione 2.5.6. Consideriamo $A \subset \mathbb{R}^n$ aperto e una funzione $f : A \to \mathbb{R}^m$, ovvero una funzione $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$ a m componenti. Se in un punto x di A

esistono tutte le derivate parziali $\frac{\partial f_i}{\partial x_j}$, definiamo la *matrice jacobiana* di f in x come la matrice $(\mathrm{J}f)(x) \in \mathcal{M}_{m\times n}(\mathbb{R})$ il cui coefficiente di posto (i,j) è $\frac{\partial f_i}{\partial x_j}(x)$.

Teorema 2.5.7. *Siano dati $A \subset \mathbb{R}^n$ e $B \subset \mathbb{R}^m$ aperti e funzioni $f : A \to \mathbb{R}^m$ tale che $f(A) \subset B$ e $g : B \to \mathbb{R}^k$. Se f e g ammettono tutte le derivate parziali in ciascun punto del proprio dominio, allora per ogni $x \in A$ si ha*

$$(\mathrm{J}(g \circ f))(x) = (\mathrm{J}g)(f(x)) \cdot (\mathrm{J}f)(x),$$

ovvero

$$\frac{\partial (g\circ f)_p}{\partial x_j}(x) = \sum_{i=1}^{m} \frac{\partial g_p}{\partial y_i}(f(x)) \cdot \frac{\partial f_i}{\partial x_j}(x).$$

Per funzioni scalari $f : A \to \mathbb{R}$ la matrice jacobiana $(\mathrm{J}f)(x)$ in un punto x di A, un vettore riga $1 \times n$, è anche detto *differenziale* o *gradiente* di f in x, ma a volte viene considerato come vettore colonna invece che riga.

Capitolo 3

Sistemi lineari, rango, determinante

In questo capitolo sfruttiamo le nozioni introdotte nei precedenti per descrivere la struttura dell'insieme delle soluzioni di un sistema lineare. Quindi definiamo il rango di una matrice, il quale, seppure implicitamente, consente di stabilire se un sistema lineare ha soluzioni o meno. Introduciamo poi la nozione di determinante di una matrice quadrata e descriviamo alcune sue applicazioni interne all'algebra lineare (calcolo dell'inversa di una matrice, risoluzione dei sistemi quadrati con matrice invertibile, determinazione del rango di una matrice). Infine diamo una sommaria introduzione all'integrazione in più variabili e mostriamo come il determinante intervenga in questo ambito nella formula di cambiamento di variabile.

3.1 Sistemi lineari

Un *sistema lineare* di m equazioni nelle n incognite $x_1, \ldots, x_n$ è uno del tipo

$$\begin{cases} a_{11} \cdot x_1 + \ldots + a_{1n} \cdot x_n &= b_1 \\ & \vdots \\ a_{m1} \cdot x_1 + \ldots + a_{mn} \cdot x_n &= b_m, \end{cases}$$

dove le a_{ij} e b_i sono costanti reali assegnate. Introducendo ora la matrice $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \in \mathcal{M}_{m\times n}(\mathbb{R})$ e il vettore $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$, e interpretando le incognite $x_1, \ldots, x_n$ come un vettore $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, possiamo riscrivere il sistema nella forma

$$A \cdot x = b.$$

Chiameremo A la *matrice incompleta* del sistema e $(A\ b)$ *matrice completa* del sistema.

Ricordando che la matrice A rappresenta anche un'applicazione da $\mathbb{R}^n$ a $\mathbb{R}^m$, la teoria già discussa comporta i seguenti semplici fatti:

- Il sistema $A \cdot x = b$ ha soluzione se e soltanto se $b \in \mathrm{Im}(A)$;
- Se $m = n$ ed esiste la matrice A^{-1} inversa di A, allora la soluzione esiste ed è unica, ed è data da $x = A^{-1} \cdot b$.

Emergono quindi in modo naturale le questioni seguenti:

- Come determinare se un dato vettore appartiene all'immagine dell'applicazione associata a una matrice assegnata?
- Come determinare se una matrice assegnata è invertibile e, se lo è, come trovarne l'inversa?

Queste domande avranno nelle prossime pagine risposte complete e di natura "algoritmica" (cioè facilmente traducibili in programmi di un calcolatore).

3.2 Rango e soluzioni di un sistema

Se $f : V \to W$ è un'applicazione lineare tra spazî vettoriali e $\mathrm{Im}(f)$ ha dimensione finita, chiamiamo tale dimensione *rango* di f e la indichiamo con $\mathrm{rank}(f)$. Notiamo che se almeno uno tra V e W ha dimensione finita, automaticamente anche $\mathrm{Im}(f)$ ha dimensione finita, dunque si può definire $\mathrm{rank}(f)$.

Se A è una matrice, chiamiamo *rango* di A e indichiamo con $\mathrm{rank}(A)$ il rango dell'applicazione lineare associata ad A. Notiamo ora che i vettori che costituiscono le colonne di A sono certamente un sistema di generatori per $\mathrm{Im}(A)$. Dal procedimento di estrazione di una base da un insieme di generatori deduciamo allora la seguente semplice caratterizzazione del rango di una matrice:

Proposizione 3.2.1. *Il rango di una matrice A è il numero massimo di colonne di A che costituiscono un sistema di vettori linearmente indipendenti.*

È ora semplice provare il seguente risultato, detto Teorema di Rouché-Capelli:

Teorema 3.2.2. *Un sistema lineare ha soluzione se e solo se la sua matrice completa e la sua matrice incompleta hanno lo stesso rango.*

Dimostrazione. Supponiamo che il sistema sia $A \cdot x = b$ con $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ e $b \in \mathbb{R}^m$. Ognuno dei fatti seguenti è facilmente equivalente al successivo:

- Esiste una soluzione;
- b appartiene all'immagine di A;
- b è combinazione lineare delle colonne di A;
- Aggiungendo b alle colonne di A la dimensione del generato rimane la stessa;
- A e $(A\ b)$ hanno lo stesso rango.

□

Osservazione 3.2.3. La matrice completa di un sistema lineare può avere rango o uguale a quello della matrice incompleta oppure superiore di 1 a quest'ultimo.

Introduciamo ora un ulteriore elemento di terminologia. Un sistema lineare si dice *omogeneo* se è del tipo $A \cdot x = 0$, ovvero se il vettore b dei *termini noti* è nullo. Naturalmente un sistema omogeneo ammette sempre la soluzione $x = 0$, ovvero $x_1 = \ldots = x_n = 0$. Più esattamente, usando la formula della dimensione, si verifica che *l'insieme delle soluzioni di un sistema lineare omogeneo $A \cdot x = 0$ in n incognite è lo spazio vettoriale* $\mathrm{Ker}(A)$ *di dimensione* $n - \mathrm{rank}(A)$.

Il prossimo risultato si può esprimere nel modo seguente: *se un sistema lineare ammette soluzioni, la sua generica soluzione è la somma di una qualsiasi fissata soluzione particolare e della generica soluzione del sistema lineare omogeneo avente la stessa matrice incompleta.*

Proposizione 3.2.4. *Dati $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, $b \in \mathbb{R}^m$ e $x_0 \in \mathbb{R}^n$ per i quali vale $A \cdot x_0 = b$, si ha*

$$\big\{x \in \mathbb{R}^n :\ A \cdot x = b\big\} \;=\; \big\{x_0 + u :\ A \cdot u = 0\big\}.$$

Dimostrazione. Dato $x \in \mathbb{R}^n$ si ha:

$$\begin{aligned} A \cdot x = b \quad &\Leftrightarrow \quad A \cdot x = A \cdot x_0 \\ &\Leftrightarrow \quad A \cdot (x - x_0) = 0 \\ &\Leftrightarrow \quad x - x_0 = u \quad \text{con } A \cdot u = 0 \\ &\Leftrightarrow \quad x = x_0 + u \quad \text{con } A \cdot u = 0. \end{aligned}$$

□

Un sistema lineare di m equazioni in n incognite si dice:

- *quadrato* se $m = n$;
- *sovradeterminato* se $m > n$;
- *sottodeterminato* se $m < n$.

Dall'ultimo risultato e la discussione che lo precede sui sistemi omogenei deduciamo ora che:

- Un sistema lineare non omogeneo quadrato oppure sovradeterminato può avere una sola soluzione, infinite soluzioni, oppure nessuna soluzione;
- Un sistema lineare non omogeneo sottodeterminato può avere infinite soluzioni oppure nessuna soluzione.

Per l'effettiva risoluzione di un sistema lineare sono disponibili semplici algoritmi, che non descriviamo qui esplicitamente. Essi sono basati sul fatto che da un sistema se ne possono ricavare altri *equivalenti* (cioè aventi le stesse soluzioni) con operazioni del tipo seguente:

- Riordino delle equazioni;
- Sostituzione dell'equazione i-esima con una combinazione lineare delle equazioni nella quale la i-esima abbia coefficiente non nullo.

3.3 Determinante

La teoria illustrata finora lascia aperte due questioni:

- Come stabilire se una matrice quadrata sia invertibile e, se lo è, come calcolarne l'inversa?

- Come calcolare il rango di una matrice qualsiasi?

La risposta ad entrambe si ottiene con l'introduzione di una funzione, denominata *determinante*, che ad ogni matrice quadrata associa un numero. Vale il fatto seguente, che si può dimostrare per induzione su n (ma non è facile):

Teorema 3.3.1. *Per ogni $n \geqslant 1$ esiste una ed una sola funzione*

$$\det_n : \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathbb{R}$$

con le seguenti proprietà:

- $\det_n$ *è lineare in ciascuna colonna fissate le altre, cioè per $j = 1, \ldots, n$, fissati $w_1, \ldots, w_{j-1}, w_{j+1}, \ldots, w_n \in \mathbb{R}^n$, la funzione*

  $$\mathbb{R}^n \ni x \mapsto \det_n(w_1, \ldots, w_{j-1}, x, w_{j+1}, \ldots, w_n) \in \mathbb{R}$$

  è lineare;

- *Se B si ottiene da A scambiando tra loro due colonne, allora*

  $$\det_n(B) = -\det_n(A);$$

- $\det_n(\mathrm{I}_n) = 1$*.*

Dando per buono questo fatto, si prova facilmente il seguente:

Proposizione 3.3.2. *Se $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ non è invertibile, allora* $\det_n(A) = 0$*.*

Dimostrazione. Dalla seconda proprietà del determinante segue subito che $\det(C) = 0$ se C ha due colonne uguali tra loro. Se $A = (w_1, \ldots, w_n)$ non è invertibile allora una delle colonne è combinazione linare delle altre, ovvero

$$w_j = t_1 \cdot w_1 + \ldots t_{j-1} \cdot w_{j-1} + t_{j+1} \cdot w_{j+1} + \ldots + t_n \cdot w_n$$

per qualche $j \in \{1, \ldots, n\}$ e per costanti t_k opportune. Allora

$$\begin{aligned}
\det_n(A) \quad = \quad & t_1 \cdot \det_n(w_1, \ldots, w_{j-1}, w_1, w_{j+1}, \ldots, w_n) \\
& + \ldots \\
& + t_{j-1} \cdot \det_n(w_1, \ldots, w_{j-1}, w_{j-1}, w_{j+1}, \ldots, w_n) \\
& + t_{j+1} \cdot \det_n(w_1, \ldots, w_{j-1}, w_{j+1}, w_{j+1}, \ldots, w_n) \\
& + \ldots \\
& + t_n \cdot \det_n(w_1, \ldots, w_{j-1}, w_n, w_{j+1}, \ldots, w_n) \\
= \quad & 0 + \ldots + 0 + 0 + \ldots + 0 \\
= \quad & 0.
\end{aligned}$$

□

In realtà, ma non è altrettanto semplice provarlo, vale anche il viceversa:

Proposizione 3.3.3. *Se $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ e $\det_n(A)) = 0$ allora A non è invertibile.*

Ammettendo questo fatto possiamo dimostrare il seguente risultato dovuto a Binet:

Teorema 3.3.4. *Date $A, B \in \mathcal{M}_{n\times n}(\mathbb{R})$ si ha $\det_n(A\cdot B) = \det_n(A)\cdot\det_n(B)$.*

Dimostrazione. Se $\det_n(B) = 0$, cioè se B non è invertibile, certamente non è invertibile $A \cdot B$ (perché come applicazione lineare non è iniettiva), dunque $\det_n(A \cdot B) = 0$ e il teorema vale. Supponiamo invece che $\det_n(B) \neq 0$ e consideriamo l'applicazione

$$D_B : \mathcal{M}_{n\times n}(\mathbb{R}) \ni C \mapsto \frac{\det_n(C\cdot B)}{\det_n(B)} \in \mathbb{R}.$$

È immediato verificare che la D_B soddisfa le proprietà caratterizzanti della funzione $\det_n$ elencate sopra, dunque $D_B = \det_n$ e la conclusione segue subito calcolando $D_B(A) = \det_n(A)$. □

Abbiamo finora enunciato l'esistenza di una quantità chiamata determinante di una matrice quadrata, senza dare metodi effettivi di calcolo, a cui ora invece ci dedichiamo. Per cominciare, per una matrice $A \in \mathcal{M}_{1\times 1}(\mathbb{R})$, ovvero $A = (a_{11})$ con $a_{11} \in \mathbb{R}$, abbiamo certamente, grazie alla prima e alla terza proprietà, che

$$\det_1(a_{11}) = a_{11}.$$

Ora per $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{R})$, possiamo osservare che A è invertibile se e soltanto se l'area del parallelogrammo che ha per lati le colonne di A è non nulla. Inoltre tale area vale 1 per $A = \mathrm{I}_2$. Tuttavia tale area in sé e per sé non può essere il determinante di A, in quanto l'area è sempre non negativa, dunque non può essere una funzione lineare delle colonne di A. Eseguendo il calcolo si vede però che l'area del parallelogrammo che ha per lati le colonne di A vale

$$|a_{11} \cdot a_{22} - a_{12} \cdot a_{21}|$$

il che suggerisce che valga l'uguaglianza

$$\det_2 \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21},$$

ed effettivamente è così. Osserviamo che $a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$ si può interpretare come un'area *con segno* del parallelogrammo che ha per lati le colonne di A: il segno è positivo se l'angolo minore di quello piatto che va dalla prima colonna alla seconda è in verso antiorario intorno all'origine, negativo altrimenti.

Passando al caso $n = 3$, possiamo di nuovo prevedere che $\det_3(A)$ rappresenti un *volume con segno* del parallelepipedo che ha per lati le colonne di A, e in questo modo come nel caso $n = 2$ si ottiene

$$\begin{aligned} & \det_3 \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\ = \quad & a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} \\ & - a_{13} \cdot a_{22} \cdot a_{31} - a_{11} \cdot a_{23} \cdot a_{32} - a_{12} \cdot a_{21} \cdot a_{33} \end{aligned}$$

detta *regola di Sarrus.* Un metodo mnemonico per ricordarla è il seguente: data $A \in \mathcal{M}_{3\times3}(\mathbb{R})$ si ricopiano le prime due colonne di A alla sua destra,

$$A = \left(a_{ij}\right)_{i,j=1,2,3} \quad \longrightarrow \quad \begin{matrix} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{matrix} ,$$

quindi si prendono, con segno positivo, i prodotti dei termini sulle tre diagonali "nord-ovest/sud-est" e, con segno negativo, i prodotti dei termini sulle diagonali "nord-est/sud-ovest"

Dall'analisi delle funzioni $\det_n$ costruite esplicitamente per $n = 1, 2, 3$ emerge quanto segue:

- $\det_n(A)$ è la somma di $n!$ addendi;
- Ogni tale addendo è $\pm$ il prodotto di n coefficienti di A;
- In ogni tale prodotto i diversi coefficienti sono su righe e colonne diverse;
- Tutti i prodotti di n coefficienti di A su righe e colonne diverse compaiono, preceduti da un segno $\pm$, come addendi di $\det_n(A)$.

Questa descrizione di fatto vale anche per $n > 3$ e fornisce quasi una formula esplicita per il calcolo di $\det_n(A)$: l'unica difficoltà è quella di attribuire correttamente i segni $\pm$ agli addendi, un argomento che qui però non affrontiamo. Dalla formula esplicita (con i segni corretti) seguono poi dei metodi di calcolo efficaci, che ora descriviamo. Per farlo, dati $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ e $i, j \in \{1, \dots, n\}$, indichiamo d'ora in poi con

$$A^{ij} \in \mathcal{M}_{(n-1)\times(n-1)}(\mathbb{R})$$

la matrice ottenuta da A cancellando la riga i-esima e la colonna j-esima. Vale la seguente:

Proposizione 3.3.5. *Per ogni* $\in \mathcal{M}_{n\times n}(\mathbb{R})$ *si ha:*

- *Fissato* $j \in \{1, \dots, n\}$

$$\det_n(A) = \sum_{i=1}^{n} (-1)^{i+j} \cdot a_{ij} \cdot \det_{n-1}(A^{ij})$$

  (formula di sviluppo di Laplace del determinante lungo la j*-esima colonna);*

- *Fissato* $j \in \{1, \dots, n\}$

$$\det_n(A) = \sum_{i=1}^{n} (-1)^{j+i} \cdot a_{ji} \cdot \det_{n-1}(A^{ji})$$

  (formula di sviluppo di Laplace del determinante lungo la j*-esima riga);*

- $\det_n({}^{\mathrm{t}}A) = \det_n(A)$.

L'efficacia delle formule di sviluppo si esplica pienamente se combinata con il fatto seguente:

Proposizione 3.3.6. *Il determinante di una matrice non cambia sostituendo una colonna oppure una riga con sé stessa più una combinazione lineare delle altre.*

Dimostrazione. Per le colonne il risultato segue subito dalle proprietà caratterizzanti della funzione $\det_n$ elencate sopra. Ma siccome il determinante non cambia trasponendo, ciò che vale per le colonne vale per le righe. $\square$

La strategia di calcolo di $\det_n(A)$ per n grande sarà dunque la seguente:

- Individuare una colonna oppure una riga di A e tramite operazioni sulle righe oppure sulle colonne ottenere su di essa molti coefficienti nulli (tipicamente: tutti tranne uno);
- Eseguire lo sviluppo di Laplace lungo quella colonna o quella riga.

D'ora in poi scriveremo soltanto $\det(A)$ invece che $\det_n(A)$.

Inversa di una matrice Abbiamo visto sopra che una matrice quadrata $n \times n$ è invertibile se e soltanto se $\det(A) \neq 0$. In realtà il determinante consente non solo di stabilire se A^{-1} esiste, ma anche di calcolarla. Si ha infatti:

Proposizione 3.3.7. *Se* $A \in \mathcal{M}_{n\times n}$ *e* $\det(A) \neq 0$ *si ha*

$$\left(A^{-1}\right)_{ij} = (-1)^{i+j} \cdot \frac{\det(A^{ji})}{\det(A)}.$$

Usando questo risultato non è difficile dedurre la *regola di Cramer* per la risoluzione dei sistemi lineari quadrati:

Proposizione 3.3.8. *Se* $A \in \mathcal{M}_{n\times n}$ *e* $\det(A) \neq 0$ *allora la soluzione del sistema* $A \cdot x = b$ *è data da*

$$x_j = \frac{\det(A_j)}{\det(A)}$$

dove A_j *è la matrice ottenuta da* A *sostituendo la* j*-esima colonna con* b.

Rango di una matrice Data $B \in \mathcal{M}_{m\times n}(\mathbb{R})$, chiamiamo *sottomatrice* A di B una matrice ottenuta da B cancellando alcune righe e colonne. Se B ha una sottomatrice $A \in \mathcal{M}_{k\times k}(\mathbb{R})$ con $\det(A) \neq 0$, allora le colonne di A sono linearmente indipendenti, da cui segue che le k colonne di B che contengono quelle di A sono linearmente indipendenti, e dunque che B ha rango almeno k. Ciò conduce, con un pochino di lavoro in più, alla seguente:

Proposizione 3.3.9. *Il rango di una matrice* $B \in \mathcal{M}_{m\times n}(\mathbb{R})$ *è il più grande intero* k *tale che esiste una sottomatrice* $A \in \mathcal{M}_{k\times k}(\mathbb{R})$ *di* B *con* $\det(A) \neq 0$.

3.4 Integrali multipli e cambiamento di parametro

In questo paragrafo accenniamo all'estensione del concetto di integrale per funzioni di più variabili, e mostriamo che la nozione di determinante interviene nella formula di cambiamento di parametro per questo integrale. Per semplicità di notazione scriviamo qui gli elementi di $\mathbb{R}^n$ come vettori riga.

Iniziamo dal caso di una funzione reale f limitata e non negativa definita su un rettangolo $R = [a, b] \times [c, d]$ del piano $\mathbb{R}^2$. Analogamente al caso di una variabile, vogliamo definire l'integrale di f su R, indicato con $\int\limits_R f$ oppure con

$$\int\limits_R f(x, y)\, \mathrm{d}x\, \mathrm{d}y,$$

in modo che rappresenti il volume del sottografico di f, ovvero

$$\{(x,y,z)\in\mathbb{R}^3:\ (x,y)\in R,\ 0\leqslant z\leqslant f(x,y)\}.$$

Per farlo chiamiamo *pluriparallelepipedo* P il dato di:

- Una suddivisione di $[a,b]$ come $a=t_0<t_1<\ldots<t_{N-1}<t_N=b$;
- Una suddivisione di $[c,d]$ come $c=s_0<s_1<\ldots<s_{M-1}<s_M=d$;
- La scelta per ogni $i=1,\ldots,N$ e $j=1,\ldots,M$ di un numero $h_{ij}\geqslant 0$.

Inoltre diciamo:

- Che il *volume* di P è $\mathcal{V}(P)=\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{M}h_{ij}\cdot(t_i-t_{i-1})\cdot(s_j-s_{j-1})$;
- Che $P\leqslant f$ se $h_{ij}\leqslant f(x,y)$ per ogni $(x,y)\in[t_{i-1},t_i]\times[s_{j-1},s_j]$;
- Che $f\leqslant P$ se $f(x,y)\leqslant h_{ij}$ per ogni $(x,y)\in[t_{i-1},t_i]\times[s_{j-1},s_j]$;
- Che f è *integrabile* su R se

  $$\sup\{\mathcal{V}(P):\ P\leqslant f\}=\inf\{\mathcal{V}(P):\ f\leqslant P\}$$

  e in tal caso chiamiamo integrale di f su R questo comune valore.

Per una funzione $f:R\to\mathbb{R}$ limitata di segno qualsiasi poniamo

$$f_+=\max\{f,0\}\qquad f_-=\max\{-f,0\},$$

di modo che $f=f_+-f_-$ e $f_\pm\geqslant 0$, e poniamo

$$\int\limits_R f=\int\limits_R f_+-\int\limits_R f_-$$

(se questi ultimi due integrali esistono).

Si può provare il seguente (non facile):

Proposizione 3.4.1. *Se $f:R\to\mathbb{R}$ è continua allora esiste $\int\limits_R f$. Inoltre*

$$\int\limits_R f=\int\limits_c^d\left(\int\limits_a^b f(x,y)\,\mathrm{d}x\right)\mathrm{d}y=\int\limits_a^b\left(\int\limits_c^d f(x,y)\,\mathrm{d}y\right)\mathrm{d}x$$

dove tutti quelli che compaiono sono integrali di funzioni continue.

Per una funzione $f : S \to \mathbb{R}$ definita su un sottoinsieme S di $\mathbb{R}^2$ limitato ma più complicato di un rettangolo, si può definire $\int\limits_S f$ con questo trucco:

- Si prende un qualsiasi rettangolo R contenente S;
- Si estende f a una funzione $\widetilde{f}$ su R ponendola 0 su $R \setminus S$;
- Si prova che l'esistenza di $\int\limits_R \widetilde{f}$ non dipende da R, né lo fa il valore di tale integrale se esiste (questo è facile).

Dunque si pone $\int\limits_S f = \int\limits_R \widetilde{f}$ per R e $\widetilde{f}$ come sopra (se il secondo integrale esiste). Tuttavia anche per una f continua la corrispondente $\widetilde{f}$ sarà tipicamente discontinua, quindi questa definizione non consente di utilizzare la proposizione precedente per effettuare il calcolo di $\int\limits_S f$. Per estendere gli insiemi su cui sia possibile calcolare in pratica gli integrali diamo le seguenti definizioni:

- Diciamo che $A \subset \mathbb{R}^2$ è *semplice* se è del tipo
$$\{(x,y) : \ x \in [a,b], \ Y_1(x) \leqslant y \leqslant Y_2(x)\}$$
oppure
$$\{(x,y) : \ y \in [c,d], \ X_1(y) \leqslant x \leqslant X_2(y)\}$$
con Y_1, Y_2 continue su $[a,b]$ tali che $Y_1 \leqslant Y_2$ su $[a,b]$, oppure X_1, X_2 continue su $[c,d]$ tali che $X_1 \leqslant X_2$ su $[c,d]$;

- Definiamo il *bordo* di un tale A come
$$\begin{aligned} \partial A \ = \ & \{(a,y) : \ Y_1(a) \leqslant y \leqslant Y_2(a)\} \\ & \cup \{(b,y) : \ Y_1(b) \leqslant y \leqslant Y_2(b)\} \\ & \cup \{(x, Y_1(x)) : \ x \in [a,b]\} \\ & \cup \{(x, Y_2(x)) : \ x \in [a,b]\} \end{aligned}$$
oppure
$$\begin{aligned} \partial A \ = \ & \{(x,c) : \ X_1(c) \leqslant x \leqslant X_2(c)\} \\ & \cup \{(x,d) : \ X_1(d) \leqslant x \leqslant X_2(d)\} \\ & \cup \{(X_1(y), y) : \ y \in [c,d]\} \\ & \cup \{(X_2(y), y) : \ y \in [c,d]\}. \end{aligned}$$

- Diciamo che $S \subset \mathbb{R}^2$ è *regolare* se si ha $S = A_1 \cup \ldots \cup A_k$ con A_j semplice per ogni j e $A_i \cap A_j \subset (\partial A_i) \cap (\partial A_j)$ per $i \neq j$.

Abbiamo allora:

Proposizione 3.4.2. • *Se A è semplice e $f : A \to \mathbb{R}$ è continua allora esiste $\int_A f$ e si ha, con la notazione come nella definizione di sopra,*

$$\int_A f = \int_a^b \left(\int_{Y_1(x)}^{Y_2(x)} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

oppure

$$\int_A f = \int_c^d \left(\int_{X_1(y)}^{X_2(y)} f(x,y)\,\mathrm{d}x \right) \mathrm{d}y;$$

- *Se S è regolare e $f : S \to \mathbb{R}$ è continua allora esiste $\int_S f$ e si ha, con la notazione come nella definizione di sopra,*

$$\int_S f = \left(\int_{A_1} f \right) + \ldots + \left(\int_{A_k} f \right).$$

La definizione di integrale si estende senza particolari difficoltà a funzioni definite su sottoinsiemi di $\mathbb{R}^n$ con $n > 2$: per funzioni non negative definite su prodotti di intervalli si definisce (se esiste) il volume del sottografico approssimandolo da sopra e da sotto con unioni di parallelepipedi n-dimensionali, e poi si procede come sopra. L'analogo n-dimensionale di insieme semplice è uno del tipo

$$A = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \quad : \quad (x_1, \ldots, x_{n-1}) \in B, \\ h(x_1, \ldots, x_{n-1}) \leqslant x_n \leqslant H(x_1, \ldots, x_{n-1})\},$$

dove $B \subset \mathbb{R}^{n-1}$ è un insieme opportuno e $h, H : B \to \mathbb{R}$ sono funzioni tali che $h \leqslant H$ su B. E in ipotesi ragionevoli continua a valere la formula di integrazione successiva:

$$\begin{aligned} &\int_A f(x_1,\ldots,x_n)\,\mathrm{d}x_1\ldots\,\mathrm{d}x_n \\ = &\int_B \left(\int_{h(x_1,\ldots,x_{n-1})}^{H(x_1,\ldots,x_{n-1})} f(x_1,\ldots,x_n)\,\mathrm{d}x_n \right) \mathrm{d}x_1\ldots\,\mathrm{d}x_{n-1}. \end{aligned}$$

Concludiamo il paragrafo enunciando la formula di cambiamento di variabile. Se $S, T \subset \mathbb{R}^n$ sono insiemi ragionevoli e $\varphi : S \to T$ è una bigezione tale che $\mathrm{J}\varphi$ esiste in ogni punto di S, per ogni $f : T \to \mathbb{R}$ continua si ha

$$\begin{aligned} &\int_T f(y_1,\ldots,y_n)\,\mathrm{d}y_1\ldots\,\mathrm{d}y_n \\ = &\int_S f(\varphi(x_1,\ldots,x_n)) \cdot |\det(\mathrm{J}\varphi(x_1,\ldots,x_n))|\,\mathrm{d}x_1\ldots\,\mathrm{d}x_n. \end{aligned}$$

Esempio 3.4.3. Se $B = \{(x,y) \in \mathbb{R}^2 : \ x^2+y^2 \leqslant 1\}$ e $f : B \to \mathbb{R}$ è definita da $f(x,y) = 1 - \sqrt{x^2+y^2}$ allora il sottografico di f è il cono retto con raggio di base 1 e altezza 1. Dunque $\int_B f$ è il volume di tale cono. Possiamo calcolarlo tramite la formula di integrazione successiva come

$$\int_{-1}^{1} \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \left(1 - \sqrt{x^2+y^2}\right) \mathrm{d}y \right) \mathrm{d}x,$$

ma i calcoli sono alquanto elaborati. Se invece utilizziamo le coordinate polari (ρ, ϑ), ovvero poniamo (qui conviene tornare a usare la notazione in colonna per i vettori)

$$\varphi\begin{pmatrix} \rho \\ \vartheta \end{pmatrix} = \begin{pmatrix} \rho\cdot\cos(\vartheta) \\ \rho\cdot\sin(\vartheta) \end{pmatrix}$$

abbiamo che $\varphi : [0,1]\times[0,2\pi] \to B$ è "quasi" una bigezione (fanno eccezione il segmento $\{0\}\times[0,2\pi]$ che φ manda interamente in 0, e i segmenti $[0,1]\times\{0\}$ e $[0,1]\times\{2\pi\}$ che hanno la stessa immagine, ma essi non contribuiscono all'area di $[0,1]\times[0,2\pi]$, dunque si possono trascurare). Ora abbiamo

$$\begin{aligned} \mathrm{J}\varphi\begin{pmatrix} \rho \\ \vartheta \end{pmatrix} &= \begin{pmatrix} \cos(\vartheta) & -\rho\cdot\sin(\vartheta) \\ \sin(\vartheta) & \rho\cdot\cos(\vartheta) \end{pmatrix} \\ \Rightarrow \det\left(\mathrm{J}\varphi\begin{pmatrix} \rho \\ \vartheta \end{pmatrix}\right) &= \rho. \end{aligned}$$

Inoltre $f\left(\varphi\begin{pmatrix}\rho\\ \vartheta\end{pmatrix}\right) = 1-\rho$, dunque

$$\int\limits_B f = \int\limits_0^{2\pi}\left(\int\limits_0^1 (1-\rho)\cdot\rho\,\mathrm{d}\rho\right)\,\mathrm{d}\vartheta = \int\limits_0^{2\pi}\frac{1}{6}\,\mathrm{d}\vartheta = \frac{\pi}{3}.$$

Capitolo 4

Il prodotto scalare dello spazio euclideo

In questo capitolo introduciamo il prodotto scalare tra due vettori dello spazio $\mathbb{R}^n$ e definiamo la norma e la distanza ed esso associate. Quindi definiamo sistemi ortogonali e ortonormali di vettori, spieghiamo come calcolare le coordinate di un vettore rispetto a una base ortogonale e come fabbricare una base ortonormale a partire da una qualsiasi. Successivamente discutiamo le decomposizioni di $\mathbb{R}^n$ in somma diretta ortogonale, introducendo le proiezioni ortogonali su sottospazi e illustrando come calcolarle. Infine parliamo di presentazioni parametriche e cartesiane di sottospazi, interpretando le seconde in termini del prodotto scalare e spiegando come passare dalle une alle altre, il che nel caso di rette e piani in $\mathbb{R}^3$ ci conduce in modo naturale all'introduzione del prodotto vettoriale tra due vettori.

4.1 Prodotto scalare, norma, distanza

Nel piano euclideo $\mathbb{R}^2$ calcoliamo la distanza tra $0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ tramite il teorema di Pitagora, ottenendo

$$d(0, x) = \sqrt{x_1^2 + x_2^2},$$

dunque la distanza tra due punti qualsiasi x e y vale

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

Dal teorema di Carnot segue allora che il coseno dell'angolo formato nell'origine da due vettori non nulli x e y vale

$$\begin{aligned} & \frac{d(0,x)^2 + d(0,y)^2 - d(x,y)^2}{2 \cdot d(0,x) \cdot d(0,y)} \\ = \quad & \frac{x_1^2 + x_2^2 + y_1^2 + y_2^2 - \left(x_1^2 - 2x_1y_1 + y_1^2 + x_2^2 - 2x_2y_2 + y_2^2\right)}{2 \cdot \sqrt{x_1^2 + x_2^2} \cdot \sqrt{y_1^2 + y_2^2}} \\ = \quad & \frac{x_1 \cdot y_1 + x_2 \cdot y_2}{\sqrt{x_1^2 + x_2^2} \cdot \sqrt{y_1^2 + y_2^2}}. \end{aligned}$$

Questo suggerisce di porre $x \bullet y = x_1 \cdot y_1 + x_2 \cdot y_2$, ovvero di introdurre una funzione

$$\begin{array}{rcl} \mathbb{R}^2 \times \mathbb{R}^2 & \to & \mathbb{R} \\ (x,y) & \mapsto & x \bullet y \end{array}$$

definita da $x \bullet y = x_1 \cdot y_1 + x_2 \cdot y_2$, in modo che la distanza tra 0 e x sia

$$\sqrt{x \bullet x}$$

e che il coseno dell'angolo tra x e y sia

$$\frac{x \bullet y}{\sqrt{x \bullet x} \cdot \sqrt{y \bullet y}}.$$

Questa definizione si estende facilmente a $\mathbb{R}^n$, introducendo

$$\begin{array}{rcl} \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R} \\ (x,y) & \mapsto & x \bullet y \end{array}$$

definita da

$$x \bullet y = x_1 \cdot y_1 + x_2 \cdot y_2 + \ldots + x_n \cdot y_n$$

che si può anche scrivere come

$$x \bullet y = {}^{\mathrm{t}}x \cdot y.$$

Si verifica facilmente che questa funzione, che chiameremo *prodotto scalare* di $\mathbb{R}^n$, soddisfa le proprietà seguenti:

- È bilineare, ovvero è lineare in ciascuno dei due argomenti fissato l'altro:

$$(a \cdot x + b \cdot y) \bullet z = a \cdot x \bullet z + b \cdot y \bullet z \qquad \forall a,b \in \mathbb{R},\ \forall x,y,z \in \mathbb{R}^n$$

$$x \bullet (a \cdot y + b \cdot z) = a \cdot x \bullet z + b \cdot x \bullet z \qquad \forall a,b \in \mathbb{R},\ \forall x,y,z \in \mathbb{R}^n;$$

- È simmetrica:
$$x \bullet y = y \bullet x \qquad \forall x, y \in \mathbb{R}^n;$$
- È definita positiva:
$$0 \bullet 0 = 0, \qquad x \bullet x > 0 \qquad \forall x \in \mathbb{R}^n,\ x \neq 0.$$

Possiamo allora definire la *norma associata* al prodotto scalare di $\mathbb{R}^n$ come la funzione

$$\begin{array}{rcl} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & \|x\| \end{array}$$

data da $\|x\| = \sqrt{x \bullet x}$, intesa come l'unica radice non negativa del numero non negativo $x \bullet x$. Si verifica allora facilmente che

$$\|a \cdot x\| = |a| \cdot \|x\| \qquad \forall a \in \mathbb{R},\ x \in \mathbb{R}^n.$$

Inoltre vale la seguente *disuguaglianza triangolare*, di dimostrazione meno immediata:
$$\|x + y\| \leqslant \|x\| + \|y\| \qquad \forall x, y \in \mathbb{R}^n.$$
Ne discende che se si definisce la distanza tra due punti

$$d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

ponendo
$$d(x, y) = \|x - y\|$$
si hanno le proprietà

- $d(x, x) = 0,\ d(x, y) > 0 \qquad \forall x, y \in \mathbb{R}^n,\ x \neq y;$
- $d(y, x) = d(x, y) \qquad \forall x, y \in \mathbb{R}^n;$
- $d(x, z) \leqslant d(x, y) + d(y, z) \qquad \forall x, y, z \in \mathbb{R}^n.$

Sistemi ortogonali e ortonormali di vettori Diciamo che:

- $v, w \in \mathbb{R}^n$ sono *ortogonali* tra loro, e scriviamo $v \perp w$, se $v \bullet w = 0$;
- $(v_1, \ldots, v_n)$ è un *sistema ortogonale* se $v_j \perp v_k$ per ogni $j \neq k$;
- $v \in \mathbb{R}^n$ è unitario se $\|v\| = 1$;

- $(v_1, \ldots, v_n)$ è un *sistema ortonormale* se è ortogonale e ogni v_j è unitario.

Si può verificare il fatto seguente:

Proposizione 4.1.1. *Se $\mathcal{B} = (v_1, \ldots, v_n)$ è una base ortogonale di $\mathbb{R}^n$ allora per ogni $v \in \mathbb{R}^n$ si ha*

$$v = \sum_{j=1}^{n} \frac{v \bullet v_j}{\|v_j\|^2} \cdot v_j,$$

ovvero la j-esima coordinata di v rispetto a $\mathcal{B}$ vale

$$\frac{v \bullet v_j}{\|v_j\|^2}.$$

Questa proposizione comporta che la determinazione delle coordinate di un vettore rispetto a una base ortogonale di $\mathbb{R}^n$ si esegue con un calcolo, senza la necessità di risolvere un sistema lineare. Vedremo nel prossimo paragrafo che la conoscenza di una base ortogonale di un sottospazio di $\mathbb{R}^n$ consente di calcolare in modo semplice la *proiezione ortogonale* su tale sottospazio. Emerge dunque in modo naturale l'esigenza di costruire basi ortogonali, a cui risponde il seguente:

Teorema 4.1.2. *Se $(w_1, \ldots, w_m)$ è una base di un sottospazio W di $\mathbb{R}^n$, il seguente procedimento produce una base ortonormale $(u_1, \ldots, u_n)$ di W:*

- *Si pone $u_1 = \frac{w_1}{\|w_1\|}$;*
- *Se $1 \leqslant k < m$, supposti costruiti $u_1, \ldots, u_k$, si pone*

$$v_{k+1} = w_{k+1} - \sum_{i=1}^{k} w_{k+1} \bullet u_i \cdot u_i, \qquad u_{k+1} = \frac{v_{k+1}}{\|v_{k+1}\|}.$$

Vale la pena di esplicitare i passi del procedimento appena descritto per $k = 1$ e $k = 2$:

- $v_2 = w_2 - w_2 \bullet u_1 \cdot u_1, \qquad u_2 = \frac{v_2}{\|v_2\|}$;
- $v_3 = w_3 - w_3 \bullet u_1 \cdot u_1 - w_3 \bullet u_2 \cdot u_2, \qquad u_3 = \frac{v_3}{\|v_3\|}$.

4.2 Proiezioni ortogonali

Dato $X \subset \mathbb{R}^n$, chiamiamo *sottospazio ortogonale* a X l'insieme

$$X^\perp = \{y \in \mathbb{R}^n : \; y \bullet x = 0 \; \forall x \in X\}.$$

È piuttosto semplice verificare i fatti seguenti:

- $X^\perp$ è sempre un sottospazio vettoriale di $\mathbb{R}^n$;
- $X^\perp = (\mathrm{Span}(X))^\perp$.

Meno facile è invece la seguente:

Proposizione 4.2.1. *Se $W \subset \mathbb{R}^n$ è un sottospazio, si ha $\mathbb{R}^n = W \oplus W^\perp$.*

Chiameremo *proiezione ortogonale* su W, indicandola con p_W, la proiezione su W associata alla decomposizione $\mathbb{R}^n = W \oplus W^\perp$. Di essa si può dare la seguente caratterizzazione geometrica:

Proposizione 4.2.2. *Dato $x \in \mathbb{R}^n$, il punto $p_W(x)$ è il punto di W più vicino a x, ovvero, posto $w_0 = p_W(x)$, si ha $d(w, x) > d(w_0, x)$ per ogni $w \in W$ diverso da w_0.*

Nel Capitolo 2 abbiamo visto che le matrici delle proiezioni associate a decomposizioni in somma diretta si caratterizzano come quelle matrici A tali che $A \cdot A = A$. Abbiamo ora la seguente specializzazione di questo risultato:

Proposizione 4.2.3. *Una matrice $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ rappresenta una proiezione ortogonale su un sottospazio se e soltanto se $A \cdot A = {}^{\mathrm{t}}A = A$.*

Veniamo ora alla manifestazione dell'utilità delle basi ortogonali già annunciata nel paragrafo precedente:

Proposizione 4.2.4. *Se $w_1, \ldots, w_m$ è una base ortogonale di un sottospazio W di $\mathbb{R}^n$ allora per ogni $x \in \mathbb{R}^n$ si ha*

$$p_W(x) = \sum_{j=1}^{m} \frac{x \bullet w_j}{\|w_j\|^2} \cdot w_j.$$

Concludiamo il paragrafo notanto che dalla definizione di proiezione ortogonale segue facilmente che per ogni $x \in \mathbb{R}^n$ si ha

$$x = p_W(x) + p_{W^\perp}(x) \quad \Rightarrow \quad p_W(x) = x - p_{W^\perp}(x).$$

Questo fatto è particolarmente utile quando si conosce una base ortogonale di $W^\perp$. Notiamo che se un sottospazio ha dimensione 1, ogni suo generatore ne costituisce una base ortogonale.

4.3 Equazioni cartesiane e parametriche di sottospazi

Per un sottospazio vettoriale W di $\mathbb{R}^n$ chiamiamo:

- *Presentazione parametrica* di W l'espressione di W come il generato di alcuni vettori, ovvero la scrittura
$$W = \operatorname{Span}(w_1, \ldots, w_m)$$
dove $w_1, \ldots, w_m$ sono certi vettori;
- *Presentazione cartesiana* di W l'espressione di W come l'insieme delle soluzioni di un sistema lineare omogeneo, ovvero la scrittura
$$W = \{x \in \mathbb{R}^n : \ A \cdot x = 0\}$$
dove $A \in \mathcal{M}_{h\times n}(\mathbb{R})$ è una certa matrice; notiamo che h è il numero di equazioni della presentazione.

È facile vedere che se W ha dimensione k allora ha sempre una presentazione parametrica con k vettori (basta prendere come generatori i vettori di una sua base), e che qualsiasi sua presentazione parametrica richiede almeno k vettori. Inoltre si può vedere che W ha sempre una presentazione cartesiana con $n-k$ equazioni, e che qualsiasi sua presentazione cartesiana ne richiede almeno $n-k$.

Osserviamo che una presentazione parametrica si può esprimere in molti modi; ad esempio le seguenti scritture sono equivalenti fra loro:

- $W = \operatorname{Span}\left(\begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ -7 \end{pmatrix}\right)$;
- $W = \left\{\begin{pmatrix} 2t+4s \\ -3t+s \\ 5t-7s \end{pmatrix} : \ t, s \in \mathbb{R}\right\}$;
- $W : \begin{cases} x = 2t+4s \\ y = -3t+s \\ z = 5t-7s. \end{cases}$

Poiché ogni sottospazio vettoriale di $\mathbb{R}^n$ ammette presentazioni sia parametriche sia cartesiane, si pone in modo naturale il problema di passare da una di un tipo a una dell'altro. A questo proposito notiamo quanto segue:

- Passare da una presentazione cartesiana $W = \{x : \ A \cdot x = 0\}$ a una parametrica significa risolvere il sistema $A \cdot x = 0$, ovvero trovare un insieme di generatori (meglio se è in realtà una base) del sottospazio W delle sue soluzioni;

- Per passare da una presentazione parametrica a una cartesiana si può usare l'ultima scrittura sopra esemplificata di presentazione parametrica e progressivamente eliminare i parametri, rimanendo con equazioni nelle sole coordinate. Nell'esempio questo si traduce in questi passaggi:

$$\begin{aligned}
&\begin{cases} x = 2t + 4s \\ y = -3t + s \\ z = 5t - 7s \end{cases} \\
\rightarrow \quad &\begin{cases} s = y + 3t \\ x = 2t + 4(y + 3t) \\ z = 5t - 7(y + 3t) \end{cases} \\
\rightarrow \quad &\begin{cases} s = y + 3t \\ x = 14t + 4y \\ z = -16t - 7y \end{cases} \\
\rightarrow \quad &\begin{cases} s = y + 3t \\ t = \frac{1}{14}(x - 4y) \\ z = -16 \cdot \frac{1}{14}(x - 4y) - 7y \end{cases} \\
\rightarrow \quad &\begin{cases} s = y + 3t \\ t = \frac{1}{14}(x - 4y) \\ 7z = -8x + 32y - 49y \end{cases} \\
\rightarrow \quad &\begin{cases} s = y + 3t \\ t = \frac{1}{14}(x - 4y) \\ 8x + 17y + 7z = 0 \end{cases}
\end{aligned}$$

  e dunque alla conclusione che W ha equazione cartesiana

$$8x + 17y + 7z = 0.$$

Proseguiamo notando che una equazione lineare omogenea per un punto x di $\mathbb{R}^n$, ovvero una del tipo

$$a_1 \cdot x_1 + \ldots + a_n \cdot x_n = 0$$

può essere scritta come $v \bullet x = 0$, con $v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n$. Dunque una presentazione cartesiana di un sottospazio W di $\mathbb{R}^n$ del tipo $W = \{x : \ A \cdot x = 0\}$, con $A \in \mathcal{M}_{h\times n}(\mathbb{R})$, si può scrivere come

$$W = (\mathrm{Span}(v_1, \ldots, v_h))^{\perp},$$

dove $v_j \in \mathbb{R}^n$ è il vettore ottenuto scrivendo come colonna la j-esima riga di A.

4.4 Equazioni di rette e piani

Nel piano $\mathbb{R}^2$ una retta (sottospazio di dimensione 1) ha presentazione parametrica minimale data da 1 generatore, e cartesiana minimale data da $2 - 1 = 1$ equazione. Dunque si tratta di presentazioni del tipo

$$\ell = \mathrm{Span}\begin{pmatrix} a \\ b \end{pmatrix} \qquad \text{oppure} \qquad \ell : \ \alpha \cdot x + \beta \cdot y = 0.$$

E passare da un tipo all'altro è immediato:

$$\begin{aligned} \ell = \mathrm{Span}\begin{pmatrix} a \\ b \end{pmatrix} \quad &\Rightarrow \quad \ell : \ b \cdot x - a \cdot y = 0, \\ \ell : \ \alpha \cdot x + \beta \cdot y = 0 \quad &\Rightarrow \quad \ell = \mathrm{Span}\begin{pmatrix} \beta \\ -\alpha \end{pmatrix}. \end{aligned}$$

Rivolgiamoci ora allo spazio $\mathbb{R}^3$. Qui abbiamo come sottospazi non banali le rette (dimensione 1) e i piani (dimensione 2). Inoltre per una retta una presentazione parametrica minimale è data da 1 generatore e una cartesiana minimale da $3 - 1 = 2$ equazioni, mentre per un piano una presentazione parametrica minimale è data da 2 generatori e una cartesiana minimale da $3 - 2 = 1$ equazione.

Poniamoci ora il problema di come passare da una tipologia di presentazione all'altra. I passaggi da retta parametrica a retta cartesiana e da piano cartesiano a piano parametrico sono tra loro analoghi e facili:

$$\ell = \mathrm{Span}\begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \Rightarrow \quad \ell : \begin{cases} b \cdot x - a \cdot y = 0 \\ c \cdot x - a \cdot z = 0 \\ c \cdot y - b \cdot z = 0, \end{cases}$$

$$\pi : \ \alpha \cdot x + \beta \cdot y + \gamma \cdot z = 0 \quad \Rightarrow \quad \pi = \mathrm{Span}\left(\begin{pmatrix} \beta \\ -\alpha \\ 0 \end{pmatrix}, \begin{pmatrix} \gamma \\ 0 \\ -\alpha \end{pmatrix}, \begin{pmatrix} 0 \\ \gamma \\ -\beta \end{pmatrix}\right).$$

Osserviamo che in entrambi i casi otteniamo presentazioni non minimali (3 equazioni cartesiane invece di 2 per una retta, 3 generatori invece di 2 per un piano), tuttavia si può vedere che nel primo caso una delle equazioni e nel secondo uno dei generatori si può sempre scartare.

Occupiamoci ora dei due passaggi opposti, ovvero quelli da retta cartesiana a retta parametrica e da piano parametrico a piano cartesiano. Nel primo caso abbiamo già osservato alla fine del paragrafo precedente che una presentazione cartesiana di una retta ℓ si può interpretare come la scrittura

$$\ell = (\mathrm{Span}(v_1, v_2))^{\perp}$$

con $v_1, v_2 \in \mathbb{R}^3$, dunque dare una presentazione parametrica di ℓ significa trovare un vettore ortogonale a v_1 e v_2. Una presentazione parametrica di un piano è invece una del tipo

$$\pi = \mathrm{Span}(w_1, w_2)$$

e trovare una sua presentazione parametrica significa esprimerlo come

$$\pi = (\mathrm{Span}(v))^{\perp},$$

dunque di nuovo si tratta, dati vettori $w_1, w_2 \in \mathbb{R}^3$, di trovare $v \in \mathbb{R}^3$ ortogonale a entrambi. Il prossimo risultato illustra come compiere questa operazione in modo sistematico:

Proposizione 4.4.1. *Dati vettori*

$$v_1 = \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}$$

il vettore

$$v_1 \times v_2 = \begin{pmatrix} b_1 \cdot c_2 - c_1 \cdot b_2 \\ -(a_1 \cdot c_2 - c_1 \cdot a_2) \\ a_1 \cdot b_2 - b_1 \cdot c_2 \end{pmatrix}$$

ha le proprietà seguenti:

- *È ortogonale a* v_1 *e* v_2*;*
- *Ha norma uguale all'area del parallelogrammo di lati* v_1 *e* v_2 *(in particolare: è nullo se e solo se* v_1 *e* v_2 *sono linearmente dipendenti);*
- *Ha verso tale che si possano disporre le dita pollice, indice e medio della mano destra lungo i vettori* v_1*,* v_2 *e* $v_1 \times v_2$ *rispettivamente.*

Chiamiamo $v_1 \times v_2$ il *prodotto vettoriale* di v_1 e v_2. Notiamo che il coefficiente sulla riga j di $v_1 \times v_2$ è il determinante della sottomatrice 2×2 della matrice (v_1, v_2) ottenuta cancellando la riga j, cambiato di segno per $j = 2$.

Esempio 4.4.2. Se $\pi = \operatorname{Span}\left(\begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix}, \begin{pmatrix} 7 \\ 3 \\ -2 \end{pmatrix}\right)$, essendo

$$\begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix} \times \begin{pmatrix} 7 \\ 3 \\ -2 \end{pmatrix} = \begin{pmatrix} -13 \\ 39 \\ 13 \end{pmatrix} = -13 \cdot \begin{pmatrix} 1 \\ -3 \\ -1 \end{pmatrix}$$

abbiamo che π ha equazione cartesiana $x - 3y - z = 0$.

Esempio 4.4.3. Se $\ell : \begin{cases} 3x + 2y - 7z = 0 \\ 4x - y + 5z = 0, \end{cases}$ essendo

$$\begin{pmatrix} 3 \\ 2 \\ -7 \end{pmatrix} \times \begin{pmatrix} 4 \\ -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \\ -43 \\ -11 \end{pmatrix},$$

abbiamo che $\ell = \operatorname{Span}\begin{pmatrix} 3 \\ -43 \\ -11 \end{pmatrix}$.

Capitolo 5

Diagonalizzazione

In questo capitolo introduciamo la nozione di diagonalizzabilità per matrici e illustriamo i criteri per stabilire se una matrice sia diagonalizzabile o meno, enunciando il fatto che lo è se è simmetrica.

5.1 Autovalori e polinomio caratteristico

Supponiamo di avere un corpo materiale rigido C (per semplicità, bidimensionale) soggetto a forze ma in equilibrio, cioè tale che la risultante di tutte le forze sul suo baricentro sia nulla. Se introduciamo un sistema di riferimento cartesiano con origine nel baricentro di C, abbiamo che C diventa un sottoinsieme di $\mathbb{R}^2$ e la forza agente su un punto x di C è un vettore $F(x) \in \mathbb{R}^2$. Dunque abbiamo una funzione $F : C \to \mathbb{R}^2$ tale che $F(0) = 0$. Supponendo che C sia non troppo grande possiamo approssimare F al primo ordine, cioè sostituire $F(x)$ con $J \cdot x$, dove $J \in \mathcal{M}_{2\times 2}(\mathbb{R})$ è la matrice jacobiana di F. Possiamo ora osservare che

$$J = S + A \qquad \text{con} \quad S = \frac{1}{2}(J + {}^{\mathrm{t}}J) \quad \text{e} \quad A = \frac{1}{2}(J - {}^{\mathrm{t}}J),$$

dunque con S simmetrica e A antisimmetrica. L'effetto di A, che ha la forma

$$A = \begin{pmatrix} 0 & k \\ -k & 0 \end{pmatrix}$$

è chiaramente quello di una rotazione di C intorno all'origine, dunque un movimento rigido, mentre quello di S è meno chiaro. Ad esempio se

$$S = \frac{1}{4}\begin{pmatrix} 5 & 3\sqrt{3} \\ 3\sqrt{3} & -1 \end{pmatrix}$$

non è affatto evidente a prima vista quale sia l'azione a cui C è soggetto. Si può tuttavia verificare con i calcoli che posto

$$v_1 = \frac{1}{2}\begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} \qquad v_2 = \frac{1}{2}\begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}$$

si ha che (v_1, v_2) è una base ortonormale di $\mathbb{R}^2$ e

$$S \cdot v_1 = 2v_1 \qquad S \cdot v_2 = -v_2$$

dunque abbiamo due direzioni tra loro perpendicolari tali che lungo la prima l'effetto di S è quello di una trazione diretta lungo la direzione stessa, e lungo la seconda l'effetto di S è quello di una compressione diretta lungo la direzione stessa, pertanto l'azione di S è molto ben compresa. Notiamo che se poniamo $X = (v_1, v_2)$, la formula precedente si traduce nella relazione

$$S \cdot X = X \cdot \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \quad \Rightarrow \quad X^{-1} \cdot S \cdot X = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

Questo suggerisce la seguente:

Definizione 5.1.1. Diciamo che $M \in \mathcal{M}_{n\times n}(\mathbb{R})$ è *diagonalizzabile* se esiste una base $X = (v_1, \ldots, v_n)$ di $\mathbb{R}^n$ tale che $M \cdot v_j = \lambda_j \cdot v_j$ per $j = 1, \ldots, n$ per opportuni $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, ovvero

$$X^{-1} \cdot M \cdot X = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix}$$

(dove l'ultima matrice è *diagonale*, ovvero tutti coefficienti fuori dalla diagonale principale sono nulli).

Vedremo che alcune M ma non tutte sono diagonalizzabili, e questo capitolo è interamente dedicato alle tecniche per comprendere quelle che lo sono.

La definizione precedente si può in realtà estendere a qualsiasi applicazione lineare $f : V \to V$, dove V è uno spazio vettoriale reale di dimensione n, dicendo che è diagonalizzabile se esiste una base $\mathcal{B}$ di V tale che $[f]_{\mathcal{B}}^{\mathcal{B}}$ sia diagonale. Scelta una base $\mathcal{C}$ qualsiasi di V e posto $M = [f]_{\mathcal{C}}^{\mathcal{C}}$, la diagonalizzabilità di f equivale a quella di M, infatti

$$[f]_{\mathcal{B}}^{\mathcal{B}} = X^{-1} \cdot [f]_{\mathcal{C}}^{\mathcal{C}} \cdot X$$

dove X è la matrice di cambio di base da $\mathcal{C}$ a $\mathcal{B}$. Nel resto del capitolo comunque ci riferiremo sempre al problema della diagonalizzabilità per matrici.

L'idea fondamentale per capire se $M \in \mathcal{M}_{n\times n}(\mathbb{R})$ sia diagonalizzabile è quella di cercare i vettori $(v_1, \ldots, v_n)$ della (eventuale) base diagonalizzante, e i corrispondenti $\lambda_1, \ldots, \lambda_n$, non tutti insieme ma "uno alla volta":

Definizione 5.1.2. Data $M \in \mathcal{M}_{n\times n}(\mathbb{R})$ diciamo che $\lambda \in \mathbb{R}$ è un *autovalore* di M se esiste $v \in \mathbb{R}^n$ non nullo tale che $M \cdot v = \lambda \cdot v$. Ogni tale v è detto *autovettore* di M relativo a λ.

Proposizione 5.1.3. *Posto $p_M(t) = \det(t \cdot \mathrm{I}_n - M)$, con t indeterminata, si ha:*

- *$p_M(t)$ è un polinomio di grado n e monico, ovvero t^n ha coefficiente 1;*
- *$\lambda \in \mathbb{R}$ è autovalore di M se e solo se è radice di $p_M(t)$, ovvero $p_M(\lambda) = 0$.*

Dimostrazione. Ciascun coefficiente di $t \cdot \mathrm{I}_n - M$ è un polinomio di grado 1 (sulla diagonale principale) o 0 (fuori da essa) in t, e il determinante è una somma di prodotti di coefficienti, dunque $p_M(t)$ è un polinomio in t. Inoltre il grado più alto si ottiene prendendo il prodotto di tutti gli elementi sulla diagonale principale, da cui segue che il grado è n e il coefficiente di t^n è 1.

Per la seconda affermazione abbiamo questa catena di equivalenze:

$$\begin{aligned}
& \lambda \text{ autovalore} \\
\Leftrightarrow\ & \exists v \neq 0 \text{ con } M \cdot v = \lambda \cdot v \\
\Leftrightarrow\ & \exists v \neq 0 \text{ con } \lambda \cdot v - M \cdot v = 0 \\
\Leftrightarrow\ & \exists v \neq 0 \text{ con } \lambda \cdot \mathrm{I}_n \cdot v - M \cdot v = 0 \\
\Leftrightarrow\ & \exists v \neq 0 \text{ con } (\lambda \cdot \mathrm{I}_n - M) \cdot v = 0 \\
\Leftrightarrow\ & \det(\lambda \cdot \mathrm{I}_n - M) = 0 \\
\Leftrightarrow\ & p_M(\lambda) = 0.
\end{aligned}$$

□

Questo risultato dice intanto che per una matrice M diagonalizzabile devono esistere radici reali di $p_M(t)$. Possiamo allora concludere che per

$$M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

essendo

$$p_M(t) = \det \begin{pmatrix} t & 1 \\ -1 & t \end{pmatrix} = t^2 + 1$$

il problema della diagonalizzabilità ha soluzione negativa. Abbiamo in realtà qualcosa di più:

Proposizione 5.1.4. *Se $M \in \mathcal{M}_{n\times n}(\mathbb{R})$ è diagonalizzabile allora $p_M(t)$ ha n radici reali contate con la loro molteplicità.*

Dimostrazione. Se

$$X^{-1} \cdot M \cdot X = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix}$$

abbiamo

$$\begin{aligned} p_M(t) &= \det(t \cdot \mathrm{I}_n - M) \\ &= \det(X) \cdot \det(t \cdot \mathrm{I}_n - M) \cdot \frac{1}{\det(X)} \\ &= \det(X) \cdot \det(t \cdot \mathrm{I}_n - M) \cdot \det(X^{-1}) \\ &= \det\big(X \cdot (t \cdot \mathrm{I}_n - M) \cdot X^{-1}\big) \\ &= \det\big(t \cdot X \cdot \mathrm{I}_n \cdot X^{-1} - X \cdot M \cdot X^{-1}\big) \\ &= \det\left(t \cdot \mathrm{I}_n - \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix}\right) \\ &= \det\begin{pmatrix} t-\lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & t-\lambda_n \end{pmatrix} \\ &= (t-\lambda_1)\cdots(t-\lambda_n). \end{aligned}$$

□

Un'analisi più accurata di quella che ha mostrato che $p_M(t)$ è monico di grado n conduce ai fatti seguenti:

- In $p_M(t)$ il coefficiente di t^{n-1} è $-\mathrm{tr}(M)$;
- In $p_M(t)$ il termine noto è $(-1)^n \cdot \det(M)$.

Ne segue che se M è diagonalizzabile allora:

- $\mathrm{tr}(M)$ è la somma degli autovalori di M (contati con la loro molteplicità);
- $\det(M)$ è il prodotto degli autovalori di M (contati con la loro molteplicità).

5.2 Molteplicità algebrica e geometrica, teorema spettrale

Data $M \in \mathcal{M}_{n\times n}(\mathbb{R})$, se v_1 e v_2 sono autovalori di M relativi ad autovalori λ_1 e λ_2 distinti, si verifica facilmente che v_1 e v_2 sono linearmente indipendenti: se fosse $v_2 = t \cdot v_1$, si avrebbe

$$M \cdot v_2 = M \cdot t \cdot v_1 = t \cdot M \cdot v_1 = t \cdot \lambda_1 \cdot v_1 = \lambda_1 \cdot t \cdot v_1 = \lambda_1 \cdot v_2 \neq \lambda_2 \cdot v_2.$$

In generale si può vedere che autovettori relativi ad autovalori distinti sono tra loro linearmente indipendenti, da cui segue:

Proposizione 5.2.1. *Se $M \in \mathcal{M}_{n\times n}(\mathbb{R})$ e $p_M(t)$ ha n radici reali distinte allora M è diagonalizzabile.*

Naturalmente però anche matrici con autovalori non distinti possono essere diagonalizzabili (basta prendere una matrice diagonale con coefficienti non distinti sulla diagonale). Per caratterizzare completamente le matrici diagonalizzabili introduciamo due nuove nozioni. Se λ è un autovalore di $M \in \mathcal{M}_{n\times n}(\mathbb{R})$ chiamiamo:

- Molteplicità algebrica di λ la sua molteplicità come radice di $p_M(t)$, indicata con m.a.(λ);
- Molteplicità geometrica di λ la dimensione del nucleo di $\lambda\cdot\mathrm{I}_n - M$, indicata con m.g.$(\lambda)$.

Abbiamo allora:

Teorema 5.2.2. *Sia $M \in \mathcal{M}_{n\times n}(\mathbb{R})$.*

- *Per ogni autovalore λ di M si ha*

$$1 \leqslant m.g.(\lambda) \leqslant m.a.(\lambda);$$

- *M è diagonalizzabile se e soltanto se ha n autovalori contati con la loro molteplicità algebrica, e per ciascuno λ di essi si ha*

$$m.g.(\lambda) = m.a.(\lambda).$$

Concludiamo con questo risultato, detto *teorema spettrale*, che non dimostriamo:

Teorema 5.2.3. *Se $S \in \mathcal{M}_{n\times n}(\mathbb{R})$ è simmetrica allora esiste una base ortonormale di $\mathbb{R}^n$ costituita da autovettori di S. In particolare, S è diagonalizzabile.*